# Estadística para ingenieros técnicos

Antonio Gámez Mellado - Luís M. Marín Trechera

UNIVERSIDAD DE CÁDIZ
SERVICIO DE PUBLICACIONES

# ESTADÍSTICA PARA INGENIEROS TÉCNICOS

Gámez Mellado, Antonio

Estadística para ingenieros técnicos / Antonio Gámez Mellado, Luis M. Marín Trechera. -- Cádiz: Servicio de Publicaciones de la Universidad, 2000. -- IX, 475 p.

ISBN 84-7786-685-6

1. Estadísticas para ingenieros - Tratados, manuales, etc. I. Marín Trechera, Luis M. II. Universidad de Cádiz. Servicio de Publicaciones, ed. III. Título

519.2

SERVICIO DE PUBLICACIONES DE LA UNIVERSIDAD DE CÁDIZ



I.S.B.N.: 84-7786-685-6
Diseño de Cubierta: Creasur
Imprime: Essan Graphic, S.L. Av. de la Marina, s/n.
Punta Umbría - Huelva.
Depósito Legal: H-347-2000

UNIVERSIDAD DE CÁDIZ
DEPARTAMENTO DE ESTADÍSTICA

# Estadística para ingenieros técnicos

PRESENTADO POR

ANTONIO GÁMEZ MELLADO
LUIS M. MARÍN TRECHERA

PROFESORES TITULARES DE ESCUELA UNIVERSITARIA
DEL ÁREA DE
ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

UNIVERSIDAD DE CÁDIZ
SERVICIO DE PUBLICACIONES
2000

# Presentación

Este libro nace dentro del departamento de Estadística e Investigación Operativa que la Universidad de Cádiz, y es el fruto del trabajo realizado por los autores durante más de 10 años de docencia en la Escuela Superior de Ingeniería de la Universidad de Cádiz.

Tenemos que reconocer que es el resultado del estudio, experiencia y reflexiones de los autores, así como de los comentarios y sugerencias de nuestros compañeros, con las que nos enriquecemos en nuestro contacto diario.

Es un trabajo realista y abierto, que iremos mejorando y enriqueciendo día a día a través de las reflexiones de nuestros alumnos. Trata de asumir y de abordar las necesidades de nuestros alumnos en la Escuela Superior de Ingeniería de Cádiz.

En resumen, esta obra no ha de entenderse como un plan ideal y utópico, sino como el planteamiento de nuestra forma actual de interpretar la enseñanza, lo que seguramente, el tiempo, la experiencia, y las circunstancias puntuales se encargarán de matizar, madurar y adecuar.

A continuación vamos a comentar, a grandes rasgos, el contenido de este libro y cómo lo hemos estructurado.

Consta de 6 partes que forman una única estructura en la que se ha intentado presentar además de una propuesta docente, las ideas que subyacen en su confección. Dentro de cada parte se engloban tanto los contenidos mínimos que nuestros alumnos necesitan, denominado *resumen teórico*, se presentan además un número suficiente de problemas del campo de la ingeniería en los que aparecen los comentarios necesarios para su comprensión, *problemas resueltos*, y por último detallamos un elevado número de problemas cuya solución aparece al final del libro, *problemas propuestos.*

La primera parte se dedica a la Estadística Descriptiva de una variable.

La segunda parte constituye el estudio descriptivo de dos variables.

La tercera parte trata del cálculo de probabilidades, y variable aleatoria.

En la cuarta parte se estudian las principales distribuciones estadísticas, tanto discretas como continuas.

En la quinta parte se estudia la estadística inferencial.

En la sexta parte se dedica a estimación y contraste de hipótesis.

Queremos agradecer las observaciones, sugerencias y críticas, tanto de los compañeros del Departamento de Matemáticas, del Departamento de Estadística, como de todas las personas que de una u otra forma nos han prestado su inestimable ayuda en la realización de este libro.

# Contenido

Presentación . . . . . . . . . . i

**I Síntesis de la Información Estadística. 1**

**Resumen Teórico 3**

1.1 Introducción. . . . . . . . . . 3

1.2 Variable y atributo. . . . . . . . . . 3

1.2.1 Variables discretas y continuas. . . . . . . . . . 4

1.2.2 Series estadísticas. . . . . . . . . . 4

1.3 Distribución de datos. . . . . . . . . . 5

1.4 Representaciones gráficas. . . . . . . . . . 7

1.5 Medidas de centralización. . . . . . . . . . 8

1.5.1 La media. . . . . . . . . . 9

1.5.2 La mediana. . . . . . . . . . 13

1.5.3 La moda. . . . . . . . . . 14

1.5.4 Comparación entre Media, Moda y Mediana. . . . . . 15

1.6 Medidas de posición. . . . . . . . . . 15

1.7 Medidas de dispersión. . . . . . . . . . 16

1.7.1 Varianza y Desviación Típica. . . . . . . . . . 16

1.7.2 Otras medidas de Dispersión. . . . . . . . . . . . . . . 17
1.8 Desigualdad de Tchebychev. . . . . . . . . . . . . . . . . . . . 18
1.9 Momentos de la distribución. . . . . . . . . . . . . . . . . . . 19
1.9.1 Momentos respecto al Origen. . . . . . . . . . . . . . . 19
1.9.2 Momentos respecto a la Media. . . . . . . . . . . . . . 19
1.10 Medidas de simetría. . . . . . . . . . . . . . . . . . . . . . . 20
1.11 Medidas de Curtosis. . . . . . . . . . . . . . . . . . . . . . . 21
1.12 Transformaciones. . . . . . . . . . . . . . . . . . . . . . . . . 22
1.12.1 Normalización o tipificacion. . . . . . . . . . . . . . . 22
1.13 Medidas de la desigualdad. . . . . . . . . . . . . . . . . . . . 23
1.13.1 La curva de Lorenz. . . . . . . . . . . . . . . . . . . . . 24
1.13.2 Medidas de desigualdad relacionadas con las curvas de Lorenz. . . . . . . . . . . . . . . . . . . . . . . . . . . 25
1.13.3 Comparación en desigualdad . . . . . . . . . . . . . . . 26

**Problemas Resueltos** 29

**Problemas Propuestos** 49

**II Análisis conjunto de variables estadísticas.** 71

**Resumen Teórico** 73

2.1 Conceptos Generales. . . . . . . . . . . . . . . . . . . . . . . . 73
2.1.1 Distribución conjunta de dos variables. . . . . . . . . . 73
2.1.2 Distribuciones marginales. . . . . . . . . . . . . . . . . 74
2.1.3 Distribuciones condicionadas. . . . . . . . . . . . . . . 75
2.1.4 Covarianza y Correlación. . . . . . . . . . . . . . . . . 75
2.1.5 Vector de medias. . . . . . . . . . . . . . . . . . . . . . 78

2.1.6 Matriz de varianzas y covarianzas. . . . . . . . . . . . 78

2.2 Ajuste y regresión en $I\!R^2$. . . . . . . . . . . . . . . . . . . . . 79

2.2.1 Introducción. . . . . . . . . . . . . . . . . . . . . . . . 79

2.2.2 Ajuste, criterio de los mínimos cuadrados. . . . . . . . 80

2.2.3 Análisis de la bondad del ajuste. . . . . . . . . . . . . 84

2.3 Regresión. Método de regresión a la media. . . . . . . . . . . 86

2.4 Análisis de la bondad de la regresión. . . . . . . . . . . . . . 88

2.5 Análisis de atributos. . . . . . . . . . . . . . . . . . . . . . . 91

**Problemas Resueltos** **95**

**Problemas Propuestos** **113**

**III Probabilidad. Variable Aleatoria.** **131**

**Resumen Teórico** **133**

3.1 Evolución histórica. . . . . . . . . . . . . . . . . . . . . . . . 133

3.2 Distintas definiciones del concepto de probabilidad. . . . . . 134

3.3 Propiedades de la función de probabilidad. . . . . . . . . . . 136

3.4 Probabilidad condicionada. Independencia. . . . . . . . . . . 137

3.5 Teorema de la probabilidad total. Teorema de Bayes. . . . . 138

3.6 Anexo: Combinatoria. . . . . . . . . . . . . . . . . . . . . . . 140

3.7 Variable aleatoria. . . . . . . . . . . . . . . . . . . . . . . . . 141

3.7.1 Variables discretas y continuas. . . . . . . . . . . . . . 144

3.8 Variable aleatoria Unidimensional. . . . . . . . . . . . . . . . 145

3.8.1 Funciones que caracterizan globalmente a una Variable Aleatoria. . . . . . . . . . . . . . . . . . . . . . . . 145

3.8.2 La función esperanza matemática: $E[X]$. Propiedades. 149

3.8.3 Funciones que caracterizan globalmente a una Variable Aleatoria. . . . . 150

3.8.4 La función característica y la función generatriz de momentos. . . . . 152

3.8.5 Cambio de variable. . . . . 153

3.8.6 La desigualdad de Tchebychev. . . . . 153

3.9 Variables aleatorias $n$-dimensionales. . . . . 154

3.9.1 Distribución conjunta y marginales. . . . . 154

3.9.2 Distribuciones condicionadas. Independencia. . . . . 157

3.9.3 Esperanza. Propiedades. Covarianza y correlación. . . 158

3.9.4 Cambio de variables. . . . . 160

**Problemas Resueltos** **161**

**Problemas Propuestos** **183**

**IV Distribuciones de Probabilidad.** **221**

**Resumen Teórico** **223**

4.1 Distribuciones discretas más usuales. . . . . 223

4.1.1 Distribución de Bernoulli. . . . . 223

4.1.2 Distribución binomial. . . . . 224

4.1.3 Distribución Hipergeométrica. . . . . 226

4.1.4 Distribución geométrica o de Pascal. . . . . 228

4.1.5 Distribución Binomial Negativa. . . . . 228

4.1.6 Distribución de Poisson. . . . . 229

4.2 Distribuciones continuas más usuales. . . . . 231

4.2.1 Distribución exponencial negativa. . . . . 231

4.2.2 Distribución Gamma. . . . . 233

4.2.3 Distribución uniforme. . . . . 235

4.3 Distribución Normal. . . . . 236

4.3.1 Propiedades de la distribución Normal. . . . . 237

4.3.2 Teorema Central del límite. . . . . 239

4.3.3 Distribución Continua Truncada. . . . . 240

**Problemas Resueltos** **241**

**Problemas Propuestos** **271**

**V Fundamentos de la Inferencia Estadística. Estimación. 291**

**Resumen Teórico** **293**

5.1 Introducción a la Inferencia. . . . . 293

5.1.1 Muestreo aleatorio simple. . . . . 294

5.2 Estadísticos y Estimadores. . . . . 294

5.3 Propiedades de los estimadores. . . . . 296

5.4 Obtención de estimadores. . . . . 298

5.4.1 Método de los momentos. . . . . 298

5.4.2 Método de máxima verosimilitud. . . . . 298

5.5 Distribuciones asociadas a la Normal. . . . . 299

5.5.1 Distribución $\chi^2$ de Pearson. . . . . 299

5.5.2 Distribución t de Student. . . . . 300

5.5.3 Distribución F de Fisher-Snedecor. . . . . 300

5.6 Distribuciones muestrales. . . . . 301

5.6.1 Distribución muestral de la media de una población con varianza conocida. . . . . 301

5.6.2 Distribución muestral de la diferencia de medias de dos poblaciones con varianza conocida. . . . . . . . . . 302

5.6.3 Distribución muestral de $(n-1)S^2/\sigma^2$. . . . . . . . . 302

5.6.4 Distribución muestral de la media de una población con varianza desconocida. . . . . . . . . . . . . . . . . . 303

5.6.5 Distribución muestral de la diferencia de medias de dos poblaciones con varianza desconocida. . . . . . . . 304

5.6.6 Distribución muestral del cociente de varianzas. . . . . 304

**Problemas Resueltos** **307**

**Problemas Propuestos** **335**

**VI Intervalos de Confianza y Contraste de Hipótesis.** **343**

**Resumen Teórico** **345**

6.1 Tipos de estimación . . . . . . . . . . . . . . . . . . . . . 345

6.2 Resumen de estimadores . . . . . . . . . . . . . . . . . . . 346

6.3 Intervalos de Confianza para la media en poblaciones normales 347

6.3.1 Con varianza conocida . . . . . . . . . . . . . . . . . . 347

6.3.2 Con varianza desconocida . . . . . . . . . . . . . . . . 348

6.4 Intervalos de Confianza para la diferencia de medias en poblaciones normales . . . . . . . . . . . . . . . . . . . . . . . 349

6.5 Intervalos de Confianza para la varianza en poblaciones normales . . . . . . . . . . . . . . . . . . . . . . . . . . . . 349

6.6 Intervalos de Confianza para el cociente entre varianzas en poblaciones normales . . . . . . . . . . . . . . . . . . . . . 350

6.7 Contraste de Hipótesis. . . . . . . . . . . . . . . . . . . . . 350

6.8 Contraste de Hipótesis para la media en una población normal. 352

6.8.1 Con varianza conocida . . . . . 353

6.8.2 Con varianza desconocida . . . . . 353

6.9 Contrastes sobre diferencias de medias . . . . . 354

6.10 Selección del tamaño de la muestra para el contraste de medias. 356

6.11 Contrastes relacionados con proporciones . . . . . 357

6.12 Contrastes sobre varianzas. . . . . . 358

**Problemas Resueltos** **361**

**Problemas Propuestos** **387**

Soluciones de los Problemas Propuestos. **402**

**1 Soluciones de la Unidad Temática 1** **405**

**2 Soluciones de la Unidad Tematica 2** **421**

**3 Soluciones de la Unidad Temática 3** **429**

**4 Soluciones de la Unidad Temática 4** **457**

**5 Soluciones de la Unidad Temática 5** **465**

**6 Soluciones de la Unidad Temática 6** **469**

Anexo I: Tablas Estadísticas. **475**

# Unidad Temática I

# Síntesis de la Información Estadística.

# Resumen Teórico

## 1.1 Introducción.

Los datos constituyen la materia prima de la Estadística, pudiéndose establecer distintas clasificaciones en función de la forma en que éstos vengan dados. Obtenemos datos al realizar cualquier tipo de prueba, experimento, valoración, medición, etc.

## 1.2 Variable y atributo.

Tendríamos una primera clasificación en función de que las observaciones resultantes del experimento realizado sean de tipo cualitativo o cuantitativo, en el primero de los casos sería un atributo y en el segundo una variable.

Como ejemplos de atributos podríamos considerar el color del pelo de un colectivo de personas, su raza o el idioma(s) que hablan y como variables su estatura, peso o edad.

Para poder operar con un atributo es necesario asignar a cada una de las clases que toma un valor numérico, con lo que lo transformamos en una variable, esta asignación se hará de forma que los resultados que se obtengan al final del estudio a realizar sean fácilmente interpretables. Si el atributo en estudio fuera el sexo de un grupo de personas asignaríamos el valor 0 a la clase varón y el 1 a la clase mujer. Si la media aritmética diera igual a 0,51 deduciríamos que el 51% de los individuos son mujeres y el 49% restante varones. También podríamos haber asignado el valor 1,372 al varón y el $-0{,}513$ a la mujer, pero un resultado para la media de 0,917 sería difícil de interpretar.

### 1.2.1 Variables discretas y continuas.

Dentro del conjunto de las variables podemos distinguir entre Discretas y Continuas. Decimos que una variable es Discreta cuando entre dos valores consecutivos ésta no toma valores intermedios y que es Continua cuando puede tomar cualquier valor dentro de un intervalo. La estatura de un grupo de personas sería una varible Continua, mientras que el número de cabellos que tienen en la cabeza sería una variable Discreta. En la práctica todas las variables son discretas debido a la limitación de los aparatos de medida, y así, en el ejemplo de las estaturas, quizás podríamos detectar una diferencia de una cienmilésima de metro, o, a lo sumo, de una millonésima, pero dados dos individuos que se diferencien en una millonésima no detectaremos otro que tenga una estatura intermedia. De todas formas, en general trataremos a las variables "teóricamente" Continuas como tales.

Si la ocasión lo requiere tenemos la posibilidad de transformar una variable Discreta en Continua o viceversa. Para convertir una variable Discreta en Continua, una vez ordenados los valores asignaremos a cada uno de ellos un intervalo que tenga por extremos el punto medio respecto al valor anterior y el punto medio respecto al valor siguiente: por ejemplo la variable Discreta que toma los valores 1 , 2 , 4 , 7 y 9 se transformaría en una variable Continua que tomaría valores en los intervalos $(0{,}5; 1{,}5]$, $(1{,}5; 3]$, $(3; 5{,}5]$, $(5{,}5; 8]$ y $(8, 10)$. Para transformar una variable Continua en Discreta basta con hacer corresponder a cada uno de los intervalos su punto medio o Marca de Clase.

El cuadro que podemos observar en la figura 1.1 recoge los distintos tipos de caracteres estadísticos y las relaciones existentes entre ellos.

### 1.2.2 Series estadísticas.

Además de por su naturaleza podemos realizar distintas clasificaciones del conjunto de los datos o Serie Estadística atendiendo a:

1. Por su número.
   (a) Finitas. Las que tienen un número finito de elementos.
   (b) Infinitas. Cuando tienen infinitos elementos.
2. Por su obtención.

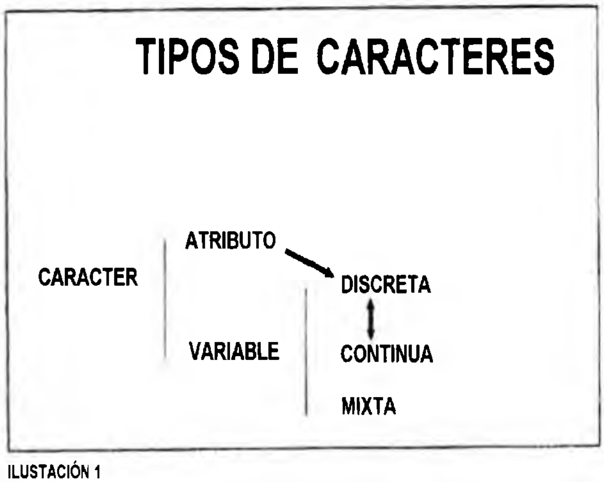



Figura 1.1: Diagrama de caracteres estadísticos

(a) Objetiva. Obtenida con métodos exactos de medición.

(b) Subjetiva. Obtenidas con apreciaciones personales.

3. Por su dimensión.

(a) Unidimensional. $x_1; x_2; x_3; ...; x_n$

(b) Bidimensional. $(x_1; y_1), (x_2; y_2), ..., (x_n; y_n)$

(c) n-dimensional $(x_1^1; x_2^1; ...; x_n^1), ..., (x_1^k; x_2^k; x_n^k)$

4. Por su dependencia temporal.

(a) Temporales. Los valores se toman en períodos de tiempo.

(b) Atemporales. No depende de ningún soporte temporal.

## 1.3 Distribución de datos.

Es importante organizar los datos de forma que tengamos un fácil acceso a ellos y podamos trabajar lo más cómodamente posible. Ello va a depender del número de observaciones distintas que se tengan y de las veces que se repitan cada una de ellas. En base a lo anterior podemos estructurar los datos de tres maneras distintas:

1. TIPO I. Cuando tengamos un número pequeño de observaciones casi todas distintas, éstas se darán por Extensión.

   Ejemplo. 2, 3, 5, 7, 7, 8, 11, 14, 16, 19.

2. TIPO II. Cuando tengamos un gran número de observaciones pero muy pocas distintas, las organizaremos en una Tabla de Frecuencias, es decir, cada uno de los valores acompañado de la frecuencia con la que se presenta.

   Ejemplo.

   | Valor | Frecuencia |
   |---|---|
   | 2 | 3 |
   | 4 | 4 |
   | 5 | 1 |

   Que indica que el valor 2 se repite 3 veces, el valor 4 se repite 4 veces y el valor 5 sólo aparece una vez.

3. TIPO III. Cuando tengamos muchas observaciones, la mayoría de ellas distintas, las dispondremos agrupándolas en intervalos e indicando el número de observaciones que caen dentro de cada intervalo.

   Ejemplo.

   | Intervalo | Frecuencia |
   |---|---|
   | (2;3] | 4 |
   | (3;7] | 6 |
   | (7;12] | 12 |
   | (12;21] | 8 |

Que nos dice que en el intervalo (2;3] hay 4 observaciones, que en el (3;7] hay 6, etc.

En cualquiera de los tres casos anteriores tenemos una Distribución de Frecuencias. A la variable que genera dicha Distribución la llamaremos genéricamente X, a cada uno de los valores que toma la variable lo denotaremos como $x_i$ , y a la frecuencia con que toma dicho valor $n_i$. Para evitar confusiones es aconsejable ordenar los valores de la variable de menor a mayor. Para efectuar cálculos, y sea cual sea el Tipo de Distribución que tengamos dispondremos los datos de la siguiente forma:

| $x_i$ | $n_i$ | $N_i$ | $f_i$ | $F_i$ |
|---|---|---|---|---|
| $x_1$ | $n_1$ | $N_1$ | $f_1$ | $F_1$ |
| $x_2$ | $n_2$ | $N_2$ | $f_2$ | $F_2$ |
| . | . | . | . | |
| . | . | . | . | |
| . | . | . | . | . |
| $x_k$ | $n_k$ | $N_k = N$ | $f_k$ | $F_k = 1$ |

Donde N representa el número total de observaciones y será igual a

$$N = \sum_{i=1}^{k} n_i.$$

Llamaremos $f_i$ es la frecuencia relativa, definida como $f_i = n_i/N$, asimismo, denotaremos $N_i$ la frecuencia absoluta acumulada, que se obtiene como

$$N_i = \sum_{j=1}^{i} n_j$$

y $F_i$, la frecuencia relativa acumulada, que viene dada por

$$F_i = \sum_{j=1}^{i} f_j \ .$$

Obsérvese que si la Distribución es de Tipo I cada una de las frecuencias absolutas es igual a 1; y si la Distribución es de Tipo III los valores $x_i$ representan a las marcas de clase.

## 1.4 Representaciones gráficas.

En función de la naturaleza de los datos y de la forma en que éstos se presenten existen distintos tipos de representaciones.

Si tenemos un Atributo podemos utilizar bien un Diagrama de Tarta o Diagrama de Sectores, como puede verse en la figura 1.2.

También puede representarse mediante un Pictograma, como se ve en la figura 1.3.

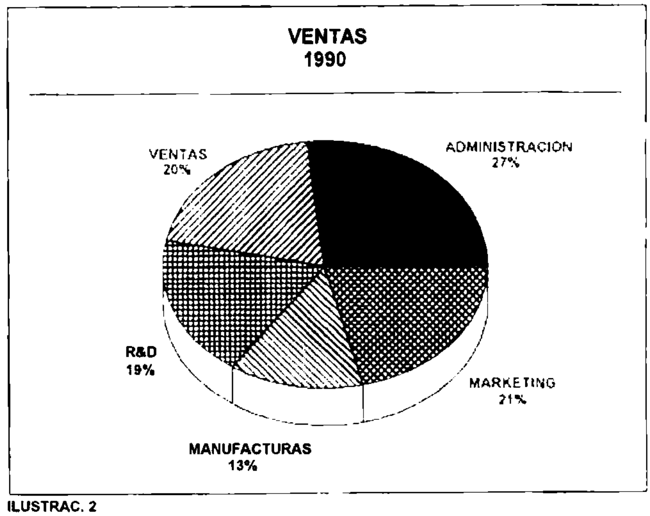



Figura 1.2: Diagrama de sectores.

Si tenemos una Distribución dada por extensión, la representaremos mediante un Diagrama de Puntos, en una dimensión (ver figura 1.4).

Y en dos dimensiones el Diagrama de Puntos queda como vemos en la figura 1.5.

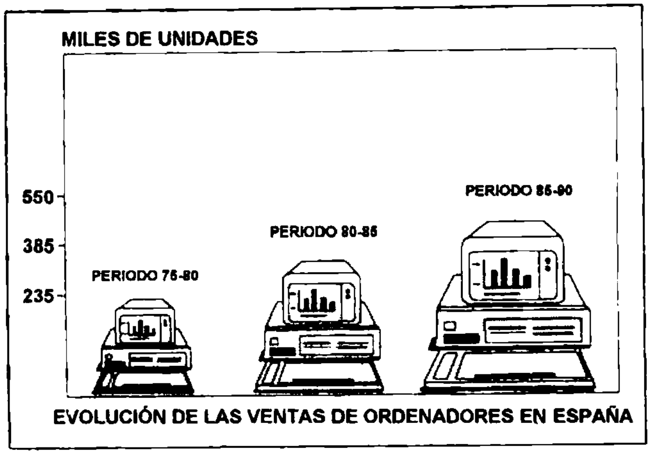
Si trabajamos con una Distribución del Tipo II, dibujaremos un Diagrama de Barras (ver figura 1.6).

Por último, si tenemos una Distribución del Tipo III, utilizaremos un Histograma (ver figura 1.7).

Aparte de las anteriores representaciones gráficas o gráficos estadísticos, existen otros tipos de representaciones que pueden ilustrar infinidad de situaciones, destacaremos entre otros el Diagrama de Área y el de Columnas. Éstos pueden verse, respectivamente, en la figura 1.8 y figura 1.9.

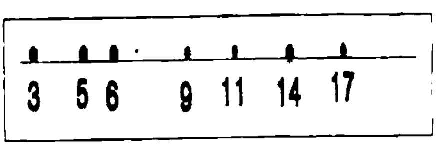
## 1.5 Medidas de centralización.

Una vez organizados los datos en su correspondiente Distribución de Frecuencias, procederemos a dar una serie de medidas que resuman toda esa información y que, "de alguna manera" representen a la Distribución.




ILUSTRAC. 3

Figura 1.3: Pictograma



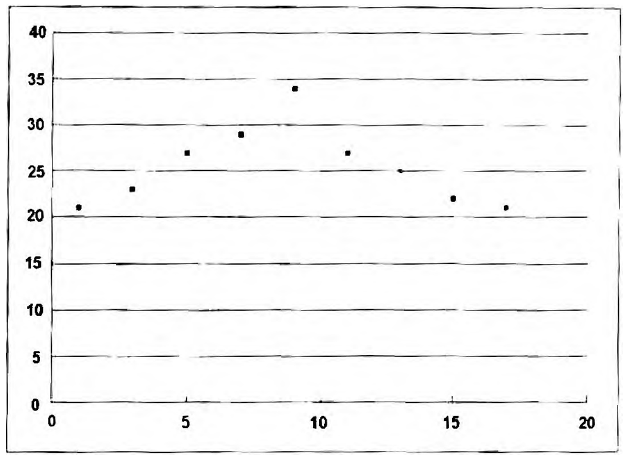

Ilustrac. 4

Figura 1.4: Diagrama de puntos en dimensión uno.

### 1.5.1 La media.

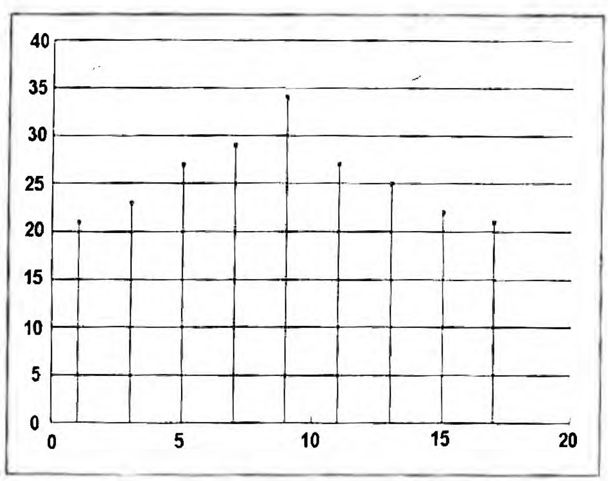
La media es una medida de representación central que necesariamente debe cumplir tres requisitos: el primero que para su obtención deben utilizarse todas las observaciones, el segundo que debe ser un valor comprendido entre el menor y el mayor de los valores de la Distribución y el tercero que debe venir expresada en la misma unidad que los datos. Entre todas las funciones que verifican estas tres propiedades destacamos a la Media Aritmética, a partir de ahora Media simplemente, que definimos de la siguiente manera:

$$\bar{X} = \frac{\sum_i x_i n_i}{N} = \sum_i f_i \, x_i \, .$$



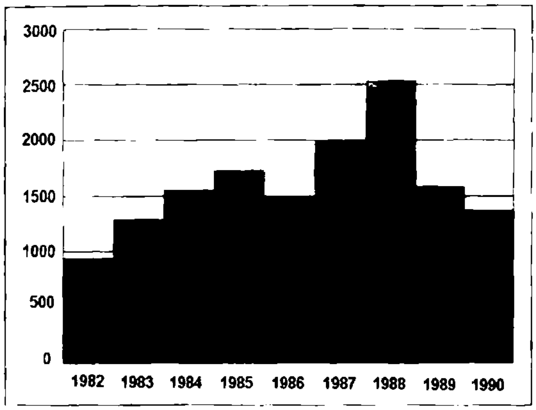

ILUSTRAC. 5

Figura 1.5: Diagrama de puntos en dos dimensiones.



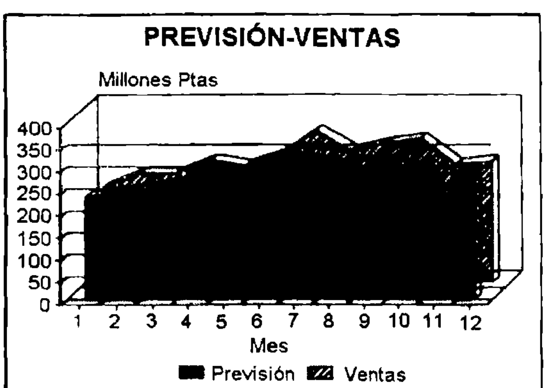

ILUSTRAC 6

Figura 1.6: Diagrama de barras.



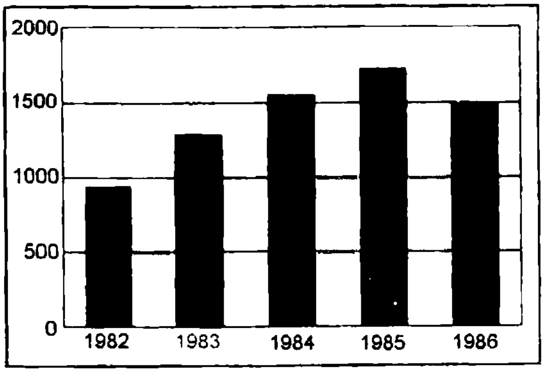

ILUSTRAC. 7

Figura 1.7: Histograma.




ILUSTRAC .8

Figura 1.8: Diagrama de Área.




ILUSTRAC. 9

Figura 1.9: Diagrama de barras.

Es decir, la suma extendida a todos los valores de la variable dividida por el número total de observaciones. A veces es más cómoda la expresión siguiente de la media, prescidiendo de las frecuencias y extendiendo la suma a todos los valores.

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}.$$

Con el mismo esquema también podemos definir la Media geométrica como la raíz $N$-ésima del producto de todos los valores de la Distribución, esto es:

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N}$$

Cuando tengamos que hacer un promedio de un grupo de razones o cocientes utilizaremos la Media armónica, definida como

$$H = \frac{N}{\sum_i \frac{1}{x_i} \cdot n_i}.$$

**Propiedades de la Media.**

Analizaremos a continuación una serie de propiedades de la Media que hacen de ésta una medida óptima de representación:

1. La suma de las desviaciones de los valores de la Distribución respecto a la Media es igual a cero, es decir:

$$\sum_{i=1}^{N}(x_i - \bar{X}) = 0.$$

2. Si a cada observación de una Distribución le sumamos una constante $k$ (traslación), tendremos una nueva variable $Y = X + k$ con Media igual a la de $X$ mas dicha constante.

3. Si multiplicamos una variable por una constante $k$(homotecia), la variable resultante $Y$ tendrá Media igual a $k$ por la Media de $X$. Estas dos propiedades se pueden resumir de la siguiente forma:

$$\text{Si} \quad Y = a \cdot X + b \qquad \longrightarrow \qquad \overline{Y} = a \cdot \overline{X} + b.$$

4. La Media es el valor que hace mínima la expresión:

$$\sum_{i=1}^{N}(x_i - \Phi)^2.$$

Precisamente ese mínimo será la Varianza de $X$, Medida de Dispersión que estudiaremos a continuación. Por otra parte, esta propiedad de la Media garantiza su bondad como medida de representación, como veremos más adelante.

### 1.5.2 La mediana.

La Mediana es un valor que, previa ordenación, deja la mitad de las observaciones en la recta real a la izquierda y la otra mitad a la derecha. Es decir, el 50% de los datos son menores que la Mediana y el otro 50% mayores. Para su cálculo y suponiendo que los valores están ordenados procederemos de la siguiente manera:

1. Si los datos vienen dados por extensión, y hay un número impar de ellos la Mediana es el elemento que se encuentra en el centro. Si el número de datos fuera par habría dos elementos centrales y la Mediana se obtendría como Media de ambos.

2. Si tenemos una Distribución de Tipo II, construiremos la columna de Frecuencias Absolutas Acumuladas, calcularemos el valor de $N/2$, nos deslizaremos por la columna de $N_i$, hasta detectar la primera frecuencia mayor o igual que $N/2$. Si dicha frecuencia es estrictamente mayor que $N/2$ la Mediana coincide con la observación que la ostenta, si por el contrario $N/2$ coincidiera con algún $N_i$ la Mediana valdría $(x_i + x_{i+1})/2$.

3. Por último, si la Distribución viene agrupada en intervalos, construiríamos también la columna de $N_i$, para fijar el intervalo donde se halla la Mediana, éste queda determinado porque es el primero que verifica que $N/2$ es estrictamente mayor que la frecuencia acumulada del intervalo anterior. Una vez fijado el intervalo, la Mediana adopta la siguiente expresión:

$$M_e = L_{i-1} + \frac{N/2 - N_{i-1}}{n_i} \cdot a_i$$

donde $L_{i-1}$ es el extremo inferior del intervalo y $a_i$ su amplitud. Esta misma expresión puede darse usando las frecuencias relativas, entonces quedaría como sigue:

$$M_e = L_{i-1} + \frac{0{,}5 - F_{i-1}}{f_i} \cdot a_i \, .$$

**Propiedad de La Mediana.**

La Mediana es el valor que hace mínima la expresión:

$$\sum_{i=1}^{N} \mid x_i - \Phi \mid .$$

### 1.5.3 La moda.

La Moda es el valor de la Distribución que más veces se repite, por tanto se puede obtener fácilmente cuando los datos vienen dados en forma puntual. Si las observaciones vienen agrupadas en intervalos distinguiremos dos casos:

1. Intervalos de igual amplitud. Fijaremos el intervalo que tenga mayor frecuencia, correspondiendo la Moda al valor:

$$M_o = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} \cdot a_i$$

2. Intervalos de distinta amplitud. En este caso el "intervalo modal" será aquel que tenga mayor Altura de Histograma $h_i$ definida como $n_i/a_i$ es decir la densidad de observaciones del intervalo, siendo la expresión de la Moda ...

$$M_o = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i \; .$$

### 1.5.4 Comparación entre Media, Moda y Mediana.

Salvo en casos muy específicos, la Media es la mejor de las Medidas de Representación, pues la Moda es bastante inestable y un pequeño cambio en las observaciones puede afectarle mucho, mientras que la Mediana es insensible al tamaño de los datos, permaneciendo constante si, por ejemplo, alteramos arbitrariamente y en cierto sentido las observaciones extremas. Por otra parte, si disponemos de las Modas y Medianas de dos Distribuciones necesitamos conocer cada uno de los datos de éstas para calcular la Moda y Mediana de la Distribución conjunta. La Media por el contrario es sensible a las alteraciones de los datos, al tamaño de éstos y si nos dan las Medias de dos conjuntos de datos, basta con conocer los tamaños de ambos grupos para calcular la Media global.

## 1.6 Medidas de posición.

Llamaremos Medidas de Posición o Cuantiles a aquellas que dividen a la Distribución en trozos, de tal forma que en cada uno de esos trozos haya el mismo número de elementos. La Mediana es un caso particular de Cuantil. De entre todas ellas destacaremos a los Cuartiles, los Deciles y los Percentiles. Los Cuartiles dividirán a la Distribución en cuatro partes iguales, los Deciles en diez y los Percentiles en cien. Habrá por tanto tres Cuartiles $(Q_1, Q_2, Q_3)$, nueve Deciles $(D_1, D_2, ..., D_9)$ y noventa y nueve Percentiles $(P_1, P_2, ..., P_{99})$. El segundo Cuartil, el quinto Decil y el quincuagésimo Percentil son iguales y coinciden con la Mediana. La forma general de cálculo es la siguiente:

$$Q_{r/k} = L_{i-1} + \frac{\frac{r \cdot N}{k} - N_{i-1}}{n_i} \cdot a_i \qquad \text{ó} \qquad X_\alpha = L_{i-1} + \frac{\alpha - F_{i-1}}{f_i} \cdot a_i$$

donde el intervalo i-ésimo es aquel que verifica que $N_i \geq \frac{r \cdot N}{k}$.

## 1.7 Medidas de dispersión.

Estudiaremos a continuación una serie de medidas que nos indicarán el nivel de concentración de los datos que se están analizando y, como consecuencia de ello, nos informarán sobre la bondad de los parámetros de centralización calculados como representantes del conjunto de datos.

### 1.7.1 Varianza y Desviación Típica.

La Varianza, $\mathrm{Var}(X)$, y su raiz cuadrada positiva, la Desviación Típica $\sigma_x$, son las más importantes Medidas de Dispersión, a la vez que están íntimamente ligadas a la Media. La Varianza viene dada por:

$$\mathrm{Var}(X) = \sigma^2 = \sum_i (x_i - \overline{X})^2 \cdot f_i \,.$$

Notas.

1. Relacione la ecuación anterior de la varianza con la Propiedad 4 de la Media que veíamos antes.

2. La suma de las desviaciones respecto de la Media no puede utilizarse como dispersión pues, según la Propiedad 1 de la media aritmética, siempre es nula.

El dar dos expresiones para un mismo concepto se explica porque la Varianza es un término de más fácil manejo, mientras que la Desviación Típica viene dada en la misma unidad que la variable. Tanto una como la otra son siempre positivas y valen cero sólo en el caso de que todos los valores coincidan con la Media (representatividad absoluta).

**Propiedades de la Varianza.**

Además de lo comentado en el párrafo anterior destacaremos las siguientes propiedades:

1. Si le sumamos una constante a una variable, la Varianza de la nueva variable no cambia.
2. Si multiplicamos una variable por una constante, la Varianza de la nueva variable es igual a la de la antigua multiplicada por la constante al cuadrado.

   Estas dos propiedades pueden resumirse en la siguiente expresión :

$$\text{Si} \quad Y = a \cdot X + b \longrightarrow \text{Var}(Y) = a^2 \cdot \text{Var}(X) .$$

### 1.7.2 Otras medidas de Dispersión.

Además de las anteriores medidas de dispersión podemos destacar las siguientes:

**El Recorrido y el Rango.**

Se define el primero como la diferencia entre el mayor y el menor de los valores y el segundo como el intervalo cuyos extremos son el mínimo y el máximo de la Distribución. Tienen la ventaja de que son fáciles de calcular, aunque cuando hay valores aislados en las colas de la Distribución dan una medida distorsionada de la dispersión de ésta.

**La Desviación absoluta.**

Definida respecto a la Media como

$$D_{\bar{X}} = \sum_i \mid x_i - \overline{X} \mid f_i .$$

Aunque también puede definirse respecto a la Mediana, valor que como vimos antes minimiza la expresión. Estas expresiones que acabamos de ver expresan la dispersión de la Distribución en términos absolutos, necesitamos definir a partir de ellas, otras que hagan posible la comparación entre varias variables y que tengan en cuenta el tamaño de las observaciones. Obsérvese que la distribución formada por los elementos 0,1; 0,2; 0,3; 0,4 y 0,5 y la que constituyen 1000,1; 1000,2; 1000,3; 1000,4 y 1000,5 tienen la misma Varianza y, sin embargo, es evidente que en el primero de los casos los elementos están muy dispersos y en el segundo bastante concentrados, esto es consecuencia de la primera de las propiedades de la Varianza. Para evitar estas situaciones introducimos las siguientes Medidas:

**Coeficiente de Variación.**

Se define como el cociente entre la Desviación Típica y el valor absoluto de la Media. Observemos que es una medida abstracta, sin dimensión, que tiene en cuenta el rango de valores en el que nos movemos, que nos permite comparar la dispersión de varias Distribuciones, además es invariante respecto a homotecias y sensible frente a traslaciones.

**Recorrido Intercuartílico. Desviación cuartílica.**

Que vienen dados por:

$$R_I = Q_3 - Q_1 \qquad ; \qquad D_C = \frac{Q_3 - Q_1}{2}$$

Estos parámetros pueden ser de utilidad cuando queramos que determinadas observaciones extremas no intervengan dando una visión sesgada de la variabilidad de la Distribución. Como inconveniente principal tiene que en su confección sólo intervienen directamente el 50% de los valores centrales.

## 1.8 Desigualdad de Tchebychev.

Esta desigualdad relaciona a la Media y a la desviación típica y tiene la expresión:

$$\text{frecuencia relativa}\left(\mid x_i - \overline{X} \mid \leq k \cdot \sigma_x\right) \geq 1 - \frac{1}{k^2}$$

que justifica el caracter de medida de dispersión de la desviación típica. Así, en un intervalo de centro la media y radio 3 veces la desviación típica se encuentra, al menos, el 89% de la distribución.

## 1.9 Momentos de la distribución.

Las medidas que hemos visto hasta ahora nos van dando visiones parciales de la Distribución, pretendemos dar ahora una herramienta eficaz que generalice esa idea, de tal forma que la mayoría de las características se puedan expresar utilizando esta herramienta: nos estamos refiriendo a los momentos de la distribución.

### 1.9.1 Momentos respecto al Origen.

Definimos el Momento de orden $k$ respecto al Origen como:

$$a_k = \frac{\sum_i x_i^k \cdot n_i}{N}$$

Es evidente que $a_0$ es igual a 1, que $a_1$ es la media aritmética, etc.

### 1.9.2 Momentos respecto a la Media.

El Momento de orden $k$ respecto a la Media viene dado por

$$m_k = \frac{\sum_i (x_i - \overline{X})^k \cdot n_i}{N}$$

Podemos comprobar que $m_o$ es igual a 1, que $m_1$ es cero y que $m_2$ es la Varianza.

Es posible, por otra parte expresar los Momentos respecto a la Media en función de los Momentos respecto al Origen, y así obtenemos

$$m_2 = \sigma^2 = a_2 - a_1^2 \quad ;$$

ó también

$$m_3 = a_3 - 3a_2a_1 + 2a_1^3.$$

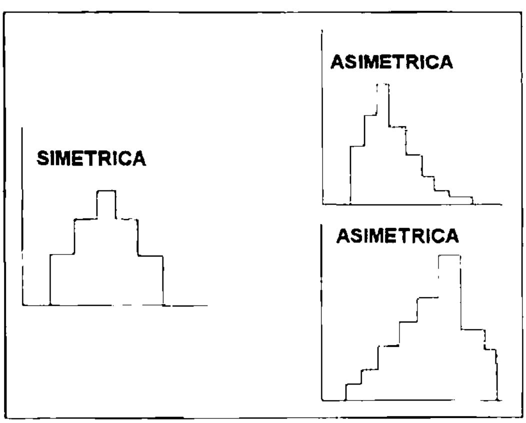
## 1.10 Medidas de simetría.

Trataremos de estudiar nuestra variable desde otro enfoque distinto, y así, en este epígrafe y en el siguiente nos detendremos a analizar la "forma" de la Distribución. En primer lugar examinaremos la Simetría y a continuación el Aplastamiento. Los coeficientes de Simetría que veremos nos indicarán si la Distribución es simétrica y caso de no serlo el tamaño y tendencia de su asimetría. Para ello consideraremos dos tipos de Distribuciones, las que tienen forma de campana y las que no la tienen, y utilizaremos distintas expresiones para su cálculo.

1. Si tiene forma de campana, como en los ejemplos que se acompañan en la figura 1.10 utilizaremos la expresión:

$$A_s = \frac{\overline{X} - M_o}{\sigma_x} .$$

   De tal forma que cuando $A_s$ es igual a cero la Distribución es simétrica, si es menor asimétrica negativa y si es mayor asimétrica positiva.




ILUSTRAC. 10

Figura 1.10: Asimetría en distribuciones acampanadas.

2. Si la distribución no tiene forma de campana, como las que siguen en la figura 1.11 calcularemos la Simetría mediante el coeficiente de simetría de Fisher:

$$g_1 = \frac{m_3}{\sigma_x^3} .$$

Siendo la discusión igual a la del caso anterior. Este coeficiente de simetría también puede utilizarse si la distribución tiene forma acampanada.

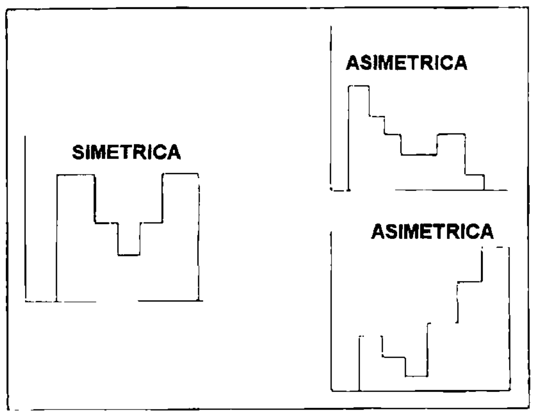



Figura 1.11: Asimetría en distribuciones no acampanadas.

Cuando la Distribución es Simétrica coinciden la Media y la Mediana y si además tiene forma de campana éstas son iguales a la Moda.

## 1.11 Medidas de Curtosis.

Examinaremos ahora el grado de aplastamiento de la Distribución a través del coeficiente de Curtosis, para lo cual la compararemos con la distribución Normal Tipificada o $N(0,1)$ que más adelante estudiaremos. Señalaremos, no obstante, que tiene forma de campana y que viene dada por la fórmula:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} .$$

El coeficiente de Curtosis toma la expresión:

$$g_2 = \frac{m_4}{\sigma_x^4} - 3 .$$

Cuando dicho coeficiente vale cero coincide con el de la $N(0,1)$ y decimos que la Distribución es Mesocúrtica (igual apuntamiento que la normal), si es negativo se dirá que es Platicúrtica (más aplastada que la normal) y si es positivo Leptocúrtica (más picuda que la normal). El gráfico de la figura 1.12 ilustra las tres situaciones anteriores.

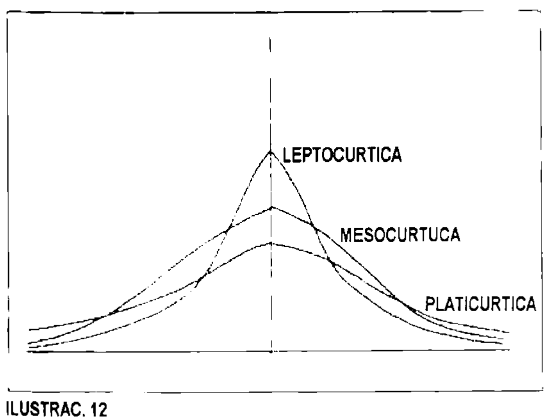



Figura 1.12: Aplastamiento / Apuntamiento de las distribuciones.

## 1.12 Transformaciones.

A veces nos encontramos con el inconveniente de que la Distribución que debemos estudiar presenta muchas irregularidades, como asimetrías acentuadas, valores extremos, etc. En otras ocasiones debemos comparar la posición de dos elementos que pertenecen a poblaciones con características distintas o del mismo elemento en situaciones distintas. En estos casos es recomendable efectuar una transformación que haga mas regular nuestra Distribución y, por tanto, con mejores condiciones para su estudio. Particular importancia tiene la Tipificación de una variable que estudiamos a continuación.

### 1.12.1 Normalización o tipificacion.

Dada una variable $X$ con Media $\bar{X}$ y Desviación Típica $\sigma_x$, la tipificación consiste en realizar la siguiente transformación:

$$Z = \frac{X - \bar{X}}{\sigma_x} .$$

A la nueva variable Z se le llama variable normalizada o tipificada y tiene Media 0 y Desviación Típica 1.

Otras transformaciones usuales son la del Logaritmo y la de la Raíz Cuadrada que consiguen una mayor simetría y concentración de los valores de la Distribución.

## 1.13 Medidas de la desigualdad.

Uno de los contextos en los que el concepto de desigualdad juega un papel relevante es en el de la Economía, y más concretamente en el estudio comparativo de distribuciones de renta o ingresos, aspecto asociado intimamente al de bienestar económico.

Dalton (1920) establece una serie de principios que, a su entender, debía cumplir toda medida de desigualdad de ingresos. De los cuatro principios, sólo el primero de ellos ha recibido una amplia y casi universal aceptación.

**Primer principio o principio de transferencia.**

"Cualquiera que sea el número de individuos y la cantidad a la que ascienda su renta, una transferencia de ingresos entre cualquier par de individuos, del más rico al más pobre, tal que no invierta el orden de su riqueza relativa, hace disminuir la desigualdad".

**Segundo principio o principio de aumentos proporcionales de ingresos.**

"Si todas las rentas aumentan de forma proporcional, disminuye la desigualdad".

**Tercer principio o principio de aumentos iguales en los ingresos.**

"Aumentos iguales en todos los ingresos hacen disminuir la desigualdad".

**Cuarto principio o principio de aumentos proporcionales de individuos.**

"La desigualdad permanece invariante si el número de individuos que poseen una determinada renta se multiplica por una cantidad igual para todas las rentas".

Es necesario aclarar, que la desviación típica no es una medida de desigualdad sino de dispersión. Si es una medida de desigualdad el coeficiente de variación de Pearson, aunque se conoce como medida de dispersión relativa.

### 1.13.1 La curva de Lorenz.

Lorenz, en 1905 propuso la siguiente técnica gráfica para describir la concentración o desigualdad de una distribución de ingresos: a cada número $t$ entre cero y cien (porcentaje) le asocia el porcentaje total de ingresos que acumula el $t\%$ de individuos más pobres de la población. El lugar geométrico de esos puntos, completado por interpolación lineal en el caso de una población finita, es una curva en forma de arco. Una distribución uniforme de la riqueza daría lugar a la diagonal $OP$ del cuadrado de lado 100, de la figura 1.13, mientras que cuanto mayor fuese la curvatura del arco, mayor es la concentración o desigualdad de la distribución de ingresos.

Una medida de la desigualdad que viene sugerida de forma natural a partir de la curva de Lorenz sería el área determinada por el segmento $OP$ y la curva.

Las curvas de Lorenz permiten distinguir entre distribuciones diferentes con un mismo valor de la medida de desigualdad. Así en la figura 1.14, están representadas dos curvas de Lorenz para las que el área entre ellas y el segmento $OP$, como medida de desigualdad, es la misma para las dos.

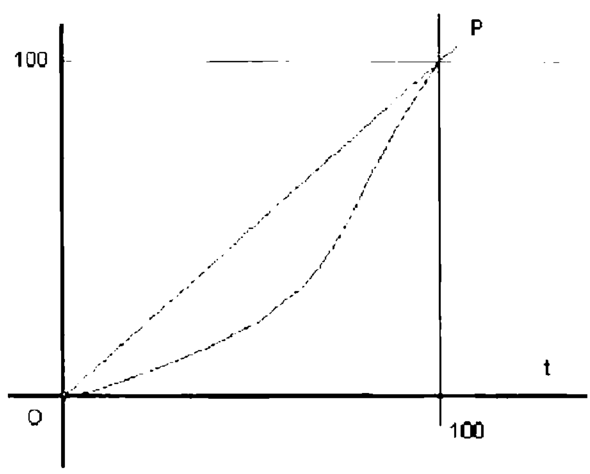



Figura 1.13: Curva de Lorenz.

**Propiedades de las curvas de Lorenz.**

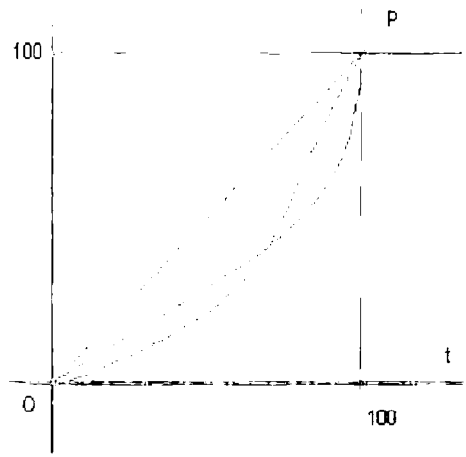



Figura 1.14: Curvas de Lorenz con igual área.

Si designamos por $L(u)$ la curva de Lorenz, verifica las siguientes propiedades:

(a) La curva de Lorenz es una función continua en $[0, 1]$, con $L(0) = 0$ y $L(1) = 1$.

(b) $L(u)$ es una función convexa.

(c) La curva de Lorenz determina la distribución mediante una transformación de escala.

### 1.13.2 Medidas de desigualdad relacionadas con las curvas de Lorenz.

#### Índice de Gini.

El índice de Gini es una de las medidas de concentración o desigualdad más utilizado y se define como el doble del área de concentración, es decir como el doble del área delimitada por la curva de Lorenz y la bisectriz del primer cuadrante. Por tanto se cumple que $0 \leq I_G \leq 1$.

**Índice de Pietra.**

Se define como la máxima distancia vertical entre la curva de Lorenz y la linea de igualdad (segmento $OP$). De aquí se sigue que será un valor comprendido entre cero y uno.

**Índice de Kakwani.**

Este autor propone en 1980 la utilización de la longitud de la curva de Lorenz como medida de la desigualdad de la distribución. Y por tanto será un valor comprendido entre $\sqrt{2}$ y 2.

### 1.13.3 Comparación en desigualdad

Dadas dos variables aleatorias no negativas $X$ e $Y$, ambas con media finita, notaremos por $L_x$ y $L_y$ a las respectivas funciones de Lorenz.

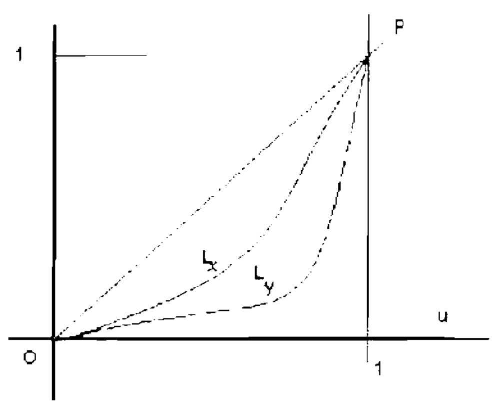



Figura 1.15: Comparación de curvas de Lorenz.

Decimos que $X$ no presenta más desigualdad que $Y$, en el sentido de Lorenz (ver figura 1.15), y lo notaremos

$$X \overset{L}{\leq} Y, \quad \text{si} \quad L_x(u) \geq L_y(u) \quad \text{para todo} \quad u \in [0,1].$$

Esta definición significa que una distribución $X$ es más desigual que otra distribución $Y$ si la curva de Lorenz asociada a $Y$ domina a la curva de Lorenz asociada a $X$.

# Problemas Resueltos

## Problema 1.1

**Se realiza una encuesta sobre el número de hermanos que tienen los 40 alumnos de una clase de primaria. Obteniéndose que 13 no tenían ningún hermano, 9 tenían un hermano, 8 tenían 2 hermanos, 6 tenían 3 hermanos, 3 tenían 4 y 1 tenía 5 hermanos. Se pide:**

1. **Reflejar la tabla de frecuencias.**
2. **Representar los datos en un diagrama de barras.**
3. **Si las familias con 3 o más hijos son consideradas familias numerosas, hallar el porcentaje de niños que pertenecen a familias numerosas.**

1. La tabla de frecuencias coorespondiente a estos datos es la siguiente:

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| 0 | 13 | 0,325 | 13 | 0,325 |
| 1 | 9 | 0,225 | 22 | 0,55 |
| 2 | 8 | 0,2 | 30 | 0,75 |
| 3 | 6 | 0,15 | 36 | 0,9 |
| 4 | 3 | 0,075 | 39 | 0,975 |
| 5 | 1 | 0,025 | 40 | 1 |

2. El diagrama de barras puede verse en la figura 1.16.

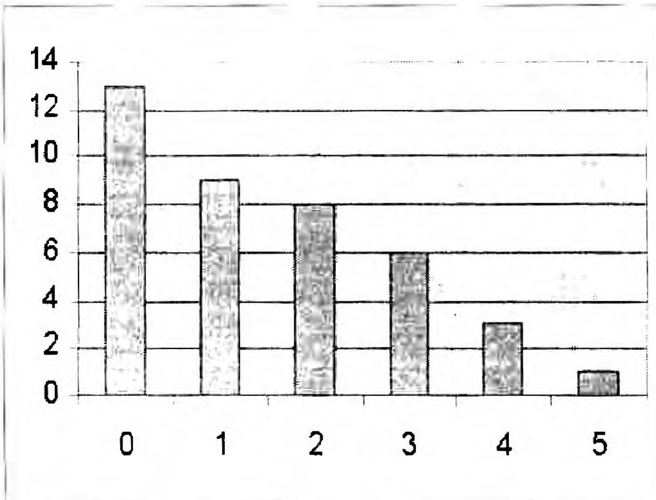



Figura 1.16: Diagrama de barras

3. Habrá que calcular el porcentaje de alumnos que pertenecen a familias con 3 o más hijos, es decir, aquellos alumnos que tienen 2 o más hermanos. En total hay 18 alumnos en estas circunstancias, que representan un porcentaje del 45%.

# Problema 1.2

**Hallar la media y la varianza del número de alumnos del problema anterior. Hallar el coeficiente de simetría de Fisher y el coeficiente de curtosis.**

| $x_i$ | $n_i$ | $n_i x_i$ | $x_i - \overline{x}$ | $n_i(x_i - \overline{x})^2$ | $n_i(x_i - \overline{x})^3$ | $n_i(x_i - \overline{x})^4$ |
|---|---|---|---|---|---|---|
| 0 | 13 | 0 | −1,5 | 29,25 | −43,875 | 65,813 |
| 1 | 9 | 9 | −0,5 | 2,25 | −1,125 | 0,563 |
| 2 | 8 | 16 | 0,5 | 2,00 | 1,000 | 0,500 |
| 3 | 6 | 18 | 1,5 | 13,50 | 20,250 | 30,375 |
| 4 | 3 | 12 | 2,5 | 18,75 | 46,875 | 117,188 |
| 5 | 1 | 5 | 3,5 | 12,25 | 42,875 | 150,063 |
| | 40 | 60 | 6,0 | 78,00 | 66,000 | 364,500 |

Con lo que tendremos que el valor de la media es:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{60}{40} = 1{,}5$$

La varianza valdrá:

$$\text{var}(x) = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \frac{78}{40} = 1{,}95$$

El momento de orden 3 será:

$$m_3 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^3 = \frac{66}{40} = 1{,}65$$

Y el momento de orden 4:

$$m_4 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^4 = \frac{364{,}5}{40} = 9{,}1125$$

Calculando el coeficiente de simetría de Fisher tendremos:

$$g_1 = \frac{m_3}{\sigma^3} = \frac{1{,}65}{1{,}3964240004^3} = 0{,}60594335$$

con lo que deducimos que la población está sesgada a la derecha.

El coeficiente de curtosis será:

$$g_2 = \frac{m_4}{\sigma^4} - 3 = \frac{9{,}1125}{1{,}3964240004^4} - 3 = -0{,}603550296$$

con lo que tenemos que la población es platicúrtica, más aplastada por tanto que la distribución normal.

# Problema 1.3

**Se realiza un estudio sobre la potencia de los vehículos producidos en una fábrica, obteniéndose los siguientes resultados:**

| **potencia (en C.V.)** | **número de vehículos** |
|---|---|
| **80 — 100** | **20** |
| **100 — 140** | **80** |
| **140 — 160** | **60** |
| **160 — 180** | **22** |
| **180 — 200** | **18** |

**Se pide:**

1. **Obtener la tabla de frecuencias correpondiente.**
2. **Obtener la representación gráfica más apropiada para estos datos.**
3. **Calcular la media, mediana y moda.**

---

1. La tabla de frecuencias correpondiente, donde llamamos $x_i$ a la marca de clase del intervalo y $a_i$ a la amplitud del mismo, sería:

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $F_i$ | $a_i$ | $h_i$ | $x_i n_i$ |
|---|---|---|---|---|---|---|---|
| 90 | 20 | 0,1 | 20 | 0,1 | 20 | 1 | 1800 |
| 120 | 80 | 0,4 | 100 | 0,5 | 40 | 2 | 9600 |
| 150 | 60 | 0,3 | 160 | 0,8 | 20 | 3 | 9000 |
| 170 | 22 | 0,11 | 182 | 0,91 | 20 | 1,1 | 3740 |
| 190 | 18 | 0,09 | 200 | 1 | 20 | 0,9 | 3420 |

2. Al venir los datos en una tabla de tipo III, la representación gráfica más apropiada sería mediante un histograma como el que puede verse en la figura 1.17.

3. Calculemos la media:

$$\overline{x} = \frac{1}{n}\sum x_i n_i = \frac{27560}{200} = 137{,}8$$

Para hallar la mediana tenemos que determinar en primer lugar en qué intervalo se encuentra. Será aquel en el que por primera vez la frecuencia relativa acumulada supere o iguale al valor 0,5. Esto ocurre en el intervalo [100; 140]. La frecuencia relativa acumulada hasta el valor 140 coincide con 0,5. Al ocurrir esto, el valor de la mediana será justamente 140. Por tanto $Me = 140$.

Para hallar la moda tenemos que determinar en primer lugar en qué intervalo se encuentra. Será aquel intervalo con mayor altura. En nuestro caso será el [140; 160] Aplicando la fórmula:

$$M_o = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

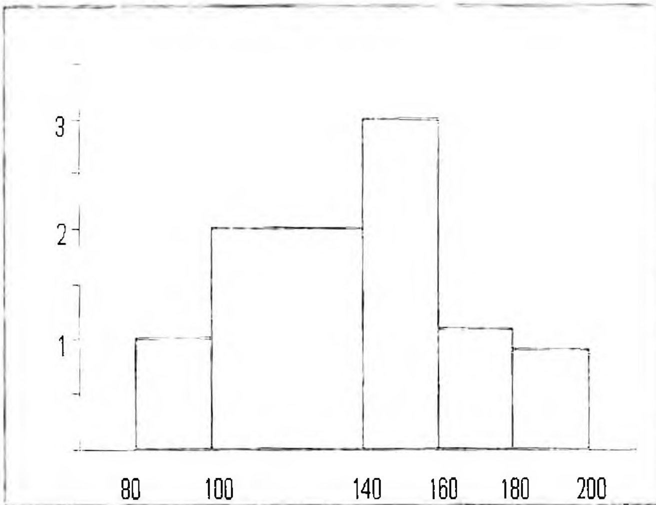



Figura 1.17: Histograma

sustituyendo:

$$M_o = 140 + \frac{1{,}1}{2 + 1{,}1} \cdot 20 = 140 + 7{,}097 = 147{,}097.$$

# Problema 1.4

**En el problema anterior, determinar gráficamente sobre el histograma el valor del primer cuartil. Hallar su valor.**

El primer cuartil será aquel valor que deje a la izquierda en el histograma un área que sea la cuarta parte del área total contenida bajo el histograma. La representación gráfica de este hecho puede verse en la figura 1.18. Para hallar el valor del primer cuartil tenemos que determinar en primer lugar en qué intervalo se encuentra. Será aquel en el que por primera vez la frecuencia relativa supere o iguale al valor 0,25. Esto ocurre en el intervalo [100; 140]. Aplicando la fórmula:

$$Q_1 = X_\alpha = L_{i-1} + \frac{\alpha - F_{i-1}}{f_i} \cdot a_i$$

en nuestro caso:

$$Q_1 = X_{0,25} = 100 + \frac{0{,}25 - 0{,}1}{0{,}4} \, 20 = 107{,}5.$$

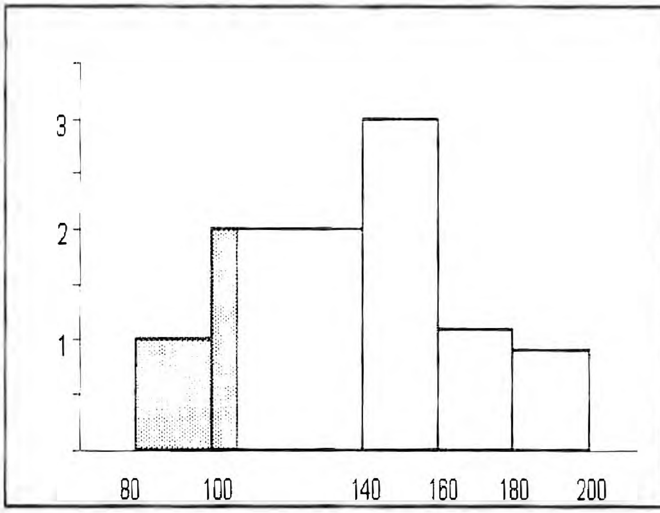



Figura 1.18: Área hasta el primer cuartil.

# Problema 1.5

**Se realiza un estudio sobre el sueldo anual (en millones de pesetas) de los empleados de una fábrica, obteniéndose los siguientes resultados:**

| **sueldo** | **empleados** |
|---|---|
| **0-2** | **17** |
| **2-4** | **36** |
| **4-6** | **38** |
| **6-8** | **9** |

**Se pide:**

1. **Obtener la tabla de frecuencias correpondiente a estos datos.**
2. **Calcular la media, la desviación típica, coefientes de simetría y curtosis.**

---

1. Al ser una tabla de tipo III, calculemos para cada intervalo su marca de clase, que denominaremos $x_i$, con lo que podremos obtener la tabla

de frecuencias correpondiente a estos datos:

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| 1 | 17 | 0,17 | 17 | 0,17 |
| 3 | 36 | 0,36 | 53 | 0,53 |
| 5 | 38 | 0,38 | 91 | 0,91 |
| 7 | 9 | 0,09 | 100 | 1 |

2.

| $x_i$ | $n_i$ | $n_i x_i$ | $x_i - \overline{x}$ | $n_i(x_i - \overline{x})^2$ | $n_i(x_i - \overline{x})^3$ | $n_i(x_i - \overline{x})^4$ |
|---|---|---|---|---|---|---|
| 1 | 17 | 17 | −2,78 | 131,383 | −365,244 | 1015,379 |
| 3 | 36 | 108 | −0,78 | 21,902 | −17,084 | 13,325 |
| 5 | 38 | 190 | 1,22 | 56,559 | 69,002 | 84,183 |
| 7 | 9 | 63 | 3,22 | 93,316 | 300,476 | 967,533 |
| | 100 | 378 | 0,88 | 303,160 | −12,850 | 2080,420 |

y, por lo tanto:

| | |
|---|---|
| media | 3,78 |
| varianza | 3,0316 |
| $\sigma$ | 1,741149046 |
| $m_3$ | −0,128496 |
| $g_1$ | −0,024343429 |
| $m_4$ | 20,80420432 |
| $g_2$ | −0,73636024 |

# Problema 1.6

**En el problema anterior, sin considerar el 5% de trabajadores que ganan menos y el 5% de trabajadores que ganan más, hallar un intervalo en el que estén contenidos los sueldos del 90% de los trabajadores restantes.**

Lo que tendremos que hallar será el percentil 5 y el percentil 95. El intervalo pedido será el comprendido entre estos dos valores. Para determinar los percentiles en primer lugar deberemos saber en qué intervalo se encuentra cada uno. El percentil 5 estará en el primer intervalo, ya que en este intervalo están aquellos valores en los que la frecuencia relativa acumulada está entre 0 y 0,17. El percentil 95 estará en el último intervalo, ya que en este intervalo están aquellos valores en los que la frecuencia relativa acumulada está entre 0,91 y 1.

Aplicando la fórmula:

$$X_\alpha = L_{i-1} + \frac{\alpha - F_{i-1}}{f_i} a_i$$

en nuestro caso:

$$X_{0,05} = 0 + \frac{0{,}05 - 0}{0{,}17} \cdot 2 = 0{,}588$$

$$X_{0,95} = 6 + \frac{0{,}95 - 0{,}91}{0{,}09} \cdot 2 = 6{,}889$$

Por tanto el intervalo pedido sería el (0,588; 6,889).

## Problema 1.7

**Los trabajadores de una empresa tienen una sueldo medio anual de 3,78 millones de pesetas, siendo la desviación típica de 1,74. Hallar un intervalo en el que estén contenidos los sueldos del 90% de los trabajadores utilizando la desigualdad de Tchebychev. Comparar los resultados con los obtenidos en el problema anterior.**

Según la desigualdad de Tchebychev:

$$\text{frecuencia relativa } (| x_i - \overline{x} | \leq k \cdot \sigma_x) \geq 1 - \frac{1}{k^2}$$

En nuestro caso, como lo que pretendemos es que el intervalo contenga al menos el 90% de la población, el valor de $k$ será aquel que cumpla:

$$1 - \frac{1}{k^2} = 0{,}9 \quad \Rightarrow \quad k = 3{,}1622$$

Por tanto el intervalo pedido será:

$$(\overline{x} - k \cdot \sigma_x, \overline{x} - k \cdot \sigma_x)$$

o lo que es igual :

$$(3{,}78 - 3{,}1622 \cdot 1{,}74 \ ; \ 3{,}78 - 3{,}1622 \cdot 1{,}74) = (-1{,}722 \ ; \ 9{,}2822)$$

Obsérvese que, aunque la media y la desviación típica utilizada coincidían con los de la población del ejercicio anterior, en este caso se ha utilizado mucha menos información que antes. Por este motivo en el ejemplo anterior se obtenía un intervalo mucho más preciso.

# Problema 1.8

**En un estudio sobre el tipo sanguíneo de una población se obtuvieron los siguientes datos: del tipo A eran 71 personas, del tipo B eran 56 personas, del tipo AB eran 31 personas, y del tipo O eran 42 personas.**

**Obtener la representación gráfica más apropiada para estos datos.**

Al ser los datos de tipo cualitativo, la representación gráfica más apropiada sería mediante un diagrama de sectores circulares como el que puede verse en la figura 1.19.

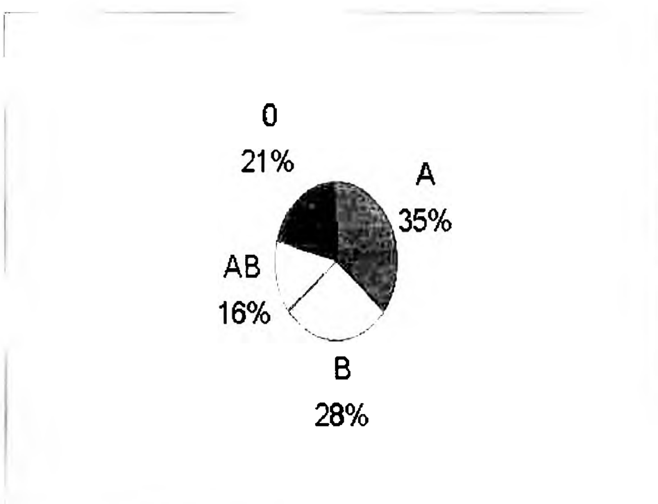



Figura 1.19: Diagrama de sectores

# Problema 1.9

**Hallar la relación entre los momentos de órdenes 3 y 4 respecto a la media $\mu$ y los momentos respecto al origen.**

$$\begin{aligned}
m_3 &= \sum_{i=1}^{k} f_i(x_i-\mu)^3 = \\
&= \sum_{i=1}^{k} f_i(x_i^3 - 3x_i^2\mu + 3x_i\mu^2 - \mu^3) = \\
&= \sum_{i=1}^{k} f_i x_i^3 - 3\mu\sum_{i=1}^{k} f_i x_i^2 + 3\mu^2\sum_{i=1}^{k} f_i x_i - \mu^3\sum_{i=1}^{k} f_i = \\
&= a_3 - 3a_1a_2 + 3a_1^3 - a_1^3 = \\
&= a_3 - 3a_1a_2 + 2a_1^3.
\end{aligned}$$

$$\begin{aligned}
m_4 &= \sum_{i=1}^{k} f_i(x_i-\mu)^4 = \\
&= \sum_{i=1}^{k} f_i(x_i^4 - 4x_i^3\mu + 6x_i^2\mu^2 - 4x_i\mu^3 + \mu^4) = \\
&= \sum_{i=1}^{k} f_i x_i^4 - 4\mu\sum_{i=1}^{k} f_i x_i^3 + 6\mu^2\sum_{i=1}^{k} f_i x_i^2 - 4\mu^3\sum_{i=1}^{k} f_i x_i + \mu^4\sum_{i=1}^{k} f_i = \\
&= a_4 - 4a_1a_3 + 6a_1^2a_2 - 4a_1^4 + a_1^4 = \\
&= a_4 - 4a_1a_3 + 6a_1^2a_2 - 3a_1^4.
\end{aligned}$$

# Problema 1.10

**Demostrar que, excepto el signo, el coeficiente de simetría de Fisher no se ve alterado por transformaciones lineales.**

Supongamos que tenemos una colección de observaciones que constituyen una variable $x$ con una media que denotaremos por $\mu_x$ y una desviación típica que denotaremos por $\sigma_x$.

Si consideramos una nueva variable $y$ obtenida a partir de $x$ mediante la transformación lineal

$$y_i = a + bx_i$$

tendremos entonces que los valores de la media y la desviación típica de la variable $y$, que denotaremos $\mu_y$ y $\sigma_y$ respectivamente, serán:

$$\mu_y = a + b\mu_x \qquad \sigma_y = |b|\sigma_x$$

Usaremos la siguiente notación:

$m_{3x}$: momento de orden 3 respecto a la media de la variable $x$

$g_{1x} = \dfrac{m_{3x}}{\sigma_x^3}$: coeficiente de simetría de Fisher de $x$.

$m_{3y}$: momento de orden 3 respecto a la media de la variable $y$

$g_{1y} = \dfrac{m_{3y}}{\sigma_y^3}$: coeficiente de simetría de Fisher de $y$.

Veamos qué relación existe entre $g_{1x}$ y $g_{1y}$. Para ello calculemos $m_{3y}$:

$$m_{3y} = \sum_{i=1}^{k} f_i(y_i - \mu_y)^3$$

sustituyendo los valores de $y_i$ y de $\mu_y$ se tendría

$$m_{3y} = \sum_{i=1}^{k} f_i(a + bx_i - a - b\mu_x)^3$$

con lo cual

$$m_{3y} = b^3 \sum_{i=1}^{k} f_i(x_i - \mu_x)^3 = b^3 m_{3x}$$

Por tanto

$$g_{1y} = \frac{m_{3y}}{\sigma_y^3} = \frac{b^3 m_{3x}}{|b|^3 \sigma_x^3} = \mathrm{sig}(b) g_{1x}.$$

# Problema 1.11

**Demostrar que el coeficiente de curtosis no se ve alterado por transformaciones lineales.**

Supongamos que tenemos una colección de observaciones que constituyen una variable $x$ con una media que denotaremos por $\mu_x$ y una desviación típica que denotaremos por $\sigma_x$.

Si consideramos una nueva variable $y$ obtenida a partir de $x$ mediante la transformación lineal

$$y_i = a + bx_i$$

tendremos entonces que los valores de la media y la desviación típica de la variable $y$, que denotaremos $\mu_y$ y $\sigma_y$ respectivamente, serán:

$$\mu_y = a + b\mu_x \qquad \sigma_y = |b|\sigma_x$$

Usemos la notación siguiente:

$m_{4x}$ momento de orden 4 respecto a la media de la variable $x$

$g_{2x} = \dfrac{m_{4x}}{\sigma_x^4} - 3$ su coeficiente de curtosis.

$m_{4y}$ momento de orden 4 respecto a la media de la variable $y$

$g_{2y} = \dfrac{m_{4y}}{\sigma_y^4} - 3$ su coeficiente de curtosis.

Veamos qué relación existe entre $g_{2x}$ y $g_{2y}$. Para ello calculemos $m_{4y}$:

$$m_{4y} = \sum_{i=1}^{k} f_i(y_i - \mu_y)^4$$

sustituyendo los valores de $y_i$ y de $\mu_y$ se tendría

$$m_{4y} = \sum_{i=1}^{k} f_i(a + bx_i - a - b\mu_x)^4$$

con lo cual

$$m_{4y} = b^4 \sum_{i=1}^{k} f_i(x_i - \mu_x)^4 = b^4 m_{4x}$$

Por tanto

$$g_{2y} = \frac{m_{4y}}{\sigma_y^4} - 3 = \frac{b^4 m_{4x}}{|b|^4\sigma_x^4} - 3 = g_{2x}$$

## Problema 1.12

**Un grupo de 20 estudiantes realiza pruebas para medir la intensidad de la corriente en un punto determinado de un circuito. Cuando contrastan sus mediciones observan que 4 han obtenido**

**que la intensidad es de 2543 mA, 9 han obtenido 2544 mA y 7 han obtenido 2545 mA.**

**Aplicar los resultados de los problemas anteriores para calcular la media, la desviación típica, el coeficiente de simetría y el de curtosis de estos datos.**

---

La tabla de frecuencias sería

| $x_i$ | $n_i$ | $f_i$ |
|---|---|---|
| 2543 | 4 | 0,20 |
| 2544 | 9 | 0,45 |
| 2545 | 7 | 0,35 |
| | 20 | 1 |

Si hacemos el cambio $y = x - 2544$ tendríamos:

| $y_i$ | $n_i$ | $f_i$ |
|---|---|---|
| $-1$ | 4 | 0,20 |
| 0 | 9 | 0,45 |
| 1 | 7 | 0,35 |
| | 20 | 1 |

Si hallamos todos los momentos centrales de $y$ tendríamos:

| $y_i$ | $n_i$ | $f_i$ | $f_iy_i$ | $f_iy_i^2$ | $f_iy_i^3$ | $f_iy_i^4$ |
|---|---|---|---|---|---|---|
| $-1$ | 4 | 0,20 | $-0{,}20$ | 0,20 | $-0{,}20$ | 0,20 |
| 0 | 9 | 0,45 | 0 | 0 | 0 | 0 |
| 1 | 7 | 0,35 | 0,35 | 0,35 | 0,35 | 0,35 |
| | 20 | 1 | 0,15 | 0,55 | 0,15 | 0,55 |

Por tanto tenemos

$$\begin{aligned}
\overline{y} &= 0{,}15 \\
\mathrm{var}(y) &= 0{,}55 - 0{,}15^2 = 0{,}55 - 0{,}0225 = 0{,}5275 \\
\sigma_y &= 0{,}7263 \\
a_{3y} &= 0{,}15 \\
a_{4y} &= 0{,}55
\end{aligned}$$

$$\begin{aligned}
m_{3y} &= a_{3y} - 3a_{1y}a_{2y} + 2a_{1y}^3 = 0{,}15 - 3(0{,}15)(0{,}55) + 2\cdot(0{,}15)^3 = \\
&= 0{,}15 - 0{,}2475 + 0{,}00675 = -0{,}09075 \\
m_{4y} &= a_{4y} - 4a_{1y}a_{3y} + 6a_{1y}^2a_{2y} - 3a_{1y}^4 = \\
&= 0{,}55 - 4(0{,}15)(0{,}15) + 6(0{,}15)^2(0{,}55) - 3(0{,}15)^4 = \\
&= 0{,}55 - 0{,}09 + 0{,}07425 - 0{,}00151875 = 0{,}5327
\end{aligned}$$

$$\begin{aligned}
g_{1y} &= -0{,}09075/(0{,}7263)^3 = -0{,}2369 \\
g_{2y} &= [0{,}5327/(0{,}7263)^4] - 3 = -1{,}0856
\end{aligned}$$

Con lo cual los parámetros estadísticos de $x$ serán:

$$\begin{aligned}
\overline{x} &= 0{,}15 + 2544 &&= 2544{,}15 \\
\mathrm{var}(x) &= \mathrm{var}(y) &&= 0{,}5275 \\
\sigma_x &= \sigma_y &&= 0{,}7263 \\
g_{1x} &= g_{1y} &&= -0{,}2369 \\
g_{2x} &= g_{2y} &&= -1{,}0856.
\end{aligned}$$

Hemos podido comprobar que al hacer el cambio de variable $y = a + bx$, con $b = 1$, se ha conseguido reducir el tiempo dedicado a operaciones, por tanto, es aconsejable en muchos problemas realizar cambios de variable adecuados.

# Problema 1.13

**El encargado de mantenimiento de una empresa realiza compras de materiales de repuesto de tres tipos, cuyos precios son 100, 120 y 140. Calcular el precio medio de los materiales en los siguientes supuestos:**

1. **Compra un material de cada tipo.**
2. **Compra un material del primer tipo, 2 del segundo y 3 del tercero.**
3. **Compra el mismo número de cada tipo de material.**

4. **Se gasta el mismo dinero en cada tipo de material.**

---

1. Si compra un material de cada tipo, entonces la tabla de frecuencias sería:

| precio | cantidad |
|---|---|
| 100 | 1 |
| 120 | 1 |
| 140 | 1 |

con lo que la media sería $\dfrac{100+120+140}{3} = 120$ ; es decir, la media aritmética de los tres precios.

2. Si compra un material del primer tipo, dos del segundo y tres del tercero, entonces la tabla de frecuencias sería:

| precio | cantidad |
|---|---|
| 100 | 1 |
| 120 | 2 |
| 140 | 3 |

con lo que la media sería $\dfrac{100 + 2 \cdot 120 + 3 \cdot 140}{6} = 126{,}667$ ; es decir, la media aritmética de los tres precios ponderada por el número de materiales de cada tipo que se han comprado.

3. Si compra el mismo número $a$ de cada tipo de material, entonces la tabla de frecuencias quedaría:

| precio | cantidad |
|---|---|
| 100 | $a$ |
| 120 | $a$ |
| 140 | $a$ |

con lo que la media sería $\dfrac{100a + 120a + 140a}{3a} = 120$ ; es decir, la media aritmética de los tres precios.

4. Si se gasta el mismo dinero $M$ en cada tipo de material, del primer tipo comprará $M/100$ unidades, del segundo tipo comprará $M/120$

unidades y del tercer tipo comprará $M/140$ unidades. Entonces la tabla de frecuencias quedaría:

| precio | cantidad |
|---|---|
| 100 | $M/100$ |
| 120 | $M/120$ |
| 140 | $M/140$ |

con lo que la media sería

$$\frac{100\dfrac{M}{100}+120\dfrac{M}{120}+140\dfrac{M}{140}}{\dfrac{M}{100}+\dfrac{M}{120}+\dfrac{M}{140}}=$$

$$=\frac{3M}{\dfrac{M}{100}+\dfrac{M}{120}+\dfrac{M}{140}}=\frac{3}{\dfrac{1}{100}+\dfrac{1}{120}+\dfrac{1}{140}}=117{,}76$$

es decir, la media armónica de los tres precios.

# Problema 1.14

**En una Escuela hay matriculados 550 alumnos en primero, 300 en segundo y 150 en tercero. El número medio de asignaturas de los alumnos de primero es de 8,2 con una varianza de 4. El número medio de asignaturas de los alumnos de segundo es de 12,4 con una varianza de 6,2. El número medio de asignaturas de los alumnos de tercero es de 6,7 con una varianza de 3,8. Hallar la media y la varianza del número de asignaturas de los alumnos de toda la Escuela.**

Calculemos en primer lugar la media.

El número total de asignaturas de los alumnos de $1^{\text{o}}$ es $8,2\cdot 550=4510$.

El número total de asignaturas de los alumnos de $2^{\text{o}}$ es $12,4\cdot 300=3720$.

El número total de asignaturas de los alumnos de $3^{\text{o}}$ es $6,7\cdot 150=1005$.

Dividiendo el número total de asignaturas entre el número total de alumnos obtenemos la media del número de asignaturas:

$$\mu=\frac{4510+3720+1005}{1000}=9{,}235$$

es decir, la media ponderada de las medias.

Para calcular la varianza total aplicaremos la fórmula

$$\sigma^2 = E[X^2] - E[X]^2 = \frac{\sum x_i^2}{n} - \mu^2$$

con lo cual la suma de los cuadrados valdrá

$$\sum x_i^2 = n(\sigma^2 + \mu^2)$$

La suma de cuadrados total será la suma de los cuadrados de los alumnos de primero, segundo y tercero. Por tanto, la suma de cuadrados total será

$$\sum x_i^2 = n_1(\sigma_1^2 + \mu_1^2) + n_2(\sigma_2^2 + \mu_2^2) + n_3(\sigma_3^2 + \mu_3^2)$$

siendo $n_1$, $n_2$, $n_3$ el número de alumnos de primero, segundo y tercero y siendo $\mu_1$, $\mu_2$, $\mu_3$, $\sigma_1^2$, $\sigma_2^2$ y $\sigma_3^2$ las medias y varianzas respectivas. La varianza total será

$$\sigma^2 = \frac{\sum x_i^2}{n} - \mu^2 = \frac{n_1(\sigma_1^2 + \mu_1^2) + n_2(\sigma_2^2 + \mu_2^2) + n_3(\sigma_3^2 + \mu_3^2)}{n} - \mu^2$$

operando:

$$\sigma^2 = \left(\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2}{n}\right) + \left(\frac{n_1\mu_1^2 + n_2\mu_2^2 + n_3\mu_3^2}{n} - \mu^2\right)$$

es decir la media de las varianzas más la varianza de las medias (este resultado es generalizable). En nuestro caso, sustituyendo tendríamos:

$$\begin{aligned}
\sigma^2 &= \left(\frac{550 \cdot 4 + 300 \cdot 6{,}2 + 150 \cdot 3{,}8}{1000}\right) + \\
&\quad + \left(\frac{550 \cdot 8{,}2^2 + 300 \cdot 12{,}4^2 + 150 \cdot 6{,}7^2}{1000} - 9{,}235^2\right) = \\
&= \left(\frac{2200 + 1860 + 570}{1000}\right) + \left(\frac{36982 + 46128 + 6733{,}5}{1000} - 85{,}285\right) = \\
&= 4{,}63 + (89{,}8435 - 85{,}285) = 9{,}1885.
\end{aligned}$$

## Problema 1.15

**En un examen la nota media de los alumnos del grupo A fue de 6, con una desviación típica de 2, mientras que la nota media de los alumnos del grupo B fue de 5, con una desviación típica de 1. Un alumno del grupo A sacó un 8 y otro del grupo B sacó un 7. ¿Cuál de los dos obtuvo una puntuación relativa más alta?**

Para poder comparar las posiciones relativas de individuos correspondientes a poblaciones distintas habrá que tipificar los datos.

La puntuación tipificada del alumno del grupo A es $\frac{8-6}{2} = 1$ mientras que la del alumno del grupo B fue de $\frac{7-5}{1} = 2$.

Por tanto el alumno que sacó el 7, en el grupo B, tiene una mejor puntuación relativa.

## Problema 1.16

**Se tienen tres números de los que se sabe que su mediana vale 20, su media aritmética vale 21 y su varianza vale 14. Hallar los tres números.**

Llamemos a los tres números $x_1, x_2, x_3$ y supondremos que están ordenados en orden creciente. Por ser la mediana 20 tenemos que $x_2 = 20$.

Por ser la media 21 tenemos que

$$\frac{1}{3}(x_1 + x_2 + x_3) = 21$$

y usando que $x_2 = 20$, tendremos que

$$x_1 + x_3 = 43.$$

Por ser la varianza 14 tenemos que

$$\frac{1}{3}\left((x_1 - 21)^2 + (x_2 - 21)^2 + (x_3 - 21)^2\right) = 14$$

lo que equivale a

$$(x_1 - 21)^2 + (x_3 - 21)^2 = 41$$

sustituyendo $x_3 = 43 - x_1$:

$$x_1^2 - 41x_1 + 441 + 484 - 44x_1 + x_1^2 = 41$$

esta ecuación tiene dos soluciones que son $x_1 = 17$ (con lo cual $x_3 = 26$) y $x_1 = 26$ (con lo cual $x_3 = 17$). Como habíamos considerado los números ordenados la solución será

$$x_1 = 17 \qquad x_2 = 20 \qquad x_3 = 26.$$

## Problema 1.17

**Dos ciclistas dan vueltas a un circuito. El primer ciclista está una hora corriendo a 20 km/h y luego otra hora corriendo a 10 km/h. El segundo ciclista está 10 vueltas corriendo a 20 km/h y luego otras 10 vueltas corriendo a 10 km/h. Hallar la velocidad media de cada uno de los dos ciclistas.**

El primer ciclista está la primera hora corriendo a 20 km/h, recorrerá por tanto, en esta hora 20 km. la segunda hora corre a 10 km/h, recorrerá por tanto, en esta segunda hora 10 km. En total ha recorrido 30 km en dos horas, siendo por tanto la velocidad media de 15 km/h, esto es, la media aritmética de 10 y 20.

El segundo ciclista está 10 vueltas corriendo a 20 km/h. Como no sabemos la longitud de estas 10 vueltas llamésmole x a esta distancia. El tiempo que tarda será x/20 horas. Las otras 10 vueltas las corre a 10 km/h, tardando un tiempo de x/10 horas. La distancia total que recorre es de 2x y el tiempo total que tarda es de x/20 + x/10. La velocidad media será 2x entre x/20 + x/10, o lo que es lo mismo, la media armónica entre 10 y 20 que es 13,333 km/h.

# Problemas Propuestos

1. Se ha revisado un lote de 1000 circuitos integrados, obteniendose el número de defectos que se indica en la tabla:

| $N^o$ de defectos | Frecuencia |
|---|---|
| 0 | 600 |
| 1 | 310 |
| 2 | 75 |
| 3 | 13 |
| 4 | 2 |

   Determinar la media y la desviación típica de la distribución. Calcular el resto de parámetros y hacer el gráfico estadístico.

2. El número de aciertos en un test correspondiente a 11 alumnos han sido: 21, 36, 19, 23, 32, 25, 28, 20, 34, 33, 31. Se pide:

   (a) Determinar la mediana.

   (b) ¿Qué porcentaje de alumnos tiene nota inferior a 32 ?

3. Representar gráficamente las siguientes distribuciones de frecuencias, relativas a tiempos de combustión de mechas para pólvora de 1 m. de longitud. Los tiempos están en segundos. La primera tabla se refiere a combustión al aire de 1000 mechas y la segunda a 1200 mechas en una atmósfera de oxígeno.

| EN EL AIRE | | EN OXIGENO | |
|---|---|---|---|
| TIEMPO | Nº DE OBSERV. | TIEMPO | Nº DE OBSERV |
| 85 | 5 | 83 | 8 |
| 86 | 23 | 84 | 19 |
| 87 | 60 | 85 | 26 |
| 88 | 137 | 86 | 60 |
| 89 | 208 | 87 | 101 |
| 90 | 225 | 88 | 156 |
| 91 | 169 | 89 | 118 |
| 92 | 103 | 90 | 215 |
| 93 | 49 | 91 | 110 |
| 94 | 15 | 92 | 129 |
| 95 | 6 | 93 | 62 |
| | | 94 | 10 |
| | | 95 | 15 |
| | | 96 | 1 |
| 1000 | | 1200 | |

4. En la fabricación de condensadores para motores, aparace un determinado número de ellos defectuosos. Se han estudiado 200 lotes diferentes de 500 piezas cada uno, obteniendose:

| Piezas defectuosas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Nº de lotes | 5 | 15 | 38 | 42 | 49 | 32 | 17 | 2 |

Representar los datos y calcular los parámetros de centralización y de dispersión. Calcular el percentil 20.

5. El número de accidentes de aviación en 1986 fue superior al de 1997. ¿Basta esto para afirmar que era más peligroso volar en 1986 que en 1996?

6. La tabla da los pesos en gramos de 50 pomelos, seleccionados aleatoriamente de una producción agraria a la que se ha aplicado un nuevo fertilizante:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 309 | 355 | 332 | 349 | 329 | 329 | 363 | 252 | 387 | 323 |
| 340 | 358 | 277 | 305 | 327 | 345 | 265 | 403 | 293 | 311 |
| 358 | 361 | 208 | 303 | 256 | 301 | 311 | 322 | 350 | 310 |
| 288 | 388 | 307 | 356 | 342 | 323 | 260 | 379 | 470 | 358 |
| 309 | 288 | 369 | 362 | 329 | 374 | 240 | 319 | 288 | 247 |

Calcular la media aritmética, moda, mediana y varianza.

Nota: Utilizar como intervalos de clase [199,5 - 219,5] ... [459,5 - 479,5]; etc.

7. Dada la siguiente distribución de frecuencias:

| $[L_{i-1}; L_i)$ | $n_i$ |
|---|---|
| 1-3 | 3 |
| 3-7 | 29 |
| 7-8 | 35 |
| 8-10 | 26 |
| 10-13 | 6 |
| 13-20 | 1 |

(a) Constrúyase una tabla en la que aparezcan las marcas de clase, las frecuencias absolutas y relativas y las frecuencias absolutas y relativas acumuladas crecientes y decrecientes.

(b) Represéntese la distribución mediante un histograma y su correspondiente polígono de frecuencias.

8. El censo, en miles de cabezas, del ganado en España en 1984, fue:

| Ganado | Nº Cabezas |
|---|---|
| Bovino | 4942 |
| Ovino | 17053 |
| Caprino | 2533 |
| Porcino | 11962 |
| Caballar | 254 |
| Mular | 145 |
| Asnal | 160 |

Representar la distribución por un diagrama de sectores y por un diagrama de rectángulos.

9. Se tiene la siguiente distribución de los matrimonios celebrados en España en 1981, por la edad y el sexo de los contrayentes.

| Edad | Varones | Mujeres |
|---|---|---|
| Menos de 20 años | 10352 | 38863 |
| De 20 a 24 años | 92970 | 112422 |
| De 25 a 29 años | 70853 | 35264 |
| De 30 a 34 años | 16205 | 7679 |
| De 35 a 39 años | 4763 | 2850 |
| De 40 a 49 años | 3215 | 2291 |
| De 50 a 59 años | 1750 | 1471 |
| De 60 y más años | 1929 | 1197 |

Representar el histograma de las dos distribuciones y compararlas. Hacer también los diagramas lineales comparativos. En ambos casos calcular los parámetros estadísticos de centralización y de dispersión.

10. La medida, en centímetros, del perímetro torácico de 1000 deportistas en la olimpiada de 1992 dio la siguiente distribución:

| X | 70-75 | 75-80 | 80-85 | 85-90 | 90-95 | 95-100 | 100-105 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 9 | 48 | 197 | 321 | 259 | 121 | 45 |

Calcular la tabla de frecuencias. Representar los datos en un diagrama. Calcular los parámetros de tendencia central y de dispersión más usuales.

11. Sabemos que, en una clase, la calificación media de un examen ha sido 5 y la desviación típica 1,5. En esa misma clase, para otro examen, la calificación media ha sido, tambien 5, y la desviación típica 1.

Si un alumno ha obtenido un 8 en el primer examen y un 7,5 en el segundo, ¿qué nota te parece más meritoria? ¿Por qué?

12. En una fábrica de tornillos se mide la longitud (en mm) de algunos de ellos y se obtiene:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 22 | 20 | 18 | 21 | 19 | 22 | 16 | 19 | 23 | 18 |
| 17 | 23 | 23 | 21 | 18 | 20 | 22 | 18 | 25 | 23 |
| 22 | 22 | 19 | 19 | 20 | 21 | 18 | 24 | 17 | 20 |
| 19 | 23 | 21 | 23 | 21 | 20 | 19 | 21 | 20 | 22 |
| 19 | 20 | 18 | 21 | 19 | 18 | 20 | 22 | 21 | 19 |

(a) Haz la tabla de frecuencias, represéntala gráficamente y calcula las medidas de centralización y de dispersión.

(b) Haz una nueva tabla de frecuencias agrupando los datos así: de 17 a 19 mm, de 20 a 22 mm y de 23 a 25 mm. Represéntala gráficamente y calcula la media, mediana y la desviación típica.

(c) ¿Qué centil corresponde a la longitud 24 mm ?

13. La estatura de 40 alumnos de una clase vienen dadas en la siguiente tabla. Calcula los parámetros de centralización y de dispersión. ¿Qué centil corresponde a una estatura de 180 cm.?

| $[L_{i-1}\ ;\ L_i)$ | $n_i$ |
|---|---|
| 158,5 - 163,5 | 1 |
| 163,5 - 168,5 | 5 |
| 168,5 - 173,5 | 11 |
| 173,5 - 178,5 | 14 |
| 178,5 - 183,5 | 6 |
| 183,5 - 188,5 | 3 |

14. Al lanzar tres dados distintos pueden darse 6x6x6 = 216 posibilidades. Imaginemos que tiramos 216 veces tres dados y se dieran todas y cada una de las 216 posibilidades distintas. En este caso, las sumas de las tres puntuaciones se distribuirían:

| $x_i$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 3 | 6 | 10 | 15 | 21 | 25 | 27 |
| $x_i$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| $n_i$ | 27 | 25 | 21 | 15 | 10 | 6 | 3 | 1 |

(a) Representa esta distribución mediante un diagrama de barras

(b) Calcula las medidas de centralización y dispersión.

(c) Calcula los intervalos $(\bar{x}-\sigma;\bar{x}+\sigma)$ y $(\bar{x}-2\sigma;\bar{x}+2\sigma)$ y el porcentaje de resultados de cada uno.

15. En una fábrica de bombillas se observaron 200 de ellas para estudiar su duración $X$ en horas, y se obtuvieron los siguientes resultados:

| $X$ | 100-200 | 200-300 | 300-400 | 400-500 | 500-600 | 600-700 | 700-800 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 2 | 7 | 16 | 49 | 62 | 41 | 23 |

(a) Representa los datos en un diagrama y calcula los parámetros de centralización y de dispersión.

(b) Obtener los centiles 10,70,90 y decir qué centiles corresponden a los valores 225 y 575.

(c) Obtener el porcentaje de bombillas cuya duración está en cada uno de los intervalos $(\bar{x} - \sigma; \bar{x} + \sigma)$ y $(\bar{x} - 2\sigma; \bar{x} + 2\sigma)$.

16. Se midió el tiempo que tardaron en imprimirse 50 archivos, y se obtuvieron los siguientes resultados, expresados en minutos:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 4 | 8 | 1 | 0 | 10 | 9 | 9 | 7 | 2 |
| 5 | 6 | 4 | 2 | 7 | 1 | 6 | 3 | 8 | 10 |
| 3 | 0 | 6 | 10 | 1 | 5 | 2 | 5 | 1 | 3 |
| 1 | 3 | 6 | 8 | 5 | 4 | 4 | 6 | 3 | 6 |
| 0 | 7 | 9 | 3 | 2 | 7 | 8 | 0 | 7 | 7 |

Construir una tabla estadística en la que aparezcan las frecuencias absolutas, las frecuencias relativas, las frecuencias absolutas acumuladas crecientes y las frecuencias relativas acumuladas decrecientes.

17. Los siguientes datos representan los litros de leche que se han consumido diariamente en una cafetería:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 35 | 30 | 12 | 46 | 49 | 54 | 56 | 59 | 60 | 70 |
| 67 | 75 | 93 | 18 | 32 | 44 | 46 | 50 | 55 | 58 |
| 21 | 60 | 72 | 67 | 77 | 36 | 45 | 47 | 73 | 69 |
| 24 | 63 | 74 | 65 | 77 | 56 | 41 | 80 | 52 | 65 |
| 103 | 45 | 48 | 56 | 58 | 74 | 80 | 71 | 24 | 54 |
| 48 | 39 | 44 | 88 | 54 | 70 | 74 | 71 | 28 | 45 |

Construir una tabla estadística, distribuyendo los datos en intervalos de amplitud 15, comenzando en el 0 - 15, en la que figuren frecuencias absolutas, frecuencias relativas y frecuencias absolutas acumuladas en sentido creciente.

18. Completar los datos que faltan en la tabla siguiente:

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| 10 | 2 | 0,05 | 2 | 0,05 |
| 13 | 4 | 0,1 | 6 | 0,15 |
| 16 | | | 16 | 0,4 |
| 19 | 15 | | | |
| 22 | 6 | 0,15 | 37 | 0,925 |
| 25 | | | | 1 |

19. Los desplazamientos de los trabajadores de la E.S.I. de Cádiz para llegar a su puesto de trabajo son, en kilómetros, los siguientes:

| $N^o$ kilómetros | $N^o$ operarios |
|---|---|
| más de 1 | 91 |
| más de 15 | 30 |
| más de 30 | 13 |
| más de 45 | 6 |
| más de 60 | 5 |
| más de 75 | 3 |
| más de 90 | 1 |

Sabiendo que el trabajador más distante realiza 93 kilómetros de viaje, obténgase la distribución de frecuencias absolutas acumuladas en sentido creciente. Indicar cuántos de estos operarios realizan un desplazamiento comprendido entre 30 y 60 kilómetros.

20. Se lanzan dos dados 80 veces y se suman las puntuaciones obtenidas en cada lanzamiento, y obtenemos los siguientes resultados:

| Suma | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frecuencia | 2 | 4 | 7 | 9 | 11 | 13 | 11 | 10 | 8 | 4 | 1 |

Hacer la representación gráfica más conveniente.

21. Dada la siguiente distribución, calcular los siguientes parámetros de centralización: Medias aritmética, geométrica, armónica y cuadrática. Comprobar la relación que existe entre ellas.

| $x_i$ | 2 | 3 | 8 | 12 | 17 |
|---|---|---|---|---|---|
| $n_i$ | 2 | 2 | 3 | 3 | 1 |

22. Medimos los recursos hidráulicos de la cuenca del guadalquivir en los últimos 27 años, y se obtienen los datos que aparecen en la tabla. Calcular la reserva media de este periodo.

| Recursos | 28432 | 28434 | 28436 | 28438 | 28440 |
|---|---|---|---|---|---|
| Años | 4 | 8 | 4 | 3 | 8 |

23. Calcular la frecuencia absoluta correspondiente al tercer intervalo, $n_3$, de la siguiente distribución; sabiendo que la media aritmética es 11,5.

| $[L_{i-1} ; L_i)$ | $n_i$ |
|---|---|
| 4 - 6 | 4 |
| 6 - 10 | 5 |
| 10 - 16 | $n_3$ |
| 16 - 20 | 3 |
| 20 - 30 | 1 |

24. Calcular la mediana de las siguientes distribuciones:

    i) En un curso, a cada alumno, se le pregunta cuántos hermanos son en su casa. Los resultados de la encuesta son los siguientes: 3, 2, 4, 5, 4, 1, 3, 3, 5, 2, 3, 6, 2, 4, 5, 3, 3, 4, 2, 4, 3, 4, 3, 3, 4, 2, 2, 4, 2, 2, 2, 4, 2, 7, 5, 4, 4, 3, 5, 4.

    Hacer el diagrama de barras que representa a los datos.

    ii) Estos mismos alumnos se han pesado, obteniéndose los siguientes pesos: 60, 60, 65, 55, 63, 48, 45, 38, 47, 65, 36, 47, 62, 63, 47, 52, 76, 74, 65, 50, 50, 59, 54, 52, 56, 57, 48, 49, 50, 50, 61, 59, 58, 45, 49, 52, 52, 52, 48, 48.

    Hacer una tabla de frecuencias agrupando los resultados en intervalos: [35,5 - 42,5); [42,5 - 49,5); [49,5 - 56,5); [56,5 - 63,5); [63,5 - 70,5); [70,5 - 79,5). Usando los intervalos anteriores, representar los datos en un histograma.

25. Medimos la estatura $X$ (en cm.) de todos los alumnos de una Escuela Taller y obtenemos:

| X | de 140 a 146 | de 146 a 152 | de 152 a 158 | de 158 a 164 | de 164 a 170 | de 170 a 176 | de 176 a 182 | de 182 a 188 |
|---|---|---|---|---|---|---|---|---|
| N$^o$ | 3 | 17 | 39 | 41 | 32 | 13 | 5 | 1 |

    (a) Dibuja el histograma correspondiente.

    (b) ¿Cuántos alumnos miden menos de 164 cm.? ¿Qué porcentaje del total son?

    (c) Calcula el porcentaje de alumnos que miden entre 152 cm. y 170 cm.

    (d) Calcula el porcentaje de los alumnos que miden menos de 190 cm.

(e) Estima el porcentaje de los que miden 167 cm. o menos.

26. La siguiente tabla muestra las superficies (expresadas en millones de $Km^2$) de los cinco océanos:

| | |
|---|---|
| Antártico | 20 |
| Ártico | 12 |
| Atlántico | 150 |
| Índico | 73 |
| Pacífico | 180 |

Representar los datos en un diagrama de barras y en un diagrama de sectores.

27. Se ha elegido una muestra de universitarios, a los cuales se les ha medido la velocidad lectora, habiéndose obtenido las siguientes puntuaciones:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 68 | 45 | 48 | 32 | 51 | 52 | 70 | 47 | 50 | 34 | 53 | 54 |
| 62 | 50 | 53 | 37 | 56 | 57 | 67 | 55 | 58 | 42 | 61 | 64 |
| 68 | 44 | 45 | 46 | 55 | 39 | 41 | 47 | 49 | 43 | 52 | 66 |
| 47 | 54 | 56 | 32 | 36 | 39 | 34 | 65 | 61 | 59 | 67 | 66 |

Se pide:

(a) Elegir la amplitud de intervalo más adecuada y hacer la distribución de frecuencias de esas puntuaciones.

(b) Trazar el polígono de frecuencias y el histograma. Describir la distribución.

28. Un gabinete de trabajo está realizando un estudio de mercado en los municipios de un área metropolitana. Entre otros datos, obtuvo la distribución de la renta per cápita por municipio, habiéndose construido una tabla completa que posteriormente se extravió quedando únicamente una fotocopia en mal estado, con muchos datos borrados. Reconstruya aquella tabla original a partir de la siguiente incompleta,

y represente el histograma correspondiente:

| Intervalo | | | $n_i$ | $N_i$ | $f_i$ | $F_i$ | $x_i$ | $a_i$ | altura |
|---|---|---|---|---|---|---|---|---|---|
| 20000 | - | 50000 | 2 | 2 | | | 35000 | | |
| | - | 60000 | | | | 0,04 | | | |
| 60000 | - | | | | | 0,12 | | 10000 | |
| | - | | | | 0,13 | | 75000 | | |
| | - | | 44 | | | | | | 0,0022 |
| 100000 | - | | | 119 | | | | 0,001 | |
| | - | 150000 | | | 0,11 | | | | |
| 150000 | - | | | 171 | | | | 50000 | |
| | - | 250000 | | | | | | | |
| | - | 850000 | 12 | 200 | | | 550000 | | |

29. En el Departamento de Personal de una fábrica se ha realizado una investigación estadística en relación a los salarios, expresados en euros, que perciben anualmente sus trabajadores por todos los conceptos. Los resultados obtenidos fueron los siguientes:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 19000 | 21000 | 20000 | 25000 | 25000 | 22000 | 30000 | 29000 |
| 24000 | 26000 | 24000 | 28000 | 26000 | 22000 | 26000 | 27000 |
| 21000 | 29000 | 23000 | 29000 | 24000 | 19000 | 30000 | 21000 |
| 21000 | 27000 | 28000 | 26000 | 27000 | 23000 | 20000 | 29000 |
| 28000 | 22000 | 25000 | 22000 | 26000 | 28000 | 25000 | 24000 |
| 19000 | 30000 | 27000 | 26000 | 20000 | 28000 | 27000 | 25000 |
| 27000 | 23000 | 23000 | 30000 | 29000 | 28000 | 27000 | 24000 |
| 26000 | 25000 | 26000 | 23000 | | | | |

Se pide:

(a) La distribución de los salarios:

i. sin agrupar en intervalos,

ii. agrupados en cuatro intervalos de igual amplitud.

(b) La representación gráfica, mediante un histograma, de la distribución obtenida en el punto (a)(ii).

(c) Porcentaje de salarios iguales o inferiores a 26500 pesetas

i. con los datos sin agrupar en intervalos,

ii. con los datos agrupados en cuatro intervalos de igual amplitud.

(d) Hallar la mediana de los salarios a partir de la distribución obtenida en el punto (a) (ii).

(e) Hallar la moda de los salarios a partir de la distribución obtenida en el punto (a) (ii).

(f) Calcular el salario medio:

i. utilizando los datos originales

ii. con los datos agrupados. Coméntense los resultados obtenidos.

30. En una población de Cádiz, la superficie útil de las viviendas sigue la siguiente distribución:

| Superficie ($m^2$) | Frecuencia relativa (%) |
|---|---|
| 50-60 | 20 |
| 60-70 | 25 |
| 70-80 | 15 |
| 80-100 | 25 |
| 100-120 | 15 |

Calcular:

(a) La superficie media por vivienda.

(b) Los tipos de vivienda que dividen la distribución en cuatro partes iguales.

(c) El tipo de vivienda más frecuente.

31. Las tres factorías de una industria informática han producido en el último año el siguiente número de ordenadores, con detalle por trimestres:

| | factoría 1 | factoría 2 | factoría 3 |
|---|---|---|---|
| $1^{er}$ trimestre | 600 | 650 | 550 |
| $2^{0}$ trimestre | 750 | 1200 | 900 |
| $3^{er}$ trimestre | 850 | 1250 | 1050 |
| $4^{0}$ trimestre | 400 | 800 | 650 |

Obtener:

(a) La producción media trimestral de cada factoría y de toda la industria.

(b) Producción media diaria de cada factoría y de toda la industria, teniendo en cuenta que durante el primer trimestre hubo 68 días laborables, en el segundo 78, en el tercero 54 y en el cuarto 74.

32. Una empresa ha pagado por un artículo: 225, 150, 300 y 200 ptas. de precio.

    Determinar el precio medio en las siguientes hipótesis:

    (a) Compraba cada vez por un mismo importe global.

    (b) Compraba por valor de 38250, 47500, 49500 y 42000 ptas. respectivamente.

    (c) Compraba 174, 186, 192 y 214 unidades respectivamente.

33. La distribución de las acciones de una determinada sociedad que cotiza en Bolsa viene agrupada en intervalos en la siguiente tabla:

| $N^o$ de acciones | $N^o$ de accionistas |
|---|---|
| 0 - 20 | 10 |
| 20 - 28 | 32 |
| 28 - 32 | 50 |
| 32 - 48 | 8 |

    Calcular:

    (a) El capital aproximado de la empresa, supuesto un valor nominal para cada título de 500 euros.

    (b) Promedio de acciones por accionista.

    (c) Paquete de acciones poseídas por un mayor número de accionistas.

    (d) En el supuesto que en la junta general de accionistas los votos se establecen en proporción al número de acciones poseídas, ¿qué mínimo de acciones debe tener un accionista para que su poder decisorio sea mayor al de la mitad de los socios?

    (e) Varianza y coeficiente de variación.

    (f) Coeficiente de asimetría de Fisher.

34. Las dos distribuciones que se dan a continuación corresponden a los

resultados logrados en un test intelectual por dos escuelas diferentes:

| Escuela A | $n_i$ | Escuela B | $n_i$ |
|---|---|---|---|
| 10-14 | 3 | 0-4 | 1 |
| 15-19 | 5 | 5-9 | 5 |
| 20-24 | 10 | 10-14 | 7 |
| 25-29 | 15 | 15-19 | 10 |
| 30-34 | 12 | 20-24 | 14 |
| 35-39 | 11 | 25-29 | 18 |
| 40-44 | 7 | 30-34 | 12 |
| 45-49 | 3 | 35-39 | 7 |
| | | 40-44 | 5 |
| | | 45-49 | 1 |

(a) ¿Qué alumno ha tenido una puntuación más elevada con respecto a su propio grupo, el que ha obtenido 36 puntos en la escuela A ó 30 en la B?

(b) ¿Cuál de los grupos ha sido más homogéneo en sus resultados?

35. Una persona realiza en automóvil el viaje Cádiz-Sevilla, a una velocidad media de 120 Km/h., y el regreso Sevilla-Cádiz a una velocidad media de 80 Km/h. Calcúlese la velocidad media para el recorrido total Cádiz-Sevilla-Cádiz.

36. En un almacén de frutas se pesan 46 cajas de cerezas que se almacenan para la venta. Los pesos, en Kg., son los siguientes:

| | | | | | | |
|---|---|---|---|---|---|---|
| 11,000 | 13,000 | 12,450 | 12,750 | 12,200 | 12,000 | 12,550 |
| 12,100 | 10,950 | 13,400 | 13,400 | 11,600 | 12,350 | 12,350 |
| 12,150 | 12,300 | 11,400 | 11,500 | 12,350 | 11,900 | 12,800 |
| 12,150 | 13,100 | 12,000 | 12,000 | 11,750 | 11,750 | 12,950 |
| 11,150 | 11,200 | 11,000 | 12,550 | 11,550 | 12,200 | 10,900 |
| 11,200 | 12,650 | 12,000 | 12,000 | 11,600 | 12,150 | 11,000 |
| 12,450 | 11,550 | 12,800 | 12,500 | | | |

(a) Calcular la desviación media.

(b) Calcular la desviación típica para los datos sin agrupar.

(c) Calcular la desviación típica para los datos agrupados en 8 intervalos como sigue a continuación: [10,900-11,225), [11,225-11,550), [11,550-11,875), [11,875-12,200), [12,200-12,525), [12,525-12,850), [12,850-13,175), [13,175-13,450). Constatar e interpretar la diferencia de valores de las desviaciones típicas.

37 Un piloto de carreras participa en una competición en la cual obtiene para los distintos recorridos, las siguientes velocidades medias:

| Recorrido | Distancia (Km) | Velocidad media (Km/h) |
|---|---|---|
| A-B | 400 | 50 |
| B-C | 600 | 60 |
| C-A | 1000 | 100 |

Calcular la velocidad media conseguida en la competición.

38. En una selección de personal, se aplicó una prueba de mecanografía, obteniéndose los siguientes resultados:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 71 | 61 | 54 | 50 | 70 | 60 | 54 | 50 |
| 69 | 59 | 54 | 49 | 69 | 58 | 54 | 47 |
| 69 | 58 | 53 | 40 | 64 | 57 | 52 | 39 |
| 64 | 56 | 52 | 34 | 63 | 55 | 51 | 30 |

Se pregunta:

(a) Formar una distribución de frecuencias con el intervalo más adecuado.

(b) Representar el polígono de frecuencias.

(c) Hallar la media.

(d) Hallar la mediana.

(e) Hallar $Q_1$ y $Q_3$.

(f) Hallar el percentil 8 y 90.

(g) Hallar la moda.

39. Un horno cerámico venía operando con temperaturas que oscilaban entre los $80\ ^0$ C y los $140^0$ C con media igual a $120^0$ C y desviación típica de $10^0$ C. Con un nuevo sistema se ha conseguido incrementar las temperaturas habituales en $9^0$ C. ¿En qué medida vendría afectado el coeficiente de variación?

40. Una empresa está organizada en dos factorías con funcionamiento y gestión independientes; la distribución de los salarios en cada una de ellas es:

| Decenas de Euros al mes | Nºtrabajadores (factoría I) | Nº trabajadores (factoría II) |
|---|---|---|
| 30-40 | 6 | 2 |
| 40-60 | 14 | 18 |
| 60-80 | 22 | 25 |
| 80-100 | 14 | 10 |
| 100-110 | 8 | 4 |
| 110-120 | 5 | 4 |
| 120-180 | 2 | 1 |

Se pide:

(a) Representación gráfica de los histogramas correspondientes.

(b) Salario total, medio y modal de la empresa, en la consideración conjunta de ambas factorías.

(c) Recorrido semiintercuartílico de los salarios de cada una de las factorías.

(d) A partir de la varianza de los salarios de las factorías I y II obtenga, sin necesidad de reagrupar las frecuencias, la varianza total de los salarios de la empresa.

41. La siguiente distribución de frecuencias corresponde a un modelo experimental del tiempo máximo que un envase de mayonesa permanece apta para el consumo si se conserva en frigorífico una vez abierta.

| Nº de días | Frecuencia |
|---|---|
| menos de 1 | - |
| 1 | 0,02 |
| 2 | 0,08 |
| 3 | 0,11 |
| 4 | 0,14 |
| 5 | 0,38 |
| 6 | 0,20 |
| 7 | 0,07 |
| más de 7 | - |

Calcular:

(a) Media, moda y mediana.

(b) Varianza y desviación típica.

(c) Coeficientes de asimetría de Fisher y Pearson.

(d) Coeficiente de Curtosis.

42. Realizada una encuesta entre fumadores se ha obtenido la siguiente tabla de frecuencias:

| Nº de cigarrillos diarios | Nº de individuos |
|---|---|
| de 4,5 a 9,5 | 10 |
| de 9,5 a 14,5 | 15 |
| de 14,5 a 19,5 | 25 |
| de 19,5 a 24,5 | 18 |
| de 24,5 a 29,5 | 22 |

Se pide:

(a) El número medio de cigarrillos fumados por individuo y día.

(b) La desviación típica.

(c) El coeficiente de variación.

(d) El porcentaje de individuos que fuman entre 12 y 22 cigarrillos diarios, ambos inclusive.

(e) El primer cuartil.

(f) El valor más frecuente de la variable.

(g) El consumo diario de tabaco para una población de 1000 personas, supuesto que el porcentaje de fumadores es del 30%, y que es válido el consumo medio por fumador y día calculado en el punto primero.

43. Un curso está dividido en cuatro grupos, de los cuales tenemos los siguientes datos:

| Grupo | Nº de alumnos | Nota media | Varianza |
|---|---|---|---|
| A | 30 | 6 | 1 |
| B | 40 | 6,5 | 1,5 |
| C | 50 | 5 | 0,81 |
| D | 60 | 4 | 0,64 |

Se pide:

(a) Calcular la nota media para todo el grupo.

(b) Calcular los coeficientes de variación de cada grupo.

(c) ¿Qué grupo resulta más homogéneo?

(d) Calcular la varianza de todas las notas del curso.

44. En una empresa metalúrgica los empleados se clasifican en tres categorías: técnicos, especialistas y administrativos. El número de empleados, el salario medio mensual y la varianza de los salarios de cada categoría en el mes de diciembre de 1998 son los que aparecen en el siguiente cuadro:

| Categoría | $N^o$ de empleados | Salario mensual medio | Varianza de los salarios |
|---|---|---|---|
| Técnicos | 20 | 200 | 400 |
| Especialistas | 100 | 120 | 49 |
| Administrativos | 40 | 100 | 25 |

(a) Calcúlese el salario medio para el conjunto de la empresa y la dispersión relativa de los salarios.

(b) En la discusión para fijar los salarios del año 1999 han sido propuestas tres alternativas:

i. La elevación de todos los salarios en un 5%.

ii. La elevación de todos los salarios en 5500 ptas. mensuales.

iii. La elevación de los salarios según el siguiente baremo: 4% a los técnicos, 5% a los especialistas y 5,5% a los especialistas y 5,5% a los administrativos.

A. Calcular los salarios medios totales que resultan de aplicar las 3 alternativas y la dispersión relativa en cada caso.

B. ¿Cuál de las tres alternativas tiene mayor efecto para reducir la dispersión relativa inicial de los salarios para el conjunto de la empresa?

45. En una compañía de seguros, el montante de las pólizas atribuibles a los agentes que figuran en la plantilla se distribuye de la siguiente

forma:

| Montante de las pólizas (miles de euros) | Nº de agentes |
|---|---|
| 0- 5 | 10 |
| 5-10 | 12 |
| 10-20 | 60 |
| 20-40 | 90 |
| 40-70 | 18 |
| 70-120 | 10 |

Calcular los parámetros de centralización, dispersión, simetría y forma de la distribución.

46. En una encuesta sobre duración media de las llamadas telefónicas urbanas se han obtenido los siguientes resultados:

| Duración media (en minutos) | Nº de llamadas |
|---|---|
| 0-1 | 60 |
| 1-2 | 90 |
| 2-3 | 150 |
| 3-6 | 200 |
| 6-9 | 160 |
| 9-12 | 40 |

(a) Calcular: Duración media, varianza, desviación típica, coeficiente de variación, mediana y moda.

(b) Si la distribución anterior es representativa del conjunto de llamadas urbanas que se registran en la ciudad, y sabiendo que diariamente se efectúan un promedio de 100000 llamadas, calcúlense los ingresos diarios que obtiene la compañía telefónica por este concepto, siendo la tarifa de 5 ptas. por llamada más 4 ptas. por cada minuto y fracción.

47. Una empresa tiene diez delegaciones para la venta de sus productos. La estadística de ventas y número de empleados para los años 1996 y

1997 es la siguiente:

| | 1996 | | 1997 | |
|---|---|---|---|---|
| Delegación | Nº empleados | Ventas | Nº empleados | Ventas |
| A | 20 | 600 | 22 | 700 |
| B | 15 | 540 | 15 | 570 |
| C | 8 | 200 | 7 | 195 |
| D | 12 | 420 | 15 | 605 |
| E | 10 | 350 | 15 | 580 |
| F | 25 | 680 | 23 | 700 |
| G | 32 | 900 | 32 | 960 |
| H | 7 | 240 | 8 | 310 |
| I | 14 | 500 | 16 | 700 |
| J | 21 | 700 | 22 | 730 |

Calcular:

(a) La cifra total de ventas de los años 1996 y 1997.

(b) Las ventas por empleado para los años 1996 y 1997.

(c) Si los precios de ventas de la empresa subieron, por término medio, en el 87 un 4% respecto a los de 1996, ¿en cuánto estimaría el incremento de productividad de la plantilla de ventas de la empresa?

48. Un grupo de 20 trabajadores de la construcción trabajando a destajo colocan un promedio de 62 ladrillos siendo la mediana del grupo 52. Otro grupo similar de 14 trabajadores colocan un promedio de 57 ladrillos siendo también la mediana igual a 57.

    (a) ¿Puede decirse que el primer grupo de trabajadores es mejor que el segundo? ¿Por qué?

    (b) Por otra parte, ¿qué interpretación estadística explica la diferencia que existe en el primer grupo entre su media y su mediana?

49. ¿Cuál es el salario medio anual en una empresa de auditorías de 323 trabajadores, sabiendo que 117 auditores perciben por término medio 62000 euros de sueldo y 206 inspectores 46500 euros?

50. Supongamos que desde hace 5 años una empresa gasta por Navidad la cantidad total de 100000 ptas. en regalos a sus clientes. Si los precios de ese regalo han sido durante los 5 años: 500, 600, 750, 800 y 1000

pesetas respectivamente, calcular el promedio del coste para el período de 5 años.

51. Se sabe que el consumo medio de gasolina de una determinada marca de automóviles es de 7 litros, con una desviación típica de 0,7 litros.

    (a) Calcular las puntuaciones tipificadas que corresponderán a los consumos de 9 litros, 6 litros y 5,5 litros.

    (b) Sabiendo que el consumo medio de otra segunda marca diferente es de 6,5 litros con una desviación típica de 0,8 litros, determinar si un automóvil de esta marca que consumiera 8 litros tendría un consumo relativo mayor a otro de la primera marca que consumiera igual.

52. Un fabricante de detergentes quiere comprar una máquina empaquetadora que realice la operación de llenar bolsas de 250 gramos.

    La casa A le ofrece una máquina con la cual realiza unas pruebas y obtiene los siguientes resultados: 250, 252, 248, 253, 247. La casa B le ofrece otra, y realizando pruebas obtiene los siguientes resultados: 258, 252, 247, 248, 245.

    Halla la media aritmética y obtiene que ambas registran una media de 250 gramos. Por ello, decide comprar la máquina que posea menos dispersión. ¿Cuál es esa máquina?

53. El Instituto Nacional de Estadística publica que el índice industrial del Estado Español es de 600 con una desviación típica de 2,5. Sabemos que Barcelona, Granada, Bilbao y Cáceres respectivamente tienen unos valores tipificados de 4,5; -2,7; 3,2 y -5. ¿Cuál es el índice industrial de estas provincias?

54. Elegida una familia tipo se comprobó que los gastos mensuales en ocio, esparcimiento, recreo y conceptos afines eran de 16000 ptas. durante todo el año laboral, mientras que en el mes de vacaciones eran 2,5 veces más.

    En la zona en la que habita la familia once meses del año y por aquellos mismos conceptos, se ha calculado un promedio de 12500 ptas./mes-familia y una desviación típica de 3800 ptas./mes-familia, mientras que en la localidad que pasan el mes de vacaciones la media y desviación típica son de 38000 y 5000 ptas. por familia respectivamente.

¿En cuál de los dos hábitats ocupa esta familia mejor posición comparada según aquel tipo de gastos?

55. Comente cada una de las siguientes frases populares.

    (a) "Hay tres clases de mentiras: la mentira, la mentira desvergonzada y la estadística".

    (b) "Si yo me como dos pollos y tú ninguno, la estadística dice que nos hemos comido uno cada uno".

    (c) "Un estadístico es aquel señor que, tendido en la mesa de la cocina, con los pies en la nevera y la cabeza en el horno, es capaz de asegurar que la temperatura ambiente es normal".

56. Una empresa clasifica a sus empleados de taller en tres grupos: peones, oficiales y especialistas; con un salario medio de 51200, 73400 y 127800 ptas., respectivamente, y 64000 para todo el conjunto.

    Calcular:

    (a) El porcentaje de empleados de cada grupo, sabiendo que hay seis peones por cada oficial.

    (b) Si está previsto un incremento salarial medio de 4000 ptas. para peones y oficiales y 9000 ptas. para los especialistas, ¿de qué modo afectará al salario medio de todos los empleados?

57 En un centro comercial se han obtenido los siguientes datos sobre las ventas de un aparato de radio para automóvil de una marca determinada en distintos comercios:

| Tienda | Precio(Euros) | Total de Ventas(Euros) |
|---|---|---|
| A | 60 | 18000 |
| B | 70 | 28000 |
| C | 80 | 16000 |
| D | 90 | 18000 |

Calcular:

(a) El precio medio por unidad pagado por el consumidor

(b) La varianza del precio.

58. A partir de la siguiente información sobre "Dependencia externa de la economía española", que figura en el Informe Económico-1986 del Banco de Bilbao:

DEPENDENCIA EXTERNA DE LA ECONOMíA ESPAÑOLA
Porcentaje sobre el PIB a precios de mercado

| Año | Importación de bienes y servicios | Exportación de bienes y servicios |
|---|---|---|
| 1981 | 18,1 | 15,8 |
| 1982 | 20,2 | 18,1 |
| 1983 | 20,6 | 18,8 |
| 1984 | 21,9 | 21,3 |
| 1985 | 21,0 | 23,3 |
| 1986 | 17,8 | 20,5 |

determínese si es más estable en el tiempo, en relación al PIB, la importación de bienes y servicios o la exportación de bienes y servicios.

59. En una compañía aérea se sabe que, por término medio, el 35% de los vuelos tiene retraso. La distribución de los vuelos retrasados es la siguiente:

| Duración del vuelo (centésimas de hora) | Nº vuelos retrasados |
|---|---|
| 0-10 | 2000 |
| 10-20 | 3000 |
| 20-30 | 2500 |
| 30-50 | 2000 |
| 50-100 | 500 |

Calcúlese:

(a) El retraso medio de los vuelos retrasados.

(b) El retraso medio del total de los vuelos.

(c) Dentro del conjunto de los vuelos retrasados, el retraso más frecuente.

(d) Dentro del conjunto de los vuelos retrasados, determinar la mediana.

# Unidad Temática II

# Análisis conjunto de variables estadísticas.

# Resumen Teórico

## 2.1 Conceptos Generales.

### 2.1.1 Distribución conjunta de dos variables.

Supongamos que de cada uno de los elementos de un colectivo se observan dos caracteres (variables, atributos o variable y atributo), nos ocuparemos en este capítulo de estudiar las relaciones de influencia entre ambos. Se pueden presentar dos casos extremos: el primero de ellos, sería aquel en que entre los caracteres existiera una relación funcional, de tal forma, que conocido el valor de uno de ellos se pudiera, mediante dicha función, obtener unívocamente el valor del otro; y en el segundo de los casos, nos encontraríamos cuando los caracteres fueran independientes y el poseer determinada información de uno no arrojara ninguna luz sobre el otro. Entre estas situaciones extremas se pueden dar una infinidad de casos, en los cuales tendríamos un cierto grado de dependencia. Nuestro objetivo será analizar el nivel de influencia existente entre los caracteres que llamaremos dependencia estadística. Entendiendo, en todo caso, que dicho análisis no establecerá la dirección de la relación, es decir, cuál es la causa y cuál es el efecto, sino sólo su intensidad.

Para proceder al estudio es conveniente organizar los datos de la distribución bidimensional en una tabla de doble entrada como la que puede verse en la figura2.20.

La distribución conjunta vendría dada por los pares de valores $(x_i, y_i)$ acompañados de su frecuencia absoluta, como consecuencia se tendrá que

$$\sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = N$$

extremo que aparece recogido en la parte inferior derecha de la Tabla.

| $X/Y$ | $y_1$ | $\cdot$ | $y_j$ | $\cdot$ | $y_s$ | $n_{i.}$ |
|---|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $\cdot$ | $n_{1j}$ | $\cdot$ | $n_{1s}$ | $n_{1.}$ |
| $\cdot$ | $\cdot$ | | $\cdot$ | $\cdot$ | $\cdot$ | $\cdot$ |
| $x_i$ | $n_{i1}$ | $\cdot$ | $n_{ij}$ | $\cdot$ | $n_{is}$ | $n_{i.}$ |
| $\cdot$ | $\cdot$ | | $\cdot$ | $\cdot$ | $\cdot$ | $\cdot$ |
| $x_r$ | $n_{r1}$ | $\cdot$ | $n_{rj}$ | $\cdot$ | $n_{rs}$ | $n_{r.}$ |
| $n_{.j}$ | $n_{.1}$ | $\cdot$ | $n_{.j}$ | $\cdot$ | $n_{.s}$ | $N$ |

Figura 2.20: Tabla de doble entrada

También podría presentarse la Tabla de doble entrada acompañada c su frecuencia relativa $f_{ij}$ que es igual al número de veces que se presenta par dividido por el total de observaciones, como consecuencia se tendrá

$$\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = 1 \quad \text{siendo} \quad f_{ij} = \frac{n_{ij}}{N}$$

extremo que aparecería recogido en la parte inferior derecha de dicha tab

Si la distribución es de atributos, la Tabla se llama de Contingencia tendrá tratamiento aparte más adelante, y si es de Variables la llamarem de Correlación. A partir de ahora limitaremos el estudio al caso en el q los caracteres sean Variables.

### 2.1.2 Distribuciones marginales.

En la Tabla I, se han sumado las frecuencias que aparecen en cada una las filas y columnas, colocándose los resultados en los "márgenes", y así

$$\sum_{i=1}^{r} n_{ij} = n_{.j} \qquad \sum_{j=1}^{s} n_{ij} = n_{i.} \qquad \sum_{i=1}^{r} f_{ij} = f_{.j} \qquad \sum_{j=1}^{s} f_{ij} = f_{i.}$$

de tal forma que la primera y última columna de la Tabla I, constituy la distrubución Marginal de $X$, y la primera y última fila la Distribuci Marginal de $Y$. Lógicamente se verifica que:

$$\sum_{i=1}^{r} n_{i.} = N \qquad \sum_{j=1}^{s} n_{.j} = N \qquad \sum_{i=1}^{r} f_{i.} = 1 \qquad \sum_{j=1}^{s} f_{.j} = 1$$

lo que garantiza que ambas sean distribuciones unidimensionales.

### 2.1.3 Distribuciones condicionadas.

Nos preguntamos ahora por lo que ocurriría con una de las Variables si fijásemos un determinado valor para la otra. La respuesta implica el concepto de Distribución Condicionada que definimos de la siguiente forma:

$$f(x_i/y_j) = \frac{n_{ij}}{n_{.j}} \qquad f(y_j/x_i) = \frac{n_{ij}}{n_{i.}}$$

que son, respectivamente, las frecuencias relativas de la distribución condicionada de $X$ para el valor $y_j$ de $Y$ $(j = 1, 2, ..., s)$ y las frecuencias relativas de la distribución condicionada de $Y$ para el valor $x_i$ de $X$ $(i = 1, 2, .., r)$.

**Independencia.** Decimos que la variable $X$ es independiente de $Y$ si, para todos los posibles valores de $i$ y de $j$, se verifica:

$$f(x_i/y_j) = f_{i.}$$

es decir, si la distribuciónes Condicionadas coinciden con la Marginales. De la misma forma se definiría la independencia de $Y$ respecto de $X$. La definición de condicionada nos da una expresión alternativa para la Independencia, y así, $X$ e $Y$ son independientes si:

$$f_{ij} = f_{i.} \cdot f_{.j} \quad \forall \quad i, j \quad \text{o equivalentemente} \quad N \cdot n_{ij} = n_{i.} \cdot n_{.j} \quad \forall \quad i, j$$

que además pone de manifiesto que la Independencia se establece en un doble sentido: $X$ es independiente de $Y$ si y solo si $Y$ lo es de $X$.

### 2.1.4 Covarianza y Correlación.

Abordaremos en este epígrafe las posibles relaciones de dependencia lineal entre las variables y la forma de cuantificarlas, para ello, y al igual que en el caso unidimensional introduciremos la noción de momento respecto al origen y respecto a la media.

**Momentos.** Definimos el Momento de Orden $h, k$ respecto al Origen como:

$$a_{hk} = \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{s} x_i^h \cdot y_j^k \cdot n_{ij}$$

Es inmediato comprobar que $a_{10}$, es la Media de $X$, que $a_{01}$ es la Media de $Y$, etc.

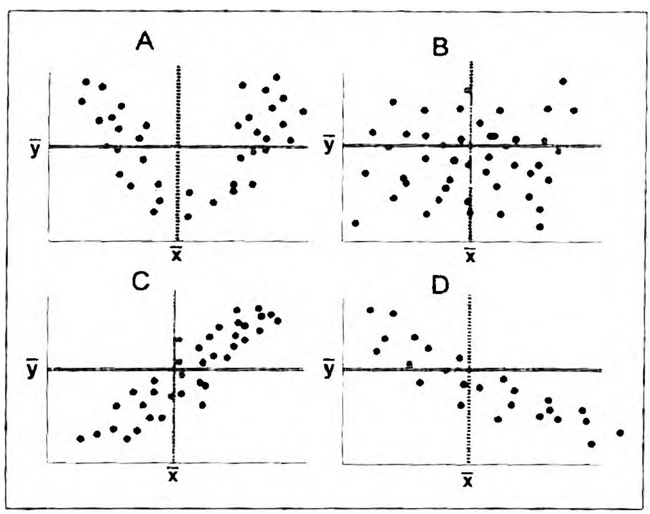
Definimos el Momento de Orden $h, k$ respecto a la Media como:

$$m_{hk} = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \overline{x})^h \cdot (y_j - \overline{y})^k \cdot n_{ij}$$

Se tiene que $m_{10}$ y $m_{01}$ valen cero. Las varianzas de $X$ e $Y$ son, res pectivamente, $m_{20}$ y $m_{02}$. Es posible expresar los Momentos respecto a l Media en función de los Momentos respecto al Origen, en particular se d la relación:

$$m_{11} = a_{11} - a_{10} \cdot a_{01}$$

**Covarianza y correlación.** A $m_{11}$ lo llamaremos Covarianza de l Distribución y se le suele notar como $S_{xy}$ o $\text{Cov}(X,Y)$. Este coeficien te jugará un importante papel en el estudio de la relación lineal entre la variables. Para analizar esta cuestión, consideremos las siguientes represen taciones gráficas que reflejan distintas situaciones, dichas representacione reciben el nombre de Nube de Puntos:




Ilustr. 1

El punto que viene determinado por la Media de $X$ y la Media de Y constituye el "centro de gravedad" de las Nubes de Puntos en todos lo casos.

Como sabemos la Covarianza viene dada por la expresión:

$$\text{Cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \overline{x}) \cdot (y_j - \overline{y}) \cdot n_{ij}$$

La covarianza es una medida simétrica y puede leerse como la suma de los productos de las desviaciones de $X$ por las desviaciones de $Y$ con respecto a sus medias respectivamente; de tal forma, que si el signo de las desviaciones de $X$ coincide con las de $Y$, como ocurre en el primer y tercer cuadrante se generan sumandos positivos de la Covarianza; y cuando el signo es distinto -segundo y cuarto cuadrante- la aportación a la Covarianza es negativa. Por tanto, la concentración de valores en los distintos cuadrantes determinará el signo y el tamaño de la covarianza. Así en (A) y en (B) la covarianza será próxima a cero, en (C) será alta y positiva y en (D) alta y negativa. Estamos en condiciones de afirmar que la covarianza detecta la relación lineal entre las variables y el sentido de ésta, pero no distingue entre la no presencia de relación (B) y la existencia de alguna dependencia no lineal (A). De todas formas aún para el estudio de relaciones lineales la covarianza adolece de ciertos problemas, como son el de venir acompañada de las unidades de las variables y el de depender del número de observaciones.

**Coeficiente de Correlación lineal.** Para obviar las carencias de $\mathrm{Cov}(X, Y)$ introducimos el coeficiente de Correlación Lineal:

$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \cdot \sigma_y}$$

que es una medida abstracta, ordinal, está acotada por el intervalo $[-1, 1]$ y tiene el signo de $\mathrm{Cov}(X, Y)$, por lo que cuando la relación lineal entre $X$ e $Y$ es exacta y directa vale 1, cuando es exacta e inversa -1 y cuando no hay relación lineal 0; con un análisis lógico para las posiciones intermedias. Cuando $r$ vale cero, se dice que las variables están incorreladas. Al cuadrado de $r$ se la llama coeficiente de determinación y se le nota como $R^2$, es una medida cardinal.

**Resultado.** La independencia implica incorrelación. El recíproco no siempre es cierto.

Este resultado es consecuencia de que la independencia se supone la descomposición de los momentos de orden $(h, k)$ (respecto al Origen o respecto a la Media) en el producto de los momentos $(h, 0)$ y $(0, k)$; así, $a_{11}=a_{10}a_{01}$ y por tanto $\mathrm{Cov}(X, Y) = a_{10}a_{01} - a_{10}a_{01} = 0$, con lo que $r = 0$ y las variables están incorreladas. Mientras que la incorrelación sólo implica esa descomposición para el momento $(1, 1)$. En cierta forma es una independencia de primer orden.

### 2.1.5 Vector de medias.

Podemos generalizar el estudio a distribuciones $k$-dimensiones ($k > 2$) de tal forma que para cada elemento del colectivo en estudio se obtienen observaciones sobre $k$ caracteres distintos que genera una variable $k$-dimensional. Dicha variable podemos interpretarla como un vector $X$ con $k$ componentes $(X_1, X_2, \ldots, X_k)$. Llamaremos Vector de Medias de la variable a:

$$\overline{X} = \begin{pmatrix} \overline{X_1} \\ \overline{X_2} \\ \cdot \\ \cdot \\ \\ \overline{X_k} \end{pmatrix}$$

### 2.1.6 Matriz de varianzas y covarianzas.

Llamaremos Matriz de Varianzas y covarianzas a la matriz cuadrada simétrica que tiene en la diagonal principal la Varianza de cada una de las variables y fuera las covarianzas entre cada par de ellas. Es decir:

$$M = \begin{pmatrix} S_1^2 & S_{12} & \cdot & \cdot & \cdot & S_{1k} \\ S_{12} & S_2^2 & \cdot & \cdot & \cdot & S_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ S_{1k} & S_{2k} & \cdot & \cdot & \cdot & S_k^2 \end{pmatrix}$$

En notación matricial lo anterior se expresa como:

$$M = \frac{1}{N} \sum (X_i - \overline{X}) \cdot (X_i - \overline{X})\,'$$

**Varianza generalizada.** Como proyección del concepto de Varianza definimos la Varianza Generalizada como el determinante de la Matriz de Varianzas y Covarianzas. Dicho determinante es siempre no negativo y para el caso de dos variables vendría dado por:

$$|M| = S_x^2 S_y^2 (1 - r^2)$$

Si las variables están incorreladas, es decir $r = 0$, $|M|$ toma su valor máximo: el producto de las varianzas.

Cuando existe cierto grado de relación lineal $0 < r^2 < 1$, la Varianza Generalizada disminuye, pues conocida una de las variables la variabilidad de la otra se reduce. Y por último, si $r^2 = 1$, conocida una de las variables queda perfectamente determinada la otra y la variabilidad conjunta es nula.

## 2.2 Ajuste y regresión en $I\!R^2$.

### 2.2.1 Introducción.

Supongamos que tenemos $N$ observaciones de una serie estadística bidimensional $(X, Y)$. Representamos en un gráfico la serie, obteniéndose una nube de puntos. El problema que nos planteamos consiste en encontrar una relación entre las variables que exprese los valores de una de ellas en función de los de la otra. Para hacer esto podemos optar por uno de los caminos siguientes: prefijar una clase funcional o trabajar sin clase funcional prefijada alguna. En el primer caso, estaremos realizando un ajuste, y en el segundo, tendremos un problema de regresión.

En el caso del ajuste, la cuestión será encontrar los parámetros que identifique la función dentro de la clase. Esto se consigue imponiendo a ésta que verifique condiciones de adaptación a la nube de puntos de la muestra de trabajo. Nosotros utilizaremos el criterio de los mínimos cuadrados que desarrollaremos en el punto siguiente.

En el segundo caso, la función de ajuste se obtiene utilizando distribuciones condicionadas. Nosotros emplearemos el método de la regresión a la media, que consiste en asociar a cada valor de la variable independientemente, el valor medio de la distribución de la variable dependiente condicionada a cada uno de dichos valores.

El objeto del ajuste es doble. Por una parte interpolador: puesto que estamos trabajando con un cierto número de observaciones, es de esperar que si éstas son representativas del fenómeno en estudio, los elementos del colectivo se comporten de forma parecida y la función de ajuste sea también válida para ellos. Y por otra parte extrapolador: bajo la suposición de que la relación funcional entre variable dependiente e independiente permanece constante, al menos en un entorno de las observaciones, es posible hacer una

previsión del valor que tomará la variable dependiente para un valor a cort
o medio plazo, volumen de cosecha en función de la lluvia caída, etc.

La elección de la familia de funciones sobre la que haremos el ajuste e uno de los problemas principales con los que habremos de enfrentarnos. E un principio, la observación de la nube de puntos puede darnos una ide de la evolución de los valores de la variable dependiente (a partir de ahor $Y$) en función de los de la independiente ($X$). Otro sistema puede consisti en efectuar un análisis de los datos, viendo las posibles permanencias d éstos. También debemos tener en cuenta las peculiaridades del fenómen estudiado.

Dividiremos esta sección en dos partes, la primera dedicada al ajuste la segunda a la regresión.

## 2.2.2 Ajuste, criterio de los mínimos cuadrados.

Fijada la familia de funciones que utilizaremos para ajustar los valores d una serie estadística bidimensional, ésta dependerá de unos parámetros. E método de los mínimos cuadrados estima parámetros imponiendo la condi ción de hacer mínima la suma de las diferencias al cuadrado entre los valore observados y los correspondientes valores ajustados.

Sean $(x_1,y_1),(x_2,y_2)...(x_N,y_N)$ los valores observados. Supongamos qu utilizamos la función de ajuste $g(x,\alpha,\beta,..,\theta)$. Obtendremos los valores d los parámetros imponiendo la condición de hacer mínima la función:

$$H(\alpha,\beta,...,\theta)=\sum_i[y_i-g(x_i,\alpha,\beta\,...,\theta)]^2$$

Para ello calculamos las parciales de $H$ respecto de cada uno de lo parámetros e igualamos a cero:

$$\frac{\partial H}{\partial\alpha}=0 \quad , \quad \frac{\partial H}{\partial\beta}=0 \quad , \quad \frac{\partial H}{\partial\theta}=0$$

con lo cual obtendremos un sistema de tantas ecuaciones como parámetros llamado sistema de ecuaciones normales, que resuelto nos da los valores d los parámetros. Pasamos a continuación a hacer un estudio más detallad para algunas funciones de uso habitual:

**Caso lineal.**

Sean $(x_1, y_1)(x_2, y_2)...(x_N, y_N)$ los valores observados y $f(x, a, b)$ la recta de ajuste de los valores de $Y$ en función de los de $X$. Obsérvese que no damos las observaciones con sus respectivas frecuencias, sino que lo hacemos por extensión para facilitar la nomenclatura.

$$f(x, a, b) = a + bx$$

obtendremos los valores $a$ y $b$ minimizando la función $G$, dada por:

$$G(a, b) = \sum_i [y_i - (a + bx_i)]^2$$

Derivamos con respecto a los dos parámetros:

$$\frac{\partial G(a, b)}{\partial a} = -2 \sum_i [y_i - (a + bx_i)]$$

$$\frac{\partial G(a, b)}{\partial b} = -2 \sum_i [y_i - (a + bx_i)]x_i$$

igualando a cero, nos queda el siguiente sistema de ecuaciones normales:

$$\left\{ \begin{array}{lcl} \sum_i y_i & = & Na + b \sum_i x_i \\ \sum_i x_i y_i & = & a \sum_i x_i + b \sum_i x_i^2 \end{array} \right.$$

o equivalentemente, dividiendo por $N$, obtenemos las ecuaciones normales expresadas en función de los momentos:

$$\left\{ \begin{array}{lcccc} a_{01} & = & a & + & b \cdot a_{10} \\ a_{11} & = & a \cdot a_{10} & + & b \cdot a_{20} \end{array} \right.$$

Si llamamos...

$$y_i^* = f(x_i) = a + bx_i \qquad \text{y} \qquad e_i = y_i - y_i^*$$

podemos expresar el sistema de ecuaciones normales de la siguiente forma:

$$\left\{ \begin{array}{lcl} \sum_i e_i & = & 0 \\ \sum_i e_i x_i & = & 0 \end{array} \right.$$

donde $e_i$ representa el residuo o error de la observación $i$-ésima. Admitiendo que se verifican las condiciones suficientes de mínimo podemos obtener fácilmente los valores de $a$ y $b$:

$$a = \overline{y} - b\overline{x}$$
$$b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

Si hubiéramos querido obtener la recta de regresión de $X$ con respecto a $Y$, llamando $a'$ y $b'$ a los coeficientes, tenemos:

$$a' = \overline{x} - b'\overline{y}$$
$$b' = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}.$$

Los valores de $b$ y $b'$ que son las pendientes de las rectas de ajuste reciben el nombre de **Coeficientes de regresión.** Hemos definido dos variables, por una parte $Y^*$, que representa a los valores ajustados de la variable $Y$ Esta nueva variable tiene por media y varianza:

$$\overline{y}^* = \overline{y}$$
$$\mathrm{Var}(Y^*) = b \cdot \mathrm{Cov}(X,Y).$$

Y por otra parte la variable residuo $e$, con media y varianza:

$$\overline{e} = 0$$
$$\mathrm{Var}(e) = \frac{1}{N}\sum_i e_i^2 = \frac{1}{N}\sum_i (y_i - y_i^*)^2$$

$\mathrm{Var}(e)$ recibe el nombre de varianza residual, y puede expresarse también de la forma:

$$\mathrm{Var}(e) = \frac{1}{N}[\sum_i y_i^2 - a\sum_i y_i - b\sum_i x_i y_i]$$

Estas dos variables son incorreladas. En efecto, multiplicando la primera ecuación normal por $a$ y la segunda por $b$, tendremos:

$$0 = a\sum_i e_i + b\sum_i e_i x_i = \sum_i e_i(a + bx_i) = \sum e_i y_i^*$$

y puesto que $\overline{e} = 0$, resulta:

$$\mathrm{Cov}(e, Y^*) = \frac{\sum e_i y_i^*}{N} - \overline{e} \cdot \overline{y}^* = 0$$

como queríamos demostrar.

**Propiedades.**

1. Las rectas de regresión de $X$ sobre $Y$ y de $Y$ sobre $X$ se cortan en el centro de gravedad de la distribución $(\bar{X}, \bar{Y})$.
2. $b, b', r$ y $\mathrm{Cov}(X, Y)$ tienen siempre el mismo signo.
3. Las dos rectas de regresión coinciden sólo en el caso de que $r = 1$ ó $r = -1$.
4. Podemos expresar las rectas de regresión de $Y$ sobre $X$ y de $X$ sobre $Y$, respectivamente, como:

$$r_{Y/X} \equiv y - \bar{y} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}(x - \bar{x})$$

$$r_{X/Y} \equiv x - \bar{x} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}(y - \bar{y})$$

Todo lo que hemos dicho hasta ahora es generalizable a funciones linealizables, sin más que hacer los cambios y transformaciones pertinentes. Por ejemplo:

| Función | Transformación | Cambio |
|---|---|---|
| $y = a \cdot x^b$ | $\ln y = \ln a + b \cdot \ln x$ | $y' = \ln y \quad ; \quad x' = \ln x$ |
| $y = a \cdot b^x$ | $\ln y = \ln a + x \cdot \ln b$ | $y' = \ln y \quad ; \quad x' = x$ |
| $y = a + b/x$ | $y = a + b \cdot x^{-1}$ | $y' = y \quad ; \quad x' = 1/x$ |

**Caso parabólico.**

Consideremos la función de ajuste $f(x, a, b, c) = a + bx + cx^2$, parábola de segundo grado, para obtener los valores de $a, b$ y $c$ utilizamos el método de los mínimos cuadrados y minimizamos la función $H$ definida por:

$$H(a, b, c) = \sum_i [y_i - (a + bx_i + cx_i^2)]^2$$

para ello calculamos las derivadas parciales de $H$ respecto a $a, b$ y $c$ e igualamos a cero:

$$\frac{\partial H}{\partial a} = -2 \sum_i [y_i - (a + bx_i + cx_i^2)] = 0$$

$$\frac{\partial H}{\partial b} = -2\sum_i [y_i - (a + bx_i + cx_i^2)]x_i = 0$$

$$\frac{\partial H}{\partial c} = -2\sum_i [y_i - (a + bx_i + cx_i^2)]x_i^2 = 0$$

Despejando los términos independientes, obtenemos el sistema de ecuaciones normales:

$$\begin{cases} \sum_i y_i &= aN + b\sum_i x_i + c\sum_i x_i^2 \\ \sum_i y_i x_i &= a\sum_i x_i + b\sum_i x_i^2 + c\sum_i x_i^3 \\ \sum_i y_i x_i^2 &= a\sum_i x_i^2 + b\sum_i x_i^3 + c\sum_i x_i^4 \end{cases}$$

O equivalentemente, dividiendo por $N$, obtenemos la expresión de las ecuaciones normales en función de los momentos bidimensionales como:

$$\begin{cases} a_{01} &= a &+ b \cdot a_{10} &+ c \cdot a_{20} \\ a_{11} &= a \cdot a_{10} &+ b \cdot a_{20} &+ c \cdot a_{30} \\ a_{21} &= a \cdot a_{20} &+ b \cdot a_{30} &+ c \cdot a_{40} \end{cases}$$

Si igual que antes definimos $y_i^* = a + bx_i + cx_i^2$ y $e_i = y_i - y_i^*$, podemos expresar el sistema de ecuaciones normales como:

$$\begin{cases} \sum_i e_i &= 0 \\ \sum_i e_i x_i &= 0 \\ \sum_i e_i x_i^2 &= 0 \end{cases}$$

La varianza residual vale ahora:

$$\mathrm{Var}(e) = \frac{\sum_i y_i^2 - a\sum_i y_i - b\sum_i x_i y_i - c\sum_i x_i^2 y_i}{N}$$

### 2.2.3 Análisis de la bondad del ajuste.

Una vez que obtenemos los valores de los parámetros, y por lo tanto la función de ajuste, vamos a dar la medida del grado de aproximación entre los valores observados y los ajustados. Según ya hemos visto anteriormente, la varianza del error o varianza residual vale:

$$\text{Var}(e) = \frac{\sum_i (y_i - y_i^*)^2}{N} = \frac{\sum_i e_i^2}{N}$$

que coincide con el error cuadrático medio (E.C.M). Al ser una suma de términos al cuadrado siempre será positivo, salvo que todos sean nulos, en cuyo caso el E.C.M. valdrá cero. Los términos serán nulos si coinciden los valores observados con los ajustados. Por lo tanto el ajuste será perfecto sólo si la varianza residual vale cero.

Si ajustamos los mismos datos a dos modelos diferentes el más adecuado será aquel que tenga un E.C.M. menor.

Relacionaremos a continuación la varianza residual, la varianza de $Y$ y la varianza de $Y^*$. De la relación $e_i = y_i - y_i^*$, que podemos expresar:

$$y_i = y_i^* + e_i$$

sabiendo que las variables $Y^*$ y $e$ son incorreladas, y aplicando que en este caso la varianza de la suma es igual a la suma de las varianzas tenemos:

$$\text{Var}(Y) = \text{Var}(Y^*) + \text{Var}(e).$$

Es decir, la varianza de $Y$ es igual a la varianza de los valores ajustados más la varianza residual. Este importante resultado podemos expresarlo también de la forma:

$$\sum_i (y_i - \overline{y})^2 = \sum_i (y_i^* - \overline{y})^2 + \sum_i (y_i - y_i^*)^2$$

sin más que multiplicar los dos miembros de la igualdad por $N$. Al ser las varianzas números positivos las relaciones siguientes son evidentes:

$$0 \leq \text{Var}(e) \leq \text{Var}(Y)$$

$$0 \leq \text{Var}(Y^*) \leq \text{Var}(Y).$$

Si $\text{Var}(e) = 0$, entonces $y_i = y_i^*$ para todo $i$, con lo que el modelo propuesto explica perfectamente las variaciones de la variable $Y$, siendo inmejorable. Si por el contrario, $\text{Var}(Y^*) = 0$, entonces $y_i^* = \overline{y}$ para todo $i$, con lo que el modelo de ajuste es constante y no explica, en forma alguna,

**Ejemplo.**

Calcularemos la regresión a la media de la siguiente distribución:

| X / Y | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Suma |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 4 | 3 | 7 | 1 | 8 | 9 | 3 | 2 | 39 |
| 2 | 3 | 6 | 1 | 8 | 5 | 9 | 6 | 4 | 3 | 45 |
| 3 | 0 | 0 | 3 | 5 | 7 | 8 | 5 | 4 | 2 | 34 |
| 4 | 0 | 0 | 0 | 2 | 4 | 6 | 7 | 4 | 3 | 26 |
| 5 | 0 | 0 | 0 | 0 | 5 | 8 | 7 | 4 | 4 | 28 |
| 6 | 0 | 0 | 0 | 2 | 8 | 9 | 9 | 6 | 6 | 40 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 8 |
| Suma | 5 | 10 | 7 | 24 | 30 | 48 | 45 | 28 | 23 | 220 |

La "curva" de regresión quedaría:

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\Phi(X)$ | 5,25 | 5,11 | 5,79 | 6,61 | 6,85 | 6,67 | 8,12 |

## 2.4 Análisis de la bondad de la regresión.

Daremos a continuación una medida de la proximidad de la curva de regresión a la distribución. Consideremos el error cuadrático medio, que en este caso viene dado por:

$$\text{E.C.M.} = \frac{\sum_i \sum_j [y_{ij} - \bar{y}_i]^2 \cdot n_{ij}}{N} = \sum_i \text{Var}(Y_i) \cdot f_{i.}$$

donde $\text{Var}(Y_i)$ es la varianza de $Y$ condicionada por $X = x_i$. El E.C.M. viene acompañado de la unidad de la variable con los consiguientes problemas. Se hace necesario, pues, dar una medida abstracta de la bondad del ajuste. La varianza de $Y$ podemos expresarla como:

$$\text{Var}(Y) = \sum_i \text{Var}(Y_i) \cdot f_{i.} + \sum_i (\bar{Y}_i - \bar{Y})^2 \cdot f_{i.}$$

quedando descompuesta en la suma de la media ponderada de las varianzas condicionadas, más la varianza ponderada de las medias condicionadas. La heterogeneidad de $Y$, resulta de la heterogeneidad debida a las distribuciones condicionadas por cada modalidad $x_i$, más la heterogeneidad existente entre todas las modalidades.

Definimos la Razón de correlación de $Y$ con respecto a $X$ como:

$$\eta^2_{y/x} = \frac{\sum_i(\bar{Y}_i - \bar{Y})^2 \cdot f_{i.}}{\mathrm{Var}(Y)} = \frac{[\mathrm{Var}(Y) - \sum_i \mathrm{Var}(Y_i)\cdot f_{i.}]}{\mathrm{Var}(Y)} = 1 - \frac{\sum_i \mathrm{Var}(Y_i)\cdot f_{i.}}{\mathrm{Var}(Y)}.$$

De forma análoga se definiría la razón de correlación de $X$ respecto a $Y$.

En general las dos razones de correlación son diferentes satisfaciendo las ecuaciones:

$$0 \leq (\eta^2_{y/x}; \eta^2_{x/y}) \leq 1.$$

La razón de correlación $\eta^2_{y/x}$ vale 0 si la varianza de las medias condicionadas es nula. Es decir, si todas las medias condicionadas son idénticas, en cuyo caso la curva de regresión es paralela al eje $X$, y decimos que $Y$ está incorrelada con $X$. El recíproco también es cierto, la ausencia de correlación de $Y$ con $X$, implica que la razón de correlación vale cero.

La razón de correlación $\eta^2_{y/x}$ vale 1 cuando la media de las varianzas condicionadas $\mathrm{Var}(Y_i)$ es cero; ahora bien, una suma de términos positivos es nula sólo si todos son nulos, lo cual implica que todas las $\mathrm{Var}(Y_i)$ son cero, es decir: al valor $x_i$ de $X$ le corresponde un único valor de $Y$, y por consiguiente, $Y$ depende funcionalmente de $X$. Recíprocamente, la dependencia funcional de $Y$ respecto a $X$, implica que la razón de correlación vale 1.

Recogemos los distintos casos que se pueden presentar en la siguiente tabla:

| | $\eta^2_{x/y} = 0$ | $0 < \eta^2_{x/y} < 1$ | $\eta^2_{x/y} = 1$ |
|---|---|---|---|
| $\eta^2_{y/x} = 0$ | Ausencia recíproca de correlación | Caso general de ausencia de correlación de $Y$ respecto a $X$ | Dependencia funcional ($Y \longrightarrow X$) Ausencia de correlación de $Y$ respecto a $X$ |
| $0 < \eta^2_{y/x} < 1$ | Caso general de ausencia de correlación de $X$ respecto a $Y$ | Caso general | Caso general de dependencia funcional no recíproca ($Y \longrightarrow X$) |
| $\eta^2_{y/x} = 1$ | Dependencia funcional ($Y \longrightarrow X$) Ausencia de correlación de $X$ respecto a $Y$ | Caso general de dependencia funcional no recíproca ($X \longrightarrow Y$) | Dependencia funcional recíproca |

*Tabla II*

De igual forma que hacíamos con el coeficiente de determinación, podemos multiplicar la razón de correlación por 100 y obtendríamos, en forma de porcentaje, la parte de la variación de $Y$ que queda explicada por la curva de regresión, que sería:

$$\frac{[\sum_i (\bar{Y}_i - \bar{Y})^2 \cdot f_{i.}]}{\mathrm{Var}(Y)} \cdot 100\%$$

el resto hasta 100, es decir:

$$\frac{\mathrm{Var}(e)}{\mathrm{Var}(Y)} \cdot 100\%$$

quedaría sin explicar por ella.

**Proposición.**

La curva de regresión es la función óptima para el criterio de los mínimos cuadrados.

De este resultado se puede deducir que el E.C.M. de la regresión de la media es menor o igual que el E.C.M. del ajuste mínimo cuadrático, y como consecuencia, que el coeficiente de determinación, es siempre menor o igual que la razón de correlación, sea cual sea la función elegida. La comparación entre $R^2$ y $\eta^2_{y/x}$, nos puede indicar si existe una función mejor para ajustar los datos.

$$R^2 \leq \eta^2_{y/x}.$$

**Notas.**

1. Decíamos antes que la regresión a la media tenía sentido cuando contábamos con un amplio número de observaciones de la variable dependiente para un valor de la independiente. Sin embargo, en ciertos casos podemos suavizar esta condición. Así, si la variable $X$ es continua, aún con un alto número de observaciones, es de esperar que haya pocos valores que se repitan, pero siempre podremos agrupar los datos en intervalos; éstos deberán ser lo suficientemente pequeños como para

que los elementos pertenecientes a un mismo intervalo puedan considerarse, a los efectos pertinentes, iguales, y lo bastante grandes como para que caiga un número mínimo de observaciones en la mayoría de las clases.

2. Cuando la nube de puntos no nos proporcione una idea de la clase funcional que debemos elegir para realizar el ajuste, una posible solución puede ser el calcular la línea de regresión y representarla gráficamente.

3. Si hemos utilizado una función para efectuar el ajuste, calcular la previsión para un valor de la variable independiente se reduce a sustituir dicho valor en la función. Si debemos apoyarnos en la línea de regresión, la previsión para una valor $x$ de $X$, sería el valor correspondiente a la clase a la que pertenece $x$. Esto hace que la previsión tenga dos matices diferentes: mientras que si utilizamos la línea de regresión, sólo podemos hacer previsiones para valores encuadrados en alguna de las clases predefinidas -en realidad sería una interpolación-; cuando nos basamos en una función de ajuste, no sólo podemos hacer estimaciones para valores intermedios, sino que podemos extrapolar y sacar conclusiones para valores exteriores; aunque, en este caso, la fiabilidad de la previsión dependerá de que las condiciones en que se realizó el ajuste, permanezcan constantes, lo que ocurrirá, normalmente, en un entorno de los puntos utilizados para realizarlo.

## 2.5 Análisis de atributos.

Cuando queremos estudiar el grado de asociación de dos variables $X$ e $Y$, siendo una de ellas, o las dos, cualitativas. Vamos a estudiar a continuación algunos coeficientes de correlación.

### Correlación biserial.

Se aplica cuando se trata de estudiar la correlación entre las dos componentes de una variable aleatoria $(X, Y)$, y es plausible considerar que su distribución conjunta es binormal. Se usa cuando una variable $Y$ es cuantitativa, mientras que la segunda variable $X$ es dicotómica, esto es, definida por dos grupos. El coeficiente se define como:

$$r_b = \frac{\bar{Y}_2 - \bar{Y}_1}{\sigma_y} \cdot \frac{pq}{z}$$

donde:

$\bar{Y}_1$ = valor medio de $Y$ condicionado al menor valor de $X$.

$\bar{Y}_2$ = valor medio de $Y$ condicionado al mayor valor de $X$.

$p$ = proporción de observaciones en la categoría mayor.

$q$ = proporción de observaciones en la categoría menor.

$\sigma_y$ = desviación típica de $Y$.

$z$ = ordenada de la función de densidad de la $N(0,1)$ correspondiente a un valor $c_0$ tal que $F(c_0) = p$.

### Correlación tetracórica.

Se denota por $r_t$, y se usa para medir el grado de asociación lineal entre dos variables $X$ e $Y$, donde ambas son dicotómicas y las verdaderas distribuciones se suponen normales. Esto es, se usa para tablas de contingencia (2x2).

$$r_t = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{\sqrt{n_{1.} \cdot n_{2.} - n_{.1} \cdot n_{.2}}}$$

### Coeficiente de contingencia.

Para estudiar el grado de asociación entre dos variables $X$ e $Y$ cuando los datos están en una tabla de contingencia r $x$ c, se define el coeficiente de contingencia:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

donde

$$\chi^2 = \sum_{i,j} \frac{n_{ij}}{n^*_{ij}} - N \qquad \text{con} \qquad n^*_{ij} = \frac{n_{i.} \cdot n_{.j}}{N}$$

Cuanto más próximo es $C^2$ a 1, mayor es el grado de correlación. Para $h = k$ (tablas de contingencia cuadradas) el valor máximo de $C^2$ es $\frac{k-1}{k}$.

## Coeficiente de correlación ordinal.

Sirven para medir el grado de acuerdo entre dos o más ordenaciones de un conjunto de $n$ objetos.

Para dos ordenaciones el más utilizado es el de Spearman:

$$r_s = 1 - \frac{6\left(\sum_{i=1}^{n} d_i^2\right)}{n(n^2-1)}$$

donde $d_i$ es la diferencia de puestos obtenidos por el objeto $i$-ésimo en las dos ordenaciones consideradas.

Si tenemos que $r_s = 1$, esto supone un acuerdo total, y si $r_s = -1$ desacuerdo pleno.

Para $k$ ordenaciones se utiliza el llamado coeficiente de concordancia

$$r_c = 3 \cdot \frac{4\left(\sum_{i=1}^{n} r_i^2\right) - n \cdot k^2(n+1)^2}{nk^2 \cdot (n^2-1)}$$

donde $r_i$ es la suma de los rangos alcanzados por el objeto $i$-ésimo en las $k$ ordenaciones.

# Problemas Resueltos

## Problema 2.1

**Se somete a una serie de alumnos a un test. Se mide el tiempo que tardaron en realizarlo (de 11 a 15 minutos) y la nota obtenida (de 1 a 5) obteniéndose los siguientes resultados:**

| | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 2 | 1 |
| 2 | 9 | 10 | 11 | 12 | 13 |
| 3 | 11 | 13 | 15 | 14 | 12 |
| 4 | 13 | 19 | 17 | 16 | 9 |
| 5 | 15 | 17 | 18 | 16 | 10 |

**Se pide:**

1. **Representar los datos en un estereograma.**

2. **Hallar la nota media obtenida.**

3. **Hallar el tiempo medio empleado por los alumnos que sacaron la máxima puntuación.**

---

1 El estereograma puede verse en la figura 2.22.

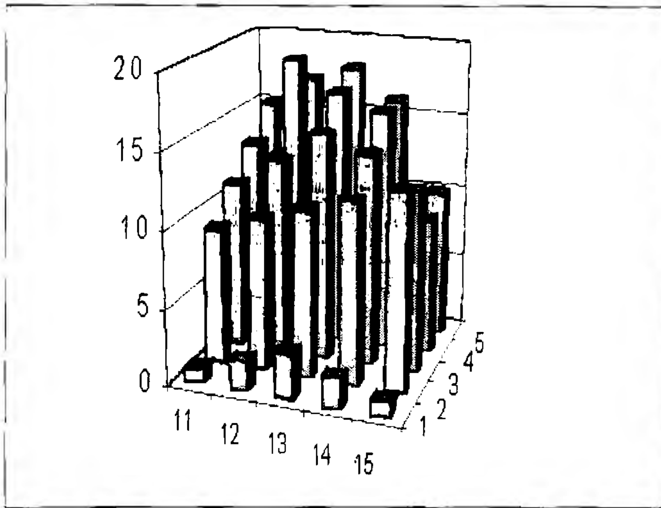



Figura 2.22: Estereograma Nota / Tiempo

2. Para hallar la nota media habrá que calcular la distribución marginal:

| | 11 | 12 | 13 | 14 | 15 | |
|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 2 | 1 | 9 |
| 2 | 9 | 10 | 11 | 12 | 13 | 55 |
| 3 | 11 | 13 | 15 | 14 | 12 | 65 |
| 4 | 13 | 19 | 17 | 16 | 9 | 74 |
| 5 | 15 | 17 | 18 | 16 | 10 | 76 |

El calculo de la nota media sería:

$$\overline{X} = \frac{1 \cdot 9 + 2 \cdot 55 + 3 \cdot 65 + 4 \cdot 74 + 5 \cdot 76}{279} = 3{,}548387097.$$

3. Para hallar el tiempo medio empleado por los alumnos que sacaron la máxima puntuación, habrá que usar la distribución condicionada.

| | 11 | 12 | 13 | 14 | 15 | |
|---|---|---|---|---|---|---|
| 5 | 15 | 17 | 18 | 16 | 10 | 76 |

El calculo del tiempo medio sería:

$$E[Y/X = 5] = \frac{11 \cdot 15 + 12 \cdot 17 + 13 \cdot 18 + 14 \cdot 16 + 15 \cdot 10}{76} = 12{,}855.$$

## Problema 2.2

**Para la distribución bidimensional que se da a continuación, calcular los siguientes momentos** $a_{10}$, $a_{01}$, $a_{20}$, $a_{02}$, $a_{11}$, $a_{21}$ **y** $a_{12}$.

| $x_i$ | $y_i$ | $n_i$ |
|---|---|---|
| 0 | 0 | 14 |
| 0 | 1 | 2 |
| 0 | 2 | 15 |
| 1 | 0 | 1 |
| 1 | 1 | 8 |
| 1 | 2 | 2 |
| 2 | 0 | 4 |
| 2 | 1 | 1 |
| 2 | 2 | 3 |

Construyamos la tabla:

| $x_i$ | $y_i$ | $n_i$ | $x_in_i$ | $y_in_i$ | $x_i^2n_i$ | $y_i^2n_i$ | $x_iy_in_i$ | $x_i^2y_in_i$ | $x_iy_i^2n_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 2 | 0 | 2 | 0 | 2 | 0 | 0 | 0 |
| 0 | 2 | 15 | 0 | 30 | 0 | 60 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 1 | 2 | 2 | 2 | 4 | 2 | 8 | 4 | 4 | 8 |
| 2 | 0 | 4 | 8 | 0 | 16 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 2 | 1 | 4 | 1 | 2 | 4 | 2 |
| 2 | 2 | 3 | 6 | 6 | 12 | 12 | 12 | 24 | 24 |
| | | 50 | 27 | 51 | 43 | 91 | 26 | 40 | 42 |

de donde

$$a_{10} = \frac{27}{50} = 0{,}54 \qquad a_{01} = \frac{51}{50} = 1{,}02$$

$$a_{20} = \frac{43}{50} = 0{,}86 \qquad a_{02} = \frac{91}{50} = 1{,}82$$

$$a_{21} = \frac{40}{50} = 0{,}8 \qquad a_{12} = \frac{42}{50} = 0{,}84$$

$$a_{11} = \frac{26}{50} = 0{,}52.$$

# Problema 2.3

**Se hace un estudio entre el precio de unos determinados componentes (X) y el tiempo transcurrido hasta la primera avería (Y) en meses. Obteniéndose los siguientes resultados:**

| **Y\ X** | **8000** | **20000** | **25000** | **30000** | **35000** |
|---|---|---|---|---|---|
| **[0,8]** | **30** | **2** | **0** | **0** | **0** |
| **(8,12]** | **0** | **23** | **1** | **0** | **0** |
| **(12,14]** | **0** | **2** | **8** | **0** | **0** |
| **(14,16]** | **0** | **0** | **2** | **4** | **0** |
| **(16,20]** | **0** | **0** | **0** | **1** | **3** |

**Se pide:**

1. **Obtener e interpretar los coeficientes de centralización, dispersión, simetría y forma del tiempo transcurrido hasta la primera avería.**

2. **Elegir la mejor clase funcional para ajustar estos datos.**

3. **Realizar una predicción del tiempo transcurrido hasta la primera avería para un componente cuyo precio es de 40000. Analizar la bondad del ajuste.**

(a)

| Y | $y_i$ | $n_i$ | $N_i$ | $h_i$ | $y_i \cdot n_i$ | $y_i^2 \cdot n_i$ | $y_i^3 \cdot n_i$ | $y_i^4 \cdot n_i$ |
|---|---|---|---|---|---|---|---|---|
| [0,8] | 4 | 32 | 32 | 4 | 128 | 512 | 2048 | 8192 |
| (8,12] | 10 | 24 | 56 | 6 | 240 | 2400 | 24000 | 240000 |
| (12,14] | 13 | 10 | 66 | 5 | 130 | 1690 | 21970 | 285610 |
| (14,16] | 15 | 6 | 72 | 3 | 90 | 1350 | 20250 | 303750 |
| (16,20] | 18 | 4 | 76 | 1 | 72 | 1296 | 23328 | 419904 |
| | | 76 | | | 660 | 7248 | 91596 | 1257456 |

$$a_1 = \frac{660}{76} = 8{,}68421 \quad , \quad a_2 = \frac{7248}{76} = 95{,}36842$$

$$a_3 = \frac{91596}{76} = 1205{,}21053 \quad , \quad a_4 = \frac{1257456}{76} = 16545{,}47368.$$

$$m_2 = a_2 - a_1^2 = 19{,}95290859 \quad , \quad m_3 = a_3 - 3a_1a_2 + 2a_1^3 = 30{,}46056$$

$$m_4 = a_4 - 4a_1a_3 + 6a_1^2a_2 - 3a_1^4 = 771{,}3174469.$$

**Coeficientes de centralización**

Media Aritmética $\bar{Y} = a_1 = 8{,}68421$ meses

Mediana $Me = 8 + \frac{38-32}{24} \cdot 4 = 8+1 = 9$ meses

Moda $Mo = 8 + \frac{5}{5+4} \cdot 4 = 8+2{,}2222=10{,}2222$ meses

Media Geométrica $G = \sqrt[76]{4^{32} \cdot 10^{24} \cdot 13^{10} \cdot 15^{6} \cdot 18^{4}} = 7{,}495$ meses

Media Armónica $H = \frac{76}{11{,}7915} = 6{,}445$ meses

Media Cuadrática $C = 9{,}76568$ meses

Cuartiles $Q_1$ =4,75 meses , $Q_2 = 9$ meses , $Q_3 = 12{,}222$ meses

Deciles $D_1$ =1,89 , $D_2 = 3{,}79$ , $D_3 = 5{,}76$ , $D_4 = 7{,}59$ , $D_5 = 9$

$D_6$ =10,2667 , $D_7$ =11,5333 , $D_8$ =12,96 , $D_9 = 14{,}8$

Percentiles $P_1$ =1,18 $\ldots P_{99} = 19{,}23$.

**Coeficientes de dispersión**

Desviación respecto a la media $D_{\bar{Y}} = 3{,}9446$ meses

Desviación respecto a la mediana $D_{Me} = 3{,}89474$ meses

Desviación cuartílica $D_Q = 3{,}725$ meses

Desviación percentílica $D_P = 6{,}45$ meses

Rango o recorrido $Re = 20 - 0 = 20$ meses

Varianza $Var(Y) = m_2 = 19{,}95290859$ meses$^2$

Desviación típica $\sigma_Y = \sqrt{Var(Y)} = 4{,}46687$ meses

Coeficiente de variación de Pearson $Cv = 0{,}514367 \cdot 100\% = 51{,}4367\%$.

### Coeficientes de simetría

Coeficiente de simetría de Pearson $A_p = \frac{\bar{Y}-Mo}{\sigma_Y} = -0{,}3438 < 0$

Ligeramente sesgada a la izquierda.

Coeficiente de simetría de Fisher $g_1 = \frac{m_3}{(\sigma_Y)^3} = 0{,}341765 > 0$

Ligeramente sesgada a la derecha.

La diferencia entre ambas medidas de simetría viene motivada por el hecho de que el signo del coeficiente de Pearson depende solamente de la media y la moda, mientras que el de Fisher depende de todos los valores de la distribución. Además, el coeficiente de Pearson sólo es útil si la distribución es unimodal y ligeramente acampanada.

### Coeficiente de curtosis

$g_2 = \frac{m_4}{(\sigma_Y)^4} - 3 = -1{,}06259 < 0.$

Distribución platicúrtica. (Más aplastada que la normal.)

(b) Para ajustar $Y$ en función de $X$, calculamos la tabla:

| $x_i$ | $y_i$ | $n_i$ | $x_in_i$ | $y_in_i$ | $x_i^2n_i$ | $y_i^2n_i$ | $x_iy_in_i$ | $x_i^2y_in_i$ |
|---|---|---|---|---|---|---|---|---|
| 8000 | 4 | 30 | 240000 | 120 | $1{,}92 \cdot 10^{9}$ | 480 | $9{,}6 \cdot 10^5$ | $7{,}68 \cdot 10^{9}$ |
| 20000 | 4 | 2 | 40000 | 8 | $8 \cdot 10^8$ | 32 | $1{,}6 \cdot 10^5$ | $3{,}2 \cdot 10^9$ |
| 20000 | 10 | 23 | 460000 | 230 | $9{,}2 \cdot 10^9$ | 2300 | $4{,}6 \cdot 10^6$ | $9{,}2 \cdot 10^{10}$ |
| 20000 | 13 | 2 | 40000 | 26 | $8 \cdot 10^8$ | 338 | $5{,}2 \cdot 10^5$ | $1{,}04 \cdot 10^{10}$ |
| 25000 | 10 | 1 | 25000 | 10 | $6{,}25 \cdot 10^8$ | 100 | $2{,}5 \cdot 10^5$ | $6{,}25 \cdot 10^9$ |
| 25000 | 13 | 8 | 200000 | 104 | $5 \cdot 10^9$ | 1352 | $2{,}6 \cdot 10^6$ | $6{,}5 \cdot 10^{10}$ |
| 25000 | 15 | 2 | 50000 | 30 | $1{,}25 \cdot 10^9$ | 450 | $7{,}5 \cdot 10^5$ | $1{,}88 \cdot 10^{10}$ |
| 30000 | 15 | 4 | 120000 | 60 | $3{,}6 \cdot 10^9$ | 900 | $1{,}8 \cdot 10^6$ | $5{,}4 \cdot 10^{10}$ |
| 30000 | 18 | 1 | 30000 | 18 | $9 \cdot 10^8$ | 324 | $5{,}4 \cdot 10^5$ | $1{,}62 \cdot 10^{10}$ |
| 35000 | 18 | 3 | 105000 | 54 | $3{,}68 \cdot 10^9$ | 972 | $1{,}9 \cdot 10^6$ | $6{,}62 \cdot 10^{10}$ |
| | | 76 | 1310000 | 660 | $2{,}78 \cdot 10^{10}$ | 7248 | $1{,}4 \cdot 10^{7}$ | $3{,}4 \cdot 10^{11}$ |

$$\bar{X} = \frac{1310000}{76} = 17236{,}84211$$

$$Var(X) = 68286010{,}87 \quad , \quad \sigma_X = 8263{,}535$$

$$\bar{Y} = \frac{660}{76} = 8{,}68421$$

$$Var(Y) = 19{,}95290859 \quad , \quad \sigma_Y = 4{,}466867.$$

$$Cov(X,Y) = \frac{14070000}{76} - (17236{,}84211)(8{,}68421) = 35443{,}2133.$$

La recta de regresión de Y sobre X es:

$$r_{y/x} \equiv (Y - \bar{Y}) = \frac{Cov(X,Y)}{Var(X)}(X - \bar{X})$$
$$r_{y/x} \equiv Y = 0{,}00051904 \cdot X - 0{,}262410623.$$

El coeficiente de correlación lineal es:

$$r = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = 0{,}9602055.$$

Y la relación lineal es directa y aceptable.

El coeficiente de determinación es $R^2 = r^2$ =0,92199; con lo que el ajuste es aceptable.

(c) Para hacer la previsión del tiempo transcurrido hasta la primera avería para un componente que cueste 40000 tendremos:

$Y^*_{40000} = 0{,}00051904 \cdot 40000 - 0{,}262410623$ =20,499189 meses

Otros ajustes son peores:

| | |
|---|---|
| Regresión Exponencial | $Y = a \cdot b^X$ |
| | $Y = 2{,}484828 \cdot 1{,}000064053^X$ |
| | r = 0,949727 |
| Regresión Logarítmica | $Y = a + b \ln X$ |
| | $Y = -67{,}0224 + 7{,}8673745 \ln X$ |
| | r = 0,937036 |
| Regresión Potencial | $Y = a \cdot X^b$ |
| | $Y = (4{,}83361 \quad 10^{-4}) \cdot X^{1{,}002716}$ |
| | r = 0,95722. |

Y el ajuste parabólico no es significativamente mejor que el lineal:

Regresión Parabólica

$$Y = 2{,}46577 \cdot 10^{-9} X^2 + 4{,}29625 \cdot 10^{-4} X + 0{,}377844$$

$$R^2 = 1 - \frac{Var(e)}{Var(Y)} = 1 - \frac{1{,}530917696}{19{,}95290859} = 0{,}92327 \quad \text{pues}$$

$$Var(e) = \frac{\sum y_i^2 - a_0 \sum y_i - a_1 \sum x_i y_i - a_2 \sum x_i^2 y_i}{N}$$

de donde

$$Var(e) = \frac{116{,}3497449}{76} = 1{,}530917696.$$

## Problema 2.4

**Dos variables $X$ e $Y$ son independientes con $\overline{Y} = 2$. Calcular la recta de regresión de $Y$ sobre $X$. ¿Cómo será la recta de regresión de $X$ sobre $Y$?**

Al ser independientes, la covarianza es nula. Por tanto el coeficiente $b$ de la recta de regresión valdrá 0. Por tanto la recta será:

$$y - \overline{y} = b(x - \overline{x}) = 0 \Rightarrow y = \overline{y} = 2.$$

La recta de regresión de $Y$ sobre $X$ será por tanto $y = 2$.

Análogamente, la recta de regresión de $X$ sobre $Y$ será del tipo $x = \overline{x}$.

## Problema 2.5

**Las ecuaciones**

$$\begin{cases} 2x + y - 3 = 0 \\ x + 3y - 4 = 0 \end{cases}$$

**representan, al realizar un estudio estadístico bidimensional, las rectas de regresión. Hallar la media de $x$ y de $y$ y el coeficiente de determinación. ¿Cuál de las dos rectas es la de $y$ sobre $x$ ?**

Para hallar la media de $x$ y la media de $y$ tenemos que hallar el punto de corte de ambas rectas. En este caso el punto de corte es el (1,1). Por tanto $\overline{x} = 1$ y $\overline{y} = 1$.

Si fuese la primera la recta de regresión de $y$ sobre $x$ tendríamos

$$y = -2x + 3$$

por tanto $b = -2$. Si la segunda fuese la de $x$ sobre $y$ entonces:

$$x = -3y + 4$$

con lo que $b' = -3$. Entonces el coeficiente de determinación sería

$$R^2 = bb' = (-2)(-3) = 6$$

lo cual no puede ocurrir, ya que el coeficiente de determinación no puede ser mayor que la unidad.

En cambio, si fuese la segunda la recta de regresión de $y$ sobre $x$ tendríamos

$$y = (-1/3)x + 4/3$$

por tanto $b = (-1/3)$. Si la primera fuese la de $x$ sobre $y$ entonces:

$$x = (-1/2)y + 3/2$$

con lo que $b' = (-1/2)$. Entonces el coeficiente de determinación sería

$$R^2 = bb' = (-1/2)(-1/3) = 1/6$$

lo cual sí es posible.

# Problema 2.6

**La recta de regresión de $y$ sobre $x$ es $y = 2x + 2$ siendo $\overline{x} = 3$ y $r^2 = 0{,}8$. Hallar la recta de regresión de $x$ sobre $y$.**

Como las dos rectas de regresión pasan por el punto $(\overline{x}, \overline{y})$ hallemos en primer lugar $\overline{y}$.

$$\overline{y} = 2(3) + 2 = 8$$

Calculemos a continuación la pendiente de la recta de regresión de $x$ sobre $y$. Como sabemos que $bb' = r^2$ entonces

$$b' = r^2/b = \frac{0{,}8}{2} = 0{,}4.$$

Por tanto la recta de regresión de $x$ sobre $y$ será aquella que pase por el punto (3,8) y tenga por pendiente 0,4. Esto es:

$$x - 3 = 0{,}4(y - 8)$$

que equivale a

$$x = 0{,}4y - 0{,}2.$$

# Problema 2.7

**La recta de regresión de $y$ sobre $x$ es $y = 2x + 2$ siendo $\sigma_x^2 = 4$. Hallar el valor mínimo que puede tomar $\sigma_y^2$.**

---

Como sabemos que

$$b = \frac{cov(x,y)}{\sigma_x^2}$$

obtenemos que

$$cov(x,y) = b \cdot \sigma_x^2 = 2(4) = 8.$$

Usemos ahora la relación

$$R^2 = \frac{cov^2(x,y)}{\sigma_x^2 \sigma_y^2} \leq 1$$

de donde

$$\sigma_y^2 \geq \frac{cov^2(x,y)}{\sigma_x^2} = \frac{64}{4} = 16.$$

Por tanto, el mínimo valor que puede tomar $\sigma_y^2$ es 16.

# Problema 2.8

**La recta de regresión de $y$ sobre $x$ es $y = 2x+2$ siendo $\overline{x} = 3$, $\sigma_x^2 = 4$ y $\sigma_y^2 = 25$. Hallar $\overline{y}$, $Cov(x,y)$, el coeficiente $r^2$ y la recta de regresión de $x$ sobre $y$.**

---

Como las dos rectas de regresión pasan por el punto $(\overline{x}, \overline{y})$ hallemos en primer lugar $\overline{y}$.

$$\overline{y} = 2(3) + 2 = 8.$$

Como sabemos que

$$b = \frac{cov(x,y)}{\sigma_x^2}$$

obtenemos que

$$cov(x,y) = b \cdot \sigma_x^2 = 2(4) = 8.$$

Calculemos $r^2$

$$r^2 = \frac{cov^2(x,y)}{\sigma_x^2\sigma_y^2} = \frac{64}{100} = 0{,}64.$$

El valor de $b'$ será

$$b' = \frac{cov(x,y)}{\sigma_y^2} = \frac{8}{25} = 0{,}32.$$

La recta de regresión de $x$ sobre $y$ será aquella que pase por el punto (3,8) y tenga por pendiente 0,32. Esto es:

$$x - 3 = 0{,}32(y - 8)$$

que equivale a

$$x = 0{,}32y + 0{,}44.$$

## Problema 2.9

**Se realiza un estudio para comprobar la relación entre la longitud de la mano y la longitud del brazo. Se realizan las mediciones en centímetros, y se obtiene una covarianza de 12. Si los datos se hubieran medido en milímetros, ¿qué covarianza se hubiera obtenido?**

Llamemos $X$ a la variable que mide la longitud de la mano en centímetros y llamemos $Y$ a la variable que mide la longitud del brazo en centímetros. Si las mediciones las realizamos en milímetros en lugar de en centímetros tendremos nuevas variables. Llamemos $X'$ a la variable que mide la longitud de la mano en milímetros y llamemos $Y'$ a la variable que mide la longitud del brazo en milímetros. Evidentemente se cumple que

$$\begin{array}{ll} X' = 10 \cdot X & \overline{X'} = 10 \cdot \overline{X} \\ Y' = 10 \cdot Y & \overline{Y'} = 10 \cdot \overline{Y} \end{array}$$

Como sabemos la Covarianza viene dada por la expresión:

$$\text{Cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \overline{X}) \cdot (y_j - \overline{Y}) \cdot n_{ij}$$

La covarianza entre $X'$ e $Y'$ será:

$$\text{Cov}(X',Y') = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{s}({x'}_i - \overline{X'}) \cdot ({y'}_j - \overline{Y'}) \cdot n_{ij}$$

sustituyendo:

$$\mathrm{Cov}(X',Y') = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{s}(10x_i - 10\overline{X})\cdot(10y_j - 10\overline{Y})\ \ n_{ij}$$

con lo que

$$\mathrm{Cov}(X',Y') = 100\cdot\mathrm{Cov}(X,Y) = 100\cdot 12 = 1200.$$

El hecho de ser una mayor que otra no significa que usando unas unidades se obtenga un mejor ajuste lineal que usando otra, ya que el ajuste será bueno o malo en función del valor del coeficiente de correlación lineal que se obtendrá al dividir la covarianza por el producto de las desviaciones típicas.

# Problema 2.10

**Un investigador introduce 5 mg de un compuesto orgánico en una suspensión y se dedica a medir el peso del compuesto según vayan pasando los días, obteniendo los siguientes resultados:**

| **días** | **peso** |
|---|---|
| **0** | **5,00** |
| **1** | **5,56** |
| **2** | **6,25** |
| **3** | **7,14** |
| **4** | **8,33** |
| **5** | **10,00** |
| **6** | **12,50** |
| **7** | **16,67** |
| **8** | **25,00** |
| **9** | **50,00** |

**Realizar un ajuste lineal de estos datos. Realizar predicciones de lo que ocurrirá en los siguientes días.**

---

Si llamamos $x$ a la variable número de días transcurridos, e $y$ al peso del compuesto; realizando los distintos cálculos obtenemos:

| | |
|---|---|
| media de $x$ | 4,5 |
| media de $y$ | 14,65 |
| $\sigma_x$ | 2,872281323 |
| $\sigma_y$ | 13,15189435 |
| $\mathrm{cov}(x,y)$ | 30,5465 |

de donde obtenemos que la recta de regresión $y = a + bx$, siendo:

| | |
|---|---|
| b | 3,702606061 |
| a | -2,016727273 |
| $R^2$ | 0,653871398 |

quedaría $y = -2,016727 + 3,702606x$; y las previsiones que realizaríamos serían:

| | |
|---|---|
| prev 10 | 35,00933333 |
| prev 11 | 38,71193939 |
| prev 12 | 42,41454545 |

Debemos decir, no obstante, que las previsiones no son demasiado buenas, ya que el coeficiente de determinación tiene un valor menor de 0,75.

# Problema 2.11

**Repetir el ejercicio anterior utilizando un ajuste exponencial del tipo $y = Ae^{Bx}$.**

Queremos realizar un ajuste del tipo

$$y = Ae^{Bx}$$

tomando logaritmos convertimos este ajuste en uno del tipo lineal:

$$y' = \log(y) = \log(A) + Bx$$

Con lo cual usaríamos la siguiente tabla:

| $x_i$ | $y'_i$ |
|---|---|
| 0 | 1,61 |
| 1 | 1,72 |
| 2 | 1,83 |
| 3 | 1,97 |
| 4 | 2,12 |
| 5 | 2,30 |
| 6 | 2,53 |
| 7 | 2,81 |
| 8 | 3,22 |
| 9 | 3,91 |

Realizando los distintos cálculos obtenemos:

| | |
|---|---|
| media de $x$ | 4,5 |
| media de $y'$ | 2,40 |
| $\sigma_x$ | 2,872281323 |
| $\sigma_{y'}$ | 0,695381674 |
| $\text{cov}(x, y')$ | 1,900706263 |

de donde obtenemos que la recta de regresión de $y'$ sobre $x$ es $y' = a + bx$: siendo $A = e^a$ y $B = b$:

| | |
|---|---|
| $b$ | 0,230388638 |
| $B$ | 0,230388638 |
| $a$ | 1,364852827 |
| $A$ | 2,623942072 |
| $R^2$ | 0,90558575 |

y el ajuste es $y = 2{,}623942072 \cdot e^{0{,}230388638 \cdot x}$; donde las previsiones que realizaríamos serían:

| | |
|---|---|
| prev 10 | 39,20244851 |
| prev 11 | 49,35938128 |
| prev 12 | 62,1478661 |

Estas previsiones son mejores que las obtenidas en el ejercicio anterior pues el coeficiente de determinación está próximo a 1.

# Problema 2.12

**Repetir el ejercicio anterior utilizando un ajuste hiperbólico.**

Queremos realizar un ajuste del tipo

$$y = \frac{1}{A + Bx}$$

invirtiendo los términos convertimos este ajuste en uno del tipo lineal:

$$y' = \frac{1}{y} = A + Bx$$

Con lo cual usaríamos la siguiente tabla:

| $x_i$ | $y'_i$ |
|---|---|
| 0 | 0,20 |
| 1 | 0,18 |
| 2 | 0,16 |
| 3 | 0,14 |
| 4 | 0,12 |
| 5 | 0,10 |
| 6 | 0,08 |
| 7 | 0,06 |
| 8 | 0,04 |
| 9 | 0,02 |

Realizando los distintos cálculos obtenemos:

| | |
|---|---|
| media de $x$ | 4,5 |
| media de $y'$ | 0,11 |
| $\sigma_x$ | 2,872281323 |
| $\sigma_{y'}$ | 0,057432921 |
| cov$(x, y')$ | -0,164963444 |

de donde obtenemos que la recta de regresión de $y'$ so bre $x$ es $y = A + B \cdot x$:

| | |
|---|---|
| $B$ | -0,019995569 |
| $A$ | 0,199974876 |
| $R^2$ | 0,99999926 |

y el ajuste es $y = 1/(0{,}199974876 - 0{,}19995569 \cdot x)$; y las previsiones que realizaríamos serían:

| | |
|---|---|
| prev 10 | 52119,78281 |
| prev 11 | -50,05911381 |
| prev 12 | -25,0175427 |

Obsérvese que ahora el valor de $R^2$ es prácticamente igual a 1. Sin embargo las previsiones que se obtienen son o muy grandes o incluso negativas, lo cual no es lógico. Esta situación se produce porque la asíntota de la hipérbola está en un valor situado entre 10 y 11. La primera previsión es muy grande al estar este valor muy próximo a la asíntota. Los siguientes son negativos al encontrarnos ya en la otra rama de la hipérbola. En estos casos extremos las previsiones dadas por el ajuste hiperbólico no son muy buenas. Son mejores las obtenidas en el ejercicio anterior.

## Problema 2.13

**Dadas la siguiente distribución bidimensional:**

| **X/Y** | **100** | **200** | **300** |
|---|---|---|---|
| **125** | **12** | **30** | **18** |
| **220** | **8** | **20** | **12** |

**Estudiar la independencia entre estas variables. Calcular la covarianza.**

---

Calculando la tabla de frecuencias relativas, obtenemos:

| $X/Y$ | 100 | 200 | 300 | |
|---|---|---|---|---|
| 125 | 0,12 | 0,30 | 0,18 | 0,6 |
| 220 | 0,08 | 0,20 | 0,12 | 0,4 |
| | 0,2 | 0,5 | 0,3 | 1 |

Y si comprobamos que la frecuencia relativa de la distribución conjunta coincide siempre con el producto de las frecuencias relativas marginales. Esto es:

$$\begin{aligned}
f_{1,1} &= 0{,}12 = 0{,}6 \cdot 0{,}2 = f_{1\cdot} \cdot f_{\cdot 1} \\
f_{1,2} &= 0{,}30 = 0{,}6 \cdot 0{,}5 = f_{1\cdot} \cdot f_{\cdot 2} \\
f_{1,3} &= 0{,}18 = 0{,}6 \cdot 0{,}3 = f_{1\cdot} \cdot f_{\cdot 3} \\
f_{2,1} &= 0{,}08 = 0{,}4 \cdot 0{,}2 = f_{2\cdot} \cdot f_{\cdot 1} \\
f_{2,2} &= 0{,}20 = 0{,}4 \cdot 0{,}5 = f_{2\cdot} \cdot f_{\cdot 2} \\
f_{2,3} &= 0{,}12 = 0{,}4 \cdot 0{,}3 = f_{2\cdot} \cdot f_{\cdot 3}
\end{aligned}$$

y por tanto las variables $X$ e $Y$ son independientes.

La covarianza entre las dos variables es nula, ya que son independientes y la independencia implica incorrelación lineal. El recíproco no es cierto, como veremos en el problema siguiente.

## Problema 2.14

**Se plantea a un grupo de 20 alumnos dos cuestiones tipo verdadero/falso. Se les califica segun los siguientes criterios: 1 si la respuesta es correcta, -1 si la respuesta es errónea y 0 si se deja sin responder. Las notas obtenidas por los 20 alumnos fueron las**

**siguientes:**

| | -1 | 0 | 1 |
|---|---|---|---|
| **-1** | **0** | **3** | **0** |
| **0** | **5** | **4** | **5** |
| **1** | **0** | **3** | **0** |

**donde la primera columna indica la nota obtenida en la primera cuestión y la primera fila la nota obtenida en la segunda.**

**Hallar la nota media y la varianza de cada una de las dos preguntas. Hallar la covarianza. ¿Son independientes ambas variables?**

---

Si llamamos $X$ a la nota obtenida en la primera pregunta, la distribución marginal de $X$ será:

| $x_i$ | $n_i$ |
|---|---|
| $-1$ | 3 |
| 0 | 14 |
| 1 | 3 |

obteniéndose una nota media de

$$\overline{X} = \frac{-1 \cdot 3 + 0 \cdot 14 + 1 \cdot 3}{20} = 0.$$

y una varianza de

$$\mathrm{var}(Y) = \frac{(-1)^2 \cdot 3 + 0^2 \cdot 14 + 1^2 \cdot 3}{20} - 0^2 = \frac{1 \cdot 3 + 0 \cdot 14 + 1 \cdot 3}{20} = \frac{6}{20}.$$

Si llamamos $Y$ a la nota obtenida en la segunda pregunta, la distribución marginal de $Y$ será:

| $y_i$ | $n_i$ |
|---|---|
| $-1$ | 5 |
| 0 | 10 |
| 1 | 5 |

obteniéndose una nota media de

$$\overline{Y} = \frac{-1 \cdot 5 + 0 \cdot 10 + 1 \cdot 5}{20} = 0.$$

y una varianza de

$$\mathrm{var}(Y) = \frac{(-1)^2 \cdot 5 + 0^2 \cdot 10 + 1^2 \cdot 5}{20} - 0^2 = \frac{1 \cdot 5 + 0 \cdot 10 + 1 \cdot 5}{20} = \frac{1}{2}$$

La covarianza sería:

$$\begin{aligned}\mathrm{Cov}(x,y) &= [1\cdot 1\cdot 0 + 1\cdot 0\cdot 3 + 1\cdot(-1)\cdot 0 + \\ &+ 0\cdot 1\cdot 5 + 0\cdot 0\cdot 4 + 0\cdot(-1)\cdot 5 + \\ &+ (-1)\cdot 1\cdot 0 + (-1)\cdot 0\cdot 3 + (-1)\cdot(-1)\cdot 0] - 0\cdot 0 = 0.\end{aligned}$$

Sin embargo, las dos variables no son independientes, puesto que si vemos la tabla de frecuencias relativas

| | $-1$ | $0$ | $1$ | |
|---|---|---|---|---|
| $-1$ | $0$ | $3/20$ | $0$ | $3/20$ |
| $0$ | $5/20$ | $4/20$ | $5/20$ | $14/20$ |
| $1$ | $0$ | $3/20$ | $0$ | $3/20$ |
| | $5/20$ | $10/20$ | $5/20$ | $1$ |

se puede comprobar que la frecuencia relativa de la distribución conjunta no coincide con el producto de las frecuencias relativas marginales. Por ejemplo:

$$f_{1,1} = 0 \neq \frac{3}{20}\frac{5}{20} = f_{1\cdot}\cdot f_{\cdot 1}.$$

# Problemas Propuestos

1. Calcular las medias y varianzas marginales.

| $X/Y$ | 10 | 20 | 30 | 40 | 50 | 60 |
|---|---|---|---|---|---|---|
| 15 | 5 | 7 | | | | |
| 25 | | 20 | 23 | | | |
| 35 | | | 30 | 47 | 2 | |
| 45 | | | 10 | 11 | 20 | 6 |
| 55 | | | | 9 | 7 | 3 |

2. Calcular la recta de regresión de Y/X y la recta de regresion de X/Y. Calcular el coeficiente de correlación lineal y comentar el resultado

| $X$ | $Y$ |
|---|---|
| 2 | 150 |
| 3 | 130 |
| 6 | 125 |
| 10 | 120 |
| 12 | 100 |

3. Calcular las rectas de regresión y el coeficiente de correlación lineal.

| $X/Y$ | 5 | 10 | 15 | 20 |
|---|---|---|---|---|
| 10 | 2 | | | |
| 20 | 5 | 4 | 1 | |
| 30 | 3 | 8 | 6 | 3 |
| 40 | | 3 | 6 | 6 |
| 50 | | | 2 | 1 |

4. Las calificaciones de 10 alumnos en Álgebra y Cálculo son las que aparecen en la tabla:

   (a) Dibujar la nube de puntos.

   (b) Calcular las rectas de regresión y el coeficiente de correlación lineal.

   (c) ¿Qué nota esperará un universitario en Cálculo, si ha obtenido 8,3 en Álgebra?

| Álgebra | Cálculo |
|---|---|
| 6 | 6,5 |
| 4 | 4,5 |
| 8 | 7 |
| 5 | 5 |
| 3,5 | 4 |
| 7 | 8 |
| 5 | 7 |
| 10 | 10 |
| 5 | 6 |
| 4 | 5 |

5. De dos variables observadas conjuntamente, se ha obtenido:

| $X/Y$ | $0-10$ | $10-20$ | $20-30$ |
|---|---|---|---|
| 2 | 3 | 5 | |
| 4 | 3 | 1 | 2 |
| 6 | 8 | 6 | 3 |
| 8 | | 1 | 3 |

Calcular la covarianza.

6. Las alturas $X$, y los pesos $Y$ de 20 hombres son:

| $X$ | $Y$ | $X$ | $Y$ |
|---|---|---|---|
| 1,72 | 63 | 1,76 | 71 |
| 1,70 | 75 | 1,70 | 70 |
| 1,70 | 68 | 1,69 | 66 |
| 1,68 | 70 | 1,66 | 60 |
| 1,75 | 74 | 1,78 | 74 |
| 1,69 | 72 | 1,74 | 69 |
| 1,71 | 67 | 1,70 | 65 |
| 1,69 | 69 | 1,69 | 71 |
| 1,67 | 70 | 1,71 | 73 |
| 1,74 | 84 | 1,78 | 69 |

Establecer la distribución correspondiente y hallar las medias aritméticas y las desviaciones típicas marginales.

7. Las variables X e Y observadas conjuntamente 100 veces, han presentado los siguientes valores:

| $x_i$ | $y_i$ | $n_i$ |
|---|---|---|
| 3 | 0 | 13 |
| 6 | 0 | 13 |
| 10 | 1 | 30 |
| 11 | 1 | 24 |
| 12 | 2 | 20 |

(a) Calcular las medias y varianzas marginales.

(b) Calcular la covarianza.

8. Ajustar la recta de regresión de $Y$ sobre $X$ por el método de los mínimos cuadrados:

| $x_i$ | $y_i$ |
|---|---|
| 2 | 150 |
| 3 | 130 |
| 6 | 125 |
| 10 | 120 |
| 12 | 100 |

9. Ajustar una parábola de regresión de $Y$ sobre $X$ por el método de los mínimos cuadrados:

| $Y/X$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 3 | 2 | | | |
| 2 | | | 4 | 4 | |
| 3 | | | | | 3 |

10. Determinar si existe dependencia entre la cantidad de lluvia y los grados de temperatura en base a las siguientes observaciones:

| $Lluvia(Litros/m^2)$ | $Temperatura(grados)$ |
|---|---|
| 14,1 | 20,8 |
| 10,2 | 15,6 |
| 8,3 | 23,9 |
| 30,3 | 24,2 |
| 32,5 | 25,0 |
| 21,4 | 10,3 |
| 34,0 | 15,2 |
| 18,6 | 17,2 |
| 19,0 | 20,1 |
| 22,7 | 15,1 |

11. Se han medido el peso y la altura de 20 personas, (X; Y), obteniéndose:

| | | | | |
|---|---|---|---|---|
| (63; 1,71) | (75; 1,70) | (67; 1,70) | (70; 1,68) | (73; 1,75) |
| (72; 1,69) | (37; 1,17) | (69; 1,96) | (70; 1,66) | (84; 1,73) |
| (71; 1,76) | (70; 1,71) | (66; 1,96) | (60; 1,67) | (73; 1,78) |
| (69; 1,74) | (65; 1,70) | (72; 1,68) | (73; 1,71) | (68; 1,78) |

Se pide:

(a) Agrupar los datos en una tabla de doble entrada. (Considerar intervalos de amplitud de 5 Kg /5 cm).

(b) Calcular la media y la desviación típica de las distribuciones marginales de X e Y.

(c) Hallar la covarianza. Interpretarla.

12. En el proceso de manufacturación de un articulo de vestir se han controlado dos características: tiempo empleado $X$ (minutos) y perfeccionamiento en el acabado $Y$ (errores cometidos). arrojando la siguiente distribución conjunta sobre una muestra de ciento veinte unidades:

| $Y$ / $X$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 0 | 2 | 5 | 10 | 12 |
| 1 | 6 | 10 | 28 | 8 |
| 2 | 15 | 12 | 6 | 6 |

Se pide:

(a) Distribuciones marginales.

(b) Media y moda de las distribuciones marginales.

(c) Desviación típica y asimetría marginales.

13. La siguiente tabla da el número de calzado $X$ y los pesos $Y$ de 55 estudiantes. Con estos datos estudiar la independencia lineal entre ambas variables.

| $X$ / $Y$ | 55 | 60 | 65 | 70 | 75 | 80 | 85 |
|---|---|---|---|---|---|---|---|
| 39 | 1 | | | | | | |
| 40 | | 3 | 3 | 4 | | | |
| 41 | | 3 | 4 | 6 | | | 1 |
| 42 | | | 8 | 8 | 7 | 2 | |
| 43 | | | 2 | | 1 | | |
| 44 | | | | | | | 2 |

14. Las seis cooperativas agrarias de una comarca presentaban las siguientes cifras correspondientes a las variables: $X$ = stock medio diario en naves de almacenamiento (miles de ptas). $Y$ = cifra comercializada diariamente (miles de ptas). $Z$ = empleados fijos en plantilla. $V$ = empleados eventuales.

| Cooperativa | $X$ | $Y$ | $Z$ | $V$ |
|---|---|---|---|---|
| $A$ | 26 | 146 | 6 | 8 |
| $B$ | 33 | 167 | 8 | 6 |
| $C$ | 12 | 92 | 6 | 8 |
| $D$ | 18 | 125 | 8 | 6 |
| $E$ | 18 | 118 | 10 | 4 |
| $F$ | 25 | 132 | 10 | 4 |

Determinar el grado de relación lineal entre las variables $(X, Y)$, $(Y, Z)$ y $(Z, V)$.

15. Una factoría de una cierta marca de refrescos ha tomado al azar 10 semanas del año, observando la temperatura media (en grados centígrados) correspondiente a cada una de ellas y la cantidad de refrescos pedidos durante cada uno de dichos períodos. La información obtenida es la siguiente:

| Temperatura media | Cantidad de refrescos |
|---|---|
| 10 | 21 |
| 28 | 65 |
| 12 | 19 |
| 31 | 72 |
| 30 | 75 |
| 19 | 39 |
| 24 | 67 |
| 5 | 11 |
| 9 | 12 |
| 15 | 24 |

Estudiar el grado de relación lineal entre ambas variables.

16. Para los datos que se representan, ¿se puede afirmar que la fluidez verbal de los alumnos está relacionada con los ingresos económicos de sus padres?

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | 13 | 35 | 10 | 17 | 33 | 15 | 22 | 20 | 28 | 20 | 25 | 23 | 23 |
| 125 | 50 | 115 | 55 | 60 | 100 | 65 | 100 | 70 | 95 | 80 | 85 | 90 | 75 |

(La primera fila corresponde a la fluidez verbal medida en número de palabras por minuto y la segunda a los ingresos, medidos en miles de pesetas al mes)

17 Las ciento treinta agencias de una entidad bancaria presentaban, en el ejercicio 1996, los siguientes datos correspondientes a las variables $X$ = cuentas a plazo/total cuentas, $Y$ = saldo medio de las cuentas a 31-XII (en millones de ptas).

| $X/Y$ | menos de 0,1 | de 0,1 a 0,3 | más de 0,3 |
|---|---|---|---|
| menos de 20 | 48 | | |
| de 20 a 50 | 21 | 11 | |
| de 50 a 100 | 14 | 8 | 2 |
| de 100 a 250 | 7 | 5 | 1 |
| más de 50 | 6 | 6 | 1 |

(a) Distribuciones marginales.

(b) Mediana de $Y$ y tercer cuartil de $X$.

(c) Distribución de las agencias, según $Y$, cuando la ratio $X$ se encuentra comprendida entre 0,1 y 0,3.

(d) Distribución, en términos relativos, de $X$ para agencias con saldo medio por encima de las 100.000 ptas.

18. El siguiente modelo teórico correspondiente al número de piezas defectuosas encontradas en dos muestras aleatorias de tres unidades. Con $X$ = defectuosas primera muestra, $Y$ = defectuosas segunda muestra, se dispone de la siguiente estructura conjunta de posibilidades:

| $X/Y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0,047 | 0,093 | 0,062 | 0,014 |
| 1 | 0,093 | 0,187 | 0,124 | 0,028 |
| 2 | 0,062 | 0,124 | 0,084 | 0,018 |
| 3 | 0,014 | 0,028 | 0,018 | 0,004 |

Comprobar que $X$ e $Y$ son dos variables independientes.

19. Se está intentando estudiar si existe relación entre el número de años que lleva un individuo afiliado a un partido político y el nivel de satisfacción con dicho partido. Para ello se parte de los datos de 10 individuos tomados al azar.

| Años | Satisfacción |
|---|---|
| 8 | 7 |
| 7 | 2 |
| 10 | 3 |
| 3 | 9 |
| 6 | 12 |
| 13 | 7 |
| 4 | 2 |
| 12 | 11 |
| 6 | 6 |
| 1 | 6 |

¿Se puede decir que los años de afiliación y la satisfacción están correlacionados?

20. En una encuesta en la que se han entrevistado a 480 familias, se han obtenido para las mismas los siguientes datos sobre ingresos mensuales $X$ y depósitos a la vista en Bancos y Cajas de ahorros $Y$:

| $X$ / $Y$ | de 0 a 200.000 | de 200.000 a 500.000 | de 500.000 a 2.000.000 | de 2.000.000 a 10.000.000 |
|---|---|---|---|---|
| de 50.000 a 100.000 | 40 | 12 | 8 | |
| de 100.000 a 150.000 | 16 | 48 | 12 | 4 |
| de 150.000 e a 250.000 | 8 | 80 | 92 | 20 |
| de 250.000 a 500.000 | 4 | 40 | 72 | 24 |

Suponiendo que las marcas de clase sean representativas de cada intervalo:

(a) Calcúlense los valores que toman $n_{1.}, n_{2.}, n_{.2}, n_{.3}$

(b) Exprésense, en tanto por uno, los valores que toman las siguientes frecuencias $f_{12}, f_{23}, f_{34}, f_{42}, f_{2.}, f_{3}$.

(c) Exprésense, en tanto por uno, los valores que toman $f(X_1/Y = 350.000)$ y $f(Y_2/X = 200.000)$.

(d) Calcúlense las medias marginales de $X$ e $Y$

21. A partir de los siguientes datos:

| X | Y | X | Y | X | Y |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 4 | 1 |
| 2 | 3 | 4 | 5 | 6 | 2 |
| 2 | 1 | 1 | 6 | 5 | 6 |
| 3 | 4 | 2 | 5 | 1 | 6 |
| 5 | 3 | 5 | 1 | 6 | 2 |
| 4 | 2 | 4 | 2 | 5 | 1 |
| 1 | 6 | 4 | 5 | 4 | 2 |
| 3 | 4 | 4 | 1 | 6 | 5 |

obtenidos al lanzar dos datos simultáneamente:

(a) Constrúyase la tabla de doble entrada.

(b) Calcúlense:

i. Las medias marginales.

ii. La covarianza.

iii. El porcentaje de valores pares obtenidos con el segundo dado, cuando se ha obtenido un 3 con el primero.

(Nota: se designan por $X$ e $Y$ los valores obtenidos con el primer y segundo dado respectivamente.)

22. Dada la siguiente distribución bidimensional de frecuencias:

| X/Y | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 100 | 2 | 4 | 6 | 10 | 8 |
| 200 | 1 | 2 | 3 | 5 | 4 |
| 300 | 3 | 6 | 9 | 15 | 12 |
| 400 | 4 | 8 | 12 | 20 | 16 |

(a) Calcúlese:

i. La media marginal de $X$ y las medias condicionadas de $X$ para cada valor de $Y$.

ii. La media marginal de $Y$ y las medias condicionadas de $Y$ para cada valor de $X$.

iii. La covarianza.

iv. La varianza marginal de $X$ y la varianza condicionada de $X$ al valor $Y = 4$.

v. La varianza marginal de $Y$ y la varianza de $Y$ condicionada a $X = 200$.

(b) ¿Son estadísticamente independientes?

23. Veintiocho tiendas de deporte han creado mancomunadamente un servicio propio de reparaciones de material deportivo; observándose, en los seis meses que lleva funcionando, las características: $X_{1i}$ = importe de las reparaciones solicitadas por la tienda $i$. $X_{2i}$ = días de retraso, en promedio en el servicio de la empresa $i$. $X_{3i}$ = cuantía del ahorro por el servicio, satisfecha por la empresa $i$.

$$\begin{vmatrix} N & \sum X_{1i} & \sum X_{2i} & \sum X_{3i} \\ 0 & \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \sum X_{1i}X_{3i} \\ 0 & 0 & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ 0 & 0 & 0 & \sum X_{3i}^2 \end{vmatrix} = \begin{vmatrix} 28 & 60 & 32 & 14 \\ 0 & 150 & 123 & 83 \\ 0 & 0 & 74 & 11 \\ 0 & 0 & 0 & 35 \end{vmatrix}$$

Obtener el rango de la matriz de varianzas y covarianzas. Interpretar el resultado.

24. Dada la siguiente distribución bidimensional:

| X | 0,7 | 1 | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 2,2 | 2,2 | 2,5 | 2,7 | 2,8 | 3 | 3,3 | 3,4 | 4 | 4 |

Se pide:

(a) Calcular las dos rectas de regresión.

(b) Comprobar que dichas rectas se cortan en el centro de gravedad.

(c) Estimar el valor de $Y$ para $X = 10$.

(d) Estimar el valor de $X$ para $Y = 4{,}6$.

(e) Intrepretar y analizar los resultados obtenidos en los dos puntos anteriores.

(f) Interpretar el significado de $b$ y de $b'$.

(g) Interpretar el significado de $1 - R^2$.

25. Dada la distribución:

| X | 1 | 1,5 | 2 | 2,5 | 3 | 3,75 | 4,5 | 5 |
|---|---|---|---|---|---|---|---|---|
| Y | 1 | 1,5 | 2,95 | 5,65 | 8,8 | 15 | 25 | 32 |

Se pide:

(a) Elegir la mejor clase funcional para ajustar la distribución y estimar sus parámetros.

(b) Establecer la bondad del ajuste.

(c) Calcular la previsión para $Y$ cuando $X = 7$. Analizar dicha previsión.

26. Dada la distribución:

| X | 2,5 | 3,75 | 5 | 7,5 | 10 | 12,5 | 20 |
|---|---|---|---|---|---|---|---|
| Y | 8 | 14 | 23,75 | 40 | 62 | 90 | 165 |

Se pide:

(a) Utilizar una función del tipo $Y = a\,X^b$ para ajustar la distribución.

(b) Dar una medida de la bondad del ajuste.

(c) Interpretar el valor de $R^2$

27 Dada la distribución :

| X | 1 | 1,5 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Y | 1 | 1,75 | 2,65 | 4,7 | 7 | 9,5 | 12 | 15 |

Se pide:

(a) Ajustar la distribución utilizando una función del tipo $Y = a\,X^b$.

(b) Analizar la bondad del ajuste.

(c) Interprete el significado del coeficiente $a$.

28. Dada la distibución:

| X | 5 | 6 | 8 | 10 | 13 | 18 | 20 |
|---|---|---|---|---|---|---|---|
| Y | 1,5 | 1,25 | 0,93 | 0,7 | 0,46 | 0,23 | 0,15 |

Se pide:

(a) Estimar los parámetros de la clase funcional $Y = a \cdot b^{-0,2X}$ para ajustar la distribución.

(b) Estudiar la bondad del ajuste.

(c) Sería posible plantear un ajuste del tipo $Y = a \cdot b^{cX}$ Justifíquelo.

29. Dada la distribución:

| X | 5 | 8 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 3 | 3,5 | 4 | 5,6 | 8 | 11,25 | 16 | 22 | 32 | 45 | 36 |

Se pide:

(a) Estimar los parámetros de la clase funcional $Y = a \cdot b^{-0,2X}$ para ajustar la distribución.

(b) Estudiar la bondad del ajuste.

(c) Comentar los resultados del ajuste.

30. Dada la distribución:

| X/Y | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 1 | 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 5 | 6 | 3 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 2 | 7 | 5 | 2 | 2 | 0 | 0 | 0 |
| 4 | 0 | 1 | 7 | 5 | 3 | 3 | 0 | 0 | 0 |
| 5 | 0 | 1 | 5 | 8 | 5 | 4 | 3 | 0 | 0 |
| 6 | 0 | 0 | 4 | 6 | 7 | 5 | 4 | 0 | 0 |
| 7 | 0 | 0 | 2 | 3 | 8 | 8 | 7 | 4 | 0 |
| 8 | 0 | 0 | 1 | 2 | 5 | 6 | 7 | 6 | 1 |
| 9 | 0 | 0 | 1 | 2 | 3 | 4 | 6 | 4 | 5 |

Se pide:

(a) Obtener la poligonal de regresión de $Y$ respecto a $X$.

(b) Calcular la razón de correlación.

(c) A la vista de la poligonal, ajustar a la distribución una función.

(d) Calcular $R^2$, analizarlo y compararlo con la razón de correlación.

31. Contestar a las mismas cuestiones del problema anterior para la distribución:

| X/Y | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 4 | 3 | 7 | 1 | 8 | 9 | 3 | 2 |
| 2 | 3 | 6 | 1 | 8 | 5 | 9 | 6 | 4 | 3 |
| 3 | 0 | 0 | 3 | 5 | 7 | 8 | 5 | 4 | 2 |
| 4 | 0 | 0 | 0 | 2 | 4 | 6 | 7 | 4 | 3 |
| 5 | 0 | 0 | 0 | 0 | 5 | 8 | 7 | 4 | 4 |
| 6 | 0 | 0 | 0 | 2 | 8 | 9 | 9 | 6 | 6 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 3 |

32. Dada la distribución:

| X/Y | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 16 | 15 | 14 | 15 | 13 | 12 | 11 | 7 | 4 | 2 | 0 | 0 |
| 2 | 3 | 3 | 3 | 2 | 4 | 5 | 4 | 5 | 5 | 3 | 0 | 0 |
| 3 | 2 | 1 | 2 | 2 | 2 | 3 | 3 | 4 | 5 | 4 | 0 | 0 |
| 4 | 1 | 1 | 1 | 1 | 2 | 3 | 2 | 3 | 6 | 5 | 3 | 0 |
| 5 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 3 | 4 | 7 | 4 | 0 |
| 6 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 3 | 8 | 6 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 5 | 9 | 3 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 10 | 5 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 8 | 8 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 7 | 19 |

Se pide:

(a) Obtener la poligonal de regresión de $Y$ respecto a $X$ y calcular la razón de correlación.

(b) Ajustar a la distribución una función lineal, una parabólica y una exponencial.

(c) Determinar el grado de bondad de cada uno de estos ajustes y compararlos con la razón de correlación.

33. Dada la distribución:

| X/Y | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 2 | 6 | 2 | 1 | 0 | 0 | 0 |
| 5 | 4 | 7 | 5 | 0 | 0 | 0 |
| 10 | 0 | 1 | 7 | 3 | 0 | 0 |
| 15 | 0 | 0 | 6 | 3 | 1 | 0 |
| 20 | 0 | 0 | 4 | 8 | 2 | 0 |
| 25 | 0 | 0 | 3 | 8 | 5 | 0 |
| 30 | 0 | 0 | 0 | 5 | 9 | 0 |
| 35 | 0 | 0 | 0 | 1 | 7 | 4 |
| 40 | 0 | 0 | 0 | 0 | 3 | 4 |

Se pide:

(a) Obtener la poligonal de regresión de $Y$ respecto a $X$ y calcular la razón de correlación.

(b) Ajustar los datos a una recta y a una parábola y discutir los resultados.

34. La tabla que se da a continuación recoge las importaciones y exportaciones (miles de millones de ptas) españolas en los últimos años.

| Año | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 |
|---|---|---|---|---|---|---|
| Importación | 2451 | 2976 | 3475 | 4177 | 4630 | 5115 |
| Exportación | 1493 | 1890 | 2234 | 2847 | 3743 | 4109 |
| Año | 1994 | 1995 | 1996 | 1997 | 1998 | |
| Importación | 4955 | 6051 | 6989 | 8396 | 8915 | |
| Exportación | 3816 | 4210 | 4660 | 5134 | 5643 | |

Se pide:

(a) Ajustar las variables importación y exportación en función de los años. Analizar la bondad del ajuste.

(b) Realizar previsiones de Exportación / Importación para los años 1999, 2000 y 2001. Dar la validez de dichas previsiones.

(c) Considerar la serie de las diferencias entre importaciones y exportaciones y ajustarla. Analizar la validez del ajuste y compararlo con los realizados a las variables por separado.

(d) Extraer las conclusiones oportunas sobre el futuro del comercio exterior español.

35. La tabla que se adjunta nos ofrece información sobre la proporción de subsidios en el PIB en España.

| Año | % | Año | % | Año | % | Año | % |
|---|---|---|---|---|---|---|---|
| 1964 | 1 | 1971 | 1,25 | 1978 | 2,23 | 1985 | 2,85 |
| 1965 | 1,07 | 1972 | 1,23 | 1979 | 1,99 | 1986 | 2,18* |
| 1966 | 0,89 | 1973 | 1,12 | 1980 | 2,17 | 1987 | 1,95 |
| 1967 | 1,02 | 1974 | 1,13 | 1981 | 2,08 | 1988 | 2,11 |
| 1968 | 1,04 | 1975 | 1,36 | 1982 | 2,64 | 1989 | 1,96 |
| 1969 | 0,87 | 1976 | 1,55 | 1983 | 2,75 | | |
| 1970 | 1,06 | 1977 | 1,68 | 1984 | 2,99 | | |

(*) Margaret Thatcher reduce la intervención estatal en GB y es seguida por otros paises de la OCDE.

Se pide:

(a) Realizar el ajuste del los datos.

(b) Indicar el comportamiento del modelo anterior si no consideramos los años 86, 87, 88 y 89.

(c) Analizar las diferencias entre lo acontecido en los dos puntos anteriores.

(d) Realizar previsiones para los años 90 y 91 a partir de los dos ajustes y analizarlas.

36. El consumo de energía eléctrica en el período 1934-1950 se representa en la siguiente tabla: (Los datos vienen en forma de números índices con base en 1940).

| Año | 1934 | 1935 | 1936* | 1937* | 1938* | 1939*$^+$ |
|---|---|---|---|---|---|---|
| Indice | 91,3 | 98,3 | 80,5 | 96,3 | 77,7 | 80,5 |
| Año | 1940$^+$ | 1941$^+$ | 1942$^+$ | 1943$^+$ | 1944$^+$ | 1945$^+$ |
| Indice | 100 | 107,6 | 119,5 | 128,2 | 126,6 | 114,5 |
| Año | 1946 | 1947 | 1948 | 1949 | 1950 | |
| Indice | 147,5 | 160,3 | 164,1 | 153 | 188,2 | |

(*) Guerra Civil Española. ($^+$) II Guerra Mundial.

Se pide:

(a) Ajustar los datos a una recta.

(b) Efectuar la misma operación al período 1940-1950.

(c) Analizar la bondad de los ajustes y realizar previsiones para el año 1951.

37. Un grupo de 15 alumnos han sido ordenados en función de las calificaciones obtenidas en Estadística (E) e Historia del Arte (H) con el resultado que indica la tabla:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| H | 9 | 15 | 2 | 13 | 11 | 10 | 1 | 8 | 5 | 14 | 3 | 6 | 4 | 12 | 7 |

Se pide:

(a) Calcular el coeficiente de correlación lineal e interpretarlo.

(b) Si definimos $d_i = x_i - y_i$ , demostrar que el coeficiente de correlación de rangos de Spearman coincide con $r$, esto es:

$$r = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

38. En un concurso de aptitudes para acceder a una plaza de trabajo, dos psicólogos han clasificado a un conjunto de 16 candidatos de la siguiente forma:

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| P1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| P2 | 1 | 3 | 4 | 2 | 5 | 7 | 8 | 12 | 6 | 13 | 9 | 10 | 11 | 15 | 14 | 16 |

donde C corresponde al número de candidato, P1 a la clasificación hecha por el primer psicólogo y P2 a la clasificación hecha por el segundo.

¿Podemos decir que los criterios de ambos psicólogos concuerdan? Justifique la respuesta.

39. En una prueba de acceso a la Univesidad para mayores de 25 años los resultados obtenidos en función de la edad de los alumnos han sido los siguientes:

| Resultado/Edad | 25-30 | 30-40 | 40-50 | 50-60 | 60-80 |
|---|---|---|---|---|---|
| Apto | 15 | 12 | 8 | 5 | 2 |
| No Apto | 27 | 18 | 3 | 5 | 3 |

Se pide:

(a) Asignando a los valores Apto, No Apto, los valores 0 y 1 respectivamente, obtenga el coeficiente de correlación lineal e interpretelo.

(b) Demuestre que el coeficiente de correlación biserial coincide con $r$, esto es:

$$r = \frac{\overline{Y}/_{x=1} - \overline{Y}/_{x=0}}{\sigma_y} \sqrt{pq}$$

siendo $p$ y $q$ los porcentajes de aptos y no aptos, respectivamente.

40. En la misma prueba del ejercicio anterior se consideró la distribución de hombres y mujeres en relación a los resultados obtenidos:

| Resultado/Sexo | Hombres | Mujeres |
|---|---|---|
| Aptos | 34 | 22 |
| No Aptos | 25 | 17 |

Se pide:

(a) ¿Existen diferencias entre los resultados de ambos sexos?

(b) Demuestre que el coeficiente de correlación tetracórica coincide con $r$, esto es:

$$r = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.} - n_{.1}n_{.2}}}$$

# Unidad Temática III

# Probabilidad. Variable Aleatoria.

# Resumen Teórico

## 3.1 Evolución histórica.

Hay determinados experimentos en los que las situaciones o causas determinan perfectamente los resultados o efectos: por ejemplo si conocemos la altura desde la que se deja caer un objeto en el vacío podemos calcular el tiempo que tardará en impactar con la superficie. Estaríamos ante un hecho determinista. En otras experiencias, sin embargo, repitiendo el proceso varias veces en las mismas condiciones obtenemos resultados diferentes: son experimentos aleatorios. Si lanzamos un dado perfecto al aire y consideramos el "número" de la cara superior, los valores 1,2,3,4,5 y 6 se obtienen indistintamente y sin ninguna ley apreciable: dichos números son los sucesos elementales del experimento. También podemos considerar sucesos compuestos como formar parejas, tríos, etc, que se generan como unión de sucesos elementales y que se consiguen al lanzar varias veces el dado o al combinar el lanzamiento de varios de ellos.

Como en la mayoría de los descubrimientos, la noción de Probabilidad se ha ido desarrollando a lo largo del tiempo en función de la necesidad, de los recursos y de la aportación de los grandes genios que son capaces, en un momento de inspiración, de dar un paso de un siglo.

Es difícil establecer históricamente el nacimiento de la probabilidad, aunque su conceptualización como disciplina matemática es reciente: parece, no obstante, que su origen tiene relación con los juegos de azar. Algunos autores consideran que la génesis del cálculo de probabilidades se encuentra en la resolución del problema planteado por el caballero de Meré, un jugador empedernido de la Francia del siglo XVII, a su amigo y matemático Pascal, quien mantuvo una abundante correspondencia sobre dicho problema con su colega Fermat. El problema consistía en cómo deberían repartirse el dinero de las apuestas depositado en la mesa si los jugadores se veían obligados a finalizar la partida sin que existiera un ganador.

Conectando con el primer párrafo, si prolongáramos indefinidamente la sucesión de pruebas con sus resultados, llegaríamos al conjunto potencialmente infinito de todos los resultados del experimento aleatorio que se llama Universo, Población o Colectivo asociado al mismo. De hecho, en el caso de fenómenos deterministas se hace preciso el registro y constatación de ciertas permanencias, mientras que en el caso aleatorio dichas permanencias aparecen al considerar un gran número de pruebas. Por ejemplo, si lanzo $n$ veces una moneda y obtengo $k$ caras, la frecuencia del suceso cara es $k/n$. Este hecho de que la frecuencia de un suceso tiende a aproximarse a un número fijo al aumentar el número de pruebas se ha denominado Ley del azar o Ley de estabilidad de las series estadísticas.

El llegar a la conclusión anterior fue posible gracias a la aportación de Bernoulli(1654-1705), que demostró la llamada Ley de los grandes números, que enunciada de una forma sencilla dice así: "la frecuencia relativa de un suceso tiende a estabilizarse en torno a un número, a medida que el número de pruebas del experimento crece indefinidamente".

No obstante, la primera definición que se conoce del concepto de probabilidad fue enunciada por Laplace(1749-1827). Aunque de forma un tanto libre, Laplace entiende que: "en el fondo, la teoría de probabilidades es sólo sentido común expresado con números". Después de Laplace, el interés por el cálculo de probabilidades fue disminuyendo, llegando prácticamente a desaparecer como disciplina matemática durante el siglo XIX.

No es hasta principios del siglo XX cuando el matemático ruso Kolmogorov, apoyándose en la Teoría de conjuntos y en la Teoría de la medida, da una definición axiomática del Cálculo de Probabilidades, como generalización y síntesis de los conocimientos que de la Probabilidad se tenían hasta entonces.

## 3.2 Distintas definiciones del concepto de probabilidad.

De la introducción histórica que hemos elaborado se desprende la existencia de tres definiciones del concepto de Probabilidad que pasamos a definir y discutir.

### Definición Frecuentista.

La probabilidad de un suceso es el valor límite de su frecuencia relativa al repetir indefinidamente la experimentación. Los inconvenientes de definir así la probabilidad son los siguientes:

1. Desde el punto de vista del Análisis no puede interpretarse el límite anterior por la imposibilidad de fijar el número de repeticiones.
2. En algunas ocasiones no es posible realizar una experimentación indefinida.
3. Las condiciones bajo las cuales se realiza la experimentación pueden variar a lo largo del tiempo y, con ellas, las frecuencias relativas.

### Definición Clásica. (Laplace)

La probabilidad de un suceso $A$ es el cociente entre el número de casos favorables al suceso y el número de casos posibles. Inconvenientes:

1. No es válida cuando los sucesos elementales no son equiprobables.
2. A veces no es posible contar.

Para evitar todos estos inconvenientes y un gran número de paradojas y dificultades surgidas a comienzos del presente siglo, se hizo necesaria una profunda revisión del concepto de probabilidad utilizando las herramientas más precisas del momento: La Teoría de conjuntos, desarrollada principalmente por Borel, y la potente Teoría de la medida debida a Lebesgue. La probabilidad sería entonces una medida de la incertidumbre, con propiedades similares a las medidas de longitudes, tiempo, etc.

Una concepción más operativa es definir la probabilidad como una medida personal de la incertidumbre de un suceso, basada en aquellos experimentos previos, que con la información disponible, se consideren indistinguibles o intercambiables. En situaciones repetitivas, cuando exista una amplia experiencia, la probabilidad vendrá determinada por la frecuencia relativa, mientras que, en otros casos, dependerá de distintos tipos de información. Todo lo anterior llevó a Kolmogorov a introducir axiomáticamente el concepto de Probabilidad:

**Teorema.** Dados $n$ sucesos cualesquiera $A_1, A_2, \ldots, A_n$, tal que se cumple $P(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$, se tiene que:

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_n) &= P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \ \ldots \\ &\quad \cdot P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \ . \end{aligned}$$

**Dependencia e Independencia.**

Diremos que dos sucesos $A$ y $B$ son independientes si el conocimiento de la ocurrencia de uno no modifica la probabilidad de aparición del otro. O sea, $P(B/A) = P(B)$. **Resultado.** Dos sucesos $A$ y $B$ son independientes si $P(A \cap B) = P(A) \cdot P(B)$.

La definición de independencia se puede generalizar a un número finito de sucesos. Así tendremos que $A_1, A_2, ..., A_n$ son independientes si:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j) & \forall i \neq j \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k) & \forall i \neq j, j \neq k, i \neq k \\ \ldots & \ldots \ldots\ldots & \\ P(A_1 \cap A_2 \cap \ldots \cap A_n) &= \prod_{i=1}^{n} P(A_i) \ . \end{aligned}$$

## 3.5 Teorema de la probabilidad total. Teorema de Bayes.

**Probabilidades a priori y probabilidades a posteriori.**

La probabilidad de obtener un cinco en el lanzamiento de un dado perfecto es $1/6$, ya que aplicando la definición de Laplace, se tiene que sólo hay un caso favorable y seis casos posibles, todos ellos equiprobables. A estas probabilidades que se pueden determinar de antemano, sin realizar ningún tipo de comprobación experimental, se les denomina probabilidades a priori. Ahora bien, no siempre es posible establecer a priori la probabilidad de cada suceso elemental del experimento aleatorio. Cuando esto ocurra, no tendremos más remedio que estimar la probabilidad, estudiando el valor límite al que se acerca la frecuencia relativa al realizar un gran número de pruebas en análogas condiciones. Estas probabilidades obtenidas por este procedimiento se denominan probabilidades a posteriori. Mediante este método no es posible llegar a determinar con precisión la probabilidad de un determinado suceso; pero a medida que el número de pruebas aumente obtendremos una mejor estimación de la probabilidad deseada.

**Ejemplo.**

El Estudio General de Medios (E.G.M.) realizó durante los meses de octubre y noviembre de 1998, un estudio para obtener el perfil sociológico de la audiencia de las emisoras de radio. En ese estudio se determinó que el 46% de los oyentes de las emisoras de frecuencia modulada son mujeres. De este modo, podemos asignar, experimentalmente, la probabilidad de que un oyente de FM sea mujer. A esta probabilidad asignada después de una experiencia se la denomina probabilidad a posteriori, o más correctamente, probabilidad experimental.

El valor 0,46 no se puede considerar fijo e inalterable, pero es fácil que si se repitiera el estudio al cabo de pocos meses, con un número análogo de encuestados, se obtuviera un valor muy próximo a 0,46. Es importante tener en cuenta que las frecuencias relativas que se obtienen de un número reducido de pruebas no pueden servir para realizar una estimación fiable de la probabilidad de un determinado suceso.

Este método de asignación de probabilidades es totalmente imprescindible para el cálculo de tasas de natalidad, defunción, esperanza de vida, expectativa de éxito de un nuevo producto comercial, etc.

## Teorema de la probabilidad total.

Consideremos un experimento que se realiza en dos etapas, en la primera supongamos que los posibles sucesos $A_1, A_2, ..., A_n$ constituyen un sistema completo. Mientras que en la segunda etapa los resultados posibles, $B_j$, dependen de la primera. Si conocemos las probabilidades condicionadas $P(B_j/A_i)$ para cada combinación de sucesos se verifica que:

$$P(B_j) = \sum_{i=1}^{n} P(B_j/A_i)P(A_i) .$$

**Teorema de Bayes.**

En las mismas hipótesis del Teorema anterior tenemos que:

$$P(A_k/B_j) = \frac{P(B_j/A_k)}{\qquad} . \sum_{i=1}^{n} P(B_j/A_i)P(A_i)$$

## 3.6 Anexo: Combinatoria.

**Introdución.**

La Combinatoria estudia los diferentes modos en que se puede llevar a cabo una cierta tarea de ordenación o agrupación de unos cuantos objetos siguiendo unas reglas prefijadas. Por ejemplo:

¿Cómo se pueden colocar las 32 fichas de ajedrez en el tablero de modo que no haya dos en el mismo cuadro?

¿Cuántas banderas diferentes se pueden hacer con tres franjas horizontales si tenemos tela de 6 colores diferentes?

**Variaciones con repetición.**

Supongamos que tenemos $m$ elementos distintos $a_1, a_2, a_3, ..., a_m$. Pretendemos ocupar $n$ lugares con ellos de modo que cada elemento pueda ocupar más de un lugar. Las distintas posiciones se llaman variaciones con repetición de $m$ elementos tomados $n$ a $n$. El número total se denota por $VR_{m,n}$, y es igual a $m^n$. Por ejemplo:

El número de quinielas de fútbol que hay que hacer para acertar 14 con seguridad es: $VR_{3,14} = 3^{14}$.

Con las cifras 1, 2, 3, 4, 5, ¿cuántos números de dos cifras pueden formarse? $VR_{5,2} = 5^2 = 25$.

**Variaciones ordinarias.**

Sean $m$ elementos $a_1, a_2, ..., a_m$. Pretendemos ocupar $n$ lugares con ellos de modo que cada elemento sólo pueda ocupar un lugar. (En este caso ha de ser $n < m$). Las distintas disposiciones se llaman variaciones de $m$ lementos tomados $n$ a $n$. Se denotan por $V_{m,n}$ y su número es: $V_{m,n} =$ $m - 1)(m - 2) \dots (m - (n - 1)) = m(m - 1) \dots (m - n + 1)$. También se puede expresar como:

$$V_{m,n} = \frac{m!}{(m-n)!}$$

Por ejemplo: Con las cifras 1, 2, 3, 4, 5, ¿cuántos números de dos cifras distintas pueden formarse? $= 5 \cdot 4 = 20$. En una quiniela hípica hay que acertar los tres primeros resultados de una carrera en la que compiten 10 caballos. El número de quinielas que hay que hacer para asegurar el acierto es: $V_{10,3} = 10 \cdot 9 \cdot 8 = 720$.

## Permutaciones.

¿De cuántas formas pueden ordenarse $n$ elementos? A cada una de dichas ordenaciones las llamaremos permutaciones. El número total de éstas se nota por $P_n$ y son :

$$P_n = n \cdot (n-1) \cdot (n-2) \dots 3 \cdot 2 \cdot 1$$

A este número se le llama $n$ factorial o factorial de $n$ y se representa por $n!$. Así nos quedaría:

$$P_n = V_{n,n} =$$

Por ejemplo: ¿De cuántas maneras se pueden alinear en una fila cinco personas para una foto de grupo? $P_5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$.

¿De cuántas formas diferentes se pueden colocar siete libros en un estante? $P_7 = 7! = 5040$ .

## Permutaciones con repetición.

Dados los elementos $a_1, a_2, a_3,$ se llaman permutaciones con repetición de orden $(\alpha, \dots, \gamma)$ a cada uno de los grupos de $\alpha + \beta + \dots + \gamma$ elementos que se pueden formar con la condición de que haya $\alpha$ elementos iguales a $a_1$, $\beta$ elementos iguales a $a_2$, ... , y $\gamma$ elementos iguales a $a_n$. Si $\alpha + \beta + \dots + \gamma = m$, el número total se nota por $P_m^{\alpha,\beta,\dots,\gamma} = PR(\alpha, \beta, \dots, \gamma)$ y vale:

$$PR(\alpha, \beta, \dots, \gamma) = \frac{m!}{\alpha! \beta! \cdots \gamma!}$$

Por ejemplo: ¿De cuántas maneras se pueden repartir 3 relojes, 2 bicicletas y 4 pelotas entre 9 niños, de modo que cada niño reciba un regalo? $PR_9^{,2,4} = 9!/(3!\,2!\,4!) = 1260$.

## Combinaciones.

Imaginemos $m$ elementos distintos. Pretendemos seleccionar $n$ de ellos sin importarnos el lugar que le asignemos. Las posibles elecciones se llaman combinaciones de $m$ elementos tomados de $n$ en $n$. El número total de todas ellas se representa por $C_{m,n}$, siendo su valor igual a:

$$C_{m,n} = \frac{V_{m,n}}{P_n} .$$

Por ejemplo: En una carrera compiten 10 corredores y se clasifican los tres primeros para la fase siguiente, ¿de cuántas maneras puede producirse la clasificación? $C_{10,3} = 120$.

## Combinaciones con repetición.

Llamaremos combinaciones con repetición de $n$ elementos tomados $r$ a $r$ a cada uno de los grupos que pueden formarse con $r$ elementos elegidos de entre $n$ posibles, sin importar el que se repitan. Se notan por $CR_{n,r}$ y vale

$$CR_{n\,r} \qquad \binom{n+r-1}{r}$$

Por ejemplo: ¿De cuántas formas se pueden distribuir 3 bolas iguales en 5 cajas distinguibles? $CR_{5,3} = \qquad = 35$.

**Nota:** También se pueden dar las siguientes definiciones:

## Variaciones.

Se llaman variaciones de $m$ elementos tomados de $n$ en $n$, a los grupos de $n$ elementos distintos que se pueden formar, de modo que un grupo se diferencie de otro bien por los elementos que los componen, bien por el orden en que aparezcan.

### Permutaciones.

Se llaman permutaciones de $m$ elementos a las distintas ordenaciones que podemos hacer con ellas. Una permutación, se diferencia de otra sólo en el orden, pues en ambas aparecen los mismos elementos.

### Combinaciones.

Se llaman combinaciones de $m$ elementos tomados de $n$ en $n$, a los grupos de $n$ elementos distintos que se pueden formar, de modo que un grupo se diferencie de otro por los elementos que los componen, sin que importe el orden en que aparezcan. En las combinaciones no influye el orden en que aparecen los elementos. Por ejemplo, ABCD y BACD son dos variaciones distintas pero la misma combinación, puesto que son los mismos cuatro elementos.

### Números combinatorios.

A los valores de $C_{m,n}$ se les llama números combinatorios y se les designa por:

$$C_{m,n} = \binom{m}{n} = \frac{m!}{(m-n)! \cdot n!}$$

### Propiedades de los números combinatorios:

1. $$\binom{m}{0} = \binom{m}{m} = 1$$

2. $$\binom{m}{n} = \binom{m}{m-n}$$

3. $$\binom{m-1}{n-1} + \binom{m-1}{n} = \binom{m}{n}$$

## 3.7 Variable aleatoria.

Para estudiar los experimentos aleatorios es preciso introducir una función que haga corresponder a cada suceso elemental del espacio muestral un número. Así:

1. Al lanzar una moneda, le podemos asociar el valor 1 al suceso elemental "cara" y el valor 0 al suceso "cruz".
2. Al lanzar dos dados al aire se puede asociar a cada resultado la suma de los puntos obtenidos.

A cualquier función que asocie a cada suceso elemental un número se le denomina variable aleatoria (v.a.), y, por tanto, sus valores dependen del resultado del experimento aleatorio. Podríamos decir que la v.a. es una abstracción cuantificada de la realidad.

Es importante observar que la asignación de valores a los resultados del experimento no es única, de hecho basta con fijar valores distintos a resultados distintos, para obtener una infinidad de funciones. No obstante, la idea es que dicha asignación sea lo más natural posible para que una vez manipulada la v.a. los resultados sean fácilmente interpretables en términos del experimento de partida.

### 3.7.1 Variables discretas y continuas.

Una v.a. se denomina Discreta si toma valores aislados o puntuales. Por ejemplo, si $X$ ="número de varones de una familia de 2 hijos", la variable $X$ puede tomar los valores 0, 1 y 2. Si deseáramos calcular la probabilidad de que $X$ tome cada uno de sus valores posibles y suponiendo que la probabilidad de que un hijo sea varón es 0,49, tendríamos, en el supuesto de que los sucesos fueran independientes, lo siguiente: $P[X = 2] = (0,49).(0,49) = 0,2401$ . $P[X = 1] = (0,49).(0,51) + (0,51).(0,49) = 0,4998$ . $P[X = 0] = (0,51).(0,51) = 0,2601$ . Siendo la suma de estas probabilidades, lógicamente, la unidad.

Una variable aleatoria es Continua si puede tomar cualquier valor dentro de un (varios) intervalo(s) determinado(s). Así, la variable $X$ que asocia a cada individuo de un colectivo su estatura es continua; en este caso, la probabilidad de que la variable tome exactamente un valor determinado, 160 cms por ejemplo, es cero. Ello es debido a que se trata de calcular la probabilidad de que $X = 160,00000...$ cms y no cualquiera de los infinitos valores tan cercanos como se quiera a 160 cms, como por ejemplo 159,99999...cms ó 160,000...01...cms. De hecho, esa es la principal característica de la variable continua y nos obligará a darle un tratamiento especial.

Para terminar podríamos considerar una v.a. Mixta. la cual tomaría valores dentro de un (varios) intervalo(s) y algunos valores puntuales más fuera de él (ellos).

Lo dicho hasta ahora vale tanto para experimentos simples (lanzar un dado y anotar el número de la cara superior), como para experimentos complejos (medir el peso, la estatura y la edad en años de un grupo de personas). En el primer caso la variable asociada sería unidimensional (discreta en el ejemplo), y en el segundo multidimensional (tridimensional, continua para las dos primeras dimensiones y discreta para la edad). Conviene no confundir con el concepto de variable mixta dado en el párrafo anterior

## 3.8 Variable aleatoria Unidimensional.

### 3.8.1 Funciones que caracterizan globalmente a una Variable Aleatoria.

Como hemos visto, la v.a. es la proyección numérica que hacemos de los resultados de un experimento aleatorio, y puesto que cada suceso tenía una determinada probabilidad de ocurrencia trasladamos dicha probabilidad al valor correspondiente de la v.a. Si la variable es discreta y toma pocos valores distintos, como en el ejemplo de los hijos varones, es factible, e incluso conveniente, dar todos esos valores con sus probabilidades de una forma explícita. Pero si la variable es discreta y toma muchos valores diferentes (tal vez infinitos) ó si es continua, lo anterior es poco recomendable e incluso imposible. Es por ello que necesitamos apoyarnos en una serie de funciones, relacionadas íntimamente con la Probabilidad, que nos permitan resolver el problema. Estas funciones serán la F. de Probabilidad en el caso discreto, la F. de Densidad en el continuo y la F. de Distribución, la F. Característica y la F. Generatriz de Momentos, entre otras, en ambos casos. En este apartado nos centraremos en las tres primeras.

**Caso Discreto.** Si una v.a. toma valores $x_1, x_2, \ldots, x_n, \ldots$, ( finito o infinito) la regla que asocia a cada uno de ellos las probabilidades

$p_1, p_2, \ldots, p_n, \ldots$ respectivamente, donde $p_i = P[X = x_i]$ se denomina Función de Probabilidad. Como la suma de todas las probabilidades de los sucesos elementales es uno, se tiene que:

$$\sum_{i=1}^{n/\infty} p_i = 1$$

Una v.a. queda perfectamente determinada cuando se conoce su función de probabilidad, pudiéndose expresar ésta de dos formas (utilizando el ejemplo anterior):

1. Por extensión, dando una tabla que asocia cada valor a su probabilidad.

| X | $P[X = x]$ |
|---|---|
| 0 | 0,2601 |
| 1 | 0,4998 |
| 2 | 0,2401 |

2. Dando una función que relacione $X_j$ con su probabilidad $p_j$.

$$P[X = x] = \binom{2}{x} 0,49^x \, 0,51^{2-x}$$

Otra forma de caracterizar una v.a. es a través de la llamada Función de Distribución, definida por:

$$F(x) = P[X < x]$$

La función de distribución en un valor $x$, es la probabilidad de que $X$ tome valores menores o iguales a $x$. Es decir, es una función que acumula toda la probabilidad entre menos infinito y el punto donde está definida. Para el ejemplo anterior tenemos:

$$\begin{aligned}
F(-0,5) &= F[X < -0,5] = 0 \\
F(0) &= P[X = 0] = 0{,}2601 \\
F(1) &= P[X = 1] + P[X = 0] = 0{,}7599 \\
F(2) &= P[X = 0] + P[X = 1] + P[X = 2] = 1 \\
F(2,5) &= 1
\end{aligned}$$

La función de distribución está definida en toda la recta real. Es creciente y no negativa. Toma el valor cero en menos infinito y el valor uno en más infinito. Sólo tiene discontinuidades de salto (precisamente en los puntos donde la función de probabilidad es distinta de cero).

**Caso Continuo.** Como se dijo en su presentación, la característica principal de la variable continua es que la probabilidad de que tome un determinado valor es cero. Sin embargo, aun siendo ello cierto, no todos los valores tienen "la misma ocurrencia". Podemos comprobar que hay bastantes individuos que tienen estatura alrededor de 160 cm., menos alrededor de 210 cm. y ninguno en el entorno de los 350 cm. Lo anterior hace que nos preguntemos, no ya por la probabilidad de tomar exactamente un valor, sino por la probabilidad de que la variable se encuentre en el entorno de un punto, es decir, que tome valores dentro de un intervalo. Así, consideremos un intervalo pequeño centrado en 160 cm., por ejemplo:

$$[160 - \triangle/2 \quad ; \quad 160 + \triangle/2] \ ,$$

con lo que la probabilidad de que $X$ tome algún valor del intervalo será pequeña pero no nula

$$P[160 - \triangle/2 \quad < \quad X \quad < \quad 160 + \triangle/2] \ .$$

Si a continuación se levanta sobre el intervalo considerado un rectángulo de altura h(160), tal que su área sea igual al valor de la probabilidad anterior tendremos:

$$\text{Área} = h(160) \cdot \triangle = P[160 - \triangle/2 \quad < \quad X \quad < \quad 160 + \triangle/2] \ ,$$

de ahí que la altura del rectángulo pueda asociarse a una densidad de probabilidad, por ser igual al cociente entre la probabilidad del intervalo y su tamaño.

Si seleccionamos ahora valores $x_i$ de la v.a. $X$ distantes entre sí una distancia $\triangle$ y se determinan las alturas $h(x_i)$ de los infinitos intervalos así obtenidos se obtendrá un histograma y, uniendo los centros de las caras superiores de los rectángulos, el correspondiente polígono de frecuencias (ver figura 3.23).

Es evidente que la suma de las áreas de los rectángulos debe ser uno, por ser igual a la probabilidad de que $X$ tome cualquier valor. Si hacemos que la anchura de los rectángulos tienda a cero, el polígono de frecuencias se transformará en una curva como la de la figura 3.24.

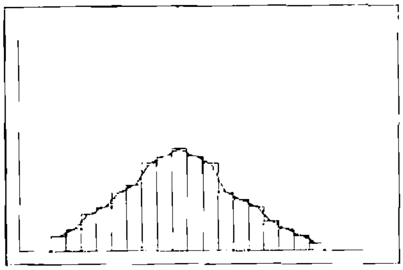



Figura 3.23: Áreas sobre un histograma.

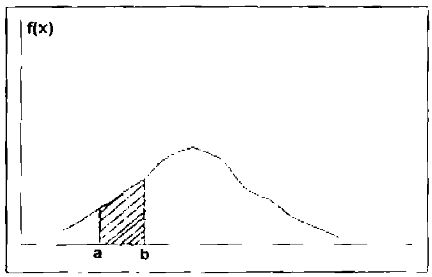



Figura 3.24: Área bajo la función de densidad entre $a$ y $b$.

Dicha curva es la representación gráfica de la función $f(x)$, denominada **Función de Densidad** de la v.a. $X$. obtenida como límite de los valores $h(x)$:

$$\lim_{\Delta\to 0} h(x) = \lim_{\Delta\to 0} \frac{P[x - \Delta/2 < X < x + \Delta/2]}{\Delta} = f(x)$$

Si deseamos calcular la probabilidad de que la variable se encuentre entre dos valores dados, (observe la figura 3.24) ésta será igual al área bajo la curva $f(x)$ entre $a$ y $b$. Es decir:

$$P[a \leq X \leq b] = \int_a^b f(x)dx$$

Lógicamente se verificará que:

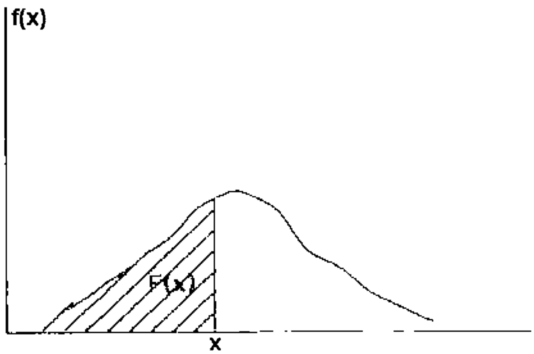
$$P[-\infty < X \leq +\infty] = \int_{-\infty}^{+\infty} f(x)dx = 1$$

Obviamente $f(x)$ es no negativa. Cualquier función no negativa que verifique la condición dada anteriormente será una función de densidad.

Al igual que en el caso discreto, se define la Función de Distribución $F(x)$ de una variable aleatoria $X$ como:

$$F(x) = \int_{-\infty}^{x} f(x)dx = P[X <$$

La función de distribución en $x$ nos da la probabilidad de que $X$ tome valores menores o iguales a $x$, gráficamente se puede ver en la figura 3.25.




ILUSTAC. 3

Figura 3.25: Área acumulada hasta el valor $x$. $F(x) = P[X < x]$

Las propiedades de la función de distribución son idénticas a las del caso discreto con la única excepción de que ahora ésta no presenta discontinuidades de salto. Para ciertas variables continuas de uso frecuente los valores de $F(x)$ se encuentran tabulados, lo cual facilita considerablemente el cálculo de probabilidades. Para terminar este epígrafe, vemos la relación existente entre las funciones definidas:

$$P[a < X < b] = P[a < X < b] = \int_a^b f(x)dx = F(b) - F(a) \ .$$

### 3.8.2 La función esperanza matemática: $E[X]$. Propiedades.

Si $X$ es una variable aleatoria discreta que toma los valores $x_1, x_2, \ldots, x_n, \ldots$ con probabilidades respectivas $p_1, p_2, \ldots . p_n \ldots$ definimos la **Esperanza** de

$X$ como:

$$E[X] = \sum_{i=1}^{n/\infty} p_i x_i$$

Si $g(X)$ es una función de la v.a. $X$, definimos la esperanza de $g(X)$ como:

$$E[g(X)] = \sum_{i=1}^{n/\infty} p_i g(x_i)$$

Si $X$ es una variable aleatoria continua con función de densidad $f(x)$, definimos la esperanza de $X$ como:

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Si $g(X)$ es una función de la v.a. $X$, definimos la esperanza de $g(X)$ como:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx .$$

En los dos últimos casos la definición será válida sólo si existe la integral. A partir de ahora y salvo que la situación así lo requiriera, nos limitaremos a expresar los resultados para variables continuas.

**Propiedades.**

Si $a$ es una constante cualquiera se verifica que:

$$1.\text{-}E[a \cdot X] = a \cdot E[X] \ .$$

$$2.\text{-}E[a + X] = a + E[X]$$

Ambas propiedades se pueden demostrar fácilmente sin más que calcular la esperanza matemática de las funciones $g(X) = a \cdot X$ y $g(X) = a + X$.

### 3.8.3 Funciones que caracterizan globalmente a una Variable Aleatoria.

A partir de la función esperanza que acabamos de definir introduciremos una serie de aplicaciones que nos ofrecerán visiones parciales de la distribución en estudio.

Si $X$ es una v.a. definimos sus momentos de orden $k$ respecto al origen y respecto a la media, respectivamente, como:

$$\alpha_k = E[X^k] \quad , \quad \mu_k = E[(X-\mu)^k]$$

## La media.

Particular importancia tienen el momento de orden 1 respecto al origen y el momento de orden 2 respecto a la media:

$$\alpha_1 = E[X] = \mu \quad , \quad Var[X] = \mu_2 = E[(X-\mu)^2] = \sigma^2 \ .$$

El primero de ellos es la Media, esperanza o valor esperado de $X$ y el segundo es la Varianza de $X$, coincidiendo su raíz cuadrada con la desviación típica.

Igual que en la Estadística Descriptiva podemos expresar los momentos respecto a la media en función de los momentos respecto al origen. En particular la varianza podría escribirse como:

$$\mathrm{Var}[X] = \mu_2 = \sigma^2 = E[X^2] - E[X]^2 = \alpha_2 - \alpha_1^2 \ .$$

Las siguientes son propiedades evidentes de la varianza (siendo $a$ constante):

$$1.\text{-}\mathrm{Var}[a \cdot X] = a^2 \cdot \mathrm{Var}[X] \ .$$

$$2.\text{-}\mathrm{Var}[a + X] = \mathrm{Var}[X] \ .$$

En este punto es interesante analizar cómo quedarían algunos de los conceptos que estudiamos en Estadística Descriptiva.

## La Moda.

Es el valor que más se repite, es decir, el de mayor probabilidad si la variable es discreta, o el de mayor densidad si es continua. En el primer caso la moda será el valor $x_i$, tal que $p_i$ es mayor o igual que $p_j$ para todo $j$ distinto de $i$.

Si la variable es continua, la moda coincidirá con el valor de $X$ que maximice la función de densidad, debiéndose utilizar los procedimientos analíticos de obtención de puntos óptimos. Como en el caso descriptivo, también aquí son válidas las nociones de modas múltiples y, por tanto, de modas relativas.

**La Mediana.**

Es el valor central de la distribución y coincidirá con el valor de $x$ tal que $F(x) = 0{,}5$. En el caso de variable discreta la similitud es total con la definición dada en Estadística Descriptiva.

**Coeficientes de Simetría y de Curtosis.**

Se definen respectivamente como:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \qquad ; \qquad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3.$$

Siendo la discusión sobre ambos coeficientes la misma que hacíamos en el caso descriptivo.

**Normalización o Tipificación.**

Decimos que una v.a. está tipificada o normalizada cuando su media vale cero y su desviación típica uno. Para tipificar una variable $X$ procederemos restándole su media y dividiéndola por su desviación típica

$$Z = \frac{X - \mu}{\sigma} .$$

### 3.8.4 La función característica y la función generatriz de momentos.

A partir de la esperanza matemática, podemos definir dos funciones que equivalen a las de probabilidad/densidad y de distribución y que, como ya vimos, caracterizan totalmente a la distribución de probabilidades; de tal forma, que conocida cualquiera de ellas quedarían perfectamente determinadas todas las demás y, por supuesto, la distribución. Estas funciones van a ser:

1. Función Característica.

$$\varphi_x(t) = E[e^{iXt}] = \int_{-\infty}^{+\infty} e^{ixt} f(x)\, dx$$

2. Función generatriz de Momentos.

$$m_x(t) = E[e^{Xt}] = \int_{-\infty}^{+\infty} e^{xt} f(x)\, dx \ .$$

Además de ser una función que caracteriza a la distribución, nos permite, como su propio nombre indica, obtener fácilmente cualquier momento de ésta. Así:

$$\alpha_k = \left[\frac{d^k m_x(t)}{dt^k}\right]_{t=0}$$

En sentido inverso, el conocimiento de todos los momentos nos daría la función generatriz de momentos.

### 3.8.5 Cambio de variable.

Sea $X$ una v.a. con función de densidad $f(x)$ y $Y = h(x)$ una función de $X$ creciente o decreciente, es decir biunívoca; entonces la función de densidad de la nueva variable $Y$ viene dada por:

$$g(y) = f(k(y)) \mid \frac{dx}{dy} \mid \ ,$$

donde $k(y)$ es la función inversa de $h(x)$.

### 3.8.6 La desigualdad de Tchebychev.

Si $X$ es una v.a. y $k$ cualquier constante positiva, se verifica que:

$$P[\mu - k\sigma \leq X \leq \mu + k\sigma] \geq 1 - \frac{1}{k^2}.$$

Esta expresión nos da la proporción mínima de valores de $X$ que se encuentran a una distancia de la media menor o igual a $k$ veces la desviación típica. Para $k$ igual a tres, por ejemplo, la desigualdad garantiza que al menos el 89% de la distribución está en el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$.

# 3.9 Variables aleatorias $n$-dimensionales.

## 3.9.1 Distribución conjunta y marginales.

Extenderemos las nociones vistas para una variable unidimensional al caso de variables $n$-dimensionales. En particular nos centraremos en el estudio de variables bidimensionales, siendo generalizable lo que aquí se diga para cualquier dimensión finita.

En un principio distinguiremos entre variables discretas y continuas, aunque más adelante sólo consideraremos las continuas.

**1.- Variable Discreta.**

Consideremos (X;Y) v.a. discreta con su correspondiente probabilidad para cada par de valores. Definimos:

$$P(X = x; Y = y) = f(x; y).$$

La función $f(x; y)$, que llamamos función de probabilidad, verifica:

1. $$f(x; y) \geq 0 \qquad \forall\,(x; y)\ .$$

2. $$\sum_x \sum_y f(x; y) = 1$$

Si $X$ toma los valores $x_1, x_2, \ldots, x_m$ e $Y$ los valores $y_1, y_2, \ldots, y_n$ entonces podemos expresar la función de probabilidad conjunta a través de la siguiente tabla:

| X / Y | $y_1$ | $y_2$ | | $y_k$ | | $y_n$ | |
|---|---|---|---|---|---|---|---|
| $x_1$ | $f(x_1; y_1)$ | $f(x_1; y_2)$ | $\ldots$ | $f(x_1; y_k)$ | $\ldots$ | $f(x_1; y_n)$ | $f_1(x_1)$ |
| $x_2$ | $f(x_2; y_1)$ | $f(x_2; y_2)$ | $\ldots$ | $f(x_2; y_k)$ | $\ldots$ | $f(x_2; y_n)$ | $f_1(x_2)$ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ |
| $x_i$ | $f(x_i; y_1)$ | $f(x_i; y_2)$ | $\ldots$ | $f(x_i; y_k)$ | $\ldots$ | $f(x_i; y_n)$ | $f_1(x_i)$ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ |
| $x_m$ | $f(x_m; y_1)$ | $f(x_m; y_2)$ | $\ldots$ | $f(x_m; y_k)$ | $\ldots$ | $f(x_m; y_n)$ | $f_1(x_m)$ |
| | $f_2(y_1)$ | $f_2(y_2)$ | $\ldots$ | $f_2(y_k)$ | | $f_2(y_n)$ | 1 |

Donde en los márgenes derecho e inferior se han obtenido las distribuciones marginales de $X$ e $Y$ respectivamente. Así, la probabilidad de que $X$ tome un valor genérico será:

$$P[X = x_i] = \sum_k P[X = x_i; Y = y_k] = \sum_k f(x_i; y_k) = f_1(x_i) \ .$$

De igual forma se haría para $Y$. Siendo $f_1$ y $f_2$ las funciones de probabilidad marginales de $X$ e $Y$ respectivamente. La función de distribución vendrá dada por:

$$y) = P[X < x \cdot Y < y] = \quad \sum_{u<x} \ \sum_{v<y} \quad v)$$

## 2.- Variable Continua.

Sea una v.a. continua $(X;Y)$, si $f$ verifica:

$$1. \qquad f(x;y) > 0 \quad \forall \quad (x;y) \ ,$$

$$2. \qquad \int_{\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x \cdot y) \, dxdy = 1 \ ,$$

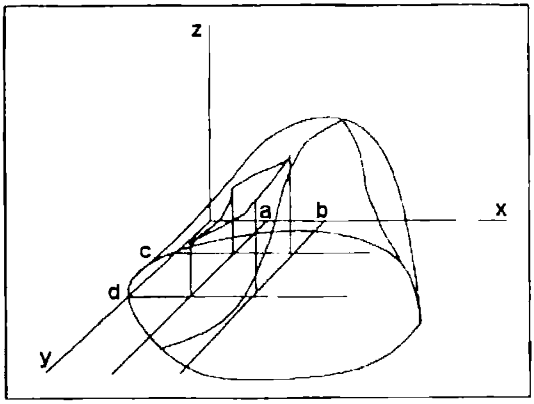
decimos que $f$ es una función de densidad conjunta. Sea $z = f(x;y)$ que representa una superficie de probabilidad o densidad, de tal forma que el área encerrada entre la superficie $z$ y el plano $XY$ vale la unidad. Puede verse ésto en la figura 3.26.

La probabilidad de que la variable aleatoria tome valores dentro del rectángulo $(a;b)$ x $(c;d)$ viene dada por:

$$P[a < x < b; c < y < d] = \int_{x=a} \int_{y=c} f(x;y) \, dxdy$$

Si $A$ representa cualquier suceso y $R_A$ la región del plano $XY$ que corresponde a $A$. Definimos su probabilidad como:

$$= \int \int_{R_A} f(x.y) \, dxdy \ .$$




ILUSTRAC. 4

Figura 3.26: Función de densidad bidimensional.

La función de distribución conjunta vendrá dada por:

$$F(x;y) = P(-\infty < X \leq x; -\infty < Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u;v)\, dudv \ .$$

La relación entre la función de distribución $F$ y la función de densidad $f$ será:

$$\frac{\partial^2 F}{\partial x \partial y} = f(x;y) \ .$$

Las funciones de distribución marginales serán:

$$F_1(x) = P[X \leq x] = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(u;v)\, dudv \ .$$

$$F_2(y) = P[Y \leq y] = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f(u;v)\, dudv \ .$$

Derivando obtenemos las correspondientes funciones de densidad marginales:

$$f_1(x) = \frac{\partial F_1(x)}{\partial x} = \int_{-\infty}^{+\infty} f(u;v)\, dv \ .$$

$$f_2(x) = \frac{\partial F_2(y)}{\partial y} = \int_{-\infty}^{+\infty} f(u;v)\, du \ .$$

### 3.9.2 Distribuciones condicionadas. Independencia.

Nos limitaremos en este epígrafe a dar los resultados para una v.a. continua. Los correspondientes a variable discreta se obtienen de forma inmediata intercambiando los conceptos homólogos.

Definimos la función de densidad condicionada de $X$ para un valor dado de $Y$ como:

$$f(x/y) = \frac{f(x;y)}{f_2(y)} ,$$

siempre que $f_2(y)$ sea distinto de cero. De la forma definida, el lector puede comprobar que efectivamente se trata de una función de densidad. Su correspondiente función de distribución tiene la forma

$$F(x/y) = \frac{\int_{-\infty}^{x} f(x;y)\, dx}{f_2(y)}$$

De igual manera obtendríamos las condicionadas de $Y$ con respecto a $X$.

De las relaciones siguientes (todas ellas inmediatas)

$$f(x;y) = f(y/x) \cdot f_1(x) ,$$

$$f(x/y) = \frac{f(y/x)}{f_2(y)} \cdot f_1(x) ,$$

$$f_2(y) = \int f(y/x) f_1(x)\, dx ,$$

podemos deducir que

$$f(x/y) = \frac{f(y/x) f_1(x)}{\int f(y/x) f_1(x) dx}$$

Que es la expresión del **Teorema de Bayes** para funciones de densidad.

#### Independencia.

Diremos que las variables $X$ e $Y$ son independientes si las distribuciones condicionadas coinciden con las marginales, es decir:

$$f(x/y) = f_1(x) \quad \equiv \quad f(y/x) = f_2(y)$$

Igual que siempre la condición de independencia se establece en un doble sentido. El que las variables sean independientes implica que el conocimiento de que una variable toma un determinado valor o está contenida en un rango de valores no da ninguna información sobre los valores de la otra variable. De la definición de distribución condicionada podemos deducir que dos variables $X$ e $Y$ son independientes si y sólo si:

$$f(x;y) = f_1(x) \cdot f_2(y) \ .$$

Esta definición de independencia se puede extender a cualquier conjunto finito de variables, y así, las variables $X_1, X_2, ..., X_n$ son independientes si se verifica:

$$f(x_1;x_2;\ldots:x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n) \ .$$

La independencia conjunta implica la de cualquier subconjunto de variables de las $n$ dadas; en contra de lo que pasaba en la relación entre sucesos, donde era necesario exigir la condición de independencia para cualquier combinación de sucesos.

### 3.9.3 Esperanza. Propiedades. Covarianza y correlación.

Sea $g(X;Y)$ una función de la v.a. $(X;Y)$ de función de densidad $f(x;y)$ y función de distribución $F(x;y)$. Definimos la esperanza matemática de $g(X;Y)$ como:

$$E[g(X;Y)] = \int_{-\infty}^{+\infty} g(x;y)\, dF(x;y) = \int_{-\infty}^{+\infty} g(x;y)f(x;y)\, dxdy \ ,$$

siempre que dicha integral exista. Especial importancia tienen los casos en que g(X;Y) toma las formas:

1. $g(X;Y) = X^r \cdot Y^s$ en donde la esperanza recibe el nombre de momento de orden $r, s$ respecto al origen y la notamos $\alpha_{r,s}$.

2. $g(X;Y) = (X - \alpha_{1,0})^r \cdot (Y - \alpha_{0,1})^s$ en cuyo caso tendremos el momento de orden $r, s$ respecto a la media. Que se identifica por $\mu_{rs}$.

Podemos comprobar sin demasiada dificultad que:

$$\alpha_{1,0} = E[X] \quad : \quad \alpha_{0,1} = E[Y]$$

$$\mu_{1,0} = \mu_{0,1} = 0$$
$$\mu_{2,0} = \sigma_x^2 \quad ; \quad \mu_{0,2} = \sigma_y^2 \ .$$

Además $\mu_{1,1} = \sigma_{x,y}$ será la covarianza de $(X;Y)$, definiéndose el coeficiente de correlación lineal como:

$$\rho = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

La definición de esperanza podemos generalizarla al caso de un número finito de variables, siendo su expresión:

$$E[g(X_1;X_2;...;X_n)] = \int ... \int g(x_1;x_2;...;x_n) f(x_1;x_2;...;x_n)\, dx_1 dx_2 ... dx_n.$$

como siempre, si existe dicha integral.

**Propiedades.**

1. $E[X_1 + X_2 + ... + X_n] = E[X_1] + E[X_2] + ... + E[X_n]$
2. Si las variables son independientes:
$$E[X_1 X_2 ... X_n] = E[X_1] \cdot E[X_2] \cdot \ldots \cdot E[X_n] \ .$$
3. Si las variables son independientes, su covarianza es nula. El recíproco no es cierto en general.
4. Si $Z = aX + b$ y $T = cY + d$, tendremos que: $\sigma_{Z,T} = a \cdot c \cdot \sigma_{X,Y}$ .
5. $| \rho_{X,Y} | \leq 1$ .
6. Si $Y = aX + b$ entonces $| \rho_{X,Y} | = 1$ , y su signo es igual al signo de $a$.
7. Si $Z = X \pm Y$, entonces $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{X,Y}$ .
   En particular, si $X, Y$ son independientes tenemos: $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

Llamaremos **Vector de medias** de la variable aleatoria $n$-dimensional $X' = (X_1;X_2;...;X_n)$ al vector cuyas componentes son las esperanzas de cada una de las $X_i$ que representaremos por $\mu = E[X]$. Llamaremos **Matriz de varianzas y covarianzas** de la variable aleatoria $n$-dimensional $X' = (X_1;X_2;...;X_n)$ a la matriz cuadrada de orden $n$ dada por:

$$M_X = E[(X - \mu) \cdot (X - \mu)' \,].$$

Obviamente dicha matriz contiene en la diagonal a las varianzas de las variables, y en el resto a las covarianzas entre ellas. La matriz de varianzas y covarianzas es, por tanto, simétrica y además semidefinida positiva.

### 3.9.4 Cambio de variables.

Sea $(X;Y)$ una v.a. con función de densidad $f(x;y)$. Sean $Z = g_1(X;Y)$ y $T = g_2(X;Y)$ una transformación continua y biunívoca definida por dos funciones $g_1$ y $g_2$ que admiten parciales continuas. Entonces la función de densidad de la nueva variable $(Z;T)$ existe en todos aquellos puntos donde el determinante:

$$J = \frac{\partial(z,t)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial t}{\partial x} & \frac{\partial t}{\partial y} \end{vmatrix}$$

sea distinto de cero. Siendo dicha función igual a:

$$g(z;t) = f(h_1(z;t); h_2(z;t)) \mid J_1 \mid .$$

Donde $h_1$ y $h_2$ son las funciones inversas de $g_1$ y $g_2$. Y $J_1$ viene definido por:

$$J_1 = \frac{\partial(x;y)}{\partial(z;t)}.$$

Es inmediata la generalización al caso de $n$ variables.

# Problemas Resueltos

## Problema 3.1

**Demostrar, a partir de la axiomática de Kolmogorov las siguientes propiedades:**

1. $P(\emptyset) = 0$
2. **Si** $A, B \in \mathcal{A}$ **con** $A \subset B$ **entonces** $P(B - A) = P(B) - P(A)$
3. **Si** $A, B \in \mathcal{A}$ **con** $A \subset B$ **entonces** $P(A) \leq P(B)$
4. $P(A) \leq 1 \qquad \forall A \in \mathcal{A}$
5. $P(\overline{A}) = 1 - P(A)$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$

Demostremos estas propiedades a partir de la axiomática de Kolmogorov:

1. $\emptyset = \emptyset \cup \emptyset$ aplicando A3 tenemos $P(\emptyset) = P(\emptyset) + P(\emptyset) \Longrightarrow P(\emptyset) = 0$.
2. $B = (B - A) \cup A \Longrightarrow P(B) = P(B - A) + P(A)$.
3. Se deduce de P2, aplicando que, por A1 se cumple que $P(B - A) \leq 0$.
4. $A \subset \Omega$, aplicando P3 se tiene que $P(A) \leq P(\Omega) = 1$.
5. $\overline{A} = \Omega - A$ y se aplica P2.
6. $A \cup B = A \cup (B - (A \cap B))$ y se aplica P2.

7. Se demuestra por inducción sobre $n$ (el caso $n = 2$ es la propiedad anterior P6).

## Problema 3.2

**A partir de restos de distintas barajas españolas se consigue formar un mazo de 20 cartas, de las que 8 son de oros y el resto de copas. Se extraen tres cartas de la baraja con reemplazamiento. Se pide:**

1. **Hallar la probabilidad de que la primera carta sea de oros.**
2. **Hallar la probabilidad de que la segunda carta sea de oros.**
3. **Hallar la probabilidad de que la segunda carta sea de oros sabiendo que la primera fue de oros.**
4. **Hallar la probabilidad de que las dos primeras cartas sean de oros.**
5. **Hallar la probabilidad de que las tres cartas sean de oros.**
6. **Hallar la probabilidad de que ninguna carta sea de oros.**
7. **Hallar la probabilidad de que al menos una sea de oros.**

**Determinar, en primer lugar, el espacio muestral $\Omega$ apropiado, determinando para los distintos sucesos que sean de interés con qué subconjuntos de $\Omega$ se corresponden.**

---

Teniendo en cuenta que extraemos tres cartas y que de ellas sólo nos interesa el hecho de que sean o no de oros, podemos considerar el siguiente espacio muestral:

$$\Omega = \{(o,o,o),(o,o,c),(o,c,o),(o,c,c),(c,o,o),(c,o,c),(c,c,o),(c,c,c)\}$$

donde (o,c,o) por ejemplo, indica que la primera carta es de oros, la segunda de copas y la tercera de oros.

Definiremos los siguientes sucesos

$$A_1 = \{\text{La primera carta es de oros}\}$$
$$A_2 = \{\text{La segunda carta es de oros}\}$$

$$A_3 = \{\text{La tercera carta es de oros}\}$$

que se corresponderán con los siguientes subconjuntos de $\Omega$

$$A_1 = \{(o,o,o),(o,o,c),(o,c,o),(o,c,c)\}$$

$$A_2 = \{(o,o,o),(o,o,c),(c,o,o),(c,o,c)\}$$

$$A_3 = \{(o,o,o),(o,c,o),(c,o,o),(c,c,o)\}$$

1. El primer apartado nos pide calcular la probabilidad de $A_1$. Como del total de 20 cartas hay 8 de oros, aplicando la regla de Laplace tendríamos que $P[A_1] = \frac{8}{20}$.

2. Lo que nos pide este segundo apartado es la probabilidad de $A_2$. Como al realizar la segunda extracción se ha vuelto a introducir la carta primera en la baraja, estamos en las mismas circunstancias del apartado anterior, y por tanto del total de 20 cartas hay 8 de oros. Aplicando de nuevo la regla de Laplace tendríamos que $P[A_2] = \frac{8}{20}$.

3. Lo que nos pide este tercer apartado es la probabilidad de $A_2$ condicionada a $A_1$. Como hemos razonado en el apartado anterior, al extraer la segunda carta estamos en las mismas circunstancias que al extraer la primera, no aportándonos ningún tipo de información adicional el conocer qué ocurrió en la extracción anterior. Tendríamos, por tanto, $P[A_2/A_1] = \frac{8}{20} = P[A_2]$. Los dos sucesos son, por tanto, independientes.

4. Si queremos calcular la probabilidad de que tanto la primera como la segunda carta sean de oros, lo que tendremos que calcular es la probabilidad del suceso $A_1 \cap A_2$ que se corresponde con los elementos del espacio muestral

   $$A_1 \cap A_2 = \{(o,o,o),(o,o,c)\}.$$

   Como los dos sucesos son independientes, podemos aplicar que la probabilidad de la intersección es el producto de las probabilidades.

   $$P[A_1 \cap A_2] = P[A_1]P[A_2] = \frac{8}{20}\frac{8}{20} = \frac{4}{25}.$$

5. Para calcular la probabilidad de que las tres cartas sean de oros habrá que determinar la probabilidad de $A_1 \cap A_2 \cap A_3 = \{(o, o, o)\}$. Como las extracciones son con reemplazamiento. las tres se realizan en idénticas condiciones, siendo por tanto los tres sucesos independientes. Se tiene, por tanto:

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3] = \frac{8}{20}\frac{8}{20}\frac{8}{20} = \frac{8}{125}.$$

6. Para hallar la probabilidad de que ninguna carta sea de oros consideramos el suceso $\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} = \{(c, c, c)\}$. como estos tres sucesos también son independietes tendríamos

$$P[\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}] = P[\overline{A_1}]P[\overline{A_2}]P[\overline{A_3}] = \frac{12}{20}\frac{12}{20}\frac{12}{20} = \frac{27}{125}.$$

7. El suceso "al menos una carta sea de oros" es el complementario del suceso "ninguna carta sea de oros". Tenemos por tanto:

$$P[\overline{\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}}] = 1 - P[\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}] = 1 - \frac{12}{20}\frac{12}{20}\frac{12}{20} = 1 - \frac{27}{125} = \frac{98}{125}$$

# Problema 3.3

**Realizar el problema anterior considerando que las extracciones se realizan sin reemplazamiento.**

1. El primer apartado nos pide calcular la probabilidad de $A_1$. Este apartado es exactamente igual que en el problema anterior. Como del total de 20 cartas hay 8 de oros, aplicando la regla de Laplace tendríamos que $P[A_1] = \frac{8}{20}$.

2. Lo que nos pide este segundo apartado es la probabilidad de $A_2$. Como al realizar la segunda extracción no se ha vuelto a introducir la carta primera en la baraja, habrá que tener en cuenta si la primera fue de oros o no para calcular la probabilidad. Considerando los sucesos $A_1$ y $\overline{A_1}$ forman un sistema exhaustivo y excluyente. Aplicando el teorema de la probabilidad total tendríamos

$$P[A_2] = P[A_2/A_1]P[A_1] + P[A_2/\overline{A_1}]P[\overline{A_1}]$$

calculemos cada una de estas probabilidades:

$$\begin{array}{llll} P[A_1] = \frac{8}{20} & & P[\overline{A_1}] = \frac{12}{20} \\ P[A_2/A_1] = \frac{7}{19} & & P[A_2/\overline{A_1}] = \frac{8}{19} \end{array}$$

(La probabilidad $P[A_2/A_1]$ se calcula en el siguiente apartado, y de manera análoga se calcula $P[A_2/\overline{A_1}]$ ).

Sustituyendo en la fórmula del teorema de la probabilidad total:

$$P[A_2] = P[A_2/A_1]P[A_1] + P[A_2/\overline{A_1}]P[\overline{A_1}] = \frac{7}{19}\frac{8}{20} + \frac{8}{19}\frac{12}{20} = \frac{8}{20}.$$

3. Lo que nos pide este tercer apartado es la probabilidad de $A_2$ condicionada a $A_1$. Si la primera carta fue de oros, al extraer la segunda quedan 19 cartas de las que 7 son de oros, por tanto, aplicando la regla de Laplace:

$$P[A_2/A_1] = \frac{7}{19}.$$

4. Si queremos calcular la probabilidad de $A_1 \cap A_2$, como los dos sucesos no son independientes, no podemos aplicar que la probabilidad de la intersección es el producto de las probabilidades. Tendremos que aplicar la siguiente fórmula:

$$P[A_1 \cap A_2] = P[A_1]P[A_2/A_1] = \frac{8}{20}\frac{7}{19} = \frac{14}{95}.$$

5. Para calcular la probabilidad de que las tres cartas sean de oros habrá que determinar la probabilidad de $A_1 \cap A_2 \cap A_3$. Como las extracciones no son con reemplazamiento, los tres sucesos no son independientes. Aplicaremos, por tanto:

$$P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2/A_1]P[A_3/A_1 \cap A_2] = \frac{8}{20}\frac{7}{19}\frac{6}{18} = \frac{14}{285}.$$

6. Para hallar la probabilidad de que ninguna carta sea de oros consideramos el suceso $\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} = \{(c, c, c)\}$. y entonces:

$$P[\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}] = P[\overline{A_1}]P[\overline{A_2}/\overline{A_1}]P[\overline{A_3}/\overline{A_1} \cap \overline{A_2}] = \frac{12}{20}\frac{11}{19}\frac{10}{18} = \frac{11}{57}.$$

7. El suceso de que al menos una carta sea de oros es el complementario del suceso de que ninguna carta sea de oros. Tenemos por tanto:

$$P[\overline{\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}}] = 1 - P[\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}] = 1 - \frac{12}{20}\frac{11}{19}\frac{10}{18} = 1 - \frac{11}{57} = \frac{46}{57}$$

# Problema 3.4

**En una Escuela hay matriculados 550 alumnos en primero, 300 en segundo y 150 en tercero. (Se cuenta cada alumno solamente en el curso inferior de todas las asignaturas que tenga). El 70% de los alumnos de primero está matriculado en más de 8 asignaturas. El 90% de los alumnos de segundo está matriculado en más de 8 asignaturas. El 30% de los alumnos de tercero está matriculado en más de 8 asignaturas. Se pide:**

1. **Hallar la probabilidad de que un alumno esté matriculado en más de 8 asignaturas.**
2. **Hallar la probabilidad de que un alumno de primero esté matriculado en más de 8 asignaturas.**
3. **Hallar la probabilidad de que un alumno esté matriculado en más de 8 asignaturas y sea de primero.**
4. **Hallar la probabilidad de que un alumno que esté matriculado en más de 8 asignaturas sea de primero.**

Definimos los siguientes sucesos:

$P = \{$El alumno es de primero $\}$.
$S = \{$El alumno es de segundo $\}$.
$T = \{$El alumno es de tercero $\}$.
$M = \{$El alumno está matriculado en más de 8 asignaturas $\}$.

Sabemos que los sucesos $P, S$ y $T$ son mutuamente excluyentes y que su unión nos da el total. También conocemos las siguientes probabilidades:

$$\begin{array}{lclclcl} \mathrm{P(P)} & = & 0{,}55 & \quad & \mathrm{P(M/P)} & = & 0{,}7 \\ \mathrm{P(S)} & = & 0{,}3 & & \mathrm{P(M/S)} & = & 0{,}9 \\ \mathrm{P(T)} & = & 0{,}15 & & \mathrm{P(M/T)} & = & 0{,}3. \end{array}$$

1. Para hallar la probabilidad de que un alumno esté matriculado en más de 8 asignaturas, aplicamos el teorema de la probabilidad total:

$$P(M) = P(M/P))P(P) + P(M/S)P(S) + P(M/T)P(T)$$

de donde:

$$P(M) = 0{,}7 \cdot 0{,}55 + 0{,}9 \cdot 0{,}3 + 0{,}3 \cdot 0{,}15 = 0{,}385 + 0{,}27 + 0{,}045 = 0{,}7.$$

2. La probabilidad de que un alumno de primero esté matriculado en más de 8 asignaturas nos la da el enunciado y es

$$P(M/P) = 0{,}7.$$

3. Para hallar la probabilidad de que un alumno esté matriculado en más de 8 asignaturas y sea de primero aplicamos

$$P(M \cap P) = P(M/P)P(P) = 0{,}7 \cdot 0{,}55 = 0{,}385.$$

4. La probabilidad de que un alumno que esté matriculado en más de 8 asignaturas sea de primero, aplicando el teorema de Bayes es

$$P(P/M) = \frac{P(M \cap P)}{P(M)} = \frac{0{,}385}{0{,}7} = 0.55.$$

# Problema 3.5

**Un estudiante contesta una pregunta que ofrece cuatro soluciones posibles en un examen de opción múltiple. Suponga que la probabilidad de que el estudiante sepa la respuesta a la pregunta es de 0,8 y la probabilidad de que tenga que contestar al azar es 0,2. Suponga además que la probabilidad de seleccionar la respuesta correcta al azar es 0,25. Si el estudiante contesta correctamente la pregunta, ¿cuál es la probabilidad de que realmente sepa la respuesta correcta?**

---

Si denotamos los sucesos $A, \bar{A}$ y $C$ como

$A=$ {saber la respuesta},

$\bar{A}=${no saber la respuesta, o responder al azar},

$C=${responder correctamente la pregunta}.

Entonces tenemos $P(A) = 0{,}8$, y por tanto $P(\bar{A}) = 0{,}2$.

Además $P(C/A) = 1$, y la probabilidad $P(C/\bar{A}) = 1/4$.

Usando la definición de probabilidad condicionada, y el teorema de Bayes, tenemos:

$$\begin{aligned} P(A/C) &= \frac{P(A \cap C)}{P(C)} = \frac{P(C/A) \cdot P(A)}{P(C/A) \cdot P(A) + P(C/\bar{A}) \cdot P(\bar{A})} = \\ &= \frac{1 \cdot 0{,}8}{1 \cdot 0{,}8 + 0{,}25 \cdot 0{,}2} = \frac{0{,}8}{0{,}85} = 0{,}94117 \end{aligned}$$

# Problema 3.6

**Hay 23 estudiantes en una clase. ¿Cuál es la probabilidad de que el cumpleaños de por lo menos dos de ellos sea en la misma fecha (día y mes)? Suponer que el año tiene 365 días.**

Si llamamos a los sucesos $A$ y $\bar{A}$ como

$A$= {coincidencia en alguna fecha de cumpleaños},

$\bar{A}$={no hay ninguna coincidencia en las fechas}.

Entonces la probabilidad del suceso $\bar{A}$ se puede obtener:

$$P(\bar{A}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-21}{365} \cdot \frac{365-22}{365}$$

lo que equivale a

$$P(\bar{A}) = \frac{365!}{(365-23)! \cdot 365^{23}} = 0{,}492702766$$

de donde

$$P(A) = 1 - 0{,}492702766 = 0{,}507297234.$$

# Problema 3.7

**Juan es el responsable de un aula de informática de la UCA y no se puede confiar en él: la probabilidad de que olvide hacer el mantenimiento a un PC en ausencia de su jefe es $\frac{2}{3}$.**

**El PC no está seguro: si Juan le hace un mantenimiento tiene la misma probabilidad de estropearse que de funcionar correctamente, pero solamente un 0,25 de probabilidad de que funcione correctamente si no le hace el mantenimiento.**

**1. ¿Cuál es la probabilidad de que un ordenador funcione correctamente a la vuelta del jefe?**

**2. A su regreso, el jefe de Juan encuentra un PC averiado, ¿cuál es la probabilidad de que no le hiciera el mantenimiento?**

---

Si llamamos a los sucesos $M$ y $F$ como

$M = \{\text{Hace el mantenimiento de PC}\}$ y $F = \{\text{El PC funciona}\}$

Entonces tenemos $P(\bar{M}) = 2/3$, y por tanto $P(M) = 1/3$.

Además $P(F/M) = P(\bar{F}/M) = 1/2$, la probabilidad

$P(F/\bar{M}) = 1/4$ y la probalilidad $P(\bar{F}/\bar{M}) = 3/4$.

1. $P(F) = P(F/M) \cdot P(M) + P(F/\bar{M}) \cdot P(\bar{M}) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{3}$
   Y por tanto $P(\bar{F}) = \frac{2}{3}$.

2. $P(\bar{M}/\bar{F}) = \dfrac{P(\bar{F}/\bar{M}) \cdot P(\bar{M})}{P(\bar{F})} = \dfrac{\frac{3}{4} \cdot \frac{2}{3}}{\frac{2}{3}} = \dfrac{3}{4}.$

# Problema 3.8

**Sea $X$ una variable aleatoria discreta con función de probabilidad**

$$P[X = k] = \alpha k \qquad k \in \{1, 2, 3, 4\}$$

**hallar el valor de $\alpha$ para que $P$ sea efectivamente una función de probabilidad. Calcular la media y la varianza de $X$.**

---

Para que se trate de una función de probabilidad debe verificarse que todos los valores sean no negativos y que la suma sea la unidad.

Para que cumplan la condición de que los valores sean no negativos debe verificarse $\alpha > 0$. Para que cumplan la otra condición debe verificarse

$$P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = 1$$

que equivale a

$$\alpha + 2\alpha + 3\alpha + 4\alpha = 1$$

de donde obtenemos $\alpha = 1/10$.

Calculemos la media:

$$E[X] = 1 \cdot P[X = 1] + 2 \cdot P[X = 2] + 3 \cdot P[X = 3] + 4 \cdot P[X = 4]$$

con lo que

$$E[X] = 1/10 + 4/10 + 9/10 + 16/10 = 3.$$

Para calcular la varianza calculemos en primer lugar $E[X^2]$:

$$E[X^2] = 1^2 \cdot P[X = 1] + 2^2 \cdot P[X = 2] + 3^2 \cdot P[X = 3] + 4^2 \cdot P[X = 4]$$

con lo que

$$E[X^2] = 1/10 + 8/10 + 27/10 + 64/10 = 10.$$

de donde

$$\text{var}(X) = E[X^2] - E[X]^2 = 10 - 3^2 = 1.$$

# Problema 3.9

**Sea $X$ una variable aleatoria discreta con función de probabilidad**

$$P[X = k] = \alpha k \qquad k = 1, 2 \ldots 100$$

**hallar el valor de $\alpha$ para que $P$ sea efectivamente una función de probabilidad. Calcular la media y la varianza de $X$.**

---

Para que se trate de una función de probabilidad debe verificarse que todos los valores sean no negativos y que la suma sea la unidad. Para que cumplan la condición de que tome valores no negativos debe verificarse $\alpha > 0$. Para que cumplan la otra condición debe verificarse

$$\sum_{k=1}^{100} P[X = k] = 1$$

que equivale a

$$\sum_{k=1}^{100} \alpha k = 1$$

calculando la suma

$$\alpha \sum_{k=1}^{100} k = \alpha \frac{100(100+1)}{2} = 1$$

de donde obtenemos $\alpha = 1/5050$.

Calculemos la media:

$$E[X] = \sum_{k=1}^{100} kP[X = k] = \alpha \sum_{k=1}^{100} k^2 = \alpha \cdot \frac{100(100+1)(200+1)}{6} = 67$$

Para calcular la varianza calculemos en primer lugar $E[X^2]$:

$$E[X^2] = \sum_{k=1}^{100} k^2 P[X = k] = \alpha \sum_{k=1}^{100} k^3 = \alpha \cdot \frac{100^2(100+1)^2}{4} = 5050.$$

de donde

$$\mathrm{var}(X) = E[X^2] - E[X]^2 = 5050 - 67^2 = 561.$$

# Problema 3.10

**Sea $X$ una variable aleatoria continua con función de densidad**

$$f(x) = \alpha x^2 \qquad x \in [0,1] \qquad (\text{ 0 en el resto})$$

**hallar el valor de $\alpha$ para que $f$ sea efectivamente una función de densidad. Hallar la función de distribución. Calcular la media y la varianza de $X$.**

Para que se trate de una función de densidad debe verificarse que tome valores no negativos y que la integral sea la unidad. Para que cumpla la condición de ser no negativa debe verificarse $\alpha > 0$. Para que cumpla la otra condición debe verificarse

$$\int_0^1 \alpha x^2 dx = 1$$

que equivale a

$$\alpha \cdot \left[\frac{x^3}{3}\right]_0^1 = 1 \qquad \Longrightarrow \qquad \alpha\left[\frac{1}{3} - 0\right] = 1 \qquad \Longrightarrow \qquad \alpha = 3.$$

La función de distribución será:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ \int_0^x f(t)dt = x^3 & \text{si} \quad 0 \leq x < 1 \\ 1 & \text{si} \quad x \geq 1 \end{cases}$$

Calculemos la media:

$$E[X] = \int_0^1 xf(x)dx = \int_0^1 3x^3 dx = \left[\frac{3x^4}{4}\right]_0^1 = 3/4.$$

Para calcular la varianza calculemos en primer lugar $E[X^2]$:

$$E[X^2] = \int_0^1 x^2 f(x)dx = \int_0^1 3x^4 dx = \left[\frac{3x^5}{5}\right]_0^1 = 3/5.$$

de donde $\quad \text{var}(X) = E[X^2] - E[X]^2 = 3/5 - (3/4)^2 = 0{,}0375.$

# Problema 3.11

**Consideremos una variable aleatoria bidimensional con función de densidad conjunta:**

$$f(x,y) = \begin{cases} 2kx + 2y & \textbf{si} \quad 0 < x < 1 \quad ; \quad 2 < y < 3 \\ 0 & \textbf{en el resto.} \end{cases}$$

**Encontrar el valor de $k$ para que $f$ sea función de densidad.**

Para calcular el valor de $k$ imponemos la condición de que la integral valga la unidad (la condición de que siempre sea positiva se comprobará después).

$$\int_2^3 \int_0^1 (2kx + 2y)dxdy = 1$$

veamos el valor de la integral

$$\begin{aligned} \int_2^3 \left[kx^2 + 2yx\right]_{x=0}^1 dy &= \int_2^3 (k + 2y)dy = \left[ky + y^2\right]_{y=2}^3 = \\ &= 3k + 9 - 2k - 4 = k + 5 \end{aligned}$$

por tanto

$$k + 5 = 1 \Rightarrow k = -4.$$

Veamos si se cumple la segunda condición, para ello deberá verificarse que

$$-8x + 2y \geq 0 \qquad \forall x \in (0,1) \quad \forall y \in (2,3)$$

pero esto no se verifica, ya que por ejemplo para $y = 2$ y $x = 1$ se tendría

$$-8x + 2y = -8 + 4 = -4 < 0.$$

Por tanto $f$ no es función de densidad para ningún valor de $k$.

# Problema 3.12

**Consideremos una variable aleatoria bidimensional con función de densidad conjunta:**

$$f(x,y) = \begin{cases} 2kx + 2ky & \textbf{si } 0 < x < 1 \quad 2 < y < 3 \\ 0 & \textbf{en el resto} \end{cases}$$

Se **pide:**

1. **Calcular el valor de $k$ para que $f$ sea una función de densidad**
2. **Calcular la función de distribución conjunta.**
3. **Calcular las funciones de densidad y de distribución marginales de $X$ y de $Y$.**
4. **Calcular las funciones de densidad y de distribución condicionadas.**
5. **Calcular las medias de $X$ y de $Y$.**
6. **Calcular las varianzas de $X$ y de $Y$.**
7. **Calcular la covarianza.**
8. **Estudiar si las dos variables son independientes o incorreladas.**

---

Calculemos en primer lugar el valor de $k$. Para ello imponemos la condición de que la integral valga la unidad (la condición de que siempre sea positiva se comprobará después).

$$\int_2^3 \int_0^1 (2kx + 2ky)dxdy = 1$$

Veamos el valor de la integral

$$\int_2^3 \left[kx^2 + 2kyx\right]_{x=0}^1 dy = \int_2^3 (k+2ky)dy = \left[ky + ky^2\right]_{y=2}^3 =$$

$$= 3k + 9k - 2k - 4k = 6k$$

por tanto

$$6k = 1 \Rightarrow k = 1/6.$$

Veamos si se cumple la segunda condición, para ello deberá verificarse que

$$x/3 + y/3 \geq 0 \qquad \forall x \in (0,1) \quad \forall y \in (2,3)$$

pero esto es evidente que siempre se verifica, ya que es la suma de cantidades positivas.

Calculemos la función de distribución conjunta:

$$F(x,y) = \begin{cases} 0 & \text{si } \; x \leq 0 & \\ 0 & \text{si } \; 0 < x \leq 1 & y \leq 2 \\ \displaystyle\int_2^y \int_0^x (x/3 + y/3)dxdy & \text{si } \; 0 < x \leq 1 & 2 < y \leq 3 \\ \displaystyle\int_2^3 \int_0^x (x/3 + y/3)dxdy & \text{si } \; 0 < x \leq 1 & y > 3 \\ 0 & \text{si } \; x > 1 & y \leq 2 \\ \displaystyle\int_2^y \int_0^1 (x/3 + y/3)dxdy & \text{si } \; x > 1 & 2 < y \leq 3 \\ 1 & \text{si } \; x > 1 & y > 3 \end{cases}$$

Haciendo las correspondientes integrales tenemos

$$F(x,y) = \begin{cases} 0 & \text{si } \; x \leq 0 & \\ 0 & \text{si } \; 0 < x \leq 1 & y \leq 2 \\ (x^2y + xy^2 - 2x^2 - 4x)/6 & \text{si } \; 0 < x \leq 1 & 2 < y \leq 3 \\ (x^2 + 5x)/6 & \text{si } \; 0 < x \leq 1 & y > 3 \\ 0 & \text{si } \; x > 1 & y \leq 2 \\ (y^2 + y - 6)/6 & \text{si } \; x > 1 & 2 < y \leq 3 \\ 1 & \text{si } \; x > 1 & y > 3 \end{cases}$$

Calculemos las funciones de densidad marginales:

$$\forall y \in (2,3) \qquad f_y(y) = \int_0^1 f(x,y)dx = \int_0^1 (x/3 + y/3)dx$$

de donde obtenemos

$$\forall y \in (2,3) \qquad f_y(y) = \left[x^2/6 + xy/3\right]_{x=0}^{1} = y/3 + 1/6$$

por tanto la función de densidad marginal de $Y$ es

$$f_y(y) = \begin{cases} y/3 + 1/6 & \text{si } 2 < y < 3 \\ 0 & \text{en el resto} \end{cases}$$

De donde obtendríamos que la función de distribución marginal es:

$$F_y(y) = \begin{cases} 0 & \text{si } y \leq 2 \\ (y^2 + y - 6)/6 & \text{si } 2 < y < 3 \\ 1 & \text{si } y \geq 3 \end{cases}$$

Y para la variable $X$ tendremos:

$$\forall x \in (0,1) \qquad f_x(x) = \int_2^3 f(x,y)dy = \int_2^3 (x/3 + y/3)dx$$

de donde obtenemos

$$\forall x \in (0,1) \qquad f_x(x) = \left[xy/3 + y^2/6\right]_{y=2}^{3} = x/3 + 5/6$$

por tanto la función de densidad marginal de $X$ es

$$f_x(x) = \begin{cases} x/3 + 5/6 & \text{si } 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

De donde obtendríamos que la función de distribución marginal es:

$$F_x(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ (x^2 + 5x)/6 & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Calculemos ahora las densidades condicionadas. Lo haremos como el cociente entre la conjunta y la marginal:

$$f_{y/x}(y/x) = \frac{x/3 + y/3}{x/3 + 5/6} \quad \text{si} \quad 0 < x < 1 \quad 2 < y < 3$$

siendo 0 en el resto. La función de distribución condicionada será:

$$F_{y/x}(y/x) = \int_2^y f_{y/x}(y/x)dy = \int_2^y \frac{x/3 + y/3}{x/3 + 5/6} dy$$

por tanto:

$$F_{y/x}(y/x) = \begin{cases} 0 & \text{si } y < 2 \\ \dfrac{xy/3 - 2x/3 + y^2/6 - 4/6}{x/3 + 5/6} & \text{si } 2 < y \leq 3 \\ 1 & \text{si } y > 3 \end{cases}$$

Del mismo modo, para $X$

$$f_{x/y}(x/y) = \frac{x/3 + y/3}{y/3 + 1/6} \quad \text{si} \quad 0 < x < 1 \quad 2 < y < 3$$

siendo 0 en el resto. La función de distribución condicionada será:

$$F_{x/y}(x/y) = \int_0^x f_{x/y}(x/y)dx = \int_0^x \frac{x/3 + y/3}{y/3 + 1/6} dx$$

por tanto:

$$F_{x/y}(x/y) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x^2/6 + xy/3}{y/3 + 1/6} & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Calculemos ahora las medias de $X$ y de $Y$. Podemos usar para ello las funciones de densidad marginales:

$$E[X] = \int_0^1 x f_x(x)dx = \int_0^1 x^2/3 + 5x/6 \; dx = 1/9 + 5/12 = 19/36.$$

$$E[Y] = \int_2^3 y f_y(y)dy = \int_2^3 y^2/3 + y/6 \; dy = \left[y^3/9 + y^2/12\right]_2^3 = 91/36.$$

Para calcular las varianzas calculemos primero los momentos de segundo orden respecto al origen:

$$E[X^2] = \int_0^1 x^2 f_x(x)dx = \int_0^1 x^3/3 + 5x^2/6 \; dx = 1/12 + 5/18 = 13/36.$$

$$E[Y^2] = \int_2^3 y^2 f_y(y)dy = \int_2 y^3/3 + y^2/6 \; dy = \left[y^4/12 + y^3/18)\right]_2^3 = 233/36.$$

Por tanto:

$$\text{var}(X) = 13/36 - (19/36)^2 = 107/1296$$

$$\text{var}(Y) = 233/36 - (91/36)^2 = 107/1296.$$

Para calcular la covarianza aplicamos

$$cov(x, y) = E[XY] - E[X] \cdot E[Y]$$

Calculemos $E[XY]$:

$$E[XY] = \int^3 \int_0^1 xy(x/3 + y/3)dxdy = \int_2^3 \int_0^1 (x^2y/3 + xy^2/3)dxdy$$

haciendo la integral, obtenemos que

$$E[XY] = \int_2^3 (y/9 + y^2/6)dy = \left[y^2/18 + y^3/18)\right]_2^3 = (9+27-4-8)/18 = 4/3.$$

Por tanto

$$cov[XY] = 4/3 - (19/36)(91/36) = -1681/36.$$

Como la covarianza es distinta de 0, no son incorreladas ni independientes.

# Problema 3.13

**Consideremos una variable aleatoria bidimensional con función de densidad conjunta:**

$$f(x, y) = \begin{cases} 4kxy & \textbf{si } 0 < x < 1 \quad 0 < y < 1 \\ 0 & \textbf{en el resto} \end{cases}$$

**Se pide:**

1. **Calcular el valor de $k$ para que $f$ sea una función de densidad**
2. **Calcular la función de distribución conjunta.**
3. **Calcular las funciones de densidad y de distribución marginales de $X$ y de $Y$.**
4. **Calcular las funciones de densidad y de distribución condicionadas.**

5. **Calcular las medias de $X$ y de $Y$.**

6. **Calcular las varianzas de $X$ y de $Y$.**

7. **Calcular la covarianza.**

8. **Estudiar si las dos variables son independientes o incorreladas.**

---

Calculemos en primer lugar el valor de $k$. Para ello imponemos la condición de que la integral valga la unidad (la condición de que siempre sea positiva se comprobará después).

$$\int_0^1 \int_0^1 (4kxy)dxdy = \int_0^1 \left[2kx^2y\right]_{x=0}^1 dy = \int_0^1 2kydy = \left[ky^2\right]_{y=0}^1 = k$$

por tanto

$$k = 1.$$

Veamos si se cumple la segunda condición, para ello deberá verificarse que

$$4xy \geq 0 \qquad \forall x \in (0,1) \quad \forall y \in (0,1)$$

pero esto es evidente que siempre se verifica, ya que es el producto de cantidades positivas.

Calculemos la función de distribución conjunta:

$$F(x,y) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ 0 & \text{si} \quad 0 < x \leq 1 \quad y \leq 0 \\ \displaystyle\int_0^y \int_0^x 4xydxdy & \text{si} \quad 0 < x \leq 1 \quad 0 < y \leq 1 \\ \displaystyle\int_0^1 \int_0^x 4xydxdy & \text{si} \quad 0 < x \leq 1 \quad y > 1 \\ 0 & \text{si} \quad x > 1 \qquad y \leq 0 \\ \displaystyle\int_0^y \int_0^1 4xydxdy & \text{si} \quad x > 1 \qquad 0 < y \leq 1 \\ 1 & \text{si} \quad x > 1 \qquad y > 1 \end{cases}$$

Haciendo las correspondientes integrales tenemos

$$F(x,y)=\begin{cases} 0 & \text{si} \quad x<0 \\ 0 & \text{si} \quad 0<x<1 \quad y<0 \\ x^2y^2 & \text{si} \quad 0<x<1 \quad 0<y<1 \\ x^2 & \text{si} \quad 0<x<1 \quad y>1 \\ 0 & \text{si} \quad x>1 \quad y<0 \\ y^2 & \text{si} \quad x>1 \quad 0<y<1 \\ 1 & \text{si} \quad x>1 \quad y>1 \end{cases}$$

Calculemos las funciones de densidad marginales:

$$\forall y \in (0,1) \qquad f_y(y) = \quad f(x,y)dx = \quad 4xy \ dx$$

de donde obtenemos

$$\forall y \in (0,1) \qquad f_y(y) = \left[2x^2y\right]_{x=0} = 2y$$

por tanto la función de densidad marginal de $Y$ es

$$f_y(y) = \begin{cases} 2y & \text{si } 0<y<1 \\ 0 & \text{en el resto} \end{cases}$$

De donde obtendríamos que la función de distribución marginal es:

$$F_y(y) = \begin{cases} 0 & \text{si } y<0 \\ y^2 & \text{si } 0<y<1 \\ 1 & \text{si } y>1 \end{cases}$$

Y para la variable $X$ tendremos:

$$\forall x \in (0,1) \qquad f_x(x) = \int_0 f(x,y)dy = \int_0 4xy \ dy$$

de donde obtenemos

$$\forall x \in (0,1) \qquad f_x(x) = \left[2xy^2\right]_{y\ 0} = 2x$$

por tanto la función de densidad marginal de $X$ es

$$f_x(x) = \begin{cases} 2x & \text{si } 0<x<1 \\ 0 & \text{en el resto} \end{cases}$$

De donde obtendríamos que la función de distribución marginal es:

$$F_x(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^2 & \text{si } 0 < x < 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Calculemos ahora las densidades condicionadas. Lo haremos como el cociente entre la conjunta y la marginal:

$$f_{y/x}(y/x) = \frac{4xy}{2x} = 2y \qquad \text{si} \quad 0 < x < 1 \quad 0 < y < 1$$

siendo 0 en el resto. La función de distribución condicionada será:

$$F_{y/x}(y/x) = \begin{cases} 0 & \text{si } y < 0 \\ y^2 & \text{si } 0 < y \leq 1 \\ 1 & \text{si } y > 1 \end{cases}$$

Del mismo modo, para $X$

$$f_{x/y}(x/y) = \frac{4xy}{2y} = 2x \quad \text{si} \quad 0 < x < 1 \quad 0 < y < 1$$

siendo 0 en el resto. La función de distribución condicionada será:

$$F_{x/y}(x/y) = \begin{cases} 0 & \text{si } x < 0 \\ x^2 & \text{si } 0 < x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Calculemos ahora las medias de $X$ y de $Y$. Podemos usar para ello las funciones de densidad marginales:

$$E[X] = \int_0^1 x f_x(x) dx = \int_0^1 2x^2 dx = \left[2x^3/3\right]_0^1 = 2/3.$$

$$E[Y] = \int_0^1 y f_y(y) dy = \int_0^1 2y^2 dy = \left[2y^3/3\right]_0^1 = 2/3.$$

Para calcular las varianzas calculemos primero los momentos de segundo orden respecto al origen:

$$E[X^2] = \int_0^1 x^2 f_x(x) dx = \int_0^1 2x^3 dx = \left[2x^4/4\right]_0^1 = 1/2.$$

$$E[Y^2] = \int_0^1 y^2 f_y(y)dy = \int_0^1 2y^3 dy = \left[2y^4/4\right]_0^1 = 1/2.$$

Por tanto:

$$\text{var}(X) = 1/2 - 4/9 = 1/18.$$

$$\text{var}(Y) = 1/2 - 4/9 = 1/18.$$

Para calcular la covarianza aplicamos

$$cov(x, y) = E[XY] - E[X] \cdot E[Y]$$

Calculemos $E[XY]$:

$$E[XY] = \int_0^1 \int_0^1 4x^2y^2 \; dxdy = \left[\left[4x^3y^3/9\right]_0^1\right]_0^1 = 4/9.$$

Por tanto

$$cov[XY] = 4/9 - (2/3)(2/3) = 0.$$

Como la covarianza es 0, son incorreladas. Para ver si son independientes habría que ver si las densidades marginales coinciden con las condicionadas. Como efectivamente ocurre, son independientes.

# Problemas Propuestos

1. De una bolsa se sacan al azar 4 canicas rojas y 2 blancas, y se colocan en hilera.

   (a) ¿Cuál es la probabilidad de que las de los extremos sean blancas?
   (b) ¿Y de que las dos estén juntas?

2. Se selecciona al azar un número entero de tres cifras.

   (a) ¿Cuál es la probabilidad de que contenga por lo menos un 1?
   (b) ¿Y de que contenga exactamente dos veces el 3?

3. Hay N personas en una sala. Si se hace una lista de sus cumpleaños (mes y día del mes). ¿Cuál es la probabilidad de que haya una o más repeticiones en la lista? ¿Cuál es la probabilidad de que dos o más personas tengan el mismo cumpleaños? Se supone que el año tiene 365 días, ignorándose el 29 de Febrero, y que cada uno de los días tiene la misma verosimilitud de ocurrencia para el cumpleaños de cualquier individuo.

4. Una señora de apellido Junquera asegura ser clarividente, y dice que si le muestran 8 naipes, 4 de color rojo y 4 de color negro, identificará correctamente por lo menos 6 sin tener necesidad de ver sus colores. Si solamente adivina y no tiene ninguna habilidad especial, ¿cuál es la probabilidad de que pueda identificar por lo menos 6 de los 8 naipes?

5. En una prisión cuya población es de 100 individuos, se van a seleccionar al azar dos personas para ponerlas en libertad. ¿Cuál es la probabilidad de que el más viejo de los presos sea 1 de los 2 elegidos? ¿Y que se seleccione la pareja formada por el más viejo y el más joven?

6. Se somete a un estudiante a un examen del tipo verdadero-falso que contiene 10 preguntas; para que apruebe debe responder correctamente a 8 preguntas o más. Si el estudiante está respondiendo al azar, ¿cuál es la probabilidad de que apruebe el examen?

7. El departamento de investigación de una fábrica de focos ha perfeccionado un recubrimiento para los filamentos capaz de prolongar la duración de aquéllos. Para comprobar las duraciones de los focos nuevos con la de los focos viejos, se seleccionan diez focos fabricados con el nuevo procedimiento y diez normales, y se forman parejas: un foco nuevo con uno viejo. Se somete los 10 pares a prueba, y se anota cuál de los focos de cada par falla primero, si el foco nuevo o el viejo. Suponiendo que el nuevo proceso realmente no prolonga la duración de los focos, ¿cuál es la probabilidad de que el foco viejo falle primero, en por lo menos 9 de los 10 pares? Si se efectúa el experimento y se encuentra que los focos viejos fallaron primero en 9 de los pares, ¿se debería adoptar el nuevo proceso de fabricación?

8. A una persona se le muestran al azar 3 tarjetas rojas y 3 blancas. La persona sabe que le van a mostrar 3 de cada color, para que las identifique. Si está adivinando, ¿cuál es la probabilidad de que identifique correctamente las 6? ¿De que identifique correctamente 5? ¿Exactamente 4?

9. Se supone que 16 equipos participan en un torneo de eliminatoria sencilla. Si todos tienen el mismo nivel de juego, ¿cuál es la probabilidad de que gane el favorito?

10. Se tiran un par de dados no cargados una vez, y se establece que los dos números que aparecen no son los mismos. Calcular la probabilidad de que la suma sea 7 o que la suma sea 4 o que la suma sea 12.

11. Un médico ha recabado la siguiente información en base a su experiencia, relativa a las enfermedades de sus pacientes: El 5% creen tener cáncer y realmente lo tienen. El 45% creen tener cáncer y no lo tienen, el 10% no creen tener cáncer y lo tienen, y el 40% creen no tenerlo y realmente no lo tienen.

   Siendo $A$={Paciente que cree tener cáncer} $B$={Paciente que tiene cáncer}, calcular las probabilidades $P(A/B)$ . $P(B/A)$ , $P(B/\bar{A})$ y $P(A/B)$.

12. Se seleccionan al azar dos bolas, sin reemplazamiento, de una urna que contiene 4 blancas y 8 negras.

    (a) Calcular la probabilidad de que ambas sean blancas.

    (b) Calcular la probabilidad de que la segunda bola sea blanca.

13. La caja 1 contiene 4 focos defectuosos y 16 en buen estado, la caja 2 contiene 1 foco defectuoso y 1 en buen estado. Se lanza un dado no cargado una sola vez. Si sale 1 ó 2 se saca al azar un foco de la caja 1, de lo contrario se saca de la caja 2. ¿Cuál es la probabilidad de que el foco seleccionado esté defectuoso?

14. Se supone que 0,95 es la probabilidad de que un jurado seleccionado para juzgar un caso criminal emita un veredicto adecuado. Esto es, si el jurado condena a un individuo culpable, es de 0,95. Recíprocamente, si el individuo es inocente, la probabilidad de que el jurado lo absuelva es de 0,95. Se supone que el cuerpo de policía realiza su labor a conciencia, de forma que el 99% de las personas que se presentan para ser juzgadas son verdaderamente culpables. Calcular la probabilidad de que el acusado sea inocente si el jurado lo encuentra inocente.

15. Suponer que la ciencia médica ha desarrollado una prueba para el diagnóstico del cáncer, que tiene 95% de exactitud, tanto en los que tienen cáncer como entre los que no lo tienen. Si 0,005 de la población realmente tiene cáncer, calcular la probabilidad de que determinado individuo tenga cáncer si la prueba dice que lo tiene.

16. En una escuela se puntúa a los alumnos según su nivel atlético en una escala de 0 a 4. El 1% de los estudiantes participa en un programa deportivo. De este grupo el 10% tiene una puntuación mayor que 3, en tanto que el 20% del resto de los estudiantes tiene una puntuación mayor que 3. ¿Qué proporción del total de estudiantes tiene puntuación mayor que 3? Si se elige un estudiante al azar y se obtiene que tiene una puntuación mayor que 3. ¿Cuál es la probabilidad de que participe en el programa deportivo?

17. Un examen contiene 8 preguntas tipo verdadero falso, en el que se requiere responder un mínimo de seis preguntas para aprobar.

    (a) Suponiendo que se está respondiendo al azar, ¿cuál es la probabilidad de aprobar el examen?

(b) Suponiendo que se conoce la respuesta correcta a la primera pregunta, y que por tanto, sólo se responde al azar desde la segunda a la octava, ¿cuál es la probabilidad de aprobar el examen?

(c) Si se conocen las respuestas a las dos primeras preguntas, ¿cuál es la probabilidad de aprobar el examen?

18. Una urna contiene 5 bolas blancas y dos negras. Se extraen 3 bolas, sin fijarnos en el color, y se dejan aparte. ¿Cuál es la probabilidad de que la siguiente que se extraiga sea blanca?

19. Extraemos 3 cartas de una baraja española de 40 cartas. ¿Cuál es la probabilidad de que todas sean figuras (sota, caballo o rey)?

20. En tres máquinas A, B, C se fabrican piezas de la misma naturaleza. El porcentaje de piezas defectuosas en cada máquina es respectivamente 1%, 2% y 3% . Se mezclan 3 piezas, 1 de cada máquina y se elige una al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada en A?

21. Las probabilidades de que cada uno de los tres aviones A, B, C cumpla su horario son de 0,7 , 0,8 y 0,9 respectivamente. El comportamiento de cada avión no depende de los otros. Calcula las probabilidades de que cumplan el horario:

    (a) los tres aviones,

    (b) al menos, dos de ellos.

22. Un jugador italiano expresó su sorpresa a Galileo, por observar que al jugar con tres dados la suma 10 aparece con más frecuencia que la suma 9. Según el jugador, los casos favorables al 9 serían: 126, 135, 144, 225, 234, 333 y al 10 serían: 136, 145, 226, 235, 244, 334. Pero Galileo en su libro *"Considerazione sopra il giuoco dei dadi"*, vio que estas combinaciones no se pueden considerar igualmente probables. Explicar por qué y calcular las probabilidades correspondientes.

23. ¿Cuál es la probabilidad de que al lanzar $n$ veces una moneda obtengamos $n$ caras?

24. Probabilidad de que al lanzar un dado $n$ veces se obtenga al menos una vez el 6.

25 Una urna contiene $a$ bolas blancas y $b$ bolas negras. Una persona saca $k$ bolas. ¿Cuál es la probabilidad de que sean $x$ blancas y $k-x$ negras?

26. Calcular la probabilidad de que al lanzar $n$ veces dos dados se obtenga al menos un 6 doble. ¿Cuántas partidas tendré que jugar para que tengamos $p' = 1/2$ de obtener un 6 doble?

    (Problema propuesto por de Mere a Pascal).

27. Se reparten entre 4 personas las 52 cartas de una baraja. ¿Cuáles son las probabilidades, para un jugador determinado, de obtener 0, 1, 2, 3, 4 ases?

28. Se lanzan 3 monedas al aire. ¿Cuál es la probabilidad de que las tres sean caras o las tres sean cruces?

29. Un suceso A tiene probabilidad 0,1. Calcular la probabilidad de que el suceso se verifique al menos 10 veces en 100 pruebas.

30. Un número $n$ de bolas idénticas se distribuyen entre $m$ compartimentos. ¿Cuál es la probabilidad de que un cierto compartimento tenga $h$ bolas?

31. ¿Cuál es la probabilidad de torpedear un barco, si se lanzan 3 torpedos y la probabilidad de hacer blanco con un torpedo es de 0,2?

32. Sean $A$ y $B$ dos sucesos cumpliendo

$$P(A) = 0{,}4 \qquad P(B) = 0{,}3 \qquad P(A \cap B) = 0{,}1$$

calcúlense las probabilidades de los siguientes sucesos:

$$\bar{A} \cap B \quad A \Delta B \quad A \cup B.$$

33. Se consideran los sucesos $A$ y $B$ del ejerccio anterior. Calcúlense las siguientes probabilidades:

    (a) $P(\bar{A}/B)$ (b) $P(A/A \cup B)$ (c) $P(A/A \cap B)$ (d) $P(\bar{A}/B)$

    (e) $P(A \Delta B/A \cup B)$

34. Demostrar que si $B \subset A$, $C \subset A$ y $P(C) > 0$, se verifica

$$P(B)/P(C) = P(B/A)/P(C/A).$$

35. Si tres sucesos $A$, $B$ y $C$ son independientes, demostrar que los sucesos $(A \cup B)$ y $C$ también lo son.

36. Dados los sucesos $A$ y $B$ independientes, demostrar que

    (a) Los sucesos $A$ y $B$ son independientes.

    (b) Los sucesos $\bar{A}$ y $B$ también son independientes.

37. Dados los sucesos $A$ y $B$ disjuntos, demostrar que:

$$P(A/A \cup B) = \frac{}{P(A) + P(B)}.$$

38. Se tiene una baraja de 40 cartas, se extraen cinco cartas una a una, devolviendo al montón la carta una vez vista. Calcular:

    (a) La probabilidad de que las cinco cartas sean de oros.

    (b) La probabilidad de que el primer oro aparezca en quinto lugar.

39. En una fábrica la máquina 1 produce piezas de buena calidad con una probabilidad igual a 0,8 y la máquina 2 con probabilidad 0,9. Se separa una pieza de cada máquina. Calcular:

    (a) Probabilidad de que ambas piezas sean defectuosas

    (b) Probabilidad de que una sea defectuosa y la otra no.

40. En una tienda con rebajas hay dos mostradores con 100 prendas cada uno. En el primero hay 25 prendas en mal estado y en el segundo 20. Un cliente pasa una prenda del primer mostrador al segundo. Calcular la probabilidad de que otro cliente compre una prenda en buen estado procedente del segundo mostrador.

41. Tenemos un dado trucado de tal forma que

$$P(X = 1) = 0{,}1 \qquad P(X = 2) = 0{,}2 \qquad P(X = 3) = 0{,}3$$

$$P(X = 4) = 0{,}1 \qquad P(X = 5) = 0{,}2 \qquad P(X = 6) = 0{,}1$$

Se arroja el dado 10 veces. Calcular las probabilidades de que se den los sucesos siguientes:

(a) Cuatro unos, dos doses, un cuatro y tres cincos.
(b) Tres unos, tres treses y dos cuatros.

42. Consideremos una moneda trucada de tal forma que la probabilidad de cara es 0,7 y la probabilidad de cruz 0,3. Si se lanza la moneda cinco veces, calcular las probabilidades de obtener:

    (a) Cinco caras
    (b) Dos cruces
    (c) Al menos tres caras
    (d) En las dos primeras dos cruces y el resto caras
    (e) Mas de una cara y menos de cuatro.

43. Durante un año las personas de una ciudad utilizan tres tipos de transportes: metro ($M$), autobús ($A$) y coche particular ($C$). Las probabilidades de que durante el año hayan usado unos u otros transportes son las siguientes:

    $P(M) = 0{,}3 \quad P(A) = 0{,}2 \quad P(C) = 0{,}15 \quad P(M \cap A) = 0{,}1$

    $P(M \cap C) = 0{,}05 \quad P(A \cap C) = 0{,}06 \quad P(M \cap A \cap C) = 0{,}01$

    Calcular las siguientes probabilidades:

    (a) Que una persona tome al menos dos medios de transporte.
    (b) Que una persona viaje en metro y no en autobús.
    (c) Que una persona viaje en metro o en coche y no en autobús.
    (d) Que viaje en metro o en autobús y en coche.
    (e) Que una persona vaya a pie.

44. Una fábrica de detergentes proyecta lanzar una nueva marca. En el mercado hay dos marcas $A$ y $B$. La probabilidad de compra de $A$ es $P(A) = 0{,}3$; la de $B$ es $P(B) = 0{,}5$ y la de comprar ambos es $P(A \cap B) = 0{,}1$. Para decidirse por la nueva marca, la fábrica necesita conocer la probabilidad de que no se compren ni $A$ ni $B$, así como la probabilidad de que sólo compren una de las dos marcas. Calcular estas probabilidades.

45. Se eligen al azar dos puntos en el intervalo $(0; 2)$. Calcular la probabilidad de que el producto de ambos sea menor que 3.

46. En un sanatorio se atienden únicamente cuatro tipos de enfermedades; las probabilidades de que un enfermo ingrese con cada una de las cuatro son: $P(A) = 0{,}2$; $P(B) = 0{,}05$; $P(C) = 0{,}6$ y $P(D) = 0{,}15$. Las probabilidades de curación para cada enfermedad son, respectivamente, 0,8; 0,75; 0,2 y 0,3.

    Calcúlese la probabilidad de curación de un enfermo que ingresa en el sanatorio sin saber cuál es su enfermedad.

47. Una empresa produce tres fungicidas en proporciones 1/6, 1/3 y 1/2. Las probabilidades de efectividad de cada uno son respectivamente: 0,9; 0,94 y 0,88. Calcúlese la probabilidad de que utilizando uno cualquiera de ellos, sea efectivo.

48. Se tiene una baraja de 40 cartas; se extraen 4 cartas una a una, no devolviendo las cartas al montón. Calcular:

    (a) Probabilidad de que la cuarta carta sea el as de bastos.

    (b) Probabilidad de que las cuatro sean los cuatro ases en el siguiente orden: oros, copas, espadas y bastos.

49. En una granja avícola existen tres razas de gallinas: $A$ , $B$ y $C$. De la raza $A$ hay 300 aves, de la $B$ 600 y de la $C$ 100. La probabilidad de que una gallina de la raza $A$ ponga huevos de calidad inferior a la tolerada es igual a 0,2; en la raza $B$ es 0,15 pero se desconoce cuál es esta probabilidad en la raza $C$. Sin embargo, se han establecido tres hipótesis sobre la probabilidad de que tomando una raza al azar y siendo el huevo de baja calidad proceda de la raza $C$; estas probabilidades son: 0,6 ; 0,5 y 0,3. Verifíquense qué hipótesis pueden ser correctas.

50. En un dado bien construido consideramos los sucesos $A$ y $B$, tal que $A$ es la obtención de una puntuación mayor o igual que 4, y $B$ obtener 3 ó 6. Determinar si se verifica que:

    (a) $A$ y $B$ son disjuntos

    (b) $A$ y $B$ son independientes.

51. Calcular en el juego del póker las probabilidades:

    (a) De póker

(b) De full

(c) De color (Baraja de 40 cartas).

52. Una estadística asegura que, de cada 1.000 habitantes, a 816 les gusta el fútbol, a 723 los toros, a 645 el tenis, y asimismo, que a 562 les gusta el fútbol y los toros, a 463 el fútbol y el tenis, y a 470 los toros y el tenis, y sólo a 310 les gustan los tres espectáculos. ¿Puede ser cierta la estadística?

53. Considerando el enunciado del problema anterior, calcule gráfica y numéricamente:

    (a) Los habitantes a los que solamente les gusta el fútbol.

    (b) Los habitantes a los que les gusta el fútbol y los toros pero no el tenis.

    (c) Los habitantes a los que les gusta el fútbol y el tenis pero no los toros.

    (d) Los habitantes a los que les gustan los tres espectáculos.

54. En una encuesta realizada entre personas de ambos sexos, nacidas en varios continentes y no todos sabiendo leer, se han obtenido los siguientes datos: Hombres 520: Mujeres 480; Europeos 402; Analfabetos 152; Hombres europeos 230; Europeos no analfabetos 324; Hombres analfabetos 70; Hombres o Europeos o analfabetos 730. Se pide:

    (a) El número de hombres europeos analfabetos.

    (b) El número de hombres europeos no analfabetos.

    (c) El número de mujeres no europeas analfabetas.

    (d) El número de hombres, desglosado de todas las formas posibles.

55. Enumerar los elementos del espacio muestral que corresponden a cada uno de los siguientes experimentos:

    (a) Observar el sexo de niños recién nacidos.

    (b) Observar el número de botellas rotas en cajas de una docena.

    (c) Observar la posición (distancia al centro) de una bala disparada sobre una diana.

    (d) Observar el número de caras que salen en dos monedas que se tiran sobre una mesa.

(e) Observar el número de accidentes de carretera en una determinada ciudad durante una semana.

(f) Observar el tiempo que tarda en quemarse un cohete.

56. Escribir los espacios muestrales de los experimentos siguientes. Usar diagramas en árbol y/o tablas de doble entrada para obtenerlos.

    (a) Observar los sexos de los dos primeros hijos de una mujer.

    (b) Observar las caras de tres monedas tiradas sobre una mesa.

    (c) Observar las caras de dos dados (uno rojo y otro verde) lanzados sobre una mesa.

57. Sean $E$, $F$ y $G$ tres sucesos cualesquiera en un espacio muestral $S$. Hallar las expresiones en lenguaje de conjuntos de los sucesos siguientes:

    (a) $E$ y $F$ ocurren a la vez pero $G$ no.

    (b) $F$ ocurre pero ni $E$ ni $G$ ocurren.

    (c) Sólo ocurre $G$.

    (d) Por lo menos ocurre uno de $E$, $F$ ó $G$

    (e) Sólo ocurre uno de ellos.

    (f) Ocurren los tres.

    (g) Por lo menos ocurren dos.

    (h) Ocurren dos exactamente.

    (i) No ocurren más de dos.

    (j) Ocurren menos de dos.

    (k) Ocurren $E$ ó $F$, pero no ambos, y $G$ no.

58.

59. En una clase hay 20 estudiantes, de los cuales 8 fuman(E), 12 llevan gafas(F) y 6 fuman y llevan gafas(EF). Expresar en palabras los significados de los sucesos $\overline{E}$, $\overline{F}$, $\overline{E} \cap \overline{F}$, $E \cup F$, $\overline{(E \cup F)}$, $(E - F)$, $F/E$, $E/\overline{F}$, $(E \cup F)/E$, $\overline{(E \cup F)}/E$. Obtener las probabilidades de cada uno de estos sucesos.

60. En un censo hecho a un grupo de 500 personas, se recogió la información siguiente sobre sexo, estado civil y posesión de coche: 300 varones, 290 casados, 415 tienen coche, 200 varones casados, 250 casados con coche, 250 varones tienen coche y 190 varones casados con coche. Demostrar que los números recogidos en las distintas categorías no son coherentes.

61. Dos chicos se han citado entre las seis y las siete de la tarde en un cierto lugar, conviniendo en no esperarse más de diez minutos. Calcular la probabilidad de que se encuentren, suponiendo que lleguen al azar e independientemente durante el período de tiempo establecido.

62. Sobre un segmento $\overline{AB}$ se toman dos puntos P y Q al azar. ¿Cual es la probabilidad de que con los tres trazos del segmento se pueda construir un triángulo?

63. Sean a y b dos números seleccionados al azar en el intervalo [0,1]. Hallar la probabilidad de que tenga raíces reales la ecuación de segundo grado: $x^2 + 2ax + b = 0$.

64. ¿Cuál es el número mínimo de veces que es preciso lanzar dos dados para que la probabilidad de obtener un seis doble sea mayor que la de no obtenerlo?

65. De una baraja de 48 cartas se extraen dos a la vez. Hallar la probabilidad de que...

    (a) ambas sean de copas,

    (b) por lo menos una sea de espadas,

    (c) una sea de copas y la otra de espadas.

66. ¿Cuál es la probabilidad de obtener 3 o más espadas en una extracción de 4 cartas de una baraja de 52 cartas?

67. Hállese la probabilidad de obtener exactamente tres caras en cinco tiradas con una moneda.

68. Se hacen seis tiradas con una moneda. Hallar la probabilidad de que haya por lo menos una racha ininterrumpida de tres caras.

69. ¿Cuál es la probabilidad de que al lanzar dos veces un dado, los resultados sean dos y tres en ese orden?

70. Calcular la probabilidad de obtener un doble seis por lo menos, en 24 tiradas con dos dados.

71. Se tira un dado ordinario cinco veces. Hállese el número más probable de seises.

72. Dos granjeros tienen cada uno 10 gallinas. Un día, 9 gallinas pasan de la granja A a la B. El granjero de B, que no es capaz de distinguirlas, devuelve 9 gallinas elegidas al azar, y se lo dice a su vecino. El granjero de A, que tampoco es capaz de distinguirlas, acepta el acuerdo, pero quiere conservar a Perlita, que pone yemas de dos huevos. ¿En cuál de las dos partes es preferible que empiece a buscar a Perlita?

73. Se lanza un dado equilibrado y el número obtenido es el número de bolas blancas que metemos en una urna. A continuación se vuelve a lanzar el dado y se repite la operación introduciendo ahora bolas negras. Por último sacamos una bola de la urna. ¿Cuál es la probabilidad de que sea blanca?

74. En una urna hay M bolas blancas y N-M bolas negras. Se sacan M bolas y quedan en la urna N-M bolas. A continuación se saca otra bola. ¿Cuál es la probabilidad de que ésta sea blanca?

75. Consideremos una urna que contiene $n$ bolas negras y $b$ bolas blancas; de ésta urna hacemos tres extracciones, si en una extracción nos sale bola negra metemos $r$ bolas más, además de la obtenida. Se pregunta por la probabilidad de sacar tres bolas negras seguidas.

76. De una urna que contiene 3 bolas blancas y 5 negras, se transfieren 4 bolas a una urna vacía, de ésta última se sacan dos bolas y resulta que ambas son blancas. Calcular la probabilidad de que la tercera bola que se saque de la misma urna sea blanca.

    (a) En el caso de que las dos bolas que se saquen en primer lugar vuelvan a introducirse en la urna antes de sacar la tercera.

    (b) Hacer el ejercicio en el caso de que las bolas extraídas no se reintegren.

77. Una caja contiene 5 tornillos defectuosos y 4 aceptables, otra contiene 4 defectuosos y 5 aceptables. Se traslada un tornillo de la primera a la segunda y a continuación se extrae un tornillo de la segunda caja. ¿Cuál es la probabilidad de que este tornillo sea aceptable?

78. La caja I contiene dos bolas blancas y dos negras. La caja II contiene tres bolas blancas y dos negras. Se trasladan dos bolas de la caja I a la caja II y a continuación se extrae de la caja II una bola que resulta ser negra. ¿Cuál es la probabilidad de que las bolas trasladadas fuesen una blanca y la otra negra?

79. Se tienen tres bolsas con las siguientes composiciones: $B_1 = \{3$ bolas blancas y 1 negra$\}$; $B_2 = \{2$ y $2\}$; $B_3 = \{$ 1 y $3\}$. Se hacen pruebas consistentes en tomar una bolsa al azar y sacar de ella una bola al azar. ¿Cuál es la bolsa que da una probabilidad mayor al suceso obtener una bola blanca?

80. Se tienen N urnas cada una de las cuales contiene 4 bolas blancas y 6 negras, y otra urna con 5 bolas blancas y 5 negras. Se toma una urna al azar entre las N+1 y se sacan dos bolas que resultan ser negras. La probabilidad de que en dicha urna queden 5 bolas blancas y 3 negras es 1/7. Calcular el valor de N.

81. La caja I contiene dos bolas blancas y dos negras y la caja II contiene tres blancas y dos negras. Se traslada una bola de I a II y a continuación se extrae una bola de II que resultó ser blanca. ¿Cuál es la probabilidad de que la bola trasladada fuese blanca?

82. Se han lanzado unos dados y la suma de los puntos obtenidos es cuatro. Calcular la probabilidad de que se haya jugado con dos dados.

83. En una lotería de 400 billetes hay 4 premios. Una persona compra 10 billetes. Hállese la probabilidad de obtener por lo menos 1 premio.

84. ¿Cuál es el menor número de personas que han de estar reunidas para poder asegurar que la probabilidad de que entre ellas se encuentren dos que celebren el mismo día su cumpleaños, sea por lo menos 0,5?

85. Una caja contiene 1000 tornillos, de los cuales 50 son defectuosos. Calcular la probabilidad de que al sacar 10 al azar

    (a) sean 5 defectuosos.

    (b) no haya ninguno defectuoso.

    (c) haya menos de tres defectuosos.

86. A y B juegan alternativamente con un par de dados corrientes. A gana si consigue 6 puntos antes de que B saque 7 y B gana si saca 7 antes

de que A saque 6. Comienza a jugar A. ¿Qué probabilidad tiene de ganar?

87. Una asociación tiene 100 socios y tienen que elegir un representante. La elección promete ser en extremo reñida. Al parecer apoyan 49 a un candidato y 51 al otro. A pesar del interés de la votación se calcula que sólo votará el 80% del censo. En esta hipótesis, obtener la probabilidad de que resulte elegido el candidato que está en minoría.

88. La experiencia de un banco permite asegurar que los clientes que tienen fondos suficientes y firman un cheque con fecha posterior es el 1%. Por el contrario, todos los clientes que no tienen fondos firman sus cheques con fecha posterior. La proporción de los cheques con fondo es del 90%. Se recibe en caja un cheque con fecha posterior. ¿Cuál es la probabilidad de que pertenezca a un cliente sin fondos?

89. Se hace un disparo con cada uno de tres cañones, siendo la probabilidad de hacer blanco 0,1; 0,2 y 0,3 respectivamente. Calcular la probabilidad de cada uno de los números posibles de blancos.

90. Un aparato esta formado por dos partes A y B. El proceso de fabricación es tal que la probabilidad de un defecto en A es 0,06 y la probabilidad de un defecto en B es 0,07. ¿Cuál es la probabilidad de que el aparato no sea defectuoso?

91. A, B y C por este orden tiran una moneda al aire. El primero que saque una cara gana. ¿Cuales son sus respectivas probabilidades de ganar?

92. En una región llueve el 40% de los días. Un estudio sobre la información meteorológica dada por la televisión descubrió que a veces fallaban. Más concretamente predijeron lluvia incorrectamente en el 30% de los casos, y día claro incorrectamente en el 10% de los casos.

    (a) Calcular la probabilidad de que la televisión realice un pronóstico erróneo en un día cualquiera.

    (b) Si un día determinado la televisión pronosticó lluvia, ¿cuál es la probabilidad de que dicho día lloviera?

93. En las bolsas españolas el 10% de los títulos son de alto riesgo y alta rentabilidad, y el 40% son de bajo riesgo y baja rentabilidad. Una persona está dispuesta a invertir en bolsa y desea títulos de bajo riesgo

y alta rentabilidad. Para ello estudia el grupo de títulos de bajo riesgo y descubre que de éstos el 70% son de baja rentabilidad.

(a) ¿Cuál es la probabilidad de encontrar títulos de bajo riesgo y alta rentabilidad?

(b) ¿Cuál es la probabilidad de encontrar un título de alto riesgo?

(c) ¿Cuál es la probabilidad de encontrar un título de alta rentabilidad?

94. Las probabilidades de sufrir un accidente automovilístico en día seco o lluvioso son respectivamente 0,005 y 0,012. La probabilidad de que llueva un cierto día es 0,1. Calcular:

    (a) Probabilidad de que un día cualquiera llueva y se tenga un accidente.

    (b) Probabilidad de que haya llovido un día en que se ha tenido un accidente.

95. Tenemos tres urnas, en la urna A hay 2 bolas blancas y 3 negras; en la urna B, 4 blancas y 1 negra y en la urna C, 3 blancas y 4 negras. Se selecciona una urna aleatoriamente, y se extrae una bola que resulta ser blanca. Hallar la probabilidad de haber escogido la urna A.

96. Una compañía financiera para la venta de automóviles a plazo, opera en tres grandes regiones de un país A, B y C. El 50% de las operaciones las realiza en la región A, el 40% en la B y el 10% en la región C. Se ha estimado, por largas experiencias, el tanto por ciento de los clientes que no efectúan el pago de las letras en cada una de las regiones. Para la región A es del 1%, para la B del 2% y para la C el 8%. Determinar el tanto por ciento de los clientes de la compañía que pagan las letras de la operación.

97. Un médico posee un test para diagnosticar el cáncer. Mediante muchos reconocimientos se ha comprobado que la probabilidad de que el test resulte positivo para una persona enferma de cáncer es de 0,95; mientras que es de 0,05 para una persona que no tenga la enfermedad. Supongamos que sólo el 1% de la población tiene la enfermedad del cáncer. ¿Cuál es la probabilidad de que una persona padezca cáncer si su test resulta positivo?

98. Se sabe que el 12% de los automóviles usa una determinada marca de recambios. También se sabe que si se realiza una encuesta, sólo el 80% de los usuarios que utilizan recambios de dicha marca lo reconocerán, mientras que el 2% de los que no la usan reconocen no usarla. ¿Cuál es la probabilidad de que un usuario, elegido al azar, mienta al hacerle la encuesta? Y si ha mentido, ¿cuál es la probabilidad de que use dicha marca?

99. En una oficina trabajan 4 secretarias que archivan documentos. Cada una de ellas archiva el 40%, 10%, 30% y 20% de los documentos. La probabilidad que cada una de ellas tiene de equivocarse es 0,01; 0,04; 0,06; 0,10 respectivamente.

    (a) ¿Cuál es la probabilidad de que un documento esté mal archivado?

    (b) Si se ha encontrado un documento mal archivado, ¿cuál es la probabilidad de que sea debido a la tercera secretaria?

100. Una empresa que se dedica a la fabricación de automóviles desea lanzar al mercado un nuevo modelo para el próximo año. Al estudiar la posible situación económica de este año se le plantean tres posibilidades: Inflación, estabilidad ó depresión. Se ha estimado que la primera alternativa es doblemente probable que las otras dos. Además, la probabilidad de que haya inflación y el modelo se lance al mercado es de 0,3; la probabilidad de que haya estabilidad y el modelo no se lance al mercado es 0,2; mientras que existiendo depresión, la probabilidad de lanzar el coche al mercado es 0,6.

    (a) Calcular la probabilidad de lanzar el nuevo modelo.

    (b) Calcular la probabilidad de que haya inflación, si se ha lanzado el modelo al mercado.

101. El 45% de la población española abandona su residencia habitual durante las vacaciones de verano, pero tan sólo el 5% de los citados veraneantes viaja mas allá de las fronteras del país. No obstante hay una proporción del 1% de ciudadanos que no estando de vacaciones viaja al extranjero en verano. Se pide:

    (a) Probabilidad de que un ciudadano viaje al extranjero en verano.

    (b) Probabilidad de que un español que se encuentra en el extranjero esté de veraneo.

102. El 20 por ciento de los ordenadores de una empresa son portátiles y el resto de sobremesa. De los portátiles, el 10 por ciento tienen una memoria RAM de 8 Megas, el 20 por ciento tiene una memoria de 4 Megas y el resto tiene una memoria de 1 Mega. De los de sobremesa, el 30 por ciento tienen 8 Megas, el 50 por ciento tiene 4 Megas y el resto tiene 1 Mega. Se pide:

    (a) Hallar la probabilidad de que un ordenador tenga 8 Megas.

    (b) Hallar la probabilidad de que un ordenador portátil tenga 8 Megas.

    (c) Hallar la probabilidad de que un ordenador tenga 8 Megas y sea portátil.

    (d) Hallar la probabilidad de que un ordenador que tenga 8 Megas sea portátil.

    (e) Calcular la memoria media de los ordenadores portátiles.

103. Supongamos que el temario de una oposición consta de 50 lecciones. Un opositor ha preparado 15 lecciones. En el examen se seleccionan aleatoriamente 5 lecciones. Calcular la probabilidad de que el opositor conozca las cinco lecciones. Si para aprobar hay que contestar al menos a tres lecciones, calcular la probabilidad de aprobar el examen. Determinar el número esperado de lecciones que conocerá este opositor en el examen, así como la varianza.

104. Pablo lanza una moneda hasta que obtiene la primera *cara*. Entonces, Ana lanza la moneda hasta que obtiene la primera *cara*. Hallar la probabilidad de que la moneda se haya lanzado un total de $n$ veces.

105. Se introducen tres bolas independientemente al azar en cinco urnas. Calcular la probabilidad de que:

    (a) las tres bolas estén en la misma urna.

    (b) no haya ninguna urna que contenga dos bolas.

106. Un temario está compuesto por 100 temas. El examen consiste en la exposición al azar de 2 temas. Para aprobar necesita responder correctamente los dos temas. ¿Cuál es el número mínimo de temas que ha de estudiar el opositor para que la probabilidad de aprobar sea superior que la probabilidad de suspender?

107. Un hombre tiene cuatro llaves en su bolsa. Como está oscuro, no puede ver cúal es la llave de su puerta, por lo que prueba, una a una, hasta encontrarla. Sea $X$ el número de llaves que prueba (contando la correcta) para abrir su puerta. ¿Cuál es la función de probabilidad para $X$?

108. Se reparten 5 naipes de una baraja de 52 cartas. Sea $Y$ el número de naipes con color rojo repartidos. ¿Cuál es la función de probabilidad para $Y$?

109. Sea $Z$ el número de días a la semana que va a la gasolinera. ¿Cuál es la función de probabilidad para $Z$?

110. Comprobar que $F(x)$ es una función de distibución, y obtener la función de probabilidad para $X$. Usar este resultado para calcular $P(-1 \leq X \leq 1)$.

$$F(x) = \begin{cases} 0 & \text{si} \quad x < -2 \\ 1/2 & \text{si} \quad -2 \leq x < 0 \\ 1 & \text{si} \quad x \geq 0 \end{cases}$$

111. Comprobar que $F(w)$ es una función de distribución, y obtener la función de probabilidad para $W$. Calcular $P(3 < W \leq 5)$.

$$F(w) = \begin{cases} 0 & \text{si} \quad w < 3 \\ 1/3 & \text{si} \quad 3 \leq w < 4 \\ 1/2 & \text{si} \quad 4 \leq w < 5 \\ 2/3 & \text{si} \quad 5 \leq w < 6 \\ 1 & \text{si} \quad w \geq 6 \end{cases}$$

112. Comprobar que $F(y)$ es una función de distribución, y obtener la función de probabilidad para $Y$. Calcular $P(Y \leq 100)$.

$$F(y) = \begin{cases} 0 & \text{si} \quad y < 0 \\ 1/4 & \text{si} \quad 0 \leq y < 5 \\ 1/3 & \text{si} \quad 5 \leq y < 7 \\ 1/2 & \text{si} \quad 7 \leq y < 100 \\ 5/6 & \text{si} \quad 100 \leq y < 102 \\ 1 & \text{si} \quad y \geq 102 \end{cases}$$

113. ¿Cuál es la función de distribución de $Z$? La variable aleatoria $Z$ tiene la función de probabilidad:

$$P_Z(x) = \begin{cases} 1/3 & \text{para} \quad x = 0, 1, 2 \\ 0 & \text{en otro caso} \end{cases}$$

114. Calcular la función de distribución de la variable aleatoria $U$ que tiene por función de probabilidad:

$$P_U(x) = \begin{cases} 1/3 & \text{para} \quad x = -3, 4 \\ 1/6 & \text{si} \quad x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

115. Comprobar que $F(x)$ es una función de distribución, y calcular la función de densidad de probabilidad para $X$. Usar esta función para calcular $P(-\frac{1}{2} < X < \frac{1}{2})$.

$$F(x) = \begin{cases} 0 & \text{si} \quad x < -1 \\ \dfrac{x+1}{2} & \text{si} \quad -1 \leq x \leq 1 \\ 1 & \text{si} \quad x > 1 \end{cases}$$

116. Comprobar que $F(y)$ es una función de distribución, y especificar la función de densidad de probabilidad para $Y$. Usar este resultado para obtener $P(-\frac{1}{4} < Y < \frac{1}{4})$.

$$F(y) = \begin{cases} 0 & \text{si} \quad y < 0 \\ \sqrt{y} & \text{si} \quad 0 \leq y \leq 1 \\ 1 & \text{si} \quad y > 1 \end{cases}$$

117 Comprobar que $F(z)$ es una función de distribución, y obtener la función de densidad para $Z$.

$$F(z) = \begin{cases} 0 & \text{si} \quad z < 0 \\ z^2 & \text{si} \quad 0 \leq z < \frac{1}{2} \\ 1 - 3(1-z)^2 & \text{si} \quad \frac{1}{2} \leq z < 1 \\ 1 & \text{si} \quad z \geq 1 \end{cases}$$

118. ¿Se puede decir que $H(t)$ es una función de distribución?

$$H(t) = \begin{cases} 0 & \text{si} \quad t < -1 \\ 1 - t^2 & \text{si} \quad -1 \leq t < \frac{1}{2} \\ \frac{1}{2} + t^2 & \text{si} \quad \frac{1}{2} \leq t < 1 \\ 1 & \text{si} \quad t \geq 1 \end{cases}$$

119. Comprobar que $G(t)$ es una función de distribución.

$$G(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ \frac{1}{2} & \text{si} \quad 0 \leq t < \frac{1}{2} \\ t & \text{si} \quad \frac{1}{2} \leq t \leq 1 \\ 1 & \text{si} \quad t > 1 \end{cases}$$

120. Dada la función de densidad $f(x)$ de la variable aleatoria $X$, obtener la función de distribución.

$$f(x) = \begin{cases} 1 & \text{si} \quad 99 < x < 100 \\ 0 & \text{en otro caso} \end{cases}$$

121. Sea $Y$ una variable aleatoria continua con función de densidad $f(y)$ Obtener la función de distribución.

$$f(y) = \begin{cases} 2(1-y) & \text{si} \quad 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

122. Sea $Z$ una variable aleatoria continua con función de densidad $f(z)$. Obtener la función de distribución.

$$f(z) = \begin{cases} 10e^{-10z} & \text{si} \quad z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

123. Tres baterias de misiles tienen como probabilidades de hacer blanco 0,2; 0,3 y 0,4 . Encontrar la función de distribución del número de blancos al disparar las tres baterias.

124. Se lanzan tres dados a la vez. Determinar la función de distribución de la suma de puntos obtenidos.

125. Se considera la función de distribución F(x).

$$F(x) = \begin{cases} 0 & \text{si} \quad x \leq 2 \\ \dfrac{(x-2)^3}{8} & \text{si} \quad 2 < x \leq 4 \\ 1 & \text{si} \quad x > 4 \end{cases}$$

Representarla gráficamente y la correspondiente función de densidad. Calcular: $P(1 < X < 3)$, $P(X \geq 3)$, $P(X < 3)$, $P(X \leq 3)$, $P(X \geq 4)$ y $P(X \leq 2)$.

126. Calcular la función de probabilidad para la variable aleatoria "diferencia en valor absoluto de los puntos obtenidos al tirar un dado dos veces".

127. Consideremos la variable aleatoria "suma de los puntos obtenidos al lanzar dos dados". Determinar:

    (a) Función de probabilidad.
    (b) Función de distribución.

128. Tres tiradores, A,B y C, tienen las siguientes probabilidades de hacer blanco $P(A) = 0{,}8$; $P(B) = 0{,}6$ y $P(C) = 0{,}7$. Encontrar la función de distribución del número de blancos obtenido al disparar una vez cada uno de los tres tiradores.

129. Se lanzan dos dados y se toma como variable aleatoria la diferencia de sus puntos (primero menos segundo). ¿Cuál es la media de la variable? ¿Cuál es la varianza?

130. Quiere diseñarse una máquina tragaperras cuyo juego equivale al siguiente experimento: de una urna con 6 bolas (3 rojas, 2 verdes y 1 negra), se extraen dos bolas sin reemplazamiento y se gana la cantidad correspondiente (1 punto por bola roja, 2 puntos por bola verde y 3 puntos por bola negra).

    Se pide:

    (a) Presentar la aplicación de $E$ sobre $R$: "puntuación ganada".
    (b) Obtener la función de distribución de la variable aleatoria definida en el apartado anterior.

131. La función de distribución asociada a un determinado componente eléctrico es del tipo:

$$F_\xi(x) = \begin{cases} 0 & \text{si} \quad x < 1 \\ \dfrac{2x^2 - 2}{30} & \text{si} \quad 1 \le x < K \\ 1 & \text{si} \quad x \ge K \end{cases}$$

Determinar $K$ para que $F_\xi(x)$ sea función de distribución.

132. Dada la función de probabilidad

$$P(X = i) = Ki \qquad \text{para } i = 1, 2, 3, \ldots, 20,$$

calcular:

(a) $P(X = 4)$.
(b) $P(3 \le X \le 10)$.
(c) $P(X^2 \le 100)$.

133. Dada la variable aleatoria $X$ con función de distribución:

$$F(x) = \begin{cases} 0 & \text{si} \quad x \le 0 \\ x^n & \text{si} \quad 0 < x \le 1 \\ 1 & \text{si} \quad x > 1 \end{cases}$$

donde $n \ge 1$, se pide:

(a) Calcular la función de densidad.
(b) Encontrar la mediana.
(c) Encontrar la media y la varianza de $X$.

134. Una variable aleatoria tiene como función de densidad:

$$f(x) = \begin{cases} 3x^2 & \text{si} \quad 0 < x < 1 \\ 0 & \quad\text{en otro caso} \end{cases}$$

Se pide:

(a) La función de distribución de $X$.

(b) Encontrar $F(2/3)$, $F(9/10)$ y $P(1/3 < X \leq 1/2)$.

(c) El valor de $a$ tal que $P(X \leq a) = 1/4$.

(d) La media y varianza de $X$.

135. Calcular el valor esperado de una apuesta a rojo o negro a la ruleta. (En la ruleta hay 18 números rojos, 18 negros y el 0)

136. La probabilidad de un tipo de accidente industrial en un año es 0,001. Una compañía de seguros propone a una empresa un seguro de accidentes cuyo coste anual es de 10.000 pesetas, comprometiéndose en caso de accidente a satisfacer una cantidad de 5 millones de pesetas en concepto de indemnización. Calcular el beneficio esperado para la compañía de seguros.

137. El tiempo de reparar una máquina en horas tiene la función de distribución:

$$F(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ x/2 & \text{si} \quad 0 \leq x \leq 1 \\ 1/2 & \text{si} \quad 1 \leq x \leq 2 \\ x/4 & \text{si} \quad 2 \leq x \leq 4 \\ 1 & \text{si} \quad x \geq 4 \end{cases}$$

(a) Dibujar la función de distribución.

(b) Obtener la función de densidad e interpretarla.

(c) Si el tiempo de reparación es superior a una 1 hora, ¿cuál es la probabilidad de que sea superior a 3,5 horas?

138. Dada la variable aleatoria discreta $X$, con función de densidad de probabilidad igual a

$$P_X(x) = \begin{cases} 0{,}1 & \text{si} \quad x = -2 \\ 0{,}2 & \text{si} \quad x = -1 \\ 0{,}4 & \text{si} \quad x = 0 \\ 0{,}2 & \text{si} \quad x = 1 \\ 0{,}1 & \text{si} \quad x = 2 \end{cases}$$

(a) Consideremos la variable aleatoria $Y = X^2$. Calcular su función de probabilidad.

(b) Representar gráficamente las funciones de probabilidad de las dos variables.

(c) Calcular $E(X^2)$.

139. Dada la función de probabilidad $P(X = n) = K(1-K)^{n-1}$, para $n = 1, 2, 3, \ldots$ y $0 < K \leq 1$, calcular el coeficiente de asimetría.

140. Una variable aleatoria $X$ tiene por función de densidad

$$f(x) = \frac{k}{e^x + e^{-x}} \quad \text{para} \quad x \geq 0.$$

Determinar el valor de $k$ y la función de distribución.

141. Dada la función $f(x) = -x^4 + 2x^3 + x^2 - 2x$, determinar en qué intervalos puede ser función de densidad.

142. La variable aleatoria $X$ tiene por función de densidad:

$$f(x) = \begin{cases} \dfrac{2x-1}{10} & \text{si} \quad 1 < x \leq 2 \\ 0{,}4 & \text{si} \quad 4 < x \leq 6 \\ 0 & \quad\text{fuera} \end{cases}$$

Calcular las probabilidades siguientes: $P(1{,}5 < X \leq 2)$ ; $P(X \leq 5)$; $P(4{,}5 < X \leq 5{,}5)$.

143. Consideremos la variable aleatoria continua $X$ definida en los intervalos $(0,\ 2)$, $(4,\ 8)$ y $(10,\ 11)$, con la siguiente función de densidad:

$$f(x) = \begin{cases} 0{,}15 & \text{si} \quad 0 < x \leq 2 \\ \dfrac{1}{16}x - \dfrac{1}{4} & \text{si} \quad 4 < x \leq 8 \\ \dfrac{-2}{5}x + \dfrac{22}{5} & \text{si} \quad 10 < x \leq 11 \end{cases}$$

Calcular:

(a) Función de Distribución.

(b) $P(X \leq 9)$.

(c) $P(1 < X \leq 10{,}5)$.

(d) $P(X \geq 6)$.

(e) Representar gráficamente las funciones de densidad y de distribución.

144. Dada la función de densidad

$$f(x) = K\, e^{-|ax|} \quad \text{si} \quad -\infty < x < \infty \quad , \quad a > 0.$$

Calcular mediante la función de distribución:

(a) $P(-2/a < X < 2/a)$.

(b) El valor de $a$ tal que $P(X \leq 3) = 0{,}6$.

145. La variable aleatoria $X$ tiene el siguiente campo de variación: los puntos $-1$ y $1$ y el intervalo $(2, 3)$, de la siguiente forma:

$P(X = -1) = 0{,}15; \qquad P(X = 1) = 0{,}25$

y en el intervalo $(2, 3)$ la función de densidad $f(x) = \frac{3}{5}(2x - 4)$.

Calcular la esperanza y la varianza de la variable aleatoria. Representar la función de densidad y de distribución.

146. Si la variable aleatoria $X$ tiene la función de densidad

$$f(x) = e^{-x} \quad \text{para} \quad x \geq 0,$$

calcular:

(a) Mediana de la distribución.

(b) Tercer cuartil.

(c) Percentil 32.

147. Si tenemos la función $f(x) = (7 - x)(x - 4)(x - 3)(x + 1)(x + 2)$, determinar en qué intervalo o intervalos, del campo de variación de la variable, puede ser función de densidad.

148. La variable aleatoria $X$ tiene por campo de variación el intervalo $(0, a)$. Su función de densidad en dicho intervalo es

$$f(x) = K\left(1 + \frac{x}{2}\right)$$

y su esperanza es igual a $7/3$. Determinar $K$ y $a$.

149. La vida en horas de un tipo de lámparas de radio es una variable aleatoria cuya función de densidad es

$$f(x) = \begin{cases} 0 & \text{si} \quad x < 100 \\ \dfrac{k}{x^2} & \text{si} \quad x \geq 100 \end{cases}$$

Determinar:

(a) El valor de $k$.

(b) La probabilidad de que de 5 lámparas ninguna tenga que ser sustituida en un aparato de radio durante las primeras 300 horas de uso.

(c) La probabilidad de que las 5 lámparas tengan que cambiarse durante las 300 horas primeras.

150. Si $X$ tiene como función de densidad

$$f(x) = 1 \quad \text{si} \quad 0 < x \leq 1,$$

calcular la función de densidad de las variables:

(a) $Y = X^2$.

(b) $Y = \sqrt{X}$.

(c) $Y = 1/X$.

151. Se tienen dos variables aleatorias $X$ e $Y$ independientes. La variable aleatoria $X$ tiene como función de densidad

$f(x) = 2(1 - x)$ para $0 \leq x \leq 1$,

y la variable $Y$ tiene función de densidad

$f(y) = e^{-y}$ para $y \geq 0$.

Calcular la probabilidad de que elegido un número al azar del campo de variabilidad de cada una de las dos variables, su cociente sea menor que 2.

152. Se eligen al azar dos números en el intervalo $(0, 2)$; calcular la probabilidad de que la suma de cuadrados sea menor que $1/4$.

153. Si la variable aleatoria $X$ tiene la función de densidad

$$f(x) = \frac{1}{243}x^2 \quad \text{si} \quad 0 \leq x \leq 9,$$

calcular $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$ y compararla con la cota de Chebychev.

154. El margen neto de cada contrato oscila aleatoriamente entre el 2 y el 6 por ciento, con densidad

$$f(x) = \begin{cases} \dfrac{x}{16} & \text{si} \quad 2 \leq x \leq 6 \\ \\ 0 & \text{en otro caso} \end{cases}$$

obtener la esperanza matemática de las variables:

(a) Margen neto que se acumula a reservas: $g_1(x)$ =0,2 $x$.
(b) Rentabilidad bruta: $g_2(x) = 1{+}1{,}1x$.

155. La cantidad producida de un cierto artículo es una variable aleatoria $X$ con función de densidad $f(x) = 3x^2/1000$, teniendo como campo de variación el intervalo $(0, 10)$. El precio del artículo se obtiene como una función de la cantidad producida: $Q = 40 - 2X$. Hállese:

(a) Función de densidad del precio.
(b) Precio medio.
(c) Función de densidad de la cantidad producida.
(d) Valor medio de la cantidad producida.

156. Dada la variable aleatoria $X$ con función de densidad $f(x) = 3x^2/2$ en el intervalo $(-1, 1)$ y $f(x) = 0$ fuera de él, calcular la función de densidad de la variable aleatoria $Y = X^2$.

157. Una empresa ha calculado que las ventas y costes unitarios están relacionados con el índice de precios de consumo $I$ a través de las siguientes relaciones:

Costes: $C = 1/3(I + 2)$ Ventas: $V = 1/3(19 - I)$

Si el índice de precios de consumo es una variable aleatoria con función de densidad $f(i) = 2i/99$; en el intervalo $1 \leq i \leq 10$.

(a) Calcular las distribuciones de los costes, de las ventas y del beneficio unitario.

(b) Calcular la probabilidad de que el beneficio sea negativo.

(c) Calcular el beneficio medio.

158. Calcular la función característica de la variable aleatoria $X$ cuya función de densidad es $f(x) = 2x$; para $0 \leq x \leq 1$.

159. Calcular la función de densidad de la variable aleatoria $Y = \sqrt{X}$, si la variable $X$ tiene por densidad

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{para} \quad -\infty < x < \infty$$

160. Una asociación benéfica lanza una lotería con 100 boletos de 500 ptas. y tres premios de 25.000, 10.000, y 5.000 ptas. ¿Cuál es la esperanza matemática de ganancia de una persona que juega dos boletos?

161. El tiempo necesario para la renovación de equipos industriales en una factoría tiene por función de densidad $f(x) = 0{,}1e^{-0{,}1x}$ si $x > 0$, y $f(x) = 0$ si $x \leq 0$ (con $x$ = número de meses). Se pide:

(a) Función característica.

(b) Los dos primeros momentos ordinarios.

(c) Esperanza matemática, media cuadrática y varianza.

162. La tasa de crecimiento industrial en una zona desarrollada tiene por función de densidad $f(x) = 30(x^4 - 2x^3 + x^2)$ si $0 \leq x \leq 1$, $f(x) = 0$ en el resto. Obtener los valores de la media, mediana y moda.

163. Se ha ofrecido a una editorial los derechos de publicación de una obra. Considerando que ello implicaría unos costes fijos de 3.000.000 ptas. y que el margen bruto de ganancia por libro vendido, según la tirada prevista, será de 500 ptas. y que la ganancia final obtenida está en función de lo que acaben comprando las 20 distribuidoras del país, para las que se ha estimado una cifra media de 420 ejemplares por distribuidora. Atendiendo al criterio de mayor esperanza de ganancia, determinar:

(a) Si deben adquirirse o no los derechos de publicación.

(b) ¿Para qué promedio de compra por distribuidora sería indistinta una u otra opción?

164. A un particular se le plantea la posibilidad de invertir en tres alternativas: $a_1$ = inmuebles, $a_2$ = préstamo, $a_3$ = bolsa. Está interesado en recuperar el capital en un plazo corto, en el que ha determinado las probabilidades de que la evolución de la economía presente las siguientes fases: $F_1$ = contracción (12%), $F_2$ = reanimación (27%), $F_3$ = expansión (41%), $F_4$ = crisis (20%).

Supóngase que se ha podido construir la siguiente matriz de resultados previsibles, dados en tanto por ciento sobre el capital arriesgado.

| | $F_1$ | $F_2$ | $F_3$ | $F_4$ |
|---|---|---|---|---|
| $a_1$ | 2,5 | 8 | 4,5 | −1,5 |
| $a_2$ | 2,5 | 2,5 | 2,5 | 2,5 |
| $a_3$ | 0 | 7,5 | 9,5 | −2 |

(a) ¿Cuál de las tres alternativas da mayor rentabilidad media?

(b) Si elige la segunda alternativa, por ser la menos arriesgada, ¿cuál será el coste de oportunidad respecto a la mejor opción?

165. Un detallista recibe la oferta de adquirir piezas de carne congelada a 223 euros cada unidad. La venta de cada pieza produce unos ingresos de 300 euros durante el primer mes, perdiendo a partir de entonces calidad, por lo que sólo podría revenderla como despojos al precio de 100 euros por unidad. Si la demanda prevista, para un mes, sigue una función de probabilidad del tipo:

| Piezas demandadas | Probabilidad |
|---|---|
| menos de 2 | 0,028 |
| 2 | 0,128 |
| 3 | 0,229 |
| 4 | 0,211 |
| 5 | 0,108 |
| 6 | 0,101 |
| más de 6 | 0,203 |

Calcular el número de piezas que debe adquirir el detallista para asegurar mayor beneficio esperado.

166. Se sacan al azar dos cartas sin reemplazamiento de una baraja de 52 cartas. Sea $X$ el número de ases que aparecen e $Y$ el número de espadas que aparecen. Obtener $P_{X,Y}(x,y)$ y calcular $P(X > Y)$.

167. En una bolsa hay tres bolas negras y 7 rojas. Se seleccionan dos bolas al azar sin reemplazamiento. Sea $X$ el número de bolas negras sacadas, y sea $Y$ el número de bolas rojas. Obtener $P_{X,Y}(x,y)$ y calcular $P(X \leq Y)$.

168. Se selecciona al azar un punto del interior del círculo centrado en el origen de radio unidad. Sean $X$ e $Y$ las coordenadas del punto elegido. Obtener $f_{X,Y}(x,y)$ ; $f_X(x)$ y $f_Y(y)$ .

169. Calcular $A$ para que $f(x)$ sea función de densidad:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{Ax}{y} & \text{si} \quad 0 < x < 1 \quad , \quad 1 < y < 2 \\ \\ 0 & \text{fuera} \end{cases}$$

170. Una familia tiene dos hijos varones. Sea $X$ la estatura del mayor de los dos cuando sea adulto e $Y$ la estatura del menor cuando sea adulto. Suponer que $(X,Y)$ tiene la misma verosimilitud de caer dentro del rectángulo con vértices en $(66,68)$, $(66,72)$, $(71,68)$, $(71,72)$. Calcular la probabilidad de que el mayor sea más alto que el menor al llegar a ser adultos.

171. Suponer que $(X,Y)$ es una variable aleatoria bidimensional con función de densidad

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{x} & \text{si} \quad 0 < y < x \quad , \quad 0 < x < 1 \\ \\ 0 & \text{fuera} \end{cases}$$

Encontrar las densidades marginales para $X$ y para $Y$.

172. Suponer que $(X,Y)$ tiene función de densidad

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & \text{si} \quad -a < x < a \quad , \quad -a < y < a \\ 0 & \text{fuera} \end{cases}$$

(a) Encontrar $a$.

(b) Calcular las densidades marginales para $X$ e $Y$.

173. Suponer que la función de densidad de $(X, Y)$ es: $f_{X,Y}(x, y) = 1/2$ en el interior del cuadrado de vértices: (a,0), (−a,0), (0,a), (0,−a) y $f_{X,Y}(x, y) = 0$ fuera de dicho cuadrado.

   (a) Calcular $a$.

   (b) Calcular las densidades marginales de $X$ e $Y$.

174. Sea $(X, Y)$ una variable aleatoria bidimensional con función de densidad conjunta:

$$f_{X,Y}(x, y) = \begin{cases} k & \text{si} \quad |x| \leq y \quad , \quad 0 \leq y \leq 1 \\ 0 & \qquad\qquad \text{fuera.} \end{cases}$$

   (a) Calcular $k$.

   (b) Calcular la función de distribución conjunta.

175. La rentabilidad de un negocio es una variable aleatoria bidimensional de componentes: $X$ rentabilidad del capital, $Y$ rentabilidad de las ventas.

   La función de densidad conjunta es:

$$f_{X,Y}(x, y) = \begin{cases} e^{-(x+y)} & \text{si} \quad x, y > 0 \\ 0 & \qquad \text{fuera.} \end{cases}$$

   (a) Calcula la función de distribución conjunta.

   (b) ¿Cuál es la probabilidad de que la rentabilidad del capital sea menor que la de las ventas? ¿Y de que ambas sean iguales?

176. Sea $(X, Y)$ una variable aleatoria bidimensional con función de probabilidad conjunta:

$$P_{X,Y}(x = i \ , \ y = j) = \frac{1}{2^{i+j}} \quad \text{si} \quad i, j = 1, 2, 3, \ldots$$

   Calcular las funciones de probabilidad marginales.

177. Calcular las distribuciones marginales y condicionadas de la variable aleatoria bidimensional $(X, Y)$ cuya función de densidad es:

$$f_{X,Y}(x, y) = \begin{cases} x \cdot y & \text{si} \quad 0 \leq x \leq 2 \quad , \quad 0 \leq y \leq 1 \\ 0 & \qquad\qquad \text{fuera.} \end{cases}$$

178. Se lanza una moneda tres veces y se consideran las variables:

$X=$ número de caras obtenidas.

$Y=$ diferencia, en valor absoluto, entre número de caras y de cruces.

(a) Calcula la función de distribución conjunta, las marginales y las condicionadas.

(b) Comprueba si $X$ e $Y$ son independientes.

179. Sea $(X, Y)$ una variable aleatoria bidimensional que representa a dos componentes del fondo de comercio de ciertas empresas. La función de densidad conjunta es:

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{si} \quad 0 < x < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Calcula las funciones de densidad marginales y condicionadas.

(b) $P[Y > \frac{1}{2} \,/\, x = \frac{1}{2}]$ , $P[X > \frac{1}{3} \,/\, y = \frac{2}{3}]$.

(c) Comprueba si $X$ e $Y$ son independientes.

180. Sean las variables aleatorias:

$X$: taquilla en cientos de miles de pesetas del cine A

$Y$: taquilla en cientos de miles de pesetas del cine B

La función de densidad conjunta es:

$$f_{X,Y}(x,y) = \begin{cases} \frac{4}{3}\left(xy + \frac{x^2}{2}\right) & \text{si} \quad 0 < x < 1 \quad , \quad 0 < y < 2 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcula $P[Y > 1 \,/\, X < 0{,}5]$.

181. El control de calidad de una factoria puede realizarse en dos secciones A y B, pudiéndose efectuar 5 inspecciones en la sección A y hasta 4 en la B. Sean:

$X=$ número de inspecciones realizadas en la sección A

$Y=$ número de inspecciones realizadas en la sección B

La función de probabilidad conjunta es:

| X \ Y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0,01 | 0 | 0,04 | 0,01 |
| 2 | 0,02 | 0,05 | 0 | 0,04 |
| 3 | 0,04 | 0 | 0 | 0,18 |
| 4 | 0,08 | 0,07 | 0,15 | 0,01 |
| 5 | 0,03 | 0,11 | 0,09 | 0,08 |

Calcular: $P[X \leq 4]$, $P[Y \geq 2]$, $P[Y = 3/X \geq 4]$, $P[Y > 2/X \geq 3]$, $P[2 < X \leq 4/Y = 1]$ y $P[X \geq 3/Y \geq 2]$.

182. La función de densidad conjunta de los indicadores $X$ e $Y$ de concentración parcelaria es:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{2} & \text{si} \quad (x,y) \in \mathcal{C} \\ \\ 0 & \text{en otro caso} \end{cases}$$

donde $\mathcal{C}$ es el cuadrado de vértices $\{(1,\ 0),\ (0,\ 1),\ (-1,\ 0),\ (0,\ -1)\}$

(a) Calcula las densidades marginales y condicionadas.

(b) Estudia la independencia de las variables $X$ e $Y$.

183. Sea $(X, Y)$ la variable aleatoria descrita por la siguiente tabla:

| Y \ X | 0 | 1 |
|---|---|---|
| 2 | 1/4 | 1/4 |
| 4 | 3/8 | 1/8 |

Calcula $E[g(X,Y)]$ , siendo $g(X,Y) = X + Y^2$.

184. Sea $(X, Y)$ una variable aleatoria continua con función de densidad:

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{si} \quad 0 \leq x \leq 1\ ,\ 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcular $E[X + Y^2]$.

185. Consideremos la variable aleatoria bidimensional discreta con función de probabilidad:

| X \ Y | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 1 | 1/14 | 0 | 2/14 | 1/14 |
| 2 | 0 | 1/14 | 3/14 | 1/14 |
| 3 | 2/14 | 1/14 | 0 | 2/14 |

(a) Calcular la funciones de probabilidad marginales.
(b) Calcular E[X] , E[Y] , E[XY] , E[$X^2$] , E[$Y^2$], Cov(X,Y).

186. Sea $(X,Y)$ una variable aleatoria continua con función de densidad:

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{si } \; |x| \leq y \, , \, 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

(a) Calcular las distribuciones marginales.
(b) Calcular E[X] , E[Y] , E[XY] , Cov(X,Y).

187. Sea una variable aleatoria bidimensional con función de densidad:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{si } \; 0 \leq y \leq x \leq 2 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Calcula las distribuciones marginales.
(b) Calcula E[X] , E[Y] , E[XY] , Var[X] , Var[Y] , Cov(X,Y).

188. Dada la variable aleatoria bidimensional $(X,Y)$ con la función de probabilidad siguiente:

| Y\X | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 4 | 0,01 | 0,04 | 0,18 | 0,01 | 0,08 |
| 3 | 0,03 | 0 | 0,15 | 0,09 | 0,09 |
| 2 | 0 | 0,05 | 0 | 0,07 | 0,11 |
| 1 | 0,01 | 0,02 | 0,04 | 0,08 | 0,03 |

Calcular las probabilidades:

(a) $P(X \leq 4)$; $P(Y \geq 2)$.
(b) $P(Y = 3/X \leq 4)$; $P(Y > 2/X \geq 3)$.

(c) $P(2 < X \leq 4/Y = 1)$; $P(X \geq 3/Y \geq 2)$.

189. Dada la función de densidad bidimensional

$$f_{X,Y}(x,y) = \begin{cases} k \cdot x^2 \cdot y^3 \cdot e^{-(5x+2y)} & \text{si} \quad x \geq 0\,, y \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular:

(a) $k$.

(b) Funciones de densidad marginales.

(c) Comprobar la independencia de las variables.

190. Dada la variable aleatoria bidimensional continua $(X, Y)$ con función de densidad conjunta

$$f_{X,Y}(x,y) = \begin{cases} k \cdot (2x + 3y^2) & \text{si} \quad 0 \leq x \leq 1 \quad , \quad 0 \leq y \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular:

(a) La constante $k$.

(b) Función de distribución conjunta.

(c) Funciones de distribución marginales.

(d) Funciones de distribución condicionales.

(e) P(X $\leq$ 0,1; Y $\leq$ 0,2).

(f) P(X $\geq$ 0,1; Y $\leq$ 0,8).

(g) P(0,1 < X $\leq$ 0,5; 0,3 < Y $\leq$ 0,8).

(h) P(X $\geq$ 0,3).

(i) Si P(0,6 < X $\leq$ a) = 0,1; calcular a.

(j) P(X $\leq$ 0,5/Y $\leq$ 0,2).

(k) P(0,8 < Y $\leq$ 0,9/0,5 < X $\leq$ 0,7).

(l) P(X $\leq$ 0,4/Y = 0,6).

(m) Si P(0,1 < Y $\leq$ 0,7/X = a) = 0,4; calcular a.

191. Dada la función de densidad tridimensional

$$f_{X,Y,Z}(x,y,z) = \begin{cases} \dfrac{1}{2} \cdot (x+y+z) & \text{si} \quad 0 \leq x,y,z \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular:

(a) P(X $\leq$ 1/2; Y $\geq$ 0,2).

(b) P(0,6 $<$ X $\leq$ 0,8 / 0 $<$ Y $\leq$ 0,2; Z $\geq$ 0,7)

(c) P(0,8 $\leq$ X; Z $\leq$ 0,5 / Y $\leq$ 0,9).

192. La variable aleatoria bidimensional $(X, Y)$ tiene como función de densidad $f(x,y) = e^{-(x+y)}$ en el primer cuadrante; y $f(x,y) = 0$ en los otros tres. Se toman al azar tres puntos en el primer cuadrante siguiendo esta distribución. Calcular la probabilidad de que uno al menos pertenezca al cuadrado: $0 \leq x \leq 1$; $0 \leq y \leq 1$.

193. En un aparato de control actúan dos variables $x_1$, $x_2$, independientes, ambas con distribución uniforme, la primera entre 1 y 9, la segunda entre 1 y $a$. El aparato funciona bien cuando $x_1 < x_2^2$. Calcular el valor de $a$ para que $P(x_1 > 4x_2^2) \leq 0{,}01$.

194. El tiempo que un camión permanece en un almacén está definido por una variable aleatoria X. Sea Y la variable tiempo de espera en la cola, y Z el tiempo de descarga (X = Y + Z). La distribución conjunta de X e Y es:

$$f(x,y) = \begin{cases} \dfrac{1}{4} e^{-x/2} & \text{si} \quad 0 \leq y \leq x < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Se pide:

(a) Calcular el tiempo medio total que permanece un camión en el almacén.

(b) Calcular el tiempo medio de descarga.

(c) Calcular el coeficiente de correlación entre el tiempo total y el tiempo de espera en la cola.

195. Un reciente estudio revela que entre los usuarios habituales de una línea de autobuses, el 80 % no tenía automóvil propio; del resto, el 90 % tenía un automóvil de propiedad y el 10 % dos automóviles,

no habiendo ningún propietario de tres o más coches que utilizara el servicio. Obtener la distribución conjunta de la variable $(X, Y)$, a partir de una muestra aleatoria de tres usuarios de la línea, donde:

$$X = \text{usuarios con un solo coche.} \quad Y = \text{usuarios sin coche.}$$

196.

197. La rentabilidad de un negocio es una variable aleatoria bivariante X de componentes $X_1$ = rentabilidad del capital y $X_2$ = rentabilidad de las ventas, con función de densidad

$$f(x_1, x_2) = \begin{cases} \frac{4}{3}(x_1 + x_1 x_2) & \text{si} \quad 0 \leq x_1 \leq 1 \quad , \leq x_2 \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Se pide:

(a) Representación gráfica.
(b) Densidades marginales y funciones de distribución marginales.
(c) Densidades condicionadas y funciones de distribución condicionadas.
(d) Indicar si son o no independientes ambas componentes.

198. Dada la función de densidad conjunta:

$$f(x, y) = \begin{cases} 3x & \text{si} \quad 0 < y < x \quad , \quad 0 < x < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Obtener las distribuciones marginales y la condicionada $f(x/y)$.

199. Para el período de rotación de los créditos $(X_1)$ y el de rotación de los débitos $(X_2)$ se ha especificado la densidad conjunta:

$$f(x_1, x_2) = \begin{cases} -\frac{300 - 2x_1 - x_2}{1.500.000} & \text{si} \quad 0 \leq x_1 \leq 100 \quad , \quad 0 \leq x_2 \leq 100 \\ 0 & \text{en otro caso.} \end{cases}$$

Se pide:

(a) Indicar si se trata o no de dos características independientes.
(b) Calcular $E[X_2 - X_1]$.

200. La producción de una factoría puede destinarse a dos mercados, con márgenes brutos de ganancia dados por las variables $X_1$ y $X_2$, tales que $\mathrm{Var}(X_1) = 32{,}56$; $\mathrm{Var}(X_2) = 21{,}97$ y $\mathrm{Cov}(X_1, X_2) = -4{,}31$. Si adoptamos como función del riesgo de las ganancias

$$\mathrm{R} = \mathrm{Var}\ [\alpha X_1 + (1 - \alpha) X_2].$$

Donde $\alpha$ indica la partición de las ventas. Calcular el valor del coeficiente $\alpha$ que minimiza dicho riesgo.

# Unidad Temática IV

# Distribuciones de Probabilidad.

# Resumen Teórico

## 4.1 Distribuciones discretas más usuales.

Para tener perfectamente caracterizada una variable aleatoria discreta tendremos que conocer su función de probabilidad, es decir, la probabilidad que tiene de tomar cada uno de los distintos valores.

### 4.1.1 Distribución de Bernoulli.

Decimos que un experimento es de tipo Bernoulli cuando sólo hay dos posibles resultados $E$ y $F$, que suelen identificarse a menudo como Éxito y Fracaso, siendo las probabilidades de ocurrencia de éstos $p$ y $q$ respectivamente. Obviamente se verifica que tanto $p$ como $q$ son estrictamente mayores que cero y además $p + q = 1$.

Definimos la variable aleatoria $X$ de la siguiente manera:

$$X(E) = 1 \qquad \text{y} \qquad X(F) = 0.$$

Es decir, $X$ tomará el valor 1 en caso de éxito y 0 en caso de fracaso. Entonces tenemos la siguiente función de probabilidad:

$$P(X = 1) = p \qquad P(X = 0) = q = 1 - p.$$

En este caso tenemos que la media es $p$ y la varianza es $pq$.

Podemos resumir las características más importantes de esta distribución en la tabla 4.1, donde $M_X(t)$ es la función generatriz de momentos y $\varphi_X(t)$ es la función característica.

Ejemplo: Lanzar una moneda al aire.

| Distribución Bernoulli | |
|---|---|
| rango | k=0,1 |
| $P(X=k)$ | $p$ si k=1<br>$q$ si k=0 |
| $E[X]$ | $p$ |
| $V[X]$ | $pq$ |
| $M_X(t)$ | $pe^t+q$ |
| $\varphi_X(t)$ | $pe^{it}+q$ |

Tabla 4.1: Distribución Bernoulli.

### 4.1.2 Distribución binomial.

Se repite $n$ veces independientes un experimento Bernoulli. La variable aleatoria $X$ va a contar el número de éxitos. Como ejemplo, imaginemos el suceso "obtener 5 caras en 15 lanzamientos de una moneda". Para obtener su distribución de probabilidad consideremos que realizamos $n$ veces el experimento Bernoulli y que han aparecido $k$ éxitos y $n-k$ fracasos, – por ejemplo en la secuencia EE...EFF...F –, la probabilidad de este suceso vendría dada por $p^k q^{n-k}$. Sin embargo, no es esa la única forma de obtener $k$ éxitos, pues habría que considerar todas las permutaciones (con repetición) de los elementos $E$ y $F$, lo que daría un total de

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

La probabilidad de obtener $k$ éxitos en $n$ pruebas sería, por tanto:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2 \ldots n.$$

| Distribución Binomial | |
|---|---|
| rango | $k = 0, 1 \ldots n$ |
| $P(X = k)$ | $\binom{n}{k} p^k (1-p)^{n-k}$ |
| $E[X]$ | $np$ |
| $V[X]$ | $npq$ |
| $M_X(t)$ | $(pe^t + q)^n$ |
| $\varphi_X(t)$ | $(pe^{it} + q)^n$ |

Tabla 4.2: Distribución Binomial.

Obsérvese que se trata de una verdadera distribución de probabilidad, pues:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1.$$

Lo cual además justifica el nombre de la distribución, pues lo anterior no es sino el desarrollo del binomio $(p+q)^n$. La variable aleatoria se dice que sigue una distribución binomial de parámetros $n$ y $p$ y se le suele denotar por $\mathrm{B}(n, p)$. Podemos interpretar una binomial como la suma de $n$ variables aleatorias independientes con distribución de Bernoulli.

Las características más importantes de esta distribución están reflejadas en la tabla 4.2.

Para el caso particular de que $n$ sea igual a 1 se obtiene la Distribución de Bernoulli.

Una distribución multidimensional asociada a la anterior es la **Distribución multinomial**, que se presenta cuando consideramos $k$ sucesos

$A_1, A_2, \ldots A_k$ que forman un sistema exhaustivo y excluyente, siendo sus probabilidades $p_1$,$p_2$, $\ldots p_k$ respectivamente, donde $p_1+p_2+\ldots+p_k = 1$.

Se realizan $n$ pruebas independientes. Se consideran los sucesos:

$X_1$ = número de veces que ocurre $A_1$

$\vdots$

$X_k$ = número de veces que ocurre $A_k$.

Entonces si $n = n_1 + n_2 + \ldots + n_k$ tenemos:

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1!n_2!\ldots n_k!} {p_1}^{n_1} {p_2}^{n_2} \ldots {p_k}^{n_k}.$$

Si $k = 2$, la distribución multinomial coincide con la binomial.

Ejemplo: Se lanza un dado 15 veces. Hallar la probabilidad de obtener 3 veces el valor uno, 1 vez el valor dos, 2 veces el valor tres, 2 veces el valor cuatro, 5 veces el valor cinco y 2 veces el valor seis.

### 4.1.3 Distribución Hipergeométrica.

La distribución binomial se aplica en el caso de tener repeticiones independientes de un experimento de Bernoulli, por tanto, la probabilidad de éxito es la misma en todas las repeticiones. Un caso típico de aplicación de la distribución binomial será el de realizar extracciones con reemplazamiento. En el caso de extracciones sin reemplazamiento usaremos la distribución Hipergeométrica.

Supongamos en una urna hay $N$ bolas de las cuales $N_1$ son blancas y $N_2$ son negras. Extraemos $n$ bolas sin reemplazamiento. La variable aleatoria $X$ que mide el número de bolas blancas extraídas seguirá una distribución Hipergeométrica. Su función de Probabilidad es:

$$P(X = k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

Si $n$ es menor que $N_1$ y que $N_2$, $k$ podrá tomar valores entre 0 y $n$. Si $N_1$ es menor que $n$, el número de bolas blancas que se pueden extraer será, como máximo $N_1$. Si $n$ es mayor que $N_2$, podrán salir como máximo $N_2$ bolas negras, con lo que saldrán, como mínimo $n - N_2$ bolas blancas. Por tanto el rango de valores que puede tomar una hipergeométrica variará entre $\alpha = \text{máx}\{0, n - N_2\}$ y $\beta = \text{mín}\{N_1, n\}$.

Los valores de la media y la varianza están reflejados en la tabla 4.3.

| Distribución Hipergeométrica | |
|---|---|
| rango | $k = \alpha, \ldots, \beta$<br>$\alpha = \text{máx}\{0, n - N_2\}$<br>$\beta = \text{mín}\{N_1, n\}$ |
| $P(X = k)$ | $\dfrac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}$ |
| $E[X]$ | $np$ |
| $V[X]$ | $npq\frac{N-n}{N-1}$ |
| $M_X(t)$ | No se emplea |
| $\varphi_X(t)$ | No se emplea |
| | siendo $p = \dfrac{N_1}{N}$ y $q = 1 - p$ |

Tabla 4.3: Distribución Hipergeométrica.

### 4.1.4 Distribución geométrica o de Pascal.

Mide el número de intentos necesarios hasta conseguir el primer éxito al realizar repeticiones independientes de un experimento Bernoulli.

$$P(X = k) = p(1-p)^{(k-1)} \qquad k = 1, 2, \ldots$$

Podemos resumir las características más importantes de esta distribución en la tabla 4.4.

| Distribución Geométrica | |
|---|---|
| rango | $k = 1, \ldots$ |
| $P(X = k)$ | $p(1-p)^{(k-1)}$ |
| $E[X]$ | $\frac{1}{p}$ |
| $V[X]$ | $\frac{q}{p^2}$ |
| $M_X(t)$ | $\frac{pe^t}{1-qe^t}$ |
| $\varphi_X(t)$ | $\frac{pe^{it}}{1-qe^{it}}$ |

Tabla 4.4: Distribución Geométrica o de Pascal.

### 4.1.5 Distribución Binomial Negativa.

Es una generalización de la distribución geométrica. Se define la variable aleatoria $X$ como el número de intentos necesarios para producir el $r$-ésimo éxito, tomando los valores $k = r, r+1, \ldots$ con función de probabilidad:

$$P(X = k) = \binom{k-1}{r-1} p^r \cdot q^{(k-r)} \qquad k = r, r+1, \ldots$$

| Distribución Binomial Negativa | |
|---|---|
| rango | $x = r, r+1, \ldots$ |
| $P(X=k)$ | $\binom{k-1}{r-1} p^r q^{k-r}$ |
| $E[X]$ | $\frac{r}{p}$ |
| $V[X]$ | $\frac{rq}{p^2}$ |
| $M_X(t)$ | $\left(\frac{pe^t}{1-qe^t}\right)^r$ |
| $\varphi_X(t)$ | $\left(\frac{pe^{it}}{1-qe^{it}}\right)^r$ |

Tabla 4.5: Distribución Binomial Negativa.

Podemos resumir las características más importantes de esta distribución en la tabla 4.5. Para $r$ igual a 1 tendríamos la distribución geométrica.

### 4.1.6 Distribución de Poisson.

Supongamos una determinada eventualidad que se produce en un **soporte continuo** (tiempo, línea, área, espacio,...), de forma **independiente** y con una cierta **estabilidad** para una determinada unidad de soporte. Como ejemplos podemos considerar el número de coches que pasan por un semáforo en un período de tiempo, el número de defectos por metro cuadrado de una pieza de tela, el número de hormigas de una cierta especie en un metro cúbico de tierra, etc. Las tres condiciones en negrita caracterizan el denominado Proceso de Poisson, y su variable aleatoria está definida por el número de sucesos que se producen en un intervalo de longitud fija. La distribución de Poisson se obtiene como límite de la Binomial cuando el número de veces que se realiza el experimento ($n$) tiende a infinito, la probabilidad de éxito ($p$) tiende a cero y el número medio de éxitos ($np$) se estabiliza alrededor de un

| Distribución de Poisson | |
|---|---|
| rango | $k = 0, 1 \ldots$ |
| $P(X = k)$ | $\frac{\lambda^k}{k!} \cdot e^{-\lambda}$ |
| $E[X]$ | $\lambda$ |
| $V[X]$ | $\lambda$ |
| $M_X(t)$ | $e^{\lambda(e^t - 1)}$ |
| $\varphi_X(t)$ | $e^{\lambda(e^{it} - 1)}$ |

Tabla 4.6: Distribución de Poisson.

número ($\lambda$) que será la media y el valor que caracterizará a la distribución. Calculando dicho límite cuando $n \to +\infty$ y $np = \lambda$ obtenemos que la fórmula de la probabilidad es:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

La media y la varianza de esta distribución valen:

$$\mathrm{E}[X] = \lambda \qquad \mathrm{Var} = \lambda$$

La distribución, que notaremos $P(\lambda)$, aparece al considerar el número de llamadas telefónicas realizadas por unidad de tiempo a un mismo teléfono, siendo $\lambda$ el número medio de llamadas. Sus principales características aparecen en la tabla 4.6.

La distribución de Poisson representa el número de sucesos aleatorios independientes que ocurren a una rapidez constante a lo largo del tiempo o del espacio.

Por ejemplo, en la Segunda Guerra Mundial se realizó un estudio sobre el número de bombas que caían sobre diversas zonas (del mismo área) de Londres. Al observar que se ajusta a una distribución de Poisson se deduce que los bombardeos eran completamente aleatorios, siendo la probabilidad de que caiga una bomba la misma para todas las áreas.

Otro ejemplo de interés puede presentarse si se estudia el número de accidentes por unidad de trabajo (del mismo número de empleados). Si la distribución es de Poisson, deducimos que todos los empleados tienen la misma probabilidad de accidentarse.

Todo lo anterior está referido a una unidad de soporte de tamaño 1, si quisiéramos generalizarlo a cualquier intervalo de tamaño $t$ la función de probabilidad sería:

$$P(X = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad k = 0, 1, 2, \ldots$$

## 4.2 Distribuciones continuas más usuales.

Si para tener perfectamente caracterizada una variable aleatoria discreta tendríamos que conocer su función de probabilidad, en el caso de las distribuciones continuas tendremos que conocer la función de densidad.

### 4.2.1 Distribución exponencial negativa.

Si consideramos de nuevo un proceso de Poisson y definimos X como el intervalo de ocurrencia entre dos sucesos consecutivos nos encontraremos ante una variable Exponencial. Obsérvese que hemos dado un salto cualitativo, puesto que la variable así definida es continua, tomando valores en el intervalo $(0, \infty)$. Puesto que se verifica que:

$$P[X > t] = P[\text{cero sucesos en}(0, t)] = e^{-\lambda t}.$$

Y por tanto

$$F(t) = P[X \leq t] = 1 - e^{-\lambda t},$$

derivando obtenemos que la función de densidad es:

| Distribución Exponencial | |
|---|---|
| rango | $x \in (0, +\infty)$ |
| $f(x)$ | $\lambda e^{-\lambda x}$ si $x \geq 0$ <br> $0$ si $x < 0$ |
| $E[X]$ | $\frac{1}{\lambda}$ |
| $V[X]$ | $\frac{1}{\lambda^2}$ |
| $M_X(t)$ | $\frac{\lambda}{\lambda - t}$ |
| $\varphi_X(t)$ | $\frac{\lambda}{\lambda - it}$ |

Tabla 4.7: Distribución Exponencial.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

Los valores de la media y la varianza son:

$$E(X) = \frac{1}{\lambda} \qquad \text{Var}(X) = \frac{1}{\lambda^2}$$

Podemos resumir las características más importantes de esta distribución en la tabla 4.7.

Si una variable aleatoria $X$ se distribuye según una exponencial de parámetro $\lambda$, se denota como $X \sim E(\lambda)$.

La distribución exponencial es muy utilizada al realizar estudios de fiabilidad, mantenimiento y procesos de espera. Una propiedad interesante es la de "pérdida de memoria":

$$P[X > t + t_0 / X > t_0] = P[X > t].$$

Si el tiempo de funcionamiento de un proceso se distribuye según una ley exponencial, la probabilidad de que se produzca un fallo no depende del tiempo que lleva el proceso en funcionamiento. La probabilidad de fallo es constante a lo largo de toda la vida del proceso, sin importar su edad.

### 4.2.2 Distribución Gamma.

Es una variable aleatoria continua, que depende de dos parámetros $\alpha$ y $\lambda$ que deben ser dos números reales estrictamente positivos. Su función de densidad es:

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

donde la función $\Gamma(\alpha)$ que aparece en el denominador es la función gamma, definida por

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

Esta función aparece en Análisis matemático como generalización al caso real del concepto de factorial. En efecto, mediante integración por partes se puede comprobar que

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

además se cumple que $\Gamma(1) = 1$, con lo que si $n$ es un número entero se tiene que

$$\Gamma(n) = (n - 1)!$$

Estas propiedades de la función $\Gamma(\alpha)$ se utilizan para resolver las integrales que aparecen para calcular la media y la varianza.

En el caso de que $\alpha$ sea un número entero, pongamos $\alpha = n$, tenemos una distribución Gamma de parámetros $n$ y $\lambda$, también llamada distribución de Erlang de parámetros $n$ y $\lambda$. La función de densidad quedaría entonces

$$f(x) = \begin{cases} \dfrac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

| Distribución Gamma | | |
|---|---|---|
| rango | $x \in (0, +\infty)$ | |
| $f(x)$ | $\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$ | si $x \geq 0$ |
| | $0$ | si $x < 0$ |
| $E[X]$ | $\frac{\alpha}{\lambda}$ | |
| $V[X]$ | $\frac{\alpha}{\lambda^2}$ | |
| $M_X(t)$ | $\left(\frac{\lambda}{\lambda - t}\right)^\alpha$ | |
| $\varphi_X(t)$ | $\left(\frac{\lambda}{\lambda - it}\right)^\alpha$ | |

Tabla 4.8: Distribución Gamma.

En este caso, la distribución gamma es una generalización de la exponencial. En efecto, si consideramos de nuevo un proceso de Poisson y definimos X como el intervalo transcurrido entre la ocurrencia $K$ y la $K + n$ nos encontraremos ante una variable Gamma (o Erlang) de parámetros $n$ y $\lambda$. Al igual que en el caso anterior, la variable así definida es continua, tomando valores en el intervalo $(0, \infty)$. Para $n = 1$ aparece como caso particular la ley exponencial. Podemos afirmar que la distribución Gamma de parámetros $n$ y $\lambda$ es la suma de $n$ variables independientes, todas ellas distribuidas según una ley $E(\lambda)$.

Podemos resumir las características más importantes de esta distribución en la tabla 4.8.

Si una variable aleatoria $X$ se distribuye según una Gamma de parámetros $\alpha$ y $\lambda$, se denota como $X \sim \Gamma(\alpha, \lambda)$.

### 4.2.3 Distribución uniforme.

Se presenta cuando se toma al azar un número en el intervalo $(a, b)$ de tal modo que dos subintervalos de la misma amplitud tengan la misma probabilidad de que el número seleccionado esté en su interior. La variable X sigue una distribución Uniforme o Rectangular en el intervalo $(a, b)$, $X \sim U(a, b)$, cuando su función de densidad viene dada de la siguiente forma:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad x \in (a, b) \\ 0 & \text{si} \quad x \notin (a, b) \end{cases}$$

Los valores de la media y la varianza son:

$$E(X) = \frac{b+a}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Podemos resumir las características más importantes de esta distribución en la tabla 4.9. La más utilizada es la U(0,1).

| Distribución Uniforme | |
|---|---|
| rango | $x \in (a, b)$ |
| $f(x)$ | $\begin{cases} \dfrac{1}{b-a} & \text{si} \quad x \in (a, b) \\ 0 & \text{si} \quad x \notin (a, b) \end{cases}$ |
| $E[X]$ | $\dfrac{b+a}{2}$ |
| $V[X]$ | $\dfrac{(b-a)^2}{12}$ |
| $M_X(t)$ | $\dfrac{e^{tb} - e^{ta}}{t(b-a)}$ |
| $\varphi_X(t)$ | $\dfrac{e^{itb} - e^{ita}}{it(b-a)}$ |

Tabla 4.9: Distribución Uniforme.

## 4.3 Distribución Normal.

Decimos que una variable aleatoria $X$ se distribuye según una Normal de media $\mu$ y de desviación típica $\sigma$, y se representará por $X \sim N(\mu, \sigma)$, si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La distribución está caracterizada por los parámetros $\mu$ y $\sigma$, cuyo significado veremos luego, siendo necesariamente positivo. Nos referiremos a esta distribución como $N(\mu, \sigma)$.

Podemos resumir las características más importantes de esta distribución en la tabla 4.10.

| Distribución Normal $N(\mu, \sigma)$ | |
|---|---|
| rango | $x \in (-\infty, +\infty)$ |
| $f(x)$ | $\frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ |
| $E[X]$ | $\mu$ |
| $V[X]$ | $\sigma^2$ |
| $M_X(t)$ | $e^{t\mu + \frac{t^2\sigma^2}{2}}$ |
| $\varphi_X(t)$ | $e^{it\mu - \frac{t^2\sigma^2}{2}}$ |

Tabla 4.10: Distribución Normal $N(\mu, \sigma)$.

La más utilizada es la N(0,1), denominada normal típica o normal tipificada, cuyas características puden verse en la tabla 4.11.

| Distribución Normal $N(0,1)$ | |
|---|---|
| rango | $x \in (-\infty, +\infty)$ |
| $f(x)$ | $\dfrac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}$ |
| $E[X]$ | 0 |
| $V[X]$ | 1 |
| $M_X(t)$ | $e^{\frac{t^2}{2}}$ |
| $\varphi_X\ (t)$ | $e^{-\frac{t^2}{2}}$ |

Tabla 4.11: Distribución Normal Tipificada $N(0,1)$.

### 4.3.1 Propiedades de la distribución Normal.

- Si $X \sim N(\mu, \sigma)$, entonces se cumple que

$$Y = aX + b \sim N(a\mu + b, |a|\sigma)$$

- Si $X \sim N(\mu, \sigma)$, entonces

$$Y = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Esta propiedad es consecuencia de la anterior, y nos permite obtener los valores de una Normal cualquiera conociendo los de la N(0,1), que es la que está tabulada.

- Si $X_1 \sim N(\mu_1, \sigma_1)$ y $X_2 \sim N(\mu_2, \sigma_2)$, son dos variables aleatorias normales independientes, entonces,

$$Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

es decir, la media es la suma de las medias y la varianza es la suma de las varianzas. Este resultado se generaliza a sumas de n variables aleatorias en la siguiente propiedad.

- Si $X_1, X_2, \ldots X_n$ son variables aleatorias normales independientes, de medias $\mu_1, \mu_2, \ldots \mu_n$ y varianzas $\sigma_1^2, \sigma_2^2, \ldots \sigma_n^2$ respectivamente. y llamamos

$$\mu = \mu_1 + \mu_2 + \ldots + \mu_n$$
$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$$

Entonces

$$Y = X_1 + X_2 + \ldots + X_n \sim N(\mu, \sigma)$$

- Si $X_1, X_2, \ldots X_n$ son variables aleatorias independientes, idénticamente distribuidas según una Normal de media $\mu$ y varianza $\sigma^2$ Entonces

$$X_1 + X_2 + \ldots + X_n \sim N(n\mu, \sigma\sqrt{n})$$

y

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \sim N(\mu, \sigma/\sqrt{n})$$

A partir de ahora, si una variable sigue una distribución N(0,1) utilizaremos la letra $Z$ para denominarla.

Si $Z \sim N(0,1)$, se denomina $z_\alpha$ al valor que cumple:

$$P(Z < z_\alpha) = \alpha$$

Será frecuente encontrar problemas en los que se tendrá que hallar la probabilidad de que una variable aleatoria $X \sim N(\mu, \sigma)$ esté comprendida entre los valores $a$ y $b$. Para resolverlo aplicaremos la fórmula

$$P(a < X < b) = P(X < a) - P(X < b)$$

Para hallar cada uno de estos dos factores tipificaremos. Por ejemplo, para el primer factor tendremos

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right),$$

y este valor lo buscaremos en la tabla de la Normal tipificada.

### 4.3.2 Teorema Central del límite.

Sean $X_1, X_2, \ldots X_n$ variables aleatorias independientes, cuyas medias son $\mu_1, \mu_2, \ldots \mu_n$ y cuyas varianzas son $\sigma_1^2, \sigma_2^2, \ldots \sigma_n^2$ respectivamente. Entonces si llamamos $\mu$ y $\sigma$ a

$$E(X_1 + X_2 + \ldots + X_n) = \mu_1 + \mu_2 + \ldots + \mu_n = \mu$$

$$\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 = \sigma^2$$

se cumple que

$$\lim_{n \to \infty} \frac{X_1 + X_2 + \ldots + X_n - \mu}{\sigma} \sim N(0,1)$$

**Consecuencias.**

El Teorema Central del Límite es especialmente importante en Inferencia Estadística, ya que, como veremos en el tema siguiente nos permitirá afirmar que la distribución del estadístico media muestral se va aproximando a la distribución de una Normal al ir creciendo el valor de n.

Una de las primeras consecuencias del Teorema es que nos permite deducir el por qué es tan frecuente la aparición de variables normales al realizar estudios reales (por ejemplo, al estudiar la altura de la población, o al estudiar los errores cometidos en determinadas observaciones). Esto es debido a que usualmente estas variables no son más que el resultado de la suma de muchas causas, con lo cual puede interpretarse como la suma de gran cantidad de variables aleatorias distintas, por lo que, en virtud del Teorema, se distribuirán de una manera muy similar a la normal.

Como una aplicación concreta del Teorema podemos citar la aproximación de la Binomial por la Normal. En efecto, una variable aleatoria binomial no es más que la suma de variables aleatorias de tipo Bernoulli independientes, por tanto, podemos aplicar el Teorema, con lo que obtendremos que, para valores grandes de n

Si $X \sim B(n,p)$, entonces $X$ se aproxima por $Y$ siendo $Y \quad N(np, \sqrt{npq})$

esta aproximación se considera buena si $0{,}1 < p < 0{,}9$ y $np > 5$ y $nq > 5$.

Del mismo modo, también podemos interpretar la distribución de Poisson como la suma de muchas variables de Poisson, con lo cual si n aumenta (lo que requiere que $\lambda$ sea grande) se aproximará a una normal. Es decir:

$$\text{Si } X \sim P(\lambda), \ \text{ entonces } X \text{ se aproxima por } Y \text{ siendo } Y \quad N(\lambda, \sqrt{\lambda})$$

esta aproximación se considera buena si $\lambda > 5$.

Recordemos también que la Distribución de Poisson se obtenía como el límite de la distribución Binomial cuando p era peqeño y n grande, siendo $np = \lambda$. Esta aproximación se considera buena si $p < 0{,}1$ y $np < 5$.

### 4.3.3 Distribución Continua Truncada.

A veces la variable no toma todos los valores de la recta real. Supongamos que X no está definida fuera del intervalo [a,b], entonces la función de densidad de la variable truncada en dicho intervalo que podemos denominar $f_{[a,b]}$ vendría dada – en función de una variable general – por:

$$f_{[a,b]}(x) = \begin{cases} \dfrac{f(x)}{F(b) - F(a)} & \text{si} \quad x \in [a,b] \\ 0 & \text{si} \quad x \notin [a,b] \end{cases}$$

# Problemas Resueltos

## Problema 4.1

**Comprobar que la distribución binomial cumple las condiciones para ser una función de probabilidad.**

Según la definición de la distribución binomial:

$$P[X = k] = \binom{n}{k} p^k q^{n-k} \qquad k = 0, \ldots, n$$

Para comprobar que efectivamente es una función de probabilidad habría que ver que todos los términos son positivos (lo cual es evidente) y que la suma de todos los términos es la unidad, es decir:

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = 1$$

Para demostrarlo haremos uso de la fórmula del desarrollo del binomio de Newton:

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

con lo cual, en nuestro caso tendríamos

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n$$

lo cual evidentemente es la unidad, al ser $p + q = 1$.

# Problema 4.2

**Calcular la media y la varianza de la distribución binomial.**

Calculemos primero la media:

$$E[X] = \sum_{k=0}^{n} kP[X = k] = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}$$

como el sumando correspondiente a $k = 0$ es nulo, la expresión quedaría

$$E[X] = \sum_{k=1}^{n} k\frac{n!}{k! \cdot (n-k)!}p^k q^{n-k}$$

simplificando

$$E[X] = \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)! \cdot (n-k)!}p^k q^{n-k}$$

que equivale a

$$E[X] = \sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)! \cdot (n-k)!}p^k q^{n-k} = \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k q^{n-k}$$

y por tanto

$$E[X] = \sum_{k=1}^{n} np\binom{n-1}{k-1}p^{k-1} q^{n-k}$$

si llamamos $l = k - 1$ la suma desde $k = 1$ hasta $n$ se correspondería a la suma desde $l = 0$ hasta $n - 1$. Por tanto:

$$E[X] = np\sum_{l=0}^{n-1} \binom{n-1}{l}p^l q^{n-1-l} = np(p+q)^{n-1} = np.$$

Para calcular la varianza haremos uso de la fórmula

$$\mathrm{Var}(X) = E[X^2] - E[X]^2$$

calculemos, por tanto, $E[X^2]$:

$$E[X^2] = \sum_{k=0}^{n} k^2 P[X = k] = \sum_{k=0}^{n} k^2\binom{n}{k}p^k q^{n-k}$$

como el sumando correspondiente a $k=0$ es nulo, la expresión quedaría

$$E[X^2]=\sum_{k=1}^{n} k^2 \frac{n!}{k!\cdot(n-k)!}p^k q^{n-k}$$

simplificando

$$E[X^2]=\sum_{k=1}^{n} k \frac{n!}{(k-1)!\cdot(n-k)!}p^k q^{n-k}$$

que equivale a

$$E[X^2]=\sum_{k=1}^{n} (k-1+1) \frac{n!}{(k-1)!\cdot(n-k)!}p^k q^{n-k}$$

Descomponiendo esta suma en dos, tendríamos

$$E[X^2]=A+B$$

siendo

$$A=\sum_{k=1}^{n} (k-1) \frac{n!}{(k-1)!\cdot(n-k)!}p^k q^{n-k}$$

$$B=\sum_{k=1}^{n} \frac{n!}{(k-1)!\cdot(n-k)!}p^k q^{n-k}$$

La expresión que hemos llamado $B$ ya nos apareció en el cálculo de la media y vimos que valía $np$. Calculemos ahora el valor de $A$. Como el primer sumando se anula, quedaría

$$A=\sum_{k=2}^{n} (k-1) \frac{n!}{(k-1)!\cdot(n-k)!}p^k q^{n-k}=\sum_{k=2}^{n} \frac{n!}{(k-2)!\cdot(n-k)!}p^k q^{n-k}$$

que operando nos lleva a

$$A=\sum_{k=2}^{n} n(n-1)p^2 \frac{(n-2)!}{(k-2)!\cdot(n-k)!}p^{k-2} q^{n-k}$$

si llamamos $l=k-2$ la suma desde $k=2$ hasta $n$ se correspondería a la suma desde $l=0$ hasta $n-2$. Por tanto:

$$A=n(n-1)p^2\sum_{l=0}^{n-2}\binom{n-2}{l}p^l q^{n-2-l}=n(n-1)p^2(p+q)^{n-2}=n(n-1)p^2.$$

Por tanto tenemos que:

$$\text{Var}(X) = E[X^2] - E[X]^2 = A + B - E[X]^2 = n(n-1)p^2 + np - (np)^2,$$

desarrollando

$$\text{Var}(X) = n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p) = npq.$$

Una manera más breve de obtener estos resultados hubiera sido considerando que la distribución binomial se obtiene al sumar $n$ distribuciones independientes de Bernouilli de parámetro $p$. Por tanto se puede aplicar que la media y la varianza de la suma de variables aleatorias independientes es igual a la suma de las medias y varianzas respectivamente. Con lo que tendríamos que la media de la binomial sería $n$ veces la media de la Bernoulli (es decir, $np$) y la varianza sería $n$ veces la varianza de la Bernoulli (es decir, $npq$).

# Problema 4.3

**Es conocido que la probabilidad de que un recién nacido sea varón es de 0,515. En una maternidad han nacido en una mañana 10 bebés (no hubo partos múltiples). Hallar la probabilidad de que nacieran exactamente 3 varones.**

Si consideramos una variable aleatoria de Bernoulli que tome el valor 1 si el recién nacido es varón y 0 si no lo es, lo que estamos haciendo es considerar 10 repeticiones independientes de dicho experimento, contando el número de veces que se produce el valor 1. Por tanto, la variable aleatoria que nos cuenta el número de varones de entre 10 nacimientos es una variable aleatoria binomial, donde $p = 0{,}515$ y $n = 10$. La probabilidad de que hayan nacido exactamente 3 varones es:

$$P[X=3] = \binom{10}{3}(0{,}515)^3(0{,}485)^7 = 0{,}0387995.$$

# Problema 4.4

**El número de llamadas telefónicas que se reciben en una empresa se distribuye según una Poisson de media 2 llamadas por minuto. Hallar la probabilidad de que en un periodo de 10 minutos se reciban 12 llamadas.**

Si el número de llamadas recibidas por minuto es una Poisson de parámetro 2, el número de llamadas recibidas en 10 minutos será una Poisson de parámetro 20. Si llamamos $X$ a esta variable aleatoria, lo que nos pide el enunciado es

$$P[X = 12] = e^{-20} \cdot \frac{20^{12}}{12!} = 0{,}0176254.$$

# Problema 4.5

**Se lanza 21 veces un dado. Hallar la probabilidad de que cada cara halla salido tantas veces como indica su puntuación.**

Si definimos 6 variables aleatorias de la siguiente manera: $X_k$ = número de veces que sale la cara $k$ ($k$=1,2,..., 6). Cada una de ellas es una binomial de parámetros $p = 1/6$ y $n = 21$. La distribución conjunta sería una multinomial, con lo cual tendríamos que la probabilidad pedida es:

$$P[X_1 = 1; X_2 = 2; X_3 = 3; X_4 = 4; X_5 = 5; X_6 = 6]$$

aplicando la fórmula de la distribución multinomial, esto vale:

$$\frac{21!}{1!2!3!4!5!6!}(1/6)^1(1/6)^2(1/6)^3(1/6)^4(1/6)^5(1/6)^6$$

que efectuando los cálculos nos queda que la probabilidad pedida es:

$$p = 0{,}0000935969 = 9{,}35969 \cdot 10^{-5}.$$

# Problema 4.6

**En una pecera hay 20 peces, de los cuales 7 son de colores. Se extraen al azar 4 peces. Hallar la probabilidad de que exactamente 2 sean de colores.**

Nos encontramos con un caso de distribución hipergeométrica. La probabilidad pedida será

$$P(X = k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}$$

que en nuestro caso es:

$$P(X=2) = \frac{\binom{7}{2}\binom{13}{2}}{\binom{20}{4}} = \frac{21 \cdot 78}{4845} = 0{,}33808.$$

## Problema 4.7

**Hallar el número medio de veces que habrá que lanzar una moneda para obtener la primera cara.**

El lanzamiento de una moneda es un experimento aleatorio de Bernoulli con $p = 1/2$. La variable aleatoria que nos cuenta el número de intentos necesarios hasta obtener el primer éxito, al realizar repeticiones independientes de un experimento Bernoulli, sigue una distribución geométrica o de Pascal. Cuya fórmula es:

$$P(X=k) = p(1-p)^{k-1} \qquad k = 1, 2, 3, \ldots$$

La media de esta distribución es

$$E[X] = \frac{1}{p} = 2$$

Por tanto, el número medio de veces que habrá que lanzar la moneda para obtener la primera cara es de dos veces.

## Problema 4.8

**Es conocido que la probabilidad de que un recién nacido sea varón es de 0,515. En una maternidad han nacido en un año 10000 bebés (no hubo partos múltiples). Hallar la probabilidad de que nacieran más niños que niñas.**

Si consideramos una variable aleatoria de Bernoulli que tome el valor 1 si el recién nacido es varón y 0 si no lo es, lo que estamos haciendo es considerar 10000 repeticiones independientes de dicho experimento, contando el número de veces que se produce el valor 1. Por tanto, la variable aleatoria que nos cuenta el número de varones de entre 10000 nacimientos es una variable aleatoria binomial, donde $p = 0{,}515$ y $n = 10000$. La probabilidad de que hayan nacido más niños que niñas es $P[X > 5000]$.

Como dicha probabilidad resulta bastante engorrosa de calcular ya que hay que calcular 5000 sumandos, cada uno de ellos con bastantes cálculos internos, usaremos la aproximación a la distribución normal:

$$P[X > 5000] \approx P[Y > 5000{,}5]$$

siendo $Y$ una variable aleatoria Normal, de media $np = 5150$ y de varianza $npq$ =2497,75. (Se ha efectuado la corrección de continuidad). Tipificando obtenemos que

$$P[Y > 5000{,}5] = P[\frac{Y - 5150}{49{,}9775} > \frac{5000{,}5 - 5150}{49{,}9775}] = P[Z > -2{,}99135]$$

siendo $Z$ una variable aleatoria Normal de media 0 y desviación típica 1. Buscando en las tablas:

$$P[Z > -2{,}99135] = P[Z < 2{,}99135] \approx 0{,}99861.$$

# Problema 4.9

**Si el 20% de los circuitos fabricados en cierta planta están defectuosos, ¿cuales son las probabilidades de que en un lote de 100, aleatoriamente escogidos para revisión**

**(a) a lo sumo 15 estén defectuosos;**

**(b) exactamente 15 estén defectuosos?**

---

Si llamamos $X$ a la variable aleatoria que cuenta el número de circuitos defectuosos, tenemos que $X$ se distribuye según una Binomial con parámetros $n = 100$ y $p = 0{,}2$ (proporción de defectuosos):

$$X \sim \mathrm{B}(100; 0{,}2)$$

(a) Nos piden calcular $P[X \leq 15]$, como $n$ es grande, y $p$ es pequeño, podemos aproximar la distribución binomial a través de la distribución de Poisson con parámetro $\lambda = n \cdot p = 100 \cdot 0{,}2 = 20$. Así tenemos:

$$P[X \leq 15] \approx P[Y \leq 15] \qquad \text{siendo} \qquad Y \sim \mathrm{P}(\lambda = 20)$$

$$P[Y \leq 15] = \frac{20^0 \cdot e^{-20}}{0!} + \frac{20^1 \cdot e^{-20}}{1!} + \ldots + \frac{20^{15} \cdot e^{-20}}{15!} = 0{,}1565.$$

Tenemos que tener en cuenta, que si hubiesemos hecho la aproximación por la distribución normal, la aproximación en algo mejor. Obtendríamos en este caso, haciendo la correspondiente corrección por continuidad, e interpolando en las tablas de la N(0; 1):

$$\begin{aligned} P[X \leq 15] &\approx P[Y' \leq 15{,}5] = P\left[\frac{Y'-20}{4} \leq \frac{15{,}5-20}{4}\right] = \\ &= P[Z \leq -1{,}125] = 0{,}1303. \end{aligned}$$

Siendo $Z$ la distribución N(0; 1), e $Y'$ la distribución N(20; 4); ya que $\mu = n \cdot p = 20$ y $\sigma^2 = n \cdot p \cdot q = 16$.

El valor exacto de la probabilidad, que puede obtenerse usando el programa estadístico Statgraphics, para $P[X \leq 15]$ es 0,128506.

(b) En este caso, sí podemos calcular la probabilidad que nos piden de forma exacta, pues

$$P[X = 15] = \binom{100}{15}(0{,}2)^{15}(0{,}8)^{85} = 0{,}04806.$$

Podríamos haber calculado esta probabilidad aproximando con la distribución de Poisson, como en el apartado anterior, entonces obtendríamos:

$$P[X = 15] \approx P[Y = 15] = \frac{20^{15} \cdot e^{-20}}{15!} = 0{,}05165.$$

que es sin duda una buena aproximación, aunque en este caso sí era fácil calcular la probabilidad exacta.

# Problema 4.10

**En la observación del número de glóbulos rojos (en millones) de los habitantes de una gran ciudad se observó que seguían aproximadamente una distribución normal de media 4,5 y desviación típica 0,5. Se pide calcular:**

**(a) Probabilidad de que un habitante tomado al azar tenga más de 5 millones de glóbulos rojos.**

**(b) Tanto por ciento de la ciudad con menos de 3,75 millones de glóbulos rojos.**

**(c) Número de glóbulos rojos del 20% más alto de la ciudad.**

**(d) Probabilidad de que al seleccionar al azar a 50 individuos de esta ciudad, haya más de 40 con más de 5 millones de glóbulos rojos. Probabilidad de que no haya ninguno.**

---

Si llamamos $X$ a la variable aleatoria que cuenta el número de glóbulos rojos de un individuo, tenemos que $X \sim N(4{,}5;\ 0{,}5)$.

(a) Nos piden $P(X > 5)$, si tipificamos y miramos en la tabla de la Normal tipificada tenemos:

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) = 1 - P\left[\frac{X - 4{,}5}{0{,}5} \leq \frac{5 - 4{,}5}{0{,}5}\right] = \\ &= 1 - P(Z \leq 1) = 1 - 0{,}8413 = 0{,}1587. \end{aligned}$$

(b) Nos piden $P(X < 3{,}75)$, si tipificamos y miramos en la tabla de la Normal tipificada y expresamos dicha probabilidad en tanto por ciento, tenemos:

$$\begin{aligned} P(X < 3{,}75) &= P\left[\frac{X - 4{,}5}{0{,}5} < \frac{3{,}75 - 4{,}5}{0{,}5}\right] = P(Z < -1{,}5) = \\ &= 1 - P(Z < 1{,}5) = 1 - 0{,}93319 = 0{,}06681. \end{aligned}$$

Por tanto el porcentaje es del 6,681%.

(c) En este apartado nos están pidiendo que calculemos el percentil 80, esto es $P[X < P_{80}] = 0,8$. Entonces:

$$P\left[\frac{X - 4{,}5}{0{,}5} < \frac{P_{80} - 4{,}5}{0{,}5}\right] = 0{,}8 \quad \Longrightarrow \quad P[Z < z_0] = 0{,}8.$$

Si miramos en la tabla de la N(0; 1) tenemos que $z_0 = 0{,}847$ y despejando obtenemos el valor del percentil 80:

$$P_{80} = 0{,}5 \cdot 0{,}847 + 4{,}5 = 4{,}9235 \text{ millones.}$$

(d) Si llamamos $Y$ a la variable aleatoria que cuenta el número de individuos de los 50 seleccionados que tienen más de 5 millones de glóbulos rojos, tenemos que $Y \sim B(50, p)$, siendo $p = 0{,}1587$ que ya habíamos calculado anteriormente.

Nos piden por tanto $P[Y > 40]$, esta probabilidad la podemos aproximar usando la aproximación por la distribución de Poisson de parámetro $\lambda = np$, que en este caso es $\lambda = 7{,}935$. Si llamamos $W \sim P(\lambda = 7{,}935)$ obtenemos:

$$P[Y > 40] \approx P[W > 40] = e^{-7{,}935}\left[\frac{7{,}935^{41}}{41!} + \ldots + \frac{7{,}935^{50}}{50!}\right] \approx 0.$$

Si calculamos ahora la probabilidad de que no haya ningún individuo de los 50 seleccionados que tenga más de 5 millones de glóbulos rojos. En este caso, sí podemos calcular la probabilidad que nos piden de forma exacta, pues

$$P[Y = 0] = \binom{50}{0}(0{,}1587)^0(0{,}8413)^{50} \approx 1{,}768 \cdot 10^{-4}.$$

Podríamos haber calculado esta probabilidad aproximando con la distribución de Poisson, como en el caso anterior, entonces obtendríamos:

$$P[Y = 0] \approx P[W = 0] = \frac{7{,}935^0 \cdot e^{-7{,}935}}{0!} = 3{,}5799 \cdot 10^{-4}.$$

Que es sin duda una buena aproximación, pero en este caso era preferible calcular la probabilidad exacta.

# Problema 4.11

**Consideremos el experimento lanzar un dado cúbico. Sea la variable aleatoria $X$ aquella que toma los valores de la cara superior. Determinar la distribución de probabilidad, la media y la varianza.**

---

La distribución de probabilidad de esta variable aleatoria es la distribución uniforme discreta que toma los valores $\{1, 2, 3, 4, 5, 6\}$ con probabilidades:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p_i$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

La media de la variable aleatoria discreta que toma $k$ valores es:

$$\mu = \sum_{i=1}^{k} p_i x_i = \frac{1+2+3+4+5+6}{6} = 3{,}5.$$

La varianza de la variable aleatoria discreta se define como:

$$\operatorname{var}(X) = \sum_{i=1}^{k} p_i (x_i - \mu)^2.$$

Por tanto la varianza vale

$$\begin{aligned}\operatorname{var}(X) &= \frac{1}{6}[(1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2 + \\ &\quad +(4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2] = \frac{35}{12}.\end{aligned}$$

# Problema 4.12

**Se elige una carta de una baraja española y se considera la variable aleatoria "número de la carta". Determinar la distribución de esta variable aleatoria.**

La distribución de esta variable aleatoria es la distribución uniforme discreta para $k = 10$. Por lo que tenemos que su función de masa de probabilidad es:

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $p_i$ | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 |

# Problema 4.13

**Supongamos que respondemos a un test de 10 preguntas, cada una con cuatro respuestas alternativas de las que sólo una es correcta. Si estamos respondiendo al azar, calcular:**

**(a) La probabilidad de acertar exactamente 7 preguntas.**

**(b) La probabilidad de acertar al menos 5 preguntas.**

**(c) Media, varianza, coeficiente de simetría y curtosis.**

La distribución de probabilidad del número de aciertos $X$ en el test puede modelarse mediante una distribución binomial con $n = 10$ y $p = 1/4$.

La función de probabilidad de la $B(10; 1/4)$ es

$$P_X(X = k) = \binom{10}{k}(0{,}25)^k \cdot (0{,}75)^{10-k} \quad \text{si} \quad k = 0, 1, \ldots, 10.$$

(a) En este apartado nos están pidiendo $P(X = 7)$, que vale:

$$P_X(X = 7) = \binom{10}{7}(0{,}25)^7 \cdot (0{,}75)^{10-7} = 0{,}0031.$$

(b) Aquí tenemos que calcular la $P[X \geq 5]$ que podemos determinarla como:

$$\begin{aligned}
P[X \geq 5] &= 1 - P[X < 5] = \\
&= 1 - \sum_{k=0}^{4}\binom{10}{k}(0{,}25)^k \cdot (0{,}75)^{10-k} = \\
&= 1 - \binom{10}{0}(0{,}25)^0 \cdot (0{,}75)^{10} - \binom{10}{1}(0{,}25)^1 \cdot (0{,}75)^9 - \\
&\quad - \binom{10}{2}(0{,}25)^2 \cdot (0{,}75)^8 - \binom{10}{3}(0{,}25)^3 \cdot (0{,}75)^7 - \\
&\quad - \binom{10}{4}(0{,}25)^4 \cdot (0{,}75)^6 = \\
&= 1 - 0{,}0563 - 0{,}1877 - 0{,}2816 - 0{,}2503 - 0{,}1460 = \\
&= 0{,}0781.
\end{aligned}$$

(c) Si queremos calcular la media, varianza y los coeficientes de simetría y curtosis, sabemos que para la distribución $B(\text{n}; \text{p})$:

| | |
|---|---|
| Media | $\mu = np$ |
| Varianza | $\sigma^2 = np(1-p)$ |
| Desviación típica | $\sigma = \sqrt{np(1-p)}$ |
| Momento central orden 3 | $\mu_3 = np(1-p)(1-2p)$ |
| Coeficiente de simetría | $g_1 = \dfrac{1-2p}{\sqrt{np(1-p)}}$ |
| Momento central orden 4 | $\mu_4 = np(1-p)\{3np(1-p) + [1-6p(1-p)]\}$ |
| Coefiente de curtosis | $g_2 = \dfrac{[1-6p(1-p)]}{np(1-p)}$ |

Por tanto, en nuestro caso, el número promedio de aciertos en el test es

$$\mu = np = 10 \cdot 0{,}25 = 2{,}5.$$

La varianza vale:

$$\mathrm{var}(X) = np(1-p) = 10 \cdot 0{,}25 \cdot 0{,}75 = 1{,}875.$$

La desviación típica es $\sigma = 1{,}3693$.

El coeficiente de simetría vale $g_1 = 0{,}3651$; lo que significa que está sesgada a la derecha (sesgo positivo).

El coeficiente de curtosis vale $g_2 = -0{,}067$ y por tanto la distribución es más aplastada que la normal (platicúrtica).

# Problema 4.14

**Supongamos que el temario de una oposición consta de 50 temas. Un opositor ha preparado 15 de ellos. En el examen se seleccionan al azar 5 temas. Calcular:**

(a) **La probabilidad de que el opositor se haya estudiado los 5 temas.**

(b) **La probabilidad de aprobar el examen. (Para aprobar es necesario contestar al menos 3 de los 5 temas seleccionados).**

(c) **Número esperado de temas que sabrá el estudiante, y la varianza.**

---

Si llamamos variable aleatoria $X$ a aquella que cuenta el número de temas que conoce el opositor de entre los que se han sorteado en el examen. Tenemos por tanto que esta situación sigue el modelo de la distribución hipergeométrica $H(50;\ 15;\ 5)$.

(a) Calculamos la probabilidad de que el opositor conozca los cinco temas:

$$P[X=5] = \frac{\binom{15}{5}\cdot\binom{35}{0}}{\binom{50}{5}} = \frac{3003}{2118760} = 0{,}0014.$$

es decir, únicamente 14 de cada 10000 opositores habrán tenido la suerte de saber los cinco temas habiendo preparado únicamente quince.

(b) La probabilidad de aprobar es

$$\begin{aligned}
P[X\geq 3] &= 1-P[X<3] = 1-P[X\leq 2] = \\
&= 1-\left[\frac{\binom{15}{0}\binom{35}{5}}{\binom{50}{5}} + \frac{\binom{15}{1}\binom{35}{4}}{\binom{50}{5}} + \frac{\binom{15}{2}\binom{35}{3}}{\binom{50}{5}}\right] = \\
&= 1-\frac{324632}{2118760}-\frac{15\cdot 52360}{2118760}-\frac{105\cdot 6545}{2118760} = \\
&= 1-0{,}153217-0{,}370688-0{,}324352 = \\
&= 0{,}151742.
\end{aligned}$$

Lo que equivale a que aproximadamente 15 de cada 100 opositores aprueban el ejercicio estudiando únicamente 15 de los 50 temas.

(c) El número esperado de temas que conoce el estudiante es

$$\mu = \frac{5 \cdot 15}{50} = 1{,}5.$$

Y la varianza es:

$$\mathrm{var}(X) = \frac{5 \cdot 15 \cdot 35 \cdot 45}{50^2 \cdot 49} = 0{,}964285.$$

## Problema 4.15

**Consideremos el experimento que consiste en lanzar un dado, llamamos éxito al resultado $\{5, 6\}$ y fracaso al suceso contrario. Calcular la probabilidad del suceso $A =$"obtener el tercer éxito en el octavo lanzamiento".**

Como los lanzamientos son independientes de tipo Bernoulli con probabilidad de éxito $p = 1/3$, entonces esta situación se modela con la distribución binomial negativa. Como la última ocurrencia debe ser un éxito, significa que en los 7 anteriores tendremos que haber contado 2 éxitos y 5 fracasos, que podrán disponerse de $C_{7,2} = 21$ formas diferentes. Como la probabilidad de todas esas disposiciones es la misma, tendremos que la probabilidad del suceso $A$ es

$$P(A) = \binom{7}{2} \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)^5 = 0{,}10242.$$

## Problema 4.16

**Supongamos que un distribuidor recibe de un fabricante un producto que viene en paquetes de 1000 unidades. El fabricante afirma que su proceso de fabricación garantiza que cada lote lleva a lo sumo un 5% de piezas defectuosas. Para comprobar esta afirmación el distribuidor abre una caja y examina aleatoriamente tres unidades. Calcular la probabilidad de que en la muestra de tres**

**haya al menos una que sea defectuosa, suponiendo que la afirmación del fabricante es cierta.**

---

La distribución hipergeométrica es útil en numerosas aplicaciones relacionadas con el control de calidad de productos que se sirven en lotes. En este caso tendremos que el número de piezas defectuosas en la muestra sigue una distribución hipergeométrica $H(1000;\, 50;\, 3)$.

Si se cumple la afirmación del fabricante, tendremos un 5% de defectuosas de las 1000 del lote, esto es 50 piezas defectuosas en la caja de 1000. Debemos observar que las extracciones se efectuan sin reemplazamiento, y por tanto el experimento se modela con la distribución hipergeométrica.

La probabilidad de que en el muestra haya al menos una que sea defectuosa es

$$\begin{aligned} P(X \geq 1) &= 1 - P[X = 0] = 1 - \frac{\binom{50}{0} \cdot \binom{950}{3}}{\binom{1000}{3}} = \\ &= 1 - \frac{142444900}{166167000} = 0{,}14276. \end{aligned}$$

Así, el distribuidor debe esperar que en 14 de cada 100 cajas se encuentre con alguna pieza defectuosa en la muestra de 3, incluso aunque el proceso de fabricación tenga garantía del 95%.

Podemos calcular la probabilidad anterior usando la aproximación de la hipergeométrica por la binomial pues $N = 1000$ es grande frente a $n = 3$. Si $Y \sim$ B(3; 0,05) obtenemos

$$\begin{aligned} P[X \geq 1] &\approx P[Y \geq 1] = 1 - P[Y = 0] = \\ &= 1 - \binom{3}{0} (0{,}05)^0 \cdot (0{,}95)^3 = 1 - 0{,}857375 = \\ &= 0{,}142625. \end{aligned}$$

Ésta es, sin duda, una buena aproximación de la hipergeométrica por la binomial.

# Problema 4.17

**La Dirección General de Tráfico quiere poner en funcionamiento un sistema de puntos de penalización por cada infracción de tráfico que puede conducir a la pérdida del permiso de conducir. Supongamos que se estima que un conductor es denunciado una de cada diez veces que comete una infracción, es decir p = 0,1 y que esta proporción se mantiene constante a lo largo del tiempo. Supongamos que la cuarta denuncia lleva consigo la pérdida del carnet de conducir. Determinar**

**(a) La función de probabilidad del número de infracciones cometidas hasta la retirada del carnet.**

**(b) El número esperado de infracciones necesarias para la pérdida del carnet.**

**(c) Varianza del número de infracciones.**

---

Consideraremos la variable aleatoria $X$ que representa el número de la infracción que le acarrea la pérdida del carnet. Entonces $X$ sigue una distribución binomial negativa con parámetros $k = 4$ y $p = 0{,}1$.

(a) Entonces la función de probabilidad de $X$ es

$$P_X(X = n) = \binom{n-1}{3} \cdot (0{,}1)^4 (0{,}9)^{n-4} \quad \text{si} \quad n = 4, 5, 6, \ldots$$

Por ejemplo, la probabilidad de que le sea retirado el carnet exactamente en la décima infracción que comete es

$$P[X = 10] = \binom{9}{3} (0{,}1)^4 (0{,}9)^6 = 0{,}0044641.$$

(b) El número esperado de infracciones que llevan a la retirada del carnet es

$$\mu = \frac{k}{p} = \frac{4}{0{,}1} = 40 \text{ infracciones.}$$

(c) Y la varianza de $X$ vale

$$\mathrm{var}(X) = \frac{k \cdot q}{p^2} = \frac{4 \cdot 0{,}9}{(0,1)^2} = 360.$$

Nota: Este sistema que está en estudio en la D.G.T. realmente está vigente en algunos paises de nuestro entorno.

# Problema 4.18

**Un jugador de baloncesto mantiene a lo largo de la temporada una estadística de aciertos en los tiros desde la línea de 6,25 m. del 47%. Determinar**

**(a) La distribución de probabilidad del número de tiros de tres puntos hasta que se produzca el primer fallo.**

**(b) La probabilidad de que el primer fallo se produzca en el quinto intento.**

**(c) El número promedio de lanzamientos que efectuará hasta que se produzca el primer fallo. Calcular la varianza.**

(a) La distribución geométrica, que es un caso particular $k = 1$ de la distribución binomial negativa, se utiliza en las situaciones en las que es importante detectar el primer fallo.

La distribución de probabilidad del número de tiros necesarios hasta conseguir el primer fallo es una distribución geométrica con $p = 0,53$:

$$P[X = x] = (0{,}53) \cdot (0{,}47)^{x-1} \qquad \text{si} \quad x = 1, 2, 3, \ldots$$

(b) La probabilidad de que el primer fallo se produzca en el quinto lanzamiento es

$$P[X = 5] = (0{,}53) \cdot (0{,}47)^4 = 0{,}0258623.$$

(c) El número medio de lanzamientos que efectuará antes de que se produzca el primer fallo es

$$\mu = \frac{1}{p} = \frac{1}{0{,}53} = 1{,}88679.$$

Y la varianza es

$$\mathrm{var}[X] = \frac{q}{p^2} = \frac{0{,}47}{(0{,}53)^2} = 1{,}6731.$$

# Problema 4.19

**Un fabricante de circuitos dice en su publicidad que las pruebas de un laboratorio independiente revelan que, cuando los circuitos se someten a una carga de trabajo equivalente a 10000 horas, el número medio de circuitos que fallan es 3. Determinar la probabilidad de que fallen 6 circuitos, y la de que no falle ninguno. Calcular la media y la varianza.**

Bajo estas suposiciones, puede admitirse que el número de circuitos que fallan cuando están en funcionamiento 10000 horas sigue un modelo de Poisson de parámetro $\lambda = 3$, siendo su función de probabilidad

$$P[X = x] = \frac{3^x \cdot e^{(-3)}}{x!} \qquad \text{si} \quad x = 0, 1, 2, \ldots.$$

Así la probabilidad de que fallen 6 circuitos es

$$P[X = 6] = \frac{3^6 \cdot e^{(-3)}}{6!} = 0{,}0504.$$

Y la probabilidad de que no falle ninguno es

$$P[X = 0] = \frac{3^0 \cdot e^{(-3)}}{0!} = 0{,}049787.$$

El número medio de circuitos que se espera fallarán en las 10000 horas es 3 con varianza tambien igual a 3.

# Problema 4.20

**En un proceso de fabricación de piezas se sabe que la proporción de piezas defectuosas es p = 0,01. Las piezas se sirven en cajas**

**de 200. Calcular la probabilidad de que en una caja haya cinco o más defectuosas.**

---

Si llamamos $X$ a la variable aleatoria que cuenta el número de piezas defectuosas en el lote de 200, tenemos que la variable X sigue un modelo Binomial con parámetros $n = 200$ y $p = 0{,}01$. Como $n$ es grande, y $p$ es pequeña, se puede aproximar por la distribución $Y$ de Poisson con parámetro $\lambda = np = 200 \cdot 0{,}01 = 2$.

Debemos tener en cuenta que según el modelo binomial tendremos que calcular

$$\begin{aligned} P[X \geq 5] &= 1 - P[X \leq 4] = 1 - \sum_{k=0}^{4} \binom{200}{k} (0{,}01)^k (0{,}99)^{200-k} = \\ &= 0{,}051747. \end{aligned}$$

Si usamos la aproximación de Poisson obtendremos:

$$\begin{aligned} P[X \geq 5] &\approx P[Y \geq 5] = 1 - P[X \leq 4] = 1 - \sum_{k=0}^{4} \frac{2^k \cdot e^{-2}}{k!} = \\ &= 0{,}052654. \end{aligned}$$

Como puede observarse la aproximación es bastante buena. Y en algunos casos, es el único camino factible para calcular dichas probabilidades binomiales.

# Problema 4.21

**Supongamos que un componente de un sistema informático tiene un tiempo de fallo en años que sigue una distribución exponencial de media 5. Determinar la probabilidad de que dicho componente siga funcionando despues de $x$ años. Particularizar para la probabilidad de que el componente siga funcionando después de 8 años.**

---

Tenemos que si $X$ sigue una distribución exponencial de parámetro $\alpha$, siendo $\alpha > 0$, su función de densidad es

$$f(x) = \begin{cases} \frac{1}{\alpha} \cdot e^{(-x/\alpha)} & \text{si} \quad x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

La función de distribución de la variable aleatoria exponencial es

$$F(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ 1 - e^{(-x/\alpha)} & \text{si} \quad x > 0 \end{cases}$$

La media y la varianza de la distribución exponencial con parámetro $\alpha$ son respectivamente $\mu = \alpha$ y $\text{var}(X) = \alpha^2$.

Entonces la probabilidad de que un componente funciones despues de $x$ años es

$$\begin{aligned} P[X \geq x] &= 1 - P[X \leq x] = 1 - F(x) = \\ &= 1 - \int_0^x \frac{1}{5} \cdot e^{(-t/5)}\, dt = \\ &= 1 - \left(1 - e^{(-x/5)}\right) = \\ &= e^{(-x/5)}. \end{aligned}$$

Si queremos calcular la probabilidad de que el componente siga funcionando después de 8 años tendremos

$$P[X \geq 8] = e^{(-8/5)} = 0{,}2019.$$

# Problema 4.22

**La vida útil de un dispositivo eléctrico puede considerarse una variable aleatoria con distribución exponencial de media 2 años. Se instalan 100 dispositivos de este tipo en diferentes sistemas. Calcular la probabilidad de que a lo sumo 30 de ellos fallen el primer año.**

La función de densidad de la variable aleatoria que mide la duración en años del dispositivo es

$$f(x) = \begin{cases} \frac{1}{2} \cdot e^{(-x/2)} & \text{si} \quad x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Entonces la probabilidad de que uno de éstos aparatos falle en el primer año vale

$$P[X \leq 1] = \int_0^1 \frac{1}{2} e^{(-x/2)}\, dx = 1 - e^{-0{,}5} = 1 - 0{,}60653 \approx 0{,}40.$$

Si se colocan 100 dispositivos en diversos sistemas, puede considerarse que el número $Y$ de aparatos que fallan de los 100 sigue una distribución binomial de parámetros $n = 100$ y $p = 0{,}4$. Por tanto la probabilidad de que fallen a lo sumo 30 de ellos es

$$P[Y \leq 30] = \sum_{k=0}^{30} \binom{100}{k} (0{,}4)^k \cdot (0{,}6)^{100-k}$$

que puede aproximarse mediante la distribución normal de parámetros

$$\mu = np = 100 \cdot 0{,}4 = 40 \qquad \sigma = \sqrt{100 \cdot 0{,}4 \cdot 0{,}6} = 4{,}90.$$

Así pues, aproximando por $W \sim N(40;\ 4{,}90)$ y haciendo la corrección por continuidad tenemos

$$\begin{aligned} P[Y \leq 30] &\approx P[W \leq 30{,}5] = P\left[\frac{W-40}{4{,}9} \leq \frac{30{,}5-40}{4{,}9}\right] = \\ &= P[Z \leq -1{,}94] = 0{,}0262 \end{aligned}$$

## Problema 4.23

**Sea $X$ una variable aleatoria normal con media $\mu$ y varianza $\sigma^2$. Entonces, dado un número real $k$, determinar $P[|X-\mu| < k\sigma]$. Hacer una tabla en la que se comparen el valor exacto y el valor que proporciona la desigualdad de Chebychev.**

Sabemos que $X \sim N(\mu, \sigma)$, y si tipificamos $Z = \dfrac{X-\mu}{\sigma} \sim N(0;\ 1)$.

Nos piden calcular

$$P[|X-\mu| < k\sigma] = P\left[\frac{|X-\mu|}{\sigma} < k\right] = P[|Z| < k].$$

Para diferentes valores de $k$ el valor exacto de $P[|Z| < k]$ puede calcularse a partir de la tabla de la distribución normal N(0; 1).

La desigualdad de Chebychev asegura que este valor siempre será superior a $1 - \dfrac{1}{k^2}$.

En la tabla siguiente se comparan los valores exactos de la probabilidad con los aproximados por la desigualdad de Chebychev.

$$P[|X - \mu| < k\sigma]$$

| | Distribución normal | Desigualdad de Chebychev |
|---|---|---|
| $k$ | $P[\lvert Z\rvert < k]$ | $1 - \dfrac{1}{k^2}$ |
| 1 | 0,6826 | 0 |
| 1,5 | 0,8664 | 0,5556 |
| 2 | 0,9546 | 0,7500 |
| 3 | 0,9974 | 0,8889 |
| 4 | 1,0000 | 0,9334 |

Como puede apreciarse en este ejercicio la cota para la probabilidad proporcionada por la desigualdad de Chebychev es más aproximada cuando aumenta $k$. Para valores de $k \leq 1$ la cota no aporta nada, ya que simplemente establece que $P[|X - \mu| < k\sigma] \geq 0$, que siempre es cierto.

Hay que destacar que la cota que proporciona la desigualdad de Chebychev es válida sin más hipótesis sobre la variable aleatoria que la existencia de media y varianza.

# Problema 4.24

**Cuando se efectúan cálculos con un ordenador tan sólo es posible retener en cada paso un número finito de dígitos significativos. Por ejemplo si utilizamos una calculadora que únicamente conserva cuatro dígitos a la derecha del punto decimal, un número fraccionario como 1/3 se representaría por 0,3333 cometiendo un error de truncamiento de $(0{,}3333\ldots)\cdot 10^{-4}$. Determinar, con una calculadora como ésta**

**(a) Error mayor y error menor que se pueden cometer.**

**(b) Error de redondeo esperado en una operación. Varianza.**

**(c) Error de redondeo si se suman 12 números.**

Como ya hemos visto, el error de truncamiento en un número fraccionario como 1/3 con esta calculadora es

$$0{,}333333333\ldots - 0{,}3333 = 0{,}000033333\ldots = (0{,}3333\ldots)\cdot 10^{-4}$$

En esta calculadora los errores menor y mayor que se pueden cometer serían respectivamente

$$-(0{,}49999\ldots)\cdot 10^{-4} \qquad \text{y} \qquad (0{,}49999\ldots)\cdot 10^{-4}$$

y cualquier número comprendido entre estos dos puede ser un error de redondeo.

(b) Supongamos que el error de redondeo en el cálculo de una operación es una variable aleatoria $X$ que se distribuye uniformemente en el intervalo $\left[-\frac{1}{2}10^{-4}; \frac{1}{2}10^{-4}\right]$. Entonces la esperanza de $X$ es

$$E[X] = \frac{a+b}{2} = \frac{-\frac{1}{2}10^{-4} + \frac{1}{2}10^{-4}}{2} = 0$$

y la varianza de $X$ es

$$\mathrm{var}[X] = \frac{(b-a)^2}{12} = \frac{\left(\frac{1}{2}10^{-4} + \frac{1}{2}10^{-4}\right)^2}{12} = \frac{(10^{-4})^2}{12}.$$

(c) Si se efectúan $n$ operaciones el error de redondeo de la suma es $Y = \sum_{i=1}^{n} X_i$, donde cada $X_i$ sigue una distribución uniforme. Si las $X_i$ son independientes entonces, usando el teorema central del límite, $Y$ puede aproximarse mediante una distribución normal con media y varianza:

$$E[Y] = n\cdot E[X] = 0 \qquad \text{y} \qquad \mathrm{var}[Y] = n\cdot \mathrm{var}[X] = n\cdot \frac{(10^{-4})^2}{12}.$$

Con esta distribución podemos calcular la probabilidad de los errores de redondeo de la suma. Por ejemplo si se suman 12 números, entonces el error de redondeo $Y$ puede aproximarse mediante una normal de media $\mu = 0$ y varianza $\sigma^2 = 10^{-8}$.

Usando las tablas de la normal puede calcularse fácilmente la probabilidad de que el error de redondeo de la suma de los doce números esté comprendido entre unos determinados límites o acotado por un determinado valor.

# Problema 4.25

**Sea $X$ una variable aleatoria que representa la duración en horas de una resistencia. Supongamos que $X$ se distribuye normalmente con media $\mu = 1000$ horas y desviación típica $\sigma = 100$ horas.**

**(a) Se compran 4 resistencias. Si se supone que las duraciones de cada resistencia son independientes, calcular la probabilidad de que las cuatro duren más de 900 horas.**

**(b) Supongamos que se utilizan de manera consecutiva en el mismo aparato. Calcular la probabilidad de que el número total de horas de funcionamiento sea superior a 3900 horas.**

---

(a) Si llamamos $X_1, X_2, X_3, X_4$ la duración de cada resistencia. Cada variable $X_i \sim N(1000, 100)$. Por la independencia

$$\begin{aligned}
P[X_i \geq 900 \; ; \; i = 1,2,3,4] &= \prod_{i=1}^{4} P[X_i \geq 900] = \\
&= \prod_{i=1}^{4} P\left(\frac{X_i - 1000}{100} \geq \frac{900 - 1000}{100}\right) = \\
&= \prod_{i=1}^{4} P[Z \geq -1] = P[Z \geq -1]^4 = \\
&= P[Z \leq 1]^4 = (0{,}8413)^4 = 0{,}5010.
\end{aligned}$$

(b) Tenemos que considerar ahora la variable que mide la duración de las cuatro resistencias usadas consecutivamente, es decir, la variable $Y = \sum_{i=1}^{4} X_i$. Sabemos que la distribución de esta variable es normal con media

$$\mu_Y = \sum_{i=1}^{4} \mu_{X_i} = 4 \cdot 1000 = 4000 \text{ horas.}$$

y varianza

$$\operatorname{var}(Y) = \sum_{i=1}^{4} \operatorname{var}(X_i) = 4 \cdot 100^2 = 40000.$$

Es decir, desviación típica $\sigma_Y = 200$ horas. Entonces

$$\begin{aligned} P[Y \geq 3900] &= P\left(\frac{Y-4000}{200} \geq \frac{3900-4000}{200}\right) = P[Z \geq -0,5] = \\ &= 0{,}6915. \end{aligned}$$

# Problema 4.26

**La altura de los estudiantes de un instituto se distribuye normalmente, con una media de 170 cm y una desviación típica de 5 cm. Se pide:**

**(a) Calcular los cuartiles de la población.**

**(b) Se seleccionan 5 individuos al azar para el equipo de baloncesto. Hallar la probabilidad de que al menos uno mida más de 170 cm.**

**(c) Hallar la probabilidad de que de 1000 estudiantes haya más de 520 que midan más de 170 cm.**

(a) Para calcular los cuartiles de la población debemos basarnos en los datos facilitados en el enunciado y en la definición de cuartil.

Si llamamos $X$ a la variable aleatoria que indica la altura de un individuo, sabemos que $X \sim N(170; 5)$. Por la definición de cuartil, $Q_1$ será el valor de la variable que deje a su izquierda al 25% de la población. Es decir:

$$P(X < Q_1) = 0{,}25$$

tipificando

$$P\left(\frac{X-170}{5} < \frac{Q_1-170}{5}\right) = 0{,}25.$$

Como $Z = \dfrac{X-170}{5} \sim N(0;\, 1)$ buscando en la tabla tenemos que

$$\frac{Q_1-170}{5} = -0{,}675 \quad \Rightarrow Q_1 = 166{,}625.$$

De modo análogo, para $Q_3$ llegaríamos a

$$\frac{Q_3 - 170}{5} = 0{,}675 \quad \Rightarrow Q_3 = 173{,}375$$

El cuartil $Q_2$ coincide siempre con la media, por lo que, en este caso tendremos que $Q_2 = 170$.

(b) La probabilidad de que un individuo mida más de 170 cm. es de 0,5. El experimento consiste en seleccionar un individuo al azar y comprobar si mide más de 170 cm. Así tenemos un experimento tipo Bernouilli con $p$ =0,5. Si se seleccionan 5 individuos al azar para el equipo de baloncesto, se están realizando repeticiones independientes del experimento anterior. La variable aleatoria $B$ que cuente el número de éxitos al seleccionar 5 individuos se distribuirá, por tanto según una binomial con $n = 5$ y $p$ =0,5. El suceso del que se pide que se calcule la probabilidad es

$A = \{$al menos un individuo mide más de 170 cm.$\}$

Para calcular la probabilidad de este suceso calculemos la de su complementario:

$\overline{A} = \{$ningún individuo mide más de 170 cm.$\}$

Para calcular la $P(\overline{A})$ observemos que ésta será la probabilidad de que en el experimento binomial mencionado anteriormente se tengan 0 éxitos. Es decir:

$$P(\overline{A}) = P(B = 0) = \binom{5}{0}(0{,}5)^0(0{,}5)^5 = \frac{31}{32} = 0{,}96875.$$

Por tanto $P(A) = 1 - P(\overline{A}) = 0{,}03125$.

(c) Si queremos hallar la probabilidad de que de 1000 estudiantes haya más de 520 que midan más de 170 cm, estamos realizando 1000 repeticiones independientes del experimento Bernouilli considerado anteriormente. Si llamamos $X$ a la variable aleatoria que cuente el número de éxitos en 1000 repeticiones, su distribución será binomial con $n = 1000$ y $p$ =0,5. Lo que pide el problema es hallar la probabilidad de que esta variable $X$ tome un valor mayor de 520. Como sería inviable el realizar los cálculos usando la binomial, usaremos el Teorema Central del Límite y aproximaremos la variable aleatoria $X$ por otra variable $Y$ que se distribuirá según una Normal de media $np$ y varianza $npq$. Además, como estamos realizando la aproximación de una variable discreta por una continua, aplicaremos la corrección de continuidad:

$$P(X > 520) \approx P(Y > 520{,}5) = P\left(\frac{Y - 500}{15{,}8113883} > \frac{520{,}5 - 500}{15{,}8113883}\right)$$

con lo cual, habrá que buscar en las tablas de la normal tipificada

$$P(Z > 1{,}296533841) = 0{,}09739588.$$

## Problema 4.27

**Una empresa tiene dos máquinas. La primera produce el 40 % de los productos y la otra el resto. La probabilidad de que un producto sea defectuoso es de 0,2 si es producido en la primera máquina y de 0,1 si es producido en la segunda.**

**(a) Se toma un lote de 3 productos producidos por la misma máquina y hay uno defectuoso. Hallar la probabilidad de que el lote haya sido producido en la primera máquina.**

**(b) Se toma un lote de 1000 productos de la primera máquina. Hallar la probabilidad de que menos de 180 sean defectuosos.**

(a) Definiremos los siguientes sucesos:

$$\begin{array}{lcl} M_1 & = & \{\text{el producto es producido en la primera máquina}\} \\ M_2 & = & \{\text{el producto es producido en la segunda máquina}\} \\ D & = & \{\text{el producto es defectuoso}\} \end{array}$$

Los datos proporcionados en el enunciado del problema son los siguientes:

$$P(M_1) = 0{,}4 \quad P(M_2) = 0{,}6 \quad P(D/M_1) = 0{,}2 \quad P(D/M_2) = 0{,}1$$

Consideremos el siguiente suceso:

$A = \{$de un lote de 3 productos producidos en la misma máquina, hay 1 defectuoso$\}$

Si realizamos el experimento que consiste en tomar un producto y comprobar si es defectuoso o no, estamos ante un experimento de Bernoulli.

Si repetimos dicho experimento 3 veces de manera independiente, la variable que cuente el número de éxitos se distribuirá según una binomial, donde $n$ valdrá 3 y $p$ valdrá 0,2 ó 0,1 según estemos en la primera o en la segunda máquina.

Por tanto, si los tres productos se han producido por la máquina 1 tendremos:

$$P(A/M_1) = \binom{3}{1}(0{,}2)^1(0{,}8)^2 = 0{,}384.$$

Análogamente, para la segunda máquina

$$P(A/M_2) = \binom{3}{1}(0{,}1)^1(0{,}9)^2 = 0{,}243.$$

Aplicando el teorema de la probabilidad total:

$$\begin{aligned} P(A) &= P(A/M_1)P(M_1) + P(A/M_2)P(M_2) \\ &= 0{,}384 \cdot 0{,}4 + 0{,}243 \cdot 0{,}6 = 0{,}2994. \end{aligned}$$

Lo que nos preguntan es la probabilidad de que el lote sea de la primera máquina sabiendo que se verifica el suceso $A$, es decir, lo que nos preguntan es $P(M_1/A)$. Aplicando el teorema de Bayes:

$$P(M_1/A) = \frac{P(A/M_1)P(M_1)}{P(A)} = \frac{0{,}1536}{0{,}2994} = 0{,}51302605.$$

(b) En este apartado tenemos una sóla máquina, y consideramos la variable aleatoria $X$ que cuenta el número de defectuosos en 1000 productos. $X$ se distribuirá según una binomial de con $n = 1000$ y $p$ =0,2. Lo que pregunta el problema es $P(X < 180)$. Usaremos la aproximación a la normal y la corrección de continuidad. La variable $X$ se aproximará por una variable $Y$ que será normal de media $np = 200$ y de varianza $npq = 160$. Con lo cual, buscando en las tablas de la normal tipificada tenemos:

$$\begin{aligned} P(X < 180) &\approx P(Y < 179{,}5) = P\left(\frac{Y - 200}{12{,}6491106} < \frac{179{,}5 - 200}{12{,}6491106}\right) = \\ &= P(Z < -1{,}620667301) = 0{,}052544497. \end{aligned}$$

# Problemas Propuestos

1. Dada la variable binomial B(4; 0,8), calcular

   (a) La función de probabilidad.

   (b) La función de distribución.

2. Se ha aplicado un test de aptitud numérica a un grupo de COU y se ha detectado que el 65% tiene una aptitud numérica que podríamos considerar inaceptable. ¿Cuál es la probabilidad de que de cinco alumnos elegidos al azar, al menos dos tengan una capacidad numérica aceptable?

3. Según un informe de la OCDE, el año pasado el 35% de la población mundial tenía menos de 15 años. Si fuera posible elegir una muestra aleatoria de la población mundial, formada por 10 personas, ¿cuál es la probabilidad de que a lo sumo haya tres individuos con edad inferior a 15 años?

4. El coste de una pieza es igual a C. El precio de venta depende del diámetro interior de la pieza, $X$, variable aleatoria con función de densidad $f(x) = k \cdot e^{-kx}$, para $x > 0$. Si el diámetro es mayor que tres o menor que uno se pierde la pieza, y si el diámetro está comprendido entre los límites 1 y 3, se vende a un precio igual a Q. Calcular el valor de $k$ que maximice el beneficio medio.

5. En una gasolinera la llegada de vehículos por minuto sigue la distribución de Poisson de parámetro 1,6. Calcular las probabilidades de los siguientes sucesos:

   (a) Que el número de vehículos que lleguen a la gasolinera en un minuto sea superior a tres.

(b) Que el número de vehículos esté comprendido entre dos y cinco.

(c) Que llegue algún vehículo.

6. Se embarca un cargamento de 50 mototores eléctricos. Un inspector elige 5 motores y los inspecciona, si ningún motor es defectuoso el cargamento es aceptado. Si se encuentra que 1 ó más son defectuosos se inspecciona el cargamento completo. Supongamos que en realidad hay tres motores defectuosos en todo el cargamento, ¿cuál es la probabilidad de que sea necesaria una inspección del 100 %?

7. Sabiendo que el número medio de enfermos recibidos cada 10 minutos en un centro sanitario entre las 10 y 15 horas es 1,8. Calcular la probabilidad de que entre las 12h 40' y 12h 50' haya:

   (a) Ningún enfermo.

   (b) Dos enfermos.

   (c) Al menos dos enfermos.

   (d) Más de dos enfermos.

8. De una estación parte un tren cada 20 minutos. Un viajero llega de imprevisto. Hallar:

   (a) Función de densidad de la variable aleatoria "tiempo de espera"

   (b) Probabilidad de que espere al tren menos de 7 minutos.

   (c) Esperanza y varianza de la variable aleatoria "tiempo de espera".

   (d) Probabilidad de que espere exactamente 12 minutos.

9. En una cierta región la probabilidad de que una tormenta con truenos ocurra en un día cualquiera durante el verano (en Julio y Agosto) es igual a 0,1. Suponiendo la independencia de un día con otro, cuál es la probabilidad de que la $1^a$ tormenta con truenos del verano ocurra el 3 de Agosto?

10. De una variable aleatoria uniformemente distribuida $U(a; b)$ se conoce su esperanza $\mu$ y su desviación típica $\sigma$. Hallar los extremos $a$ y $b$ de definición de la variable, en función de $\mu$ y $\sigma$.

11. A una congreso asisten 400 personas; se pide, asumiendo equiprobabilidad en las fechas de nacimiento (365 días):

(a) ¿Cuál es la probabilidad de que alguna persona de los asistentes naciera la primera semana de enero?

(b) ¿Cuál es la probabilidad de que más de la mitad hayan nacido entre los primeros cinco meses del año?

(c) ¿Cuántas cabe esperar que hayan nacido en verano?

(d) ¿Cuál es el número más probable de personas nacidas en el mes de marzo?

(e) ¿Cuál es la varianza de la variable aleatoria que cuenta el número de personas de las 400 que han nacido en la última quincena del año?

12. En una mano de bridge, el número total de triunfos que tienen los dos compañeros A y B es nueve. Obténgase la probabilidad de que los otros cuatro triunfos hayan sido repartidos:

    (a) Todos a uno de los contrarios.

    (b) Tres a un contrario y uno al otro.

    (c) Dos a cada uno de los contrarios.

13. Se está estudiando un nuevo plan de seguridad laboral, para dicho fin se quiere representar las frecuencias de los accidentes de trabajo. Se observaron los accidentes ocurridos a 647 mujeres en 5 meses, obteniendo la siguiente distribución:

| Nº de accidentes: | 0 | 1 | 2 | 3 | 4 | 5 ó más |
|---|---|---|---|---|---|---|
| Frecuencia observada: | 447 | 132 | 42 | 21 | 3 | 2 |

    Estudiar si esta distribución se ajusta a la binomial negativa.

14. En una ciudad de 20.000 habitantes hay una catástrofe y mueren 50 personas. Una persona de otra ciudad se entera por el periódico. ¿Cuál es la probabilidad de que hayan fallecido los 3 familiares suyos que viven en esa ciudad?

15. Una editorial desea poner en práctica un sistema de ventas alternativas a los que viene desarrollando. Consistirá en enviar publicidad postal a las 40 personas que constituyen la muestra piloto, invitándolas a conocer las últimas novedades en libros y ofrecerles algún obsequio a cambio.

Los expertos estiman que un 30% de las personas responderán a la llamada pero que, de éstas, sólo un 20% adquirirá algún libro. Calcular la probabilidad de que después de la prueba nadie haya decidido comprar algún libro.

16. En una fábrica el número de accidentes por semana sigue una ley de Poisson de parámetro $\lambda = 2$. Se pide:

    (a) Probabilidad de que en una semana haya algún accidente.

    (b) Probabilidad de que haya 4 accidentes en el transcurso de dos semanas.

    (c) Probabilidad de que haya 2 accidentes en una semana y otros 2 accidentes en la semana siguiente.

    (d) Si nos dicen que ha habido algún accidente, probabilidad de que en aquella semana no haya más de 3 accidentes.

17. Una centralita telefónica recibe unas 300 llamadas cada hora. No puede establecer más de 12 conexiones por minuto. Se pide:

    (a) Probabilidad de que quede saturada en un minuto dado.

    (b) Probabilidad de que reciba una sóla llamada en un minuto dado.

18. La probabilidad de que un trabajador de la construcción tenga un accidente en un día determinado es 0,0003. Calcular la probabilidad de que en una comunidad autónoma en la que hay 20.000 albañiles, haya 5 accidentes en un día.

19. Se estima que una máquina produce 150 piezas defectuosas al mes (suponemos un mes de 30 días). Se pide:

    (a) Probabilidad de que en un día determinado produzca 3 piezas defectuosas.

    (b) Número máximo de piezas defectuosas en un día, con una certeza aproximadamente del 90%.

20. Se lanza 1000 veces una moneda y se pide:

    (a) Probabilidad de que el número de caras esté comprendido entre 490 y 510.

(b) Intervalo $(a; b)$ centrado en 500, que verifique:

$$P(a < \text{número de caras} < b) = 0{,}95.$$

21. El número de personas que esperan ser atendidas en una oficina de empleo temporal sigue una ley de Poisson de media igual a 4 cada hora. Obtener el número de sillas que debe haber en la sala para que todos puedan estar sentados, con una probabilidad mínima del 90%.

22. Una compañía aseguradora sabe que hay una probabilidad de 0,00003 de que una persona fallezca en un año por accidente laboral. Si la empresa tiene contratados 180.000 seguros de vida por aquel concepto y la cuantía de la indemnizaciones es de 750.000 ptas. por póliza.

    (a) ¿Cuál es la probabilidad de que en un año deba pagar un mínimo de 3 millones de ptas. por aquel tipo de pólizas?
    (b) ¿Qué reservas de tesorería ha de tener en un año para poder pagar aquel tipo de pólizas con un 90% de probabilidad?

23. En una distribución N(5; 2) determinar las siguientes probabilidades:

    (a) $P(7 \leq X \leq 10)$
    (b) $P(-1 \leq X \leq 10)$
    (c) $P(-2 \leq X \leq -1)$

24. Si la variable aleatoria $Y$ es N(2; 3), calcular:

    (a) $P(Y \leq 0{,}3)$.
    (b) $P(-0{,}15 \leq Y \leq 2{,}123)$.
    (c) $P(Y^2 \geq 1{,}21)$.
    (d) Si sabemos que $P(a \leq Y \leq 2{,}3) = 0{,}45$; determinar a.

25. Una variable aleatoria X se distribuye normalmente con media $\mu = 20$. La probabilidad de que X caiga en el intervalo (25 ; 30) es igual a 0,25. ¿Cuánto vale la probabilidad $P(10 \leq X \leq 15)$?

26. Se supone que en cierta población humana el índice cefálico $i$ (anchura del cráneo expresada como porcentaje de su longitud) se distribuye normalmente entre los individuos. Sabemos que el 58% son dolicocéfalos ($i < 75$), el 38% mesocéfalos ($75 < i < 80$) y el 4% braquicéfalos ($i > 80$). Hállense la media y desviación típica de $i$.

27 Una persona de composición mediana está preocupado por su peso y consulta a su médico. Es hombre de 40 años y pesa actualmente 80 kilos, el médico consulta unas tablas y observa que respecto de su peso, edad y estatura, la media nacional es de 76,5 con una desviación típica de 3 kilos. Calcula:

(a) ¿A qué porcentaje de la población española es este individuo superior en peso?

(b) ¿Puede decirle el médico que no se preocupe que su peso no está entre el 10% de los más gruesos?

(c) ¿Cuántos kilos tiene que adelgazar para situarse en el 30% de los de menos peso de la población española?

28. Un centro de cálculo presta servicios de ordenador por los que cobra 15000 unidades monetarias a la hora. Las averías que se pueden producir en el ordenador, $X$, siguen una ley de Poisson de media 0,2 por hora y el coste de reparar $x$ averías viene dado por $2000x^2$, más 3000 de mantenimiento general a la hora. Se pide:

(a) Probabilidad de que en 5 horas de servicio no se hayan producido averías.

(b) El beneficio esperado por hora de servicio.

(c) Calcular el número de horas de prestación de servicio que maximiza el beneficio esperado.

29. Los errores aleatorios que se cometen en las pesadas de una balanza, siguen una normal de media 0 y desviación típica 2 decigramos. Hallar:

(a) Error máximo en una pesada, con una probabilidad de 0,95.

(b) Idem. en 10 pesadas.

(c) Número mínimo de pesadas, para asegurar que el error máximo cometido con el promedio sea inferior a 1 decigramo con una probabilidad de 0,95.

30. Se halló que las longitudes de los dedos medios de manos de varones en una cierta tribu, seguían la distribución normal con $\mu = 60$ mm. y $\sigma = 3$ mm. Resolver las siguientes cuestiones:

(a) Sabiendo que en la tribu había 800 hombres, ¿cuántos hombres aproximadamente tenían unos dedos:

   i. más largos de 62 mm.
   ii. más cortos de 57 mm.

   (b) Comprobar que la probabilidad de que un hombre tenga los dedos más largos que 65 mm. es 0,048. De un grupo de 40 hombres, ¿cuál es la probabilidad de que haya alguno que supere los 65 mm.?

   (c) De entre los hombres de la tribu cuyos dedos son más largos de 62 mm., ¿qué proporción supera los 65 mm.?

31. Se lanza un dado regular. Sea $X$ la variable aleatoria "número de puntos obtenidos" e $Y$ la variable aleatoria que vale 0 si sale la cara 1, 2 ó 3 y vale 1 si sale 4, 5 ó 6. Calcular la covarianza y la correlación entre $X$ e $Y$.

32. Las ventas de un artículo se distribuyen N($\alpha$, $\sigma$). Se sabe que el 20% de ellas son superiores a 1000 euros y que el 30% sobrepasan los 800 euros. Calcular:

   (a) La media y la varianza de la distribución.

   (b) Si los costes están relacionados con las ventas según la expresión $C = 350 + Y - 0{,}00015\, Y^2$, hallar el coste medio.

33. La función de densidad de probabilidad conjunta de dos variables aleatorias $X$ e $Y$ es:

$$f(x, y) = \begin{cases} 3/2(x^2 + y^2) & \text{si} \quad x \in (0,1), \quad y \in (0,1) \\ 0 & \text{en otro caso.} \end{cases}$$

   Se pide:

   (a) Calcular: medias, varianzas y covarianzas.

   (b) Coeficiente de correlación lineal.

34. Se lanza un dado dos veces. Sean las v. aleatorias $X$ = suma de puntos obtenidos, $Y$ = diferencia de puntos entre la primera y la segunda tirada. Hallar la covarianza y la correlación lineal entre $X$ e $Y$.

35. Las variables aleatorias $Y_1$, $Y_2$ son independientes con distribución N(0, 1). Dadas las variables aleatorias $X$, $Y$, siendo $X = 3 + 2Y_1 - Y_2$, $Y = 5 + Y_1 + Y_2$, se pide:

(a) Hallar la función de densidad bivariante de $(X, Y)$.

(b) ¿Cómo se distribuye la variable $Z = 3X + 4Y - 1$?

36. La variable aleatoria $(X, Y)$ tiene una distribución bivariante con función de densidad:

$$f(x,y) = \begin{cases} \dfrac{1}{\pi} e^{\dfrac{x^2+y^2}{2}} & \text{si} \quad x > 0, \quad y \leq 0 \\ & \text{o bien} \quad x \leq 0, \quad y > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Se pide:

(a) Demostrar que las variables marginales X, Y siguen una distribución N(0,1).

(b) Demostrar que X, Y son variables aleatorias dependientes.

(c) ¿Es normal bivariante la distribución conjunta de X,Y? ¿Qué se deduce de estas conclusiones?

37. La variable aleatoria $U$ es $\chi^2$ con $m$ grados de libertad, mientras que la variable aleatoria $V$ es $t$-Student con $n$ grados de libertad. Se pide:

(a) Hallar a tal que P(U > a) = 0,05 para m = 18 y m = 55.

(b) Hallar b tal que P(|V| > b) = 0,01 para n = 20 y n = 45.

38. Hallar el valor $F_{0,05}$ tal que $P(F > F_{0,05}) = 0{,}05$ para la distribución $F$ de Snedecor con $m$ y $n$ grados de libertad en los casos:

(a) m = 6, n = 17.

(b) m = 10, n = 35.

(c) Hallar también $a$ tal que $P(F < a) = 0{,}01$ para $m = 7$, $n = 20$.

39. Si la variable aleatoria $X$ sigue una $t$-Student con 36 grados de libertad, y la variable aleatoria $Y$ una Chi-cuadrado con 62 grados de libertad, hallar $x$ e $y$ tales que: $P(|X| > x) = 0{,}05 \quad ; P(|Y| > y) = 0{,}05$.

40. Dadas dos variables aleatorias $X$ e $Y$, N(0; 1) e independientes, se definen dos nuevas variables, $W$ y $T$. $W = aX + bY^2$; $T = cX^2 + dY$. Calcular las varianzas y el coeficiente de correlación lineal de $W$ y $T$.

41. El control de recepción de ciertos lotes acostumbra a efectuarse por uno de los siguientes criterios:

    - Se irán inspeccionando piezas hasta que aparezca una defectuosa; si ésta se diera antes de la vigésima extracción se devolverá el pedido.
    - Se tomarán 150 piezas al azar y si, entre ellas, hay 3 o más unidades defectuosas se devolverá el pedido.

    Determinar con cuál de los dos criterios hay más probabilidad de rechazar un pedido que contenga un 2% de unidades defectuosas.

42. La vida (en horas) de ciertos tubos electrónicos tienen una densidad:

$$f(x) = \begin{cases} 0 & \text{si} \quad x < 200 \\ ke^{-(x^2/80000)} & \text{si} \quad x \geq 200 \quad \text{(Normal truncada)} \end{cases}$$

    Un aparato contiene 100 de estos tubos y para su funcionamiento al menos 65 de los tubos deben estar activos. Calcular la probabilidad de que el aparato funcione después de 250 horas de servicio.

43. El tiempo que transcurre entre el acabado de un producto y su venta sigue una ley exponencial de media 12 días. ¿Cuál es la probabilidad de que un producto que lleva 6 días almacenado pueda venderse antes de 10 días más?

44. Los días que transcurren entre que se recibe un aviso de pedido y su expedición, son una variable aleatoria con densidad exponencial. El parámetro $\lambda$ de aquella ley es también aleatorio con distribución uniforme entre 0 y 0,2. Deducir la densidad compuesta de aquella variable inicial.

45. Se ha comprobado que el peso neto de un paquete que contiene determinado alimento precocinado sigue aceptablemente una ley normal. Los controles de calidad revelan que un tercio de los paquetes pesan menos de 870 gr. y sólo dos de cada mil paquetes pesan más de 1 Kg. Se pide:

    (a) Calcular la probabilidad de que elegido un paquete al azar pese más de 850 gr.

    (b) Si en una semana salen al mercado 40000 paquetes, ¿cuántos cabe esperar que pesen más de 900 gr.?

46. Una industria panificadora se plantea el número de unidades, del tipo de "pan de molde, 300 gr.", que ha de suministrar para el consumo de los fines de semana. El producto se vende a 69 ptas., lo que supone una ganancia neta de 11 ptas. por unidad; pero al mismo tiempo la empresa se hace cargo de las unidades sobrantes en los centros expendedores, abonando su coste y aprovechando residualmente aquellos retornos, con una pérdida unitaria de 6 ptas.

    Si la experiencia ha demostrado que la demanda real de los consumidores de este tipo de piezas sigue un modelo normal, con 11.500 unidades como cifra más probable y 320 de desviación típica, ¿qué cifra de producción debe preverse para abastecer el consumo de los fines de semana asegurando un mayor beneficio esperado?

47. Dadas cuatro variables aleatorias, independientes N(0; 5), $W$, $X$, $Y$ y $Z$ se definen las siguientes variables aleatorias

$$\begin{aligned} S &= 2W + 3X - Y + Z + 30 \\ T &= W^2 + X^2 + Y^2 + Z^2 \\ U &= \sqrt{\frac{1}{4}(W^2 + X^2 + Y^2 + Z^2)} \\ V &= \frac{W}{\sqrt{\dfrac{X^2 + Y^2 + Z^2}{3}}} \end{aligned}$$

    Calcular:

    (a) P(S $\leq$ 42).
    (b) El valor de t si P(T $\leq$ t) = 0,25.
    (c) P(U $\leq$ 6,973).
    (d) El valor de v si P(|V| $\leq$ v) = 0,7.

48. En una moneda la probabilidad de obtener cara es $p = 0{,}3$ y la de cruz $q = 0{,}7$. Calcular el número de veces que debemos lanzarla a fin de obtener al menos 40 caras con una probabilidad del 65%.

49. Tenemos una sucesión de variables aleatorias $X_1, X_2, \ldots, X_n$ independientes con la misma función de densidad $f(x) = 8 \cdot e^{-8x}$; si $x \geq 0$. Si se consideran 100 variables, calcular:

    (a) Probabilidad de que su suma sea menor o igual a diez.

(b) Probabilidad de que la suma esté comprendida entre 11 y 13, ambos incluidos.

50. Si la variable aleatoria Y se distribuye N(−5; 11), calcular las probabilidades de los siguientes sucesos, interpolando linealmente:

| | |
|---|---|
| (a) $Y \geq -2$. | (b) $-6 \leq Y \leq -3$. |
| (c) $\lvert Y\rvert \leq 11$. | (d) $\lvert Y\rvert \geq 0{,}3$. |
| (e) $Y \leq -3{,}5$. | (f) $Y^3 \leq -6$. |
| (g) $Y^2 \leq \pi$. | (h) $e^Y \geq 4{,}7$. |
| (i) $-4Y + 7 \leq -10$. | (j) $4/(Y + 7) \geq -2$. |
| (k) $7^{15Y} \leq 0{,}001$. | (l) $-4 \leq \ln Y \leq -1$. |

51. Dadas 15 variables aleatorias independientes $Y_i$ distribuidas N(1; $i$), se define una nueva variable

$$Y = \sum_{i=1}^{5} \frac{1}{i} Y_i.$$

Calcular $a$ de tal forma que $P(Y \leq a) = 0{,}23$.

52. Tenemos $n$ variables aleatorias independientes $Y_i$ distribuidas N(1; 4) y $n$ variables aleatorias independientes $Z_i$ distribuidas N(2; 3). Las variables $Y_i$ y las $Z_i$ son independientes entre sí. Se define una nueva variable

$$W = \sum_{i=i}^{n} Y_i - \sum_{i=i}^{n} Z_i.$$

Calcular el valor de $n$ para que $P(7W - 5 \geq 2) = 0{,}06$.

53. Calcular las probabilidades y el valor de la constante $a$ en los siguientes casos, interpolando linealmente:

(a) $P(\chi^2(10) \geq 5{,}031)$.

(b) $P(\chi^2(2) \leq 4{,}5)$.

(c) $P(6{,}821 \leq \chi^2(15) \leq 15{,}13)$.

(d) $P(\chi^2(18) \geq a\ ) = 0{,}1212$.

(e) $P(6{,}9 \leq \chi^2(12) \leq a\ ) = 0{,}83$.

(f) $P\left[\frac{\chi^2(12) - E(\chi^2(12))}{\sqrt{V(\chi^2(12))}} \leq a\right] = 0{,}12$.

54. La distribución $\chi^2(n)$ cuando el número de grados de libertad $n$ es elevado podemos aproximarla a una normal mediante la nueva variable $\sqrt{2\chi^2(n)}$ que se distribuye aproximadamente $N(\sqrt{2n-1}; 1)$. Utilizando esta aproximación, calcular las probabilidades siguientes y hallar el valor de la constante $a$:

    (a) P($\chi^2(150) \geq 128$)
    (b) P($50 \leq \chi^2(65) \leq 60$)
    (c) P($\chi^2(234) \leq 240$)
    (d) P($\chi^2(105) \leq$ a ) = 0'62
    (e) P($a \leq \chi^2(120) \leq 105$) = 0,1503.

55. Si tenemos veinte variables aleatorias independientes $Y_i$ distribuidas N(0; 5), calcular interpolando:

    (a) P $\left[\sum_{i=1}^{20} Y_i^2 \leq 600\right]$.
    (b) P $\left[\frac{1}{20}\sum_{i=1}^{20} Y_i^2 \geq 14\right]$.
    (c) P $\left[\sqrt{\frac{1}{20}\sum_{i=1}^{20} Y_i^2} \geq a\right]$.

56. Si la variable aleatoria W es igual a $\sqrt{\frac{Y_1^2+Y_2^2+\ldots+Y_{11}^2}{7}}$ siendo las variables $Y_i \sim$ N(0; 1/4) e independientes, calcular el valor de $a$ para que se verifique $P(4W + 2 \geq a) = 0{,}35$.

57. Calcular las siguientes probabilidades de la distribución $t$-Student y el valor de la constante $a$, interpolando linealmente:

    (a) P($t(5) \leq 1{,}239$)
    (b) P($t(14) \leq -1{,}5$)
    (c) P($|t(10)| \geq 2{,}5$)
    (d) P($|t(19)| \leq 0{,}98$)
    (e) P($1{,}83 \leq t(22) \leq 2{,}04$)
    (f) P($1{,}2 \leq t(27) \leq 2{,}1$)
    (g) P($-1 \leq t(17) \leq 2$)
    (h) P($t(16) \leq a$ ) = 0,1742
    (i) P($|t(7)| \leq a$) = 0,23

(j) $P(t(12) \leq a) = 0{,}6092$

(k) $P(7a + t(15) \geq 3 = 0{,}0047$

(l) $P(t(6) \geq \log_a 5) = 0{,}12$

(m) $P(t(15)^{3-2a} \leq 2{,}21) = 0{,}728$

(n) $P(a \leq t(25) \leq 1{,}6) = 0{,}1621$

(o) $P(-0{,}83 \leq t(40) \leq a) = 0{,}7204$.

58. Los costes de fabricación de un producto $C$, siguen una distribución N(10; 2) en el intervalo (1; 20). Los beneficios están relacionados con los costes mediante la función $B = -C^2 + 20C - 75$. Calcular:

    (a) Probabilidad de que la empresa obtenga beneficios negativos.

    (b) Probabilidad de que los beneficios sean decrecientes.

    (c) Probabilidad de que los beneficios superen los costes.

59. La variable aleatoria $Y$ es N(5; 2), definimos una nueva variable $Z$, $Z = 2Y^2 - 5Y + 29$. Calcular el coeficiente de variación de $Z$.

60. En un laboratorio, los errores de pesada de una balanza se distribuyen N(0,5 ; 0,005). El máximo error admisible es igual a ± 2% de la media. Se efectúan 90 pesadas por hora y se trabaja 200 horas al mes. Si el salario es de 300 ptas./h, calcular el coste mensual de los posibles errores de pesada.

61. En la cola de una entidad bancaria hay 180 personas para cobrar el subsidio de paro; el importe no es el mismo en cada caso, pero se estima una media por persona de 44000 ptas. y una desviación típica de 6200 ptas. ¿Qué probabilidad hay de que el cajero haya abonado en total más de 8.000.000 ptas.?

62. La venta diaria de una factoría de automóviles se adapta a una distribución uniforme entre 20 y 40 unidades.

    (a) Después de 182 días de venta, ¿cuál es la probabilidad de haber vendido más de 5600 coches, suponiendo las ventas independientes de un día a otro?

    (b) ¿Cuántos días de venta debemos considerar para asegurar, con un 67% de probabilidad, la venta de más de 6000 unidades?

63. Se ha calculado que el tiempo que pasa desde que entra un cliente hasta que entra otro a un supermercado, sigue una ley exponencial de parámetro $\theta = 4$; obtener la ley de probabilidad para la variable tiempo transcurrido para la entrada de 100 clientes.

64. Determinada medida estadística, aplicada a una muestra de contribuyentes, sigue un modelo $\chi^2$ con 100 grados de libertad. ¿Cuál es la probabilidad de que aquella medida $\theta$ tome valores inferiores a 135,81?

65. En un restaurante al que únicamente se puede ir habiendo hecho reserva previa, han comprobado que alrededor del 35% de las reservas anticipadas no son cubiertas a la hora de cenar. ¿Cuántas reservas tendrían que aceptar en un día, para con un 90% de probabilidad asegurar mesa a todos aquellos que la habían reservado previamente, teniendo en cuenta que el restaurante tiene capacidad para 120 personas?

66. La probabilidad de que un paciente se recupere de una rara enfermedad de la sangre es 0,4. Si se sabe que 15 personas han contraído esta enfermedad, determinar la probabilidad de que:

    (a) Al menos 10 sobrevivan.

    (b) Sobrevivan entre 3 y 8.

    (c) Sobrevivan exactamente 5 personas.

67. Encontrar la media y la varianza de la variable aleatoria del problema anterior.

68. Se selecciona un empleado de un grupo de 10 para supervisar un cierto proyecto, escogiendo aleatoriamente una placa de una caja que contiene 10 numeradas del 1 al 10. Encontrar la fórmula para la distribución de probabilidad de X que representa el número de la placa que se ha escogido. ¿Cuál es la probabilidad de que el número escogido sea menor que 4?

69. La rueda de una ruleta se divide en 25 sectores de igual área y se numeran del 1 al 25. Encontrar una fórmula para la distribución de probabilidad de X, que represente el número que se obtiene cuando se hace girar la ruleta.

70. Encontrar la media y la varianza de la variable aleatoria del problema número 68.

71. Un agricultor que siembra fruta afirma que 2/3 de su cosecha de duraznos ha sido contaminada por la mosca del mediterráneo. Encontrar la probabilidad de que al inspeccionar 4 duraznos:

    (a) Los cuatro estén contaminados por la mosca del mediterráneo.

    (b) Cualquier cantidad entre 1 y 3 esté contaminada.

72. Al probar una cierta clase de neumáticos para camión en un terreno escabroso se encontró que 25% de los camiones terminaban la prueba con los neumáticos dañados. De los siguientes 15 camiones probados, encuentre la probabilidad de que:

    (a) De 3 a 6 estén deteriorados.

    (b) Estén estropeados menos de 4.

    (c) Estén más de 5 en mal estado.

73. La probabilidad de que un paciente se recupere de una delicada operación de corazón es 0,9. ¿Cuál es la probabilidad de que exactamente 5 de los próximos 7 pacientes que se sometan a esta intervención sobrevivan?

74. Una investigación de los residentes de Cádiz mostró que 20% preferían un teléfono blanco que de cualquier otro color disponible. ¿Cuál es la probabilidad de que más de la mitad de los siguientes 20 teléfonos que se instalen en esta ciudad sean de color blanco?

75. En un cierto proceso de manufactura se sabe que, en promedio, 1 de cada 100 piezas está defectuosa. ¿Cuál es la probabilidad de que la quinta pieza inspeccionada sea la primera defectuosa?

76. El tablero de un conmutador telefónico es de muy poca capacidad en cuanto al tiempo de ocupación se refiere, de tal forma que las personas no pueden encontrar una línea libre para sus llamadas. Puede ser de interés saber el número de intentos necesarios que se requieren para tener una línea disponible. Suponga que sea $p = 0{,}05$ la probabilidad de tener línea durante la mayor congestión de llamadas. Se tiene interés particular en saber la probabilidad de que sean necesarios 5 intentos para lograr una comunicación.

77. Encontrar la probabilidad de que una persona que lanza al aire tres monedas obtenga ya sea sólo caras o sólo cruces por segunda ocasión en el quinto lanzamiento.

78. Tres personas lanzan una moneda y la que salga dispareja paga los cafés. Si todas las monedas caen iguales, se lanzan nuevamente. Encontrar la probabilidad de que se necesiten menos de 4 lanzamientos.

79. Una secretaria comete en promedio 2 errores por página. ¿Cuál es la probabilidad de que en la siguiente página

    (a) cometa 4 o más errores?

    (b) no cometa errores?

80. La probabilidad de que una persona que vive en una cierta ciudad posea un perro se estima en 0,3. Encontrar la probabilidad de que la décima persona entrevistada aleatoriamente en esta ciudad sea la quinta persona que posee un perro.

81. El número promedio de ratas de campo por Km$^2$ en una ciudad de 5 Km$^2$ se estima que es de 12. Encuentre la probabilidad de que menos de 7 ratas se encuentren

    (a) en un Km$^2$ de la ciudad.

    (b) en 2 de los siguientes 3 Km$^2$ inspeccionados.

82. Un restaurante prepara una ensalada que contiene en promedio 5 verduras diferentes. Encuentre la probabilidad de que la ensalada contenga más de 5 verduras

    (a) en un determinado día.

    (b) en 3 de los siguientes 4 días.

    (c) por primera vez el 5 de abril en dicho mes.

83. Queremos formar una comisión de 5 personas seleccionadas al azar de un grupo formado por 3 hombres y 5 mujeres. Encuentre la distribución de probabilidad para el número de hombres en la comisión.

84. Disponemos de lotes de 40 componentes, que se consideran aceptables si no contienen más de 3 defectuosos. El procedimiento de muestreo del lote consiste en seleccionar 5 componentes aleatoriamente y rechazar el lote si se encuentra un componente defectuoso. ¿Cuál es la probabilidad de que se encuentre exactamente 1 defectuoso en la muestra si hay 3 defectuosos en todo el lote?

85. Calcular la media y la varianza en el problema anterior.

86. Para evitar que lo descubran en la aduana, un viajero ha colocado 6 tabletas de droga en un tarro que contiene 9 píldoras de vitamina que son similares en apariencia. Si el oficial de la aduana selecciona 3 tabletas aleatoriamente para analizarlas, ¿cuál es la probabilidad de que el viajero sea arrestado por posesión ilegal de narcóticos?

87. El dueño de una casa planta 6 tallos que selecciona al azar de una caja que contiene 5 tallos de tulipán y 4 de narciso. ¿Cuál es la probabilidad de que plante 2 tallos de narciso y 4 de tulipán?

88. Una compañía está interesada en evaluar sus actuales procedimientos de inspección en el embarque de 50 artículos idénticos. El procedimiento es tomar una muestra de 5 piezas y autorizar el embarque si se encuentra que no más de 2 están defectuosas. ¿Qué proporción del 20% de embarques defectuosos serán autorizados?

89. Si a una persona se le reparten varias veces 13 cartas de un paquete común de 52, ¿cuántas cartas de corazones por mano podría esperar esta persona? ¿Entre qué dos valores podría esperarse que cayera el número de cartas de corazones durante el 75% del tiempo que está jugando?

90. De un lote de 10 proyectiles, se seleccionan 4 al azar y se disparan. Si el lote contiene 3 proyectiles defectuosos que no explotarán, ¿cuál es la probabilidad de que

    (a) los 4 exploten?

    (b) al menos dos exploten?

91. El número promedio de partículas radiactivas que pasan a través de un contador durante un milisegundo en un experimento de laboratorio es 4. ¿Cuál es la probabilidad de que entren 6 partículas al contador en un milisegundo determinado?

92. Se sabe que 10 es el número promedio de camiones frigoríficos que llegan por día a una cierta ciudad portuaria. Las instalaciones del puerto pueden atender a lo sumo a 15 camiones frigoríficos en un día. ¿Cuál es la probabilidad de que en un determinado día se tengan que esperar los camiones frigoríficos?

93. En un proceso de manufactura en el cual se producen piezas de vidrio, ocurren defectos o burbujas, ocasionando que la pieza no sea útil para la venta. Se sabe que en promedio 1 de cada 1000 piezas tiene una o más burbujas. ¿Cuál es la probabilidad de que en una muestra de 8000 piezas, menos de 7 de ellas tengan burbujas?

94. Las llamadas al servicio de emergencia 061 siguen un proceso de Poisson y en promedio entran 2,7 llamadas por minuto. Encontrar la probabilidad de que:

    (a) no entren más de 4 llamadas en un minuto cualquiera.
    (b) entren menos de 2 llamadas en un minuto cualquiera.
    (c) entren más de 10 llamadas en un período de 5 minutos.

95. Una firma de electrónica afirma que la proporción de unidades defectuosas de un cierto proceso es 5%. Un comprador tiene un proceso de inspección estándar de 15 unidades seleccionadas aleatoriamente de un gran lote. En una ocasión en particular, el comprador, encuentra 5 artículos defectuosos.

    (a) ¿Cuál es la probabilidad de esta ocurrencia, dado que la pretensión del 5% de defectos es correcta?
    (b) ¿Cuál sería su reacción si usted fuera el comprador?

96. Una compañía compra grandes lotes de un cierto tipo de dispositivo electrónico. Se utiliza un método que rechaza un lote completo si se encuentran 2 o más unidades defectuosas en una muestra aleatoria de 100 unidades.

    (a) ¿Cuál es la probabilidad de rechazar un lote que tiene un 1% de defectuosos?
    (b) ¿Cuál es la probabilidad de aceptar un lote que tiene un 5% de defectuosos?

97. La probabilidad de que una persona muera debido a cierta infección respiratoria es 0,002. Encontrar la probabilidad de que mueran menos de 5 personas de las próximas 2000 infectadas.

98. Suponga que en promedio una persona de cada 1000 comete un error numérico al preparar su declaración de la renta. Si se inspeccionan al azar 10000 de ellas, encontrar la probabilidad de que 6, 7 u 8 de éstas tengan error numérico.

99. Se ha advertido que el número de errores de impresión en un libro sigue una ley de Poisson de intesidad media 0,8 errores por página. Calcular la probabilidad de que:

    (a) En una página haya algún error.
    (b) En un capítulo de diez páginas haya más de 10 errores.
    (c) En las 500 páginas de que consta el libro haya menos de 350 errores.

100. En una distribución normal estandar N(0 ; 1) calcular:

    (a) $P(x \leq 2{,}43)$
    (b) $P(x \geq 0{,}74)$
    (c) $P(x \leq -1{,}45)$
    (d) $P(x \geq -2{,}4)$
    (e) $P(1{,}55 \leq x \leq 2{,}48)$
    (f) $P(-1{,}53 \leq x \leq 0{,}81)$.

101. En una distribución normal estandar N(0 ; 1) calcular $m$ si:

    (a) $P(x \leq m) = 0{,}9871$
    (b) $P(x \leq m) = 0{,}2358$
    (c) $P(0{,}47 \leq x \leq m) = 0{,}1680$.

102. En una distribución N(12 ; 2), calcular $P(13 \leq x \leq 16)$.

103. En una distribución N(1 ; 2), calcular las probabilidades:

    (a) $P(x \leq 3{,}14)$
    (b) $P(x \leq -3{,}14)$
    (c) $P(x > 3{,}14)$

104. En una distribución N(4 ; 2), calcular las probabilidades:

    (a) $P(6 \leq x \leq 9)$
    (b) $P(-1 \leq x \leq 9)$
    (c) $P(-2 \leq x \leq -1)$.

105. Calcular los extremos del intervalo simétrico respecto de la media que dejan fuera al 5% de las observaciones de una distribución $N(\mu;\sigma)$.

106. Calcular la media y la desviación típica de una variable aleatoria normal $x$, para la que se verifica:

$$P(x \geq 3) = 0{,}8413 \quad ; \quad P(x < 9) = 0{,}9772.$$

107. El peso neto de envasado de una bote de leche en polvo es de 465 gramos, que se considera como peso medio. Y se aceptan como pesos válidos los comprendidos entre 415 y 520 gramos. Se conoce que el peso en el envasado con la media indicada anteriormente sigue una distribución normal con desviación típica de 30 gramos. Se pide calcular, tomando una muestra aleatoria de 200 botes, qué porcentaje no pasaría el control de envasado.

108. La media de una variable aleatoria normal es 5 veces la desviación típica. Se cumple además que $P(x \leq 6) = 0{,}84134$. Calcula la media y la desviación típica.

109. Calcular los extremos del intervalo simétrico respecto de la media $\mu$ que contienen al 50% de las observaciones de la distribución $\mathrm{N}(\mu,\sigma)$.

110. Dada una variable aleatoria normal en la que se verifica: $P(x \leq 15) = 0{,}1$ , $P(x \leq 20) = 0{,}95$. Calcular:

    (a) $P(x \leq 13)$
    (b) $P(16 \leq x \leq 17)$
    (c) $a$ tal que $P(X \leq a) = 0{,}05$
    (d) $b$ tal que $P(X > y) = 0{,}5$.

111. En la asignatura de Estadística se ha determinado que las calificaciones se distribuyen según una normal N(5,5 ; 1,5). Calcular:

    (a) Entre qué valores en torno a la media se encontrará el 95% de los alumnos.
    (b) Entre qué valores en torno a la media se encontrará el 50% de los alumnos.
    (c) A partir de qué nota se encontrará el 10% de los alumnos con mejor calificación.

# Unidad Temática V

# Fundamentos de la Inferencia Estadística. Estimación.

# Resumen Teórico

## 5.1 Introducción a la Inferencia.

La Inferencia Estadística es la parte de la Estadística que se encarga de generalizar los resultados obtenidos a partir de muestras a toda la población.

Podemos distinguir de modo general dos grandes métodos dentro de la Inferencia Estadística:

**Métodos Paramétricos.-** Se supone que los datos provienen de una familia de distribuciones conocida (Normal, Poisson,...) y lo único que se desconoce es el valor concreto de alguno de los parámetros que la definen ($\mu$ y $\sigma$ para la Normal, $\lambda$ para la Poisson,...).

**Métodos No Paramétricos.-** No suponen conocida la distribución, y solamente suponen hipótesis muy generales respecto a las mismas, como puede ser la continuidad.

Evidentemente, las conclusiones que obtengamos y que generalizaremos para toda la población dependerán de los valores concretos que se hayan observado en la muestra. Esto lleva a muchas personas a manifestar su desconfianza y su recelo de la Estadística. Sin embargo, en la vida cotidiana muchos de nuestros comportamientos se basan en generalizaciones que hacemos a partir de muestras. Así, es muy frecuente que manifestemos que los productos de una determinada marca son mejores que los de la competencia. Dicha afirmación no la hacemos, evidentemente, tras un análisis exhaustivo de todos los productos de una y otra marca, sino basándonos en nuestra propia experiencia personal, que es claramente muy limitada. Es decir, generalizamos a partir de datos que observamos en muestras pequeñas.

### 5.1.1 Muestreo aleatorio simple.

Existen diversas maneras de extraer muestras. El más común de todos es el *muestreo aleatorio simple.* Decimos que una muestra es aleatoria simple cuando cumple todos los elementos de la población tienen la misma probabilidad de ser elegidos y las observaciones se realizan con reemplazamiento, de manera que la población es idéntica en todas las extracciones.

Si tenemos una población de la que observamos una variable que se distribuye según una variable aleatoria $X$ con función de densidad $f(x)$. Consideremos $X_1, X_2, \ldots X_n$ un conjunto de n observaciones aleatorias realizadas en la población $X$. Podemos considerar $X_1, X_2, \ldots X_n$ como n variables aleatorias independientes y con la misma distribución de probabilidad que $X$. Se define entonces $X_1, X_2, \ldots X_n$ como la muestra aleatoria de tamaño n de la población $X$. La función de densidad conjunta sería:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n) .$$

## 5.2 Estadísticos y Estimadores.

Si tenemos una población en la que estamos observando una característica que se distribuye según una variable aleatoria $X$, y tomamos una muestra aleatoria simple de tamaño n

$$x_1, x_2, \ldots, x_n ;$$

un *estadístico* será una función de los elementos de la muestra.

Por ejemplo, podemos tomar una muestra y calcular el siguiente estadístico $T$:

$$T = \frac{x_1 + x_2 + \ldots + x_n}{n} .$$

Evidentemente, el valor del estadístico dependerá de los valores que hayan tomado los elementos de la muestra. Si repetimos el experimento de tomar una muestra y calcularle el estadístico, obtendremos, por lo general, otro valor distinto. Tenemos por tanto que el estadístico será una variable aleatoria. La distribución que seguirá dicha variable aleatoria dependerá de la distribución de la variable $X$. En determinados casos podremos calcular la distribución del estadístico.

Un *estimador* será un estadístico que utilizaremos para estimar el valor de un determinado parámetro de la población. Es decir, si queremos estimar el valor que toma un determinado parámetro de la población $\theta$, tomaremos el valor que tome el estadístico $T$. Para indicar que $T$ es un estimador de $\theta$ se indicará

$$\hat{\theta} = T \; .$$

Vamos a definir algunos estadísticos de tendencia central:

- Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño n, entonces la media muestral se define como el estadístico

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \; .$$

- El segundo estadístico más util para medir el centro de un conjunto de datos es la mediana Me:

$$Me = \begin{cases} X_{[(n+1)/2]} & \text{si n es impar} \\ \dfrac{X_{[n/2]} + X_{[(n/2)+1]}}{2} & \text{si n es par} \end{cases}$$

  Donde los $X_{[i]}$ son los valores ordenados de la muestra.

- Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño n, entonces la Moda, Mo, es el valor de la muestra que ocurre con mayor frecuencia. La moda puede no existir y cuando existe no necesariamente es única.

Veamos a continuación algunos estadísticos que midan la variabilidad de la muestra:

- El rango de una muestra aleatoria $X_1, X_2, \ldots, X_n$ se define como el estadístico $X_{[n]} - X_{[1]}$, siendo $X_{[n]}$ y $X_{[1]}$ las observaciones mayor y menor de la muestra. El rango es una medición pobre de la variación de los datos, sobre todo si el tamaño de la muestra es grande.

- Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño n, entonces la varianza muestral o cuasivarianza de la muestra se define como el estadístico

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \; .$$

Para evitar los errores que se producen al utilizar un valor redondeado de $\overline{X}$ en la fórmula anterior, se puede utilizar la siguiente:

$$S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)} \; .$$

**Nota:** En algunos textos se utiliza el símbolo $S_c^2$ para indicar la cuasivarianza, reservándo el símbolo $S^2$ para indicar la varianza.

## 5.3 Propiedades de los estimadores.

### Centrado o insesgado.

Una de las propiedades que con más frecuencia se le exige a los estimadores es que sean *insesgados.* Decimos que el estimador $T$ es centrado o insesgado para el parámetro $\theta$ si para cualquier tamaño muestral se cumple que

$$E[T] = \theta \; ,$$

diremos que es asintóticamente insesgado si se verifica:

$$\lim_{n \to \infty} E[T] = \theta \; .$$

### Eficiencia.

Si tenemos dos estimadores $T_1$ y $T_2$ de un parámetro $\theta$, decimos que $T_1$ es más eficiente que $T_2$ si se verifica que

$$\text{var}[T_1] \leq \text{var}[T_2] \; .$$

### Consistencia.

Un estimador diremos que es consistente si cumple:

$$\lim_{n \to \infty} E[T] = \theta \qquad \text{y} \qquad \lim_{n \to \infty} \text{var}[T] = 0 \; ,$$

donde $n$ es el tamaño de la muestra.

Una formulación equivalente sería:

$$\lim_{n\to\infty} P[|\tilde{\theta}-\theta| > \epsilon] = 0 ;$$

y otra distinta:

$$\lim_{n\to\infty} P[|\tilde{\theta}-\theta| < \epsilon] = 1 .$$

Ejemplo: En una población normal, la media muestral es un estimador consistente de la media poblacional $\mu$.

## Estadístico suficiente.

Se dice que un estadístico es suficiente cuando recoge toda la información que facilita la muestra. Parece evidente que una estimación óptima de un parámetro sería la que se efectuase a través de estadísticos suficientes.

Si tenemos una población de la que se toma la muestra $(X_1, X_2, \dots, X_n)$, tratamos de estimar el parámetro $\theta$ de la población.

Llamaremos verosimilitud de una muestra a la función de densidad conjunta de la misma:

$$L(X;\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) .$$

Sea el estadístico $T = \varphi(x_1, x_2, \dots, x_n)$; la aplicación

$$x_1 \longrightarrow \varphi(x_1, x_2, \dots, x_n)$$

es biunívoca y continua y diferenciable para todos los $x_2, x_3, \dots, x_n$. Sea $g(t,\theta)$ la densidad de T y $h(x_2, x_3, \dots, x_n/t;\theta)$ la densidad condicionada de $x_2, x_3, \dots, x_n$ cuando $T = t$. Haciendo $t = \varphi(x_1, x_2, \dots, x_n)$ tendremos

$$L(X;\theta) = g(t;\theta)h(x_2, x_3, \dots, x_n/t;\theta)|\frac{\partial\varphi}{\partial x_1}| .$$

Entonces si $h$ no dependiese de $\theta$, decimos que $T$ es un estadístico suficiente respecto del parámetro $\theta$. Como $\dfrac{\partial\varphi}{\partial x_1}$ no depende de $\theta$, se puede expresar según la descomposición de Fisher-Neyman.

Por tanto, concluimos que $T$ es suficiente cuando $L$ se puede descomponer factorialmente por la descomposición de Fisher-Neyman:

$$L(X;\theta) = g(t;\theta)h(x_1, x_2, \ldots, x_n) .$$

## 5.4 Obtención de estimadores.

Nos podemos plantear el problema de cómo determinar los estimadores de manera que cumplan las propiedades estudiadas anteriormente. Hay 4 métodos principalmente:

1. Método de los momentos
2. Método de máxima verosimilitud
3. Método de mínimos cuadrados
4. Método de Bayes

### 5.4.1 Método de los momentos.

Si tenemos una población de la que se desean estimar $k$ parámetos, el método de los momentos consiste en igualar los $k$ primeros momentos respecto al origen de la población con los de la muestra, obteniéndose así un sistema de ecuaciones del que se obtienen como raíces los estimadores buscados. Los estimadores obtenidos por el método de los momentos son consistentes, pero, por lo general, no son insesgados ni de mínima varianza.

### 5.4.2 Método de máxima verosimilitud.

Anteriormente hemos definido la función de verosimilitud de una muestra como la función de densidad conjunta de la misma:

$$L(X;\theta) = L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \ldots, \theta_k) .$$

Los estimadores de máxima verosimilitud serán aquellos valores de los parámatros $\theta_i$ que hagan máxima esta función de verosimilitud.

En el caso continuo este problema se resuelve facilmente, recurriendo a los principios de optimización de funciones, Los valores que maximicen L serán los mismos que los que maximicen $\log(L)$. Derivando esta función e igualando a cero se obtendrán los valores de los estimadores.

## 5.5 Distribuciones asociadas a la Normal.

Van a aparecer al encontrar la distribución muestral de los estadísticos. Son de uso muy frecuente en intervalos de confianza y contrastes de hipótesis. Sus valores se encuentran en tablas.

### 5.5.1 Distribución $\chi^2$ de Pearson.

Sean $Z_1, Z_2, \ldots Z_n$ variables aleatorias N(0,1) independientes. Entonces la variable aleatoria suma

$$Z_1^2 + Z_2^2 + \ldots + Z_n^2 \ ;$$

se distribuye según una chi-cuadrado con $n$ grados de libertad. Se representa por $\chi_n^2$.

La media de una distribución $\chi_n^2$ es $n$ y su varianza es $2n$.

Tiene la propiedad de reproductividad, al sumar dos $\chi^2$ independientes de grados de libertad $n_1$ y $n_2$, el resultado es otra $\chi^2$ de $n_1 + n_2$ grados de libertad.

La función de densidad de una distribución $\chi^2$ con $n$ grados de libertad es

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2} \ ,$$

para valores positivos de $x$, siendo 0 si $x < 0$. Puede comprobarse que se trata de un caso particular de la distribución gamma, con $\alpha = n/2$ y $\beta = 2$.

### 5.5.2 Distribución t de Student.

La distribución t de Student con $n$ grados de libertad se define por

$$t_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}},$$

donde $Z$ es una variable aleatoria N(0,1) independiente del denominador.

La variable $t_n$ es simétrica con respecto al cero, con mayor dispersión que la normal estándar y tiende a ésta al aumentar el valor de $n$ ( prácticamente coinciden si $n > 100$)

Para valores de $n > 30$ se puede considerar que la normal da una buena aproximación de la t de Student.

La función de densidad de una t de Student es:

$$f(t) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}(1 + t^2/n)^{-(n+1)/2} \qquad \text{si} \quad t \in I\!R.$$

La media de la distribución t de Student vale 0 si $n > 1$ y su varianza $\frac{n}{n-2}$ si $n > 2$.

### 5.5.3 Distribución F de Fisher-Snedecor.

Se define como el cociente de dos $\chi^2$ independientes divididas por sus grados de libertad, asi tenemos que

$$F_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$$

es una $F$ Fischer-Snedecor con grados de libertad $n$ y $m$.

Su función de densidad es:

$$f(x) = \frac{\Gamma[(n+m)/2](n/m)^{n/2}}{\Gamma(n/2)\Gamma(m/2)} \frac{x^{(n/2)-1}}{(1+nx/m)^{(n+m)/2}} \quad \text{si} \quad x > 0 \;,$$

y es nula en otro caso.

La media y la varianza de una variable $X \sim F_{n,m}$ valen:

$$E[X] = \frac{n}{n-2} \quad \text{si} \quad n > 2 ,$$

$$\text{Var}(X) = \frac{2n^2(n+m-2)}{m(n-2)(n-4)} \quad \text{si} \quad n > 4 .$$

Por definición se cumple que $F_{n,m} = F_{m,n}^{-1}$ .

## 5.6 Distribuciones muestrales.

El campo de la inferencia estadística trata básicamente con las generalizaciones y predicciones. Dado que un estadístico es una variable aleatoria que depende únicamente de la muestra observada, tiene una distribución de probabilidad. La distribución de probabilidad de un estadístico recibe el nombre de distribución muestral.

### 5.6.1 Distribución muestral de la media de una población con varianza conocida.

Supongamos que tenemos una muestra aleatoria de tamaño $n$ de una población normal con media $\mu$ desconocida y varianza $\sigma^2$ conocida, entonces el estadístico $\overline{X}$ tiene una distribución normal con media

$$E[\overline{X}] = \frac{\mu + \mu + \ldots + \mu}{n} = \mu ,$$

y varianza

$$\text{var}(\overline{X}) = \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2} = \frac{\sigma^2}{n} .$$

Nota: La demostración se basa en la propiedad reproductiva de la normal.

Si $\overline{X}$ es la media de una muestra aleatoria de tamaño $n$ que se toma de una población con media $\mu$ y varianza $\sigma^2$ entonces

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$$

El resultado anterior es consecuencia del teorema central del límite. La aproximación es generalmente buena si $n \geq 30$, sin importar la forma de la población.

### 5.6.2 Distribución muestral de la diferencia de medias de dos poblaciones con varianza conocida.

Si se tienen dos poblaciones normales, la primera con media $\mu_1$ y varianza $\sigma_1^2$ y la segunda con media $\mu_2$ y varianza $\sigma_2^2$, podemos considerar el estadístico $\overline{X}_1$, que representa la media muestral de tamaño $n_1$ de la primera población y el estadístico $\overline{X}_2$, que representa la media muestral de tamaño $n_2$ de la segunda población. Nos planteamos encontrar la distribución de la diferencia de medias muestrales $\overline{X}_1 - \overline{X}_2$. La media sería

$$E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] = \mu_1 - \mu_2 \ ,$$

y la varianza:

$$\mathrm{var}(\overline{X}_1 - \overline{X}_2) = \mathrm{var}(\overline{X}_1) + \mathrm{var}(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \ .$$

Por tanto

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \longrightarrow N(0,1) \ .$$

Si $n_1$ y $n_2$ son ambos mayores o iguales que 30, la aproximación normal para la diferencia de medias suele ser bastante buena, sin importar la forma de las poblaciones.

### 5.6.3 Distribución muestral de $(n-1)S^2/\sigma^2$.

Si se extrae una muestra aleatoria de tamaño $n$ de una población normal con media $\mu$ y varianza $\sigma^2$ y calculamos la cuasivarianza muestral $S^2$, entonces el estadístico $(n-1)S^2/\sigma^2$ se distribuye según una chi-cuadrado con $n-1$ grados de libertad.

Para comprobar este resultado nos basamos en la siguiente relación:

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{(\overline{X} - \mu)^2}{\sigma^2/n} \ .$$

Es inmediato comprobar que el lado izquierdo de la igualdad anterior es una $\chi^2$ con $n$ grados de libertad (es la suma de los cuadrados de $n$ normales tipificadas independientes). También puede observarse que el segundo término del lado derecho de la igualdad se distribuye según una $\chi^2$ con un grado de libertad (es una normal tipificada al cuadrado). Al ser ambas variables independientes, podemos aplicar la propiedad reproductiva de la $\chi^2$, obteniéndose, por tanto, que $(n-1)S^2/\sigma^2$ es una $\chi^2$ con $n-1$ grados de libertad.

Si se conociera el valor de $\mu$, tendríamos una $\chi^2$ con $n$ grados de libertad.

### 5.6.4 Distribución muestral de la media de una población con varianza desconocida.

Si tenemos una población normal cuya varianza es desconocida y la muestra es pequeña, no podemos utilizar la distribución normal. En este caso usaremos que

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \text{y que} \qquad (n-1)S^2/\sigma^2 \sim \chi^2_{n-1} ,$$

se distribuye según una chi-cuadrado con $n-1$ grados de libertad. Dividiendo una normal entre la raiz cuadrada de una chi-cuadrado partida por sus grados de libertad, obtenemos una t de Student.

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2/\sigma^2}{(n-1)}}} = \frac{\dfrac{\sigma(\bar{X}-\mu)}{\sigma}}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t_{n-1}$$

Por tanto, para estimar la media de una población normal, con varianza desconocida, usaremos el estimador:

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} ,$$

que se distribuirá según una $t$ de Student con $n-1$ grados de libertad. En el caso de que $n$ sea mayor que 30, podemos usar la aproximación normal.

### 5.6.5 Distribución muestral de la diferencia de medias de dos poblaciones con varianza desconocida.

**Varianzas desconocidas pero iguales.**

Para estimar la diferencia de medias de dos poblaciones normales, con varianzas $\sigma_1$ y $\sigma_2$ desconocidas pero iguales, usaremos el estimador:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

que se distribuye según una $t$ de Student con $n_1 + n_2 - 2$ grados de libertad.

Siendo
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

**Varianzas desconocidas y distintas.**

Para estimar la diferencia de medias de dos poblaciones normales, con varianzas $\sigma_1$ y $\sigma_2$ desconocidas y distintas, usaremos el estimador:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}},$$

que se distribuye según una $t$ de Student con $\nu$ grados de libertad. Siendo $\nu$ el entero más próximo a:

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}.$$

### 5.6.6 Distribución muestral del cociente de varianzas.

Si se extrae una muestra aleatoria de tamaño $n_1$ de una población normal con media $\mu_1$ y varianza $\sigma_1^2$ y calculamos la cuasivarianza muestral $S_1^2$, entonces el estadístico

$$(n_1 - 1)S_1^2/\sigma_1^2$$

se distribuye según una chi-cuadrado con $n_1 - 1$ grados de libertad.

Si tenemos otra muestra aleatoria independiente de la anterior de tamaño $n_2$ de otra población normal con media $\mu_2$ y varianza $\sigma_2^2$ y calculamos la cuasivarianza muestral $S_2^2$, entonces el estadístico

$$(n_2 - 1)S_2^2/\sigma_2^2$$

se distribuye según una chi-cuadrado con $n_2 - 1$ grados de libertad.

Si dividimos dos distribuciones chi-cuadrado, cada una de ellas dividida por sus respectivos grados de libertad obtenemos una F de Snedecor. Por tanto,

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1} \ .$$

# Problemas Resueltos

## Problema 5.1

**Encontrar el estimador del parámetro p de una Bernoulli por el método de los momentos y por el método de máxima verosimilitud.**

---

Si tenemos $(x_1, x_2 \ldots x_n)$ una muestra aleatoria simple obtenida a partir de la población, el método de los momentos consiste en igualar los momentos de la muestra con los momentos de la población.

En este caso, la media de la población es $p$. Igualando a la media de la muestra obtenemos que

$$\tilde{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Para hallar el estimador por el método de máxima verosimilitud, hallamos la función de verosimilitud:

$$L(X;p) = \prod_{i=1}^{n} P(X = x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

tomando logaritmos

$$\log[L(X;p)] = \sum_{i=1}^{n} x_i \log(p) + \sum_{i=1}^{n}(1-x_i)\log(1-p)$$

derivando respecto a $p$ e igualando a 0 obtenemos:

$$\frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n} 1 - x_i = 0$$

de donde

$$(1-p)\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} 1 - x_i = 0$$

de donde despejando, obtenemos que

$$\tilde{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Es decir, en este caso coinciden los estimadores hallados por el método de los momentos y por el método de máxima verosimilitud.

## Problema 5.2

**Se sospecha que una moneda está trucada. Para comprobarlo se lanzó 1000 veces, obteniéndose que salieron 615 caras. Hallar el estimador de máxima verosimilitud de la probabilidad de sacar cara.**

El lanzamiento de una moneda es un experimento aleatorio de Bernouilli siendo la probabilidad de salir cara $p$. En este caso, al estar la moneda trucada, desconocemos el valor de $p$. Si repetimos el lanzamiento de la moneda 1000 veces, tenemos una muestra $(x_1, x_2 \ldots x_{1000})$ obtenido a partir de la población. Cada uno de los $x_i$ valdrá 1 si en ese lanzamiento se obtuvo cara y 0 en caso contrario. No sabemos el valor concreto de cada uno de los $x_i$, pero sí sabemos que su suma da 615.

Aplicando el problema anterior, el estimador de máxima verosimilitud de $p$ será:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

que en nuestro caso vale 0,615. Por tanto, el valor de $p$ que hace más verosímil la obtención de ese resultado es $p = 0{,}615$.

## Problema 5.3

**Encontrar el estimador del parámetro $\lambda$ de una Poisson por el método de los momentos y por el método de máxima verosimilitud.**

Si tenemos una muestra $(x_1, x_2 \ldots x_n)$ una muestra aleatoria simple obtenido a partir de la población, el método de los momentos consiste en igualar los momentos de la muestra con los momentos de la población.

En este caso, la media de la población es $\lambda$. Igualando a la media de la muestra obtenemos que

$$\tilde{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Para hallar el estimador por el método de máxima verosimilitud, hallamos la función de verosimilitud:

$$L(X;\lambda) = \prod_{i=1}^{n} P(X = x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$$

operando obtenemos

$$L(X;\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

tomando logaritmos

$$\log[L(X;\lambda)] = -n\lambda + \sum_{i=1}^{n} x_i \log(\lambda) - log[\prod_{i=1}^{n} x_i!]$$

derivando respecto a $\lambda$ e igualando a 0 obtenemos:

$$-n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

de donde despejando, obtenemos que

$$\tilde{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Es decir, en este caso coinciden los estimadores hallados por el método de los momentos y por el método de máxima verosimilitud.

# Problema 5.4

**Encontrar el estimador del parámetro $\mu$ de una Normal de desviación típica conocida igual a $\sigma$, por el método de los momentos y por el método de máxima verosimilitud.**

---

Si tenemos una muestra $(x_1, x_2 \ldots x_n)$ una muestra aleatoria simple obtenido a partir de la población, el método de los momentos consiste en igualar los momentos de la muestra con los momentos de la población.

En este caso, la media de la población es $\mu$. Igualando a la media de la muestra obtenemos que

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Para hallar el estimador por el método de máxima verosimilitud, hallamos la función de verosimilitud:

$$L(X;\mu) = \prod_{i=1}^{n} f(x_i;\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

de donde:

$$L(X;\mu) = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \cdot e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}}$$

tomando logaritmos obtenemos:

$$\log[L(X;\mu)] = n\log\left[\frac{1}{\sigma\sqrt{2\pi}}\right] - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2},$$

y derivando respecto a $\mu$ e igualando a 0 queda:

$$2\sum_{i=1}^{n}\frac{(x_i-\mu)}{2\sigma^2} = 0$$

de donde:

$$\sum_{i=1}^{n}(x_i-\mu) = 0\,,$$

por tanto:

$$\sum_{i=1}^{n} x_i - n\mu = 0$$

con lo que tenemos el estimador

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

# Problema 5.5

**Un proceso de manufactura produce fibras de distintas longitudes. Se supone que la longitud $X$ de una fibra es una variable aleatoria continua con función de densidad**

$$f(x) = \begin{cases} \dfrac{2\lambda \cdot e^{-(\lambda x^2)}}{x^2} & \textbf{si} \quad x > 0 \quad , \quad \lambda > 0 \\ 0 & \textbf{en otro caso} \end{cases}$$

**Supongamos que se escogen al azar $n$ fibras y se miden sus longitudes, $x_1, x_2, \ldots, x_n$.**

**(a) Calcular el estimador de máxima verosimilitud para el parámetro $\lambda$.**

**(b) Demostrar que el estadístico $T = \sum_{i=1}^{n} x_i^2$ es un estadístico suficiente para $\lambda$.**

---

(a) Para calcular el estimador por el método de la máxima verosimilitud, hallamos la función de verosimilitud:

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \lambda) &= \prod_{i=1}^{n} f(x_i; \lambda) = \\ &= \frac{(2\lambda)^n \cdot e^{-\lambda(x_1^2 + x_2^2 + \ldots + x_n^2)}}{x_1^2 \cdot x_2^2 \cdot \ldots \cdot x_n^2}. \end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; \lambda) = n \cdot \ln(2\lambda) - \lambda(x_1^2 + x_2^2 + \ldots + x_n^2) - 2\sum_{i=1}^{n} \ln x_i.$$

Derivando con respecto a $\lambda$ e igualando a cero tenemos

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - (x_1^2 + x_2^2 + \ldots + x_n^2) = 0.$$

Y despejando

$$\frac{n}{\lambda} = (x_1^2 + x_2^2 + \ldots + x_n^2).$$

Y el estimador de máxima verosimilitud para $\lambda$ es

$$\tilde{\lambda} = \frac{n}{x_1^2 + x_2^2 + \ldots + x_n^2}.$$

Efectivamente es un máximo, pues la segunda derivada evaluada para $\lambda = \tilde{\lambda}$ tiene signo negativo.

(b) Para demostrar que el estadístico $T = \sum_{i=1}^{n} x_i^2$ es suficiente tendremos que comprobar que la función de verosimilitud se puede escribir según la descomposición factorial de Fisher-Neyman.

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \lambda) &= \frac{(2\lambda)^n \cdot e^{-\lambda(x_1^2+x_2^2+\ldots+x_n^2)}}{x_1^2 \cdot x_2^2 \cdot \ldots \cdot x_n^2} \\ &= (2\lambda)^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i^2} \cdot \frac{1}{(x_1 \cdot x_2 \cdot \ldots \cdot x_n)^2}. \end{aligned}$$

Y si tomamos como funciones $g(T;\lambda)$ y $h(x_1, x_2, \ldots, x_n)$

$$g(T;\lambda) = (2\lambda)^n \cdot e^{-\lambda \cdot T}$$

$$h(x_1, x_2, \ldots, x_n) = \frac{1}{(x_1 \cdot x_2 \cdot \ldots \cdot x_n)^2}$$

obtenemos que la función de vesorimilitud admite una descomposición según el teorema de Fisher-Neyman:

$$L(x_1, x_2, \ldots, x_n; \lambda) = g(T;\lambda) \cdot h(x_1, x_2, \ldots, x_n).$$

Y por tanto el estadístico $T$ es suficiente para estimar el parámetro $\lambda$.

# Problema 5.6

**Considere la distribución exponencial cuya función de densidad es f(x). Calcular el estimador de máxima verosimilitud del parámetro $\alpha$ para muestras aleatorias de tamaño $n$.**

$$f(x) = \begin{cases} \frac{1}{\alpha} \cdot e^{(-x/\alpha)} & \textbf{si} \quad x > 0 \\ 0 & \textbf{en otro caso} \end{cases}$$

Para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned}
L(x_1, x_2, \ldots, x_n; \alpha) &= \prod_{i=1}^{n} f(x_i; \alpha) = \\
&= \frac{1}{\alpha} \cdot e^{(-x_1/\alpha)} \cdot \frac{1}{\alpha} \cdot e^{(-x_2/\alpha)} \cdot \ldots \cdot \frac{1}{\alpha} \cdot e^{(-x_n/\alpha)} = \\
&= \frac{1}{\alpha^n} \cdot e^{-\sum_{i=1}^{n} x_i/\alpha}
\end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; \alpha) = -n \ln \alpha - \frac{1}{\alpha} \sum_{i=1}^{n} x_i.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto a $\alpha$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es menor que cero.

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{n}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^{n} x_i = 0.$$

Y despejando

$$-n \cdot \alpha + \sum_{i=1}^{n} x_i = 0$$

Y el estimador de máxima verosimilitud para $\alpha$ es

$$\tilde{\alpha} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{X}.$$

Si comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $\alpha = \tilde{\alpha}$ tenemos:

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = \frac{n}{\alpha^2} - \frac{1}{\alpha^3} \sum_{i=1}^{n} x_i$$

$$\left. \frac{\partial^2 \ln L}{\partial \alpha^2} \right|_{\tilde{\alpha}} = \frac{n^3}{(\sum_{i=1}^{n} x_i)^2} - \frac{2n^3}{(\sum_{i=1}^{n} x_i)^2} < 0 \quad \forall n \in N.$$

Luego el estimador máximo verosímil de $\alpha$ es la media muestral.

# Problema 5.7

**Se preparó un programa de ordenador para elegir números al azar distribuidos según una Poisson de parámetro $\lambda$, pero se ha perdido la información sobre el valor de $\lambda$. Su ejecución en 10 ocasiones dió los siguientes resultados:**

**6 3,4 5,6 6,3 6,4 5,3 5,4 5 5,2 5,5**

**(a) Determinar una estimación del parámetro $\lambda$ de la Poisson.**

**(b) ¿Es esta estimación insesgada?**

---

(a) La estimación del parámetro de una Poisson puede hacerse a través del estimador de máxima verosimilitud, que como ya hemos visto en otro ejercicio anteriormente coincidía con la estimación por el método de los momentos. Se obtenía

$$\tilde{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}.$$

Esto es, la estimación del parámetro de la Poisson es la media muestral.

En nuestro caso como $\sum_{i=1}^{10} x_i = 49$, tenemos que $\tilde{\lambda} = 4{,}9$.

(b) Para comprobar si un estimador es centrado o insesgado debe verificarse

$$E[\tilde{\lambda}] = \lambda.$$

Esto es:

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{n\cdot\lambda}{n} = \lambda.$$

Luego $\tilde{\lambda} = \bar{X}$ es un estimador insesgado para $\lambda$.

# Problema 5.8

**En un estudio de supervivencia se considera una variable aleatoria con función de densidad:**

$$f(x) = \begin{cases} \dfrac{2}{a}\cdot x\cdot e^{(-x^2/a)} & \textbf{si} \quad x \geq 0 \quad , \quad a > 0 \\ 0 & \textbf{en otro caso} \end{cases}$$

(a) **Calcular el estimador de máxima verosimilitud del parámetro $a$.**

(b) **Con la siguiente muestra de tamaño 10, calcular la estimación de la mediana.**

| | | | | |
|---|---|---|---|---|
| **1,77** | **1,98** | **2,50** | **1,15** | **2,93** |
| **4,44** | **1,98** | **1,47** | **1,53** | **5,24** |

---

(a) Para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; a) &= \prod_{i=1}^{n} f(x_i; a) = \\ &= \prod_{i=1}^{n} \frac{2}{a} \cdot x_i \cdot e^{(-x_i^2/a)} = \\ &= \frac{2^n}{a^n} \cdot \prod_{i=1}^{n} x_i \cdot e^{-\sum_{i=1}^{n} x_i^2/a} \end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; a) = n \ln 2 - n \ln a + \sum_{i=1}^{n} \ln x_i - \frac{1}{a} \sum_{i=1}^{n} x_i^2.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $a$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial a} = -\frac{n}{a} + \frac{1}{a^2} \sum_{i=1}^{n} x_i^2 = 0.$$

Y despejando

$$-n \cdot a + \sum_{i=1}^{n} x_i^2 = 0$$

Y el estimador de máxima verosimilitud para el parámetro $a$ es

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^{n} x_i^2.$$

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $a = \tilde{a}$ tenemos:

$$\frac{\partial^2 \ln L}{\partial a^2} = \frac{n}{a^2} - \frac{2}{a^3}\sum_{i=1}^{n} x_i^2$$

$$\left.\frac{\partial^2 \ln L}{\partial a^2}\right|_{\tilde{a}} = \frac{n^3}{(\sum_{i=1}^{n} x_i^2)^2} - \frac{2n^3}{(\sum_{i=1}^{n} x_i^2)^2} < 0 \quad \forall n \in N.$$

Luego el estimador máximo verosímil de $a$ es el momento muestral de orden 2 con respecto al origen.

(b) Si utilizamos el estimador del parámetro $a$ del apartado anterior, tenemos:

$$n = 10 \qquad \Longrightarrow \qquad \tilde{a} = \frac{78{,}8041}{10} = 7{,}88041.$$

Por lo que la función de densidad de $X$ queda

$$f(x;\tilde{a}) = \begin{cases} \dfrac{2}{7{,}88041} \cdot x \cdot e^{(-x^2/7{,}88041)} & \text{si} \quad x \geq 0 \\ 0 & \text{fuera} \end{cases}$$

Tenemos que encontrar el valor de la mediana $Me$ estimado por esta muestra tal que

$$\int_0^{Me} f(x;\tilde{a})\,dx = \int_{Me}^{+\infty} f(x;\tilde{a})\,dx = \frac{1}{2}.$$

$$\begin{aligned}
\int_0^{Me} f(x;\tilde{a})\,dx &= \int_0^{Me} \frac{2}{7{,}88041} \cdot x \cdot e^{(-x^2/7{,}88041)}\,dx = \\
&= \left\{ \begin{array}{l} \dfrac{x^2}{7{,}8804} = t \\ \dfrac{2x}{7{,}88041}\,dx = dt \end{array} \right\} = \\
&= \int_0^{Me^2/7{,}88041} e^{-t}\,dt = \left. -e^{-t} \right|_0^{Me^2/7{,}88041} = \\
&= 1 - e^{-Me^2/7{,}88041}
\end{aligned}$$

Y como esta integral tiene que valer $1/2$, tenemos

$$1 - e^{-\frac{Me^2}{7{,}88041}} = \frac{1}{2} \qquad \longrightarrow$$

$$\frac{1}{2} = e^{-\frac{Me^2}{7,88041}} \qquad \longrightarrow$$

$$-\ln 2 = -\frac{Me^2}{7,88041} \qquad \longrightarrow$$

$$Me^2 = 7,88041 \cdot \ln 2 \approx 5,462284 \qquad \longrightarrow Me = \pm\sqrt{5,462284}$$

Y la estimación de la mediana es:

$$\tilde{Me} = 2,337153 \qquad \text{pues } x \geq 0.$$

# Problema 5.9

**Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria en la función de densidad:**

$$f(y;\theta) = \begin{cases} (\theta+1)\cdot y^{\theta} & \textbf{si} \quad 0 < y < 1 \quad , \quad \theta > -1 \\ 0 & \textbf{fuera} \end{cases}$$

**(a) Obtener un estimador para $\theta$ por el método de los momentos.**

**(b) Obtener asimismo el estimador de máxima verosimilitud para $\theta$. Comparar el resultado con el estimador obtenido por el método de los momentos.**

---

(a) Para obtener el estimador del parámetro $\theta$ por el método de los momentos, debemos igualar los dos primeros momentos, el primer momento de la muestra (media muestral $\bar{Y}$) con el primer momento poblacional (la media $\mu$).

Si calculamos la media $E[Y]$ tenemos:

$$\begin{aligned}
\mu = E[Y] &= \int_{-\infty}^{+\infty} y \cdot f(y)\, dy = \\
&= \int_0^1 (\theta+1)\cdot y^{\theta+1}\, dy = \\
&= (\theta+1)\int_0^1 y^{\theta+1}\, dy = \\
&= (\theta+1)\left[\frac{y^{\theta+2}}{\theta+2}\right]_0^1 = \frac{\theta+1}{\theta+2}\left[y^{\theta+2}\right]_0^1 =
\end{aligned}$$

$$= \frac{(\theta+1)}{(\theta+2)}.$$

Si ahora igualamos los dos primeros momentos obtenemos:

$$\bar{Y} = \frac{(\theta+1)}{(\theta+2)} \quad \Longrightarrow \quad (\theta+2)\cdot\bar{Y} = \theta+1 \quad \Longrightarrow \theta\cdot\bar{Y} + 2\bar{Y} = \theta+1$$

con lo que

$$\theta(\bar{Y}-1) = 1 - 2\bar{Y}$$

Por tanto el estimador del parámetro $\theta$ por el método de los momentos es

$$\tilde{\theta} = \frac{1-2\bar{Y}}{\bar{Y}-1} = \frac{1}{1-\bar{Y}} - 2.$$

(b) Si calculamos el estimador de máxima verosimilitud del parámetro $\theta$ tenemos que para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned} L(y_1, y_2, \ldots, y_n; \theta) &= \prod_{i=1}^{n} f(y_i;\theta) = \\ &= \prod_{i=1}^{n} (\theta+1)\cdot y_i^{\theta} = \\ &= (\theta+1)^n \cdot (y_1\cdot y_2 \cdot \ldots \cdot y_n)^{\theta}. \end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(y_1, y_2, \ldots, y_n;\theta) = n\ln(\theta+1) + \theta\sum_{i=1}^{n}\ln y_i.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $\theta$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial\theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n}\ln y_i = 0.$$

Y despejando

$$\frac{-n}{\theta+1}=\sum_{i=1}^{n}\ln y_i \qquad \Rightarrow \theta+1=\frac{-n}{\sum\limits_{i=1}^{n}\ln y_i}.$$

Y el estimador de máxima verosimilitud para el parámetro $\theta$ es

$$\hat{\theta}=\frac{-n}{\sum\limits_{i=1}^{n}\ln y_i}-1$$

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $\theta=\hat{\theta}$, y en efecto es menor que cero.

Podemos comprobar que los dos métodos de estimación nos proporcionan estimaciones diferentes del parámetro $\theta$, pues $\hat{\theta}\neq\tilde{\theta}$.

# Problema 5.10

**Si una variable aleatoria tiene por función de densidad**

$$f(x;a)=\begin{cases}\dfrac{2a}{1-a}\cdot x^{\left(\frac{3a-1}{1-a}\right)} & \textbf{si} \quad 0\leq x\leq 1 \quad , \quad a>0\\ 0 & \textbf{fuera}\end{cases}$$

**(a) Hallar el estimador del parámetro $a$ por el método de máxima verosimilitud en muestras aleatorias simples de tamaño $n$.**

**(b) Comparar este estimador con el obtenido por el método de los momentos.**

---

(a) Para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned}L(x_1,x_2,\ldots,x_n;a) &= \prod_{i=1}^{n}f(x_i;a)=\\ &= \prod_{i=1}^{n}\frac{2a}{1-a}\cdot x_i^{\left(\frac{3a-1}{1-a}\right)}=\\ &= \frac{2^n\cdot a^n}{(1-a)^n}\cdot(x_1\cdot\ldots\cdot x_n)^{\frac{3a-1}{1-a}}.\end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; a) = n \ln 2 + n \ln a - n \ln(1-a) + \frac{3a-1}{1-a} \sum_{i=1}^{n} \ln x_i.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $a$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} + \frac{n}{1-a} + \frac{2}{(1-a)^2} \sum_{i=1}^{n} \ln x_i = 0.$$

Y despejando

$$\frac{2}{(1-a)^2} \cdot \sum_{i=1}^{n} \ln x_i = -\frac{n}{a(1-a)}$$

$$\sum_{i=1}^{n} \ln x_i = \frac{na-n}{2a} \quad \Rightarrow \quad n = -2a \cdot \sum_{i=1}^{n} \ln x_i + na.$$

Y el estimador de máxima verosimilitud para el parámetro $a$ es

$$\tilde{a} = \frac{n}{n - 2\sum_{i=1}^{n} \ln x_i}$$

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $a = \tilde{a}$ tenemos que tiene signo negativo.

(b) Para calcular el estimador del parámetro $a$ por el método de los momentos, igualamos los dos primeros momentos.

Si calculamos la media de la variable $X$ tenemos:

$$\begin{aligned}
\mu = E[X] &= \int_0^1 x \cdot f(x;a)\, dx = \\
&= \int_0^1 x \cdot \frac{2a}{1-a} \cdot x^{\left(\frac{3a-1}{1-a}\right)} = \\
&= \frac{2a}{1-a} \int_0^1 x^{\left(\frac{2a}{1-a}\right)} = \\
&= \frac{2a}{1-a} \left[ \frac{1-a}{a+1} \cdot x^{\frac{2a}{1-a}+1} \right]_0^1 = \\
&= \frac{2a}{1-a} \cdot \frac{1-a}{a+1} = \frac{2a}{a+1}
\end{aligned}$$

Y al igualar el momento de la distribución $\mu$ con el primer momento de la muestra $\bar{x}$ obtenemos

$$\bar{X} = \frac{2a^*}{a^*+1} \qquad \Rightarrow \qquad (1+a^*)\bar{X} = 2a^*$$

y el estimador del parámetro $a$ por el método de los momentos es

$$a^* = \frac{\bar{X}}{2-\bar{X}}.$$

Y como podemos comprobar no coinciden ambos estimadores $\tilde{a} \neq a^*$.

# Problema 5.11

**Estímese la esperanza matemática de la distribución $X$ que tiene por función de densidad**

$$f(x;a) = \begin{cases} \frac{4}{a^3\sqrt{\pi}} \cdot x^2 \cdot e^{-\left(\frac{x^2}{a^2}\right)} & \textbf{si} \quad x \geq 0 \quad , \quad a > 0 \\ 0 & \textbf{fuera} \end{cases}$$

**mediante la muestra aleatoria simple de tamaño 10**

| | | | | |
|---|---|---|---|---|
| **5,141** | **4,801** | **3,220** | **6,009** | **6,168** |
| **6,402** | **2,490** | **2,205** | **5,488** | **3,746** |

**Nota: Obtener el estimador de máxima verosimilitud del parámetro $a$.**

Para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; a) &= \prod_{i=1}^{n} f(x_i; a) = \\ &= \prod_{i=1}^{n} \frac{4}{a^3\sqrt{\pi}} \cdot x_i^2 \cdot e^{-\left(\frac{x_i^2}{a^2}\right)} = \\ &= \frac{4^n}{a^{3n} \cdot \pi^{n/2}} \prod_{i=1}^{n} x_i^2 \cdot e^{-\sum x_i^2/a^2} \end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; a) = n \ln 4 - 3n \ln a - \frac{n}{2} \ln \pi + 2\sum_{i=1}^{n} \ln x_i - \frac{1}{a^2}\sum_{i=1}^{n} x_i^2.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $a$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial a} = -\frac{3n}{a} + \frac{2}{a^3}\sum_{i=1}^{n} x_i^2 = 0.$$

Y despejando

$$\frac{2}{a^3}\sum_{i=1}^{n} x_i^2 = \frac{3n}{a} \quad \Rightarrow \quad a^2 = \frac{2}{3n}\sum_{i=1}^{n} x_i^2.$$

Y el estimador de máxima verosimilitud para el parámetro $a$ es

$$\tilde{a} = \sqrt{\frac{2}{3n}\sum_{i=1}^{n} x_i^2}.$$

El signo negativo de la raíz no lo tenemos en cuenta pues al ser el parámetro $a$ mayor que cero, su estimador también ha de serlo.

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $a = \tilde{a}$ tenemos que tiene signo negativo.

En nuestro caso como $n = 10$, tenemos que $\tilde{a} = 3{,}917$, y sustituyendo el valor estimado de $a$ en la función de densidad tenemos:

$$f(x) = 0{,}03755 \cdot x^2 \cdot e^{-0{,}06518x^2} \quad \text{si} \quad x > 0.$$

Si calculamos la media de la variable aleatoria $X$ obtenemos:

$$\begin{aligned}
E[X] &= \int_0^{+\infty} x \cdot \frac{4}{a^3\sqrt{\pi}} \cdot x^2 \cdot e^{-\left(\frac{x^2}{a^2}\right)}\, dx = \\
&= \frac{4}{\sqrt{\pi}} \cdot \int_0^{+\infty} \frac{x^3}{a^3} \cdot e^{-\left(\frac{x^2}{a^2}\right)}\, dx = \\
&= \left\{ \text{hacemos el cambio } \frac{x^2}{a^2} = t \quad ; \; dx = \frac{1}{2} a t^{-1/2}\, dt \right\} = \\
&= \frac{4}{\sqrt{\pi}} \cdot \int_o^{+\infty} \frac{1}{2} a t \sqrt{t} \cdot e^{-t} \cdot t^{-1/2}\, dt = \\
&= \frac{2a}{\sqrt{\pi}} \cdot \int_o^{+\infty} t \cdot e^{-t}\, dt = \frac{2a}{\sqrt{\pi}} \left[ -(t+1)e^{-t} \right]_0^{+\infty} = \\
&= \frac{2a}{\sqrt{\pi}} \cdot [0 + e^0] = \frac{2a}{\sqrt{\pi}}.
\end{aligned}$$

Y por tanto como la media de la distribución $X$ vale $E[X] = \dfrac{2a}{\sqrt{\pi}}$ obtenemos como estimación de la esperanza matemática

$$\widetilde{E[X]} = \frac{2\tilde{a}}{\sqrt{\pi}} = \frac{2 \cdot 3{,}917}{\sqrt{\pi}} = 4{,}42.$$

## Problema 5.12

**El tiempo de espera en cola de un supermercado es una variable aleatoria $X$ con función de densidad:**

$$f(x) = \begin{cases} \frac{1}{a^2} \cdot x \cdot e^{(-x^2/a)} & \textbf{si} \quad x \geq 0 \quad , \quad a > 0 \\ 0 & \textbf{en otro caso} \end{cases}$$

**Se necesita conocer el percentil 10 de la variable. La función de densidad depende de una parámetro $a$, cuya estimación se realiza por el método de la máxima verosimilitud. Para obtener una estimación del parámetro $a$ se ha tomado una muestra aleatoria de tamaño ocho**

**1,3 2,4 2,5 1,1 3,1 2,2 1,9 1,4**

**(a) Calcular el estimador máximo verosímil del parámetro $a$.**

**(b) Calcular la estimación del percentil 10 de la distribución.**

---

(a) Para una muestra aleatoria de tamaño $n$, la función de verosimilitud es:

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; a) &= \prod_{i=1}^{n} f(x_i; a) = \\ &= \prod_{i=1}^{n} \frac{1}{a^2} \cdot x_i \cdot e^{(-x_i^2/a)} = \\ &= \frac{1}{a^{2n}} \cdot \prod_{i=1}^{n} x_i \cdot e^{-\left(\sum_{i=1}^{n} x_i^2/a\right)}. \end{aligned}$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; a) = 2n \ln a + \sum_{i=1}^{n} \ln x_i - \frac{1}{a} \sum_{i=1}^{n} x_i^2.$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $a$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial a} = \frac{2n}{a} + \frac{1}{a^2}\sum_{i=1}^{n} x_i^2 = 0.$$

Y despejando

$$2na + \sum_{i=1}^{n} x_i^2 = 0$$

Y el estimador de máxima verosimilitud para el parámetro $a$ es

$$\tilde{a} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2.$$

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $a = \tilde{a}$ tenemos:

$$\frac{\partial^2 \ln L}{\partial a^2} = \frac{-2n}{a^2} - \frac{2}{a^3}\sum_{i=1}^{n} x_i^2$$

$$\left.\frac{\partial^2 \ln L}{\partial a^2}\right|_{\tilde{a}} < 0 \quad \forall n \in N.$$

Luego el estimador máximo verosímil de $a$ es $\tilde{a} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2$.

(b) Si utilizamos el estimador del parámetro $a$ del apartado anterior, tenemos:

$$n = 8 \qquad \Longrightarrow \qquad \tilde{a} = \frac{1{,}3^2 + 2{,}4^2 + \ldots + 1{,}4^2}{16} = \frac{34{,}93}{16} = 2{,}183125.$$

Por lo que la función de densidad de $X$ queda

$$f(x;\tilde{a}) = \begin{cases} \dfrac{1}{4{,}766} \cdot x \cdot e^{(-x^2/2{,}183)} & \text{si} \quad x \geq 0 \\ 0 & \text{fuera} \end{cases}$$

Tenemos que encontrar el valor del percentil 10 estimado por esta muestra tal que

$$\int_0^{P_{10}} f(x;\tilde{a})\,dx = \frac{1}{10} \quad , \quad \int_{P_{10}}^{+\infty} f(x;\tilde{a})\,dx = \frac{9}{10}.$$

$$\begin{aligned}
\int_0^{P_{10}} f(x;\tilde{a})\,dx &= \int_0^{P_{10}} \frac{1}{\tilde{a}^2}\cdot x\cdot e^{(-x^2/\tilde{a})}\,dx = \\
&= \left\{ \begin{array}{l} \dfrac{x^2}{\tilde{a}} = t \\ \dfrac{2x}{\tilde{a}}\,dx = dt \end{array} \right\} = \\
&= \frac{1}{2\tilde{a}}\int_0^{P_{10}} e^{-t}\,dt = \frac{1}{2\tilde{a}}\left[-e^{-t}\right]_0^{P_{10}} = \\
&= \frac{1}{2\tilde{a}}\cdot\left[1-e^{-P_{10}}\right]
\end{aligned}$$

Y como esta integral tiene que valer $1/10$, tenemos

$$\frac{1}{2\tilde{a}}\cdot\left[1-e^{-P_{10}}\right] = \frac{1}{10} \qquad \longrightarrow \qquad \frac{\tilde{a}}{5} = 1-e^{-P_{10}}$$

$$e^{-P_{10}} = 1-\frac{\tilde{a}}{5} \qquad \longrightarrow \qquad -P_{10} = \ln\left[1-\frac{\tilde{a}}{5}\right]$$

$$P_{10} = -\ln\left[1-\frac{2{,}183125}{5}\right] \qquad \longrightarrow \qquad P_{10} = -\ln[0{,}563375]$$

Y la estimación del percentil 10 vale

$$\widetilde{P}_{10} = 0{,}5738.$$

# Problema 5.13

**Calcular un estimador para el parámetro $p$ de una distribución geométrica por el método de los momentos y por el método de la máxima verosimilitud.**

---

Por el método de los momentos se obtiene que para una muestra el primer momento muestral es la media $\bar{X}$, que habrá que igualarlo con la media poblacional, que vale $E[X] = \dfrac{q}{p}$. Si igualamos ambos momentos tendremos:

$$\frac{1-p}{p} = \bar{X} \qquad \rightarrow 1-p = p\cdot\bar{X} \qquad p(\bar{X}+1) = 1$$

Y por tanto el estimador de $p$ por el método de los momentos es

$$\tilde{p} = \frac{1}{1+\bar{X}}$$

Y por el método de la máxima verosimilitud se obtiene como estimador del parámetro $p$ el mismo que se obtiene por el método de los momentos pues si construimos la función de verosimilitud

$$L(x_1, x_2, \ldots, x_n; p) = p^n (1-p)^{\sum_{i=1}^{n} x_i} =$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; p) = n \ln p + \sum_{i=1}^{n} x_i \cdot \ln(1-p).$$

Buscamos el máximo de la función de verosimilitud. Derivamos una vez con respecto al parámetro $p$ e igualamos a cero, y comprobamos que para ese valor, la segunda derivada es negativa.

$$\frac{\partial \ln L}{\partial p} = \frac{n}{p} - \frac{1}{1-p} \sum_{i=1}^{n} x_i = 0.$$

Y despejando

$$\frac{n}{p} = \frac{1}{1-p} \sum_{i=1}^{n} x_i \quad \Rightarrow \quad n - np = p \cdot \sum_{i=1}^{n} x_i$$

$$n = p \cdot \sum_{i=1}^{n} x_i + np \quad \Rightarrow \quad p(\sum_{i=1}^{n} x_i + n) = n$$

dividiendo entre $n$:

$$p(\frac{1}{n} \sum_{i=1}^{n} x_i + 1) = 1$$

Y el estimador de máxima verosimilitud para el parámetro $p$ es

$$\hat{p} = \frac{1}{\bar{X}+1}.$$

Comprobamos que efectivamente es un máximo calculando la segunda derivada y evaluando en $p = \hat{p}$. Luego coinciden los dos estimadores obtenidos.

# Problema 5.14

**Se considera una población representada por la variable $X$, con función de densidad definida por**

$$f(x;a) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & \textbf{si} \quad 0 \le x \le \theta \\ 0 & \textbf{fuera} \end{cases}$$

**(a) Hallar el estimador del parámetro $\theta$ por el método de los momentos.**

**(b) ¿Es insesgado?**

---

(a)Para calcular el estimador del parámetro $\theta$ por el método de los momentos debemos igualar el primer momento de la variable $X$, (la media $\mu$), con el primer momento de la muestra (media muestral $\bar{X}$).

Para lo que calculamos la media de la variable $X$, que vale

$$\begin{aligned} E[X] &= \int_0^\theta x \cdot \frac{2(\theta - x)}{\theta^2}\, dx = \frac{2}{\theta^2}\left[\frac{\theta x^2}{2} - \frac{x^3}{3}\right]_0^\theta = \\ &= \frac{2}{\theta^2}\left[\frac{\theta^3}{2} - \frac{\theta^3}{3}\right] = 2\theta\left[\frac{1}{2} - \frac{1}{3}\right] = 2\theta \cdot \frac{1}{6} = \frac{\theta}{3} \end{aligned}$$

Y el estimador del parámetro $\theta$ por el método de los momentos vale

$$\frac{\tilde{\theta}}{3} = \bar{X} \qquad \tilde{\theta} = 3 \cdot \bar{X}.$$

(b) Para comprobar si el estimador del parámetro $\theta$ obtenido por el método de los momentos es centrado o insesgado, debemos comprobar si se verifica

$$E[\tilde{\theta}] = \theta.$$

Y efectivamente, si calculamos la media del estimador obtenemos:

$$\begin{aligned} E[\tilde{\theta}] &= E[3 \cdot \bar{X}] = 3 \cdot E[\bar{X}] = 3 \cdot E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{3}{n} E\left[\sum_{i=1}^{n} X_i\right] = \\ &= \frac{3}{n} \cdot n \cdot E[X_i] = 3 \cdot E[X_i] = 3 \cdot \frac{\theta}{3} = \theta. \end{aligned}$$

Y por tanto, como el valor esperado del estimador coincide con el valor del parámetro $\theta$, el estimador obtenido es centrado.

# Problema 5.15

**Sea una población definida por la función de densidad**

$$f(x;\theta) = \frac{1}{\theta} \cdot x^{-\left(\frac{1}{\theta}+1\right)} \qquad x \geq 1 \quad , \quad \theta > 0.$$

**(a) Obtener el estimador de máxima verosimilitud del parámetro $\theta$.**

**(b) Estudiar las propiedades de consistencia e insesgadez del estimador.**

---

(a) Para calcular el estimador por el método de la máxima verosimilitud, hallamos la función de verosimilitud:

$$L(x_1, x_2, \ldots, x_n; \theta) \quad = \quad \prod_{i=1}^{n} f(x_i; \lambda) = \frac{1}{\theta^n} \cdot (x_1 \cdot \ldots \cdot x_n)^{-\left(\frac{1}{\theta}+1\right)}.$$

Y tomando logaritmos y efectuando las operaciones obtenemos

$$\ln L(x_1, x_2, \ldots, x_n; \theta) = -n \cdot \ln\theta - \left(\frac{1}{\theta} + 1\right) \cdot \sum_{i=1}^{n} \ln x_i.$$

Derivando con respecto a $\theta$ e igualando a cero tenemos

$$\frac{\partial \ln L}{\partial \theta} = \frac{-n}{\theta} + \frac{1}{\theta^2} \cdot \sum_{i=1}^{n} \ln x_i = 0.$$

Y despejando

$$-n \cdot \theta + \sum_{i=1}^{n} \ln x_i = 0.$$

Y el estimador de máxima verosimilitud para $\theta$ es

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} \ln x_i.$$

Efectivamente es un máximo, pues la segunda derivada evaluada para $\theta = \tilde{\theta}$ tiene signo negativo.

(b) Si estudiamos la insesgadez y consistencia del estimador tenemos:

$$\begin{aligned}
E[\tilde{\theta}] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\ln x_i\right] = \frac{1}{n}\sum_{i=1}^{n}E[\ln x_i] = \\
&= E[\ln X] = \frac{1}{\theta}\int_1^{+\infty}\ln x \cdot x^{-\left(\frac{1}{\theta}+1\right)}\,dx = \\
&= \left\{u = \ln x; \quad du = \frac{1}{x}dx; \quad dv = x^{-\left(\frac{1}{\theta}+1\right)}dx; \quad v = \frac{x^{-1/\theta}}{-1/\theta}\right\} = \\
&= 0 + \frac{1}{\theta}\left[\theta\int_1^{+\infty}\frac{1}{x}x^{-1/\theta}\,dx\right] = \int_1^{+\infty}x^{-\left(\frac{1}{\theta}-1\right)}\,dx = \\
&= \left[\frac{x^{-1/\theta}}{-1/\theta}\right]_1^{+\infty} = \theta.
\end{aligned}$$

Luego el estimador obtenido por el método de los momentos es insesgado.

Si queremos estudiar la consistencia tendremos que ver que

$$\text{Si } n \to \infty \quad \Rightarrow \quad Var[\tilde{\theta}] \to 0$$

Si calculamos la varianza del estimador tenemos que

$$\begin{aligned}
Var[\tilde{\theta}] &= \frac{1}{n^2}Var\left[\sum_{i=1}^{n}\ln x_i\right] = \frac{1}{n}Var[\ln X] = \\
&= \frac{1}{n}\left\{E[(\ln X)^2] - (E[\ln X])^2\right\} = \frac{1}{n}[2\theta^2 - \theta^2] = \\
&= \frac{\theta^2}{n}.
\end{aligned}$$

Y por tanto si $n \to \infty$ entonces $Var[\tilde{\theta}] \to 0$ y el estimador es consistente.

Nota: (Puede comprobarse que $E[(\ln X)^2] = 2\theta^2$. Para ello habría que hacer una integral por partes análoga a la anterior).

# Problema 5.16

**Calcular las probabilidades siguientes, que corresponden a la distribución $\chi^2$ de Pearson con los grados de libertad indicados entre**

**paréntesis:**

$$P[\chi^2(10) > 7,267] \qquad \text{y} \qquad P[\chi^2(21) > 30]$$

$$\chi^2_{0,85}(23) \qquad \text{y} \qquad P[\chi^2(61) > 24,5]$$

---

Los valores $\chi^2_{10}$ y $\chi^2_{21}$ representan aquí las variables aleatorias Chi-cuadrado de Pearson con grados de libertad respectivamente 10 y 21. La disposición de la tabla de percentiles de la $\chi^2$ se debe a que habitualmente se usa para determinar abscisas asociadas con probabilidades dadas. Así, la abscisa $\chi^2_{0,95}(15)$, es decir, el punto de la distribución Chi-cuadrado con 15 grados de libertad deja un área a la izquierda de 0,95 (o equivalentemente, una cola a la derecha de 0,05) vale

$$\chi^2_{0,95}(15) = 24{,}9958$$

Así, directamente de las tablas y por interpolación se obtienen las probabilidades que nos han pedido

$$P[\chi^2_{10} > 7{,}267] = 0{,}7 \qquad \text{y} \qquad P[\chi^2_{21} > 30] = 0{,}091988$$

Al igual que antes, por interpolación se pueden calcular abscisas asociadas a probabilidades intermedias. Por ejemplo

$$\chi^2_{0,85}(23) = 30$$

que significa que $P[\chi^2_{23} < 30] = 0{,}85$.

Obsérvese que en la tabla sólo aparecen valores de la $\chi^2_n$ hasta $n = 50$ grados de libertad. La razón es que para valores mayores la distribución se aproxima a una normal; concretamente,

$$\sqrt{2\chi^2_n} \;\approx\; N(\sqrt{2n-1}, 1)$$

Esta aproximación permite calcular probabilidades (y buscar abscisas para grados de libertad mayores que 40. Por ejemplo:

$$P[\chi^2_{61} > 24,5] = P\left\{\sqrt{2\chi^2_{61}} \;>\; \sqrt{2\cdot 24,5}\right\} = P\left\{\sqrt{2\chi^2_{61}} \;>\; 7\right\}$$

usando la aproximación anterior y tipificando, tendremos que la probabilidad pedida será aproximadamente igual a

$$P[Z > 7 - \sqrt{2 \cdot 61 - 1}] = P[Z > 7 - 11] = P[Z > -4] = P[Z < 4]$$

donde $Z \sim N(0,1)$ y $P[Z < 4]$ se obtiene de la tabla de la distribución normal. Podemos, por tanto, concluir que

$$P[\chi^2_{61} > 24,5] \approx P[Z < 4] = 0{,}99997.$$

El valor obtenido haciendo uso del programa Statgraphics es 0,99999189. Como se puede ver, la aproximación es bastante buena.

# Problema 5.17

**De una población normal $N(\mu;\sigma)$ con $\sigma$ conocida, se obtiene una muestra de tamaño 4. Se consideran los siguientes estimadores del parámetro poblacional $\mu$.**

$$\begin{aligned}
\tilde{\mu_1} &= \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}x_3 \\
\tilde{\mu_2} &= \frac{1}{3}x_1 + \frac{1}{6}x_2 + \frac{1}{3}x_3 + \frac{1}{6}x_4 \\
\tilde{\mu_3} &= \frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4
\end{aligned}$$

**(a) Comprobar si los estimadores dados son insesgados o no lo son. En caso de no ser insesgados, calcular su sesgo.**

**(b) Calcular la varianza de los estimadores, e indicar cuál es el que tiene menor varianza.**

---

(a) Si estudiamos la insesgadez del primer estimador obtenemos:

$$\begin{aligned}
E[\tilde{\mu_1}] &= E\left[\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}x_3\right] = \\
&= \frac{1}{2}E[x_1] + \frac{1}{2}E[x_2] + \frac{1}{4}E[x_3] = \frac{1}{2}\mu + \frac{1}{2}\mu + \frac{1}{4}\mu = \\
&= \mu + \frac{\mu}{4}.
\end{aligned}$$

Luego el estimador $\tilde{\mu_1}$ es sesgado, con sesgo $\frac{\mu}{4}$.

Ahora estudiamos la insesgadez del segundo estimador, y tenemos:

$$\begin{aligned}
E[\tilde{\mu_2}] &= E\left[\frac{1}{3}x_1 + \frac{1}{6}x_2 + \frac{1}{3}x_3 + \frac{1}{6}x_4\right] = \\
&= \frac{1}{3}E[x_1] + \frac{1}{6}E[x_2] + \frac{1}{3}E[x_3] + \frac{1}{6}E[x_4] = \frac{1}{3}\mu + \frac{1}{6}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu = \\
&= \mu.
\end{aligned}$$

Luego el estimador $\tilde{\mu_2}$ es insesgado.

Para estudiar la insesgadez del estimador $\tilde{\mu_3}$ calculamos:

$$\begin{aligned}
E[\tilde{\mu_3}] &= E\left[\frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4\right] = \\
&= \frac{1}{4}E[x_1] + \frac{1}{4}E[x_2] + \frac{1}{4}E[x_3] + \frac{1}{4}E[x_4] = \frac{1}{4}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu = \\
&= \mu.
\end{aligned}$$

Luego el estimador $\tilde{\mu_3}$ es insesgado. Este estimador es precisamente la media muestral.

(b) Si calculamos la varianza de los tres estimadores tenemos:

$$\begin{aligned}
\text{Var}[\tilde{\mu_1}] &= \text{Var}\left[\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}x_3\right] = \\
&= \left(\frac{1}{2}\right)^2 \text{Var}[x_1] + \left(\frac{1}{2}\right)^2 \text{Var}[x_2] + \left(\frac{1}{4}\right)^2 \text{Var}[x_3] = \\
&= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{16}\sigma^2 = \\
&= \frac{9}{16}\sigma^2,
\end{aligned}$$

ya que la varianza poblacional es $\sigma^2$.

La varianza del estimador $\tilde{\mu_2}$ es

$$\begin{aligned}
\text{Var}[\tilde{\mu_2}] &= \text{Var}\left[\frac{1}{3}x_1 + \frac{1}{6}x_2 + \frac{1}{3}x_3 + \frac{1}{6}x_4\right] = \\
&= \left(\frac{1}{3}\right)^2 \text{Var}[x_1] + \left(\frac{1}{6}\right)^2 \text{Var}[x_2] + \left(\frac{1}{3}\right)^2 \text{Var}[x_3] + \left(\frac{1}{6}\right)^2 \text{Var}[x_4]
\end{aligned}$$

tenemos por tanto:

$$\text{Var}[\tilde{\mu_2}] = \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 = \frac{10}{36}\sigma^2 = \frac{5}{18}\sigma^2.$$

Análogamente, la varianza del estimador $\tilde{\mu_3}$ vale

$$\begin{aligned}
\mathrm{Var}[\tilde{\mu_3}] &= \mathrm{Var}\left[\frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4\right] = \\
&= \left(\frac{1}{4}\right)^2 \mathrm{Var}[x_1] + \left(\frac{1}{4}\right)^2 \mathrm{Var}[x_2] + \left(\frac{1}{4}\right)^2 \mathrm{Var}[x_3] + \left(\frac{1}{4}\right)^2 \mathrm{Var}[x_4]
\end{aligned}$$

de donde obtenenos

$$\mathrm{Var}[\tilde{\mu_3}] = \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{1}{16}\sigma^2 = \frac{1}{4}\sigma^2.$$

Se verifica

$$\frac{1}{4}\sigma^2 < \frac{5}{18}\sigma^2 < \frac{9}{16}\sigma^2,$$

y por tanto el estimador que tiene menor varianza es la media muestral $\tilde{\mu_3}$, que además es centrado o insesgado.

# Problemas Propuestos

1. Dar ejemplos de estimadores que sean (a) insesgados y eficientes, (b) insesgados y no eficientes, (c) sesgados y no eficientes.

2. Una muestra de cinco medidas del diámetro de una esfera proporcionó los siguientes valores en centímetros:

$$6{,}33 \quad 6{,}37 \quad 6{,}36 \quad 6{,}32 \quad 6{,}37$$

   Determinar estimaciones insesgadas y eficientes de (a) la verdadera media, (b) la verdadera varianza.

   **Nota:** Vease que $S = \sqrt{0{,}00055} = 0{,}023$ es una estimación de la verdadera desviación típica pero que esta estimación no es insesgada ni eficiente.

3. Supóngase que los pesos de 100 estudiantes de una Escuela de la Universidad representan una muestra al azar de los pesos de los 1546 estudiantes de la Escuela. Determinar estimadores insesgados y eficientes de (a) la verdadera media, (b) la verdadera varianza.

| Peso (en Kg) | frecuencia |
|---|---|
| 61 | 5 |
| 64 | 18 |
| 67 | 42 |
| 70 | 27 |
| 73 | 8 |

4. Dar una estimación insesgada y no eficiente del verdadero diámetro de la esfera del problema 2.

5. Las medidas de pesos de una muestra de jamones, en kilos, fueron los siguientes:

   8,3 10,6 9,7 8,8 10,2 9,4

   Determinar estimadores insesgados y eficientes de (a) la media de la población y (b) la varianza de la población. (c) Comparar la desviación típica muestral con la desviación típica de la población estimada.

6. De acuerdo con la teoría genética, en una población amplia los tipos de sangre MM, NM y NN se deberían encontrar con frecuencias relativas $\theta^2, 2\theta(1-\theta)$ y $(1-\theta)^2$, donde $\theta$ es la frecuencia desconocida de un gen.

   (a) Supongamos que en una muestra aleatoria de tamaño $n$ de la población encontramos $x_1, x_2, x_3$ de cada uno de los tres tipos. Hallar una expresión de $\widehat{\theta}$.

   (b) Si en una muestra de tamaño 100, las frecuencias observadas fueron 32, 46 y 22. Hallar $\widehat{\theta}$ y las frecuencias esperadas de los tres tipos de sangre proporcionadas por el modelo.

7. Una muestra de 10 tubos de televisión producidos por una compañía electrónica, dieron una duración media de 1200 horas y una desviación típica $S = 100$ horas. Estimar (a) la media y (b) la desviación típica de la población de todos los tubos de televisión producidos por esa compañía.

8. (a) Hacer el problema anterior si se obtienen los mismos resultados para muestras de 30, 50 y 100 tubos de televisión. (b) ¿Qué se puede deducir de la relación entre las desviaciones típicas muestrales y las estimaciones de las desviaciones típicas de la población para los diferentes tamaños muestrales?

9. Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población distribuida normalmente de la cual se desconoce la media y se conoce la varianza. Hallar el estimador de máxima verosimilitud de la media.

10. Si en el problema anterior se conoce la media pero se desconoce la varianza, hallar el estimador máximo verosímil de la varianza.

11. Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población distribuida según una Poisson con parámetro $\lambda$ desconocido. Hallar el estimador de máxima verosimilitud para $\lambda$.

12. Si se toma una muestra aleatoria de tamaño $n$ de una población que tiene una función de densidad dada por:

$$f(x;k) = \begin{cases} (k+1)\cdot x^k & \text{si } \; 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Calcular el estimador de máxima verosimilitud del parámetro $k$.

13. Una población tiene una función de densidad dada por $f(x;\alpha)$. Si se toman $n$ observaciones $x_1, x_2, \cdots, x_n$ de esta población, hallar el estimador de máxima verosimilitud de $\alpha$.

$$f(x;\alpha) = \; 2\alpha\sqrt{\alpha/\pi}x^2e^{-\alpha x^2} \quad \text{si} \quad -\infty < x < \infty$$

14. Demuéstrese que $\bar{X}$ es un estadístico suficiente e insesgado para la media de una distribución de Poisson con parámetro $\lambda$.

15. Sean $X_1, X_2, \cdots, X_n$ variables independientes distribuidas uniformemente en el intervalo $(-\theta, \theta)$. Obténgase un estadístico suficiente para el parámetro $\theta$.

16. Cierto componente electrónico puede fallar instantáneamente en cualquier momento. Sin embargo, los componentes no se deterioran con la edad, y por tanto la posibilidad de que se averíe en un periodo determinado no depende de su edad. El tiempo de vida de un componente de este tipo sigue una distribución exponencial, con función de densidad

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & \text{si } \; x > 0 \\ 0 & \text{fuera} \end{cases}$$

donde $\theta$ es el tiempo de vida esperado del componente. Se observan diez componentes independientemente. Sus tiempos de vida, ajustados a dias enteros, son

70  11  66  5  20  4  35  40  29  8.

¿Qué valor de $\theta$ es viable a la luz de los datos?

17. Un proceso de fabricación produce fibras de distintas longitudes. La longitud de una fibra es una variable continua con función de densidad

$$f(x;\theta) = \begin{cases} \theta^{-2}xe^{-x/\theta} & \text{si } \; x > 0 \\ 0 & \text{fuera} \end{cases}$$

donde $\theta > 0$ es un parámetro desconocido. Supongamos que se seleccionan $n$ fibras aleatoriamente y tienen longitudas $x_1, x_2, \cdots, x_n$. Hallar el EMV de $\theta$.

18. Un proceso de manufactura produce fibras de distintas longitudes. Se supone que la longitud X de una fibra es una variable continua con función de densidad

$$f(x;\lambda) = \begin{cases} 2\lambda x^{-2} e^{-\lambda x^2} & \text{si} \quad x > 0 \\ 0 & \text{fuera} \end{cases}$$

donde $\lambda > 0$. Supongamos que se escogen al azar $n$ fibras y se miden sus longitudes, $x_1, x_2, \cdots, x_n$. Demostrar que el estadístico $T = \sum x_i^2$ es un estadístico suficiente para $\lambda$.

19. Se sabe que un programa infectado por un virus informático a veces borra los ficheros del sistema. Si el virus tiene una probabilidad constante de realizar el borrado, independientemente del número de veces que se haya ejecutado antes, el número de ejecuciones que se necesitan para borrar los ficheros sigue una distribución geométrica:

$$p(x;\theta) = \theta^{x-1}(1-\theta) \quad \text{para} \quad x = 1, 2, 3, \ldots$$

El resultado de ejecutar el programa en 200 ordenadores de prueba fue el siguiente:

| Número de ejecuciones necesitadas | 1 | 2 | 3 | $\geq 4$ | Total |
|---|---|---|---|---|---|
| Número de ordenadores | 112 | 36 | 22 | 30 | 200 |

Hallar el estimador de máxima verosimilitud de $\theta$ y calcular la frecuencias esperadas.

20. La función de densidad de una distribución exponencial con periodo de garantia $c$ es

$$f(x;\lambda,c) = \begin{cases} \lambda e^{-\lambda(x-c)} & \text{si} \quad x \geq c \\ 0 & \text{fuera} \end{cases}$$

Hallar el EMV de $\lambda$ y $c$ basadas en observaciones independientes, $x_1, x_2, \cdots, x_n$, de esta distribución.

21. Se supone que el consumo por minuto de ciertos equipos sigue una distribución normal con una desviación típica de 12,7 w. Si se elijen 10 equipos al azar, calcular la probabilidad de que la media de esa muestra difiera de la poblacional en más de 4,4 w.

22. Si en el ejemplo anterior no se hubiera conocido la varianza poblacional y la muestra tuviera como desviación típica muestral $S = 12$ w, calcular la probabilidad de que la media muestral difiera de la poblacional en más de 4,4 w.

23. Con objeto de estimar el tiempo medio de ejecución de un algoritmo frente a un problema con diversos datos, se realizaron 100 pruebas, de las que se obtuvo una desviación típica muestral de 15 segundos. Calcular la probabilidad de que la media muestral difiera de la media poblacional en más de 3 segundos.

24. Se quiere estimar la proporción de circuitos impresos defectuosos producidos por cierto procedimiento. Una vez probados 150 de ellos, 25 fueron defectuosos. Calcular el estimador de máxima verosimilitud de $p$. ¿Cuál es la probabilidad de haber subestimado la proporción de defectuosos en más de un 5%?

25. En un recuento de errores cometidos por una empresa de grabación de datos, se han examinado 10 registros elegidos al azar. Si se supone que el número de errores por registro tiene distribución normal. Calcular la probabilidad de que la cuasivarianza muestral sobreestime a la varianza poblacional en más de un tercio de su valor.

26. Se supone que las cotizaciones de un cierto valor bursátil en dos bolsas del país tienen distribución normal. Se observaron las cotizaciones durante $n_1 = 10$ y $n_2 = 12$ días elegidos al azar, en cada una de ellas respectivamente, y se obtuvieron como cuasivarianzas muestrales los valores $S_1^2 = 955$ y $S_2^2 = 415{,}2$. Si las varianzas poblacionales fuesen en realidad iguales, ¿qué probabilidad habría de haber observado un desequilibrio entre las cuasivarianzas superior al obtenido?

27. Con objeto de estudiar si existe diferencia en la duración de dos componentes electrónicos antes de fallar, se tomó una muestra aleatoria simple de cada una de ellos. Los resultados en horas de funcionamiento, fueron los siguientes:

$$\begin{array}{lll} n_1 = 65 \quad , & \bar{X}_1 = 75 \quad , & S_1^2 = 225 \\ n_2 = 35 \quad , & \bar{X}_2 = 79 \quad , & S_2^2 = 195 \end{array}$$

Calcular la probabilidad de que la diferencia de medias muestrales estime la diferencia de medias poblacionales con un error mayor de 2 horas.

28. Se trata de averiguar si existe diferencia entre los rendimientos de dos compiladores de cierto lenguaje.

    Con este propósito se compilaron $n_1 = 10$ programas de características estándar, elegidos al azar, con el primer compilador y otros $n_2 = 10$, de características similares, con el segundo. Las cuasivarianzas muestrales de los tiempos de compilación fueron $S_1^2 = 45$ y $S_2^2 = 43$ segundos al cuadrado, respectivamente.

    Se supone que el tiempo de compilación de un programa elegido al azar se distribuye en ambos casos con distribución normal. Calcular la probabilidad de que la diferencia de las medias muestrales estime la diferencia de medias poblacionales con un error mayor de 2 segundos.

29. Se diseña un cierto procesador con objeto de que haga, en media, más de 10 millones de operaciones aritméticas por segundo. El número de operaciones realizadas en un segundo se supone que tiene distribución normal con desviación típica 0,8 millones. ¿Durante cuántos segundos debe probarse el procesador para que el número medio de operaciones aritméticas por segundo observadas difiera del valor medio real en menos de 0,3 millones de operaciones con probabilidad 0,95?

30. X toma los valores 1 ó 0 con probabilidades $p$ y $1-p$ respectivamente. Si se observa una muestra aleatoria de tamaño $n$, calcular el estimador de máxima verosimilitud de $p$.

31. Se preparó un programa de ordenador para elegir números al azar entre 0 y $\theta$, pero se ha perdido la información sobre el valor de $\theta$. Su ejecución en 10 ocasiones dió los siguientes resultados:

    $$2{,}3 \quad 7{,}4 \quad 0{,}2 \quad 1{,}2 \quad 3{,}4 \quad 4{,}6 \quad \sqrt{2} \quad 5{,}7 \quad 6 \quad 5$$

    Determinar una estimación de $\theta$.

32. Con objeto de comparar el rendimiento de los alumnos de dos escuelas, se sometió a 10 de cada una de ellas a un determinado test. Las puntuaciones obtenidas dieron cuasivarianzas $S_1^2 = 9$ y $S_2^2 = 8$ respectivamente.

    Suponiendo que, en ambos casos, las puntuaciones tienen distribución normal con varianzas distintas, calcular la probabilidad de que la diferencia de medias muestrales estime la diferencia de medias poblacionales con un error mayor de 4 puntos.

33. Para probar la eficiencia de un coprocesador se ha medido el ahorro de tiempo en la ejecución de 30 programas, antes y después de instalarlo en un equipo. Si se admite que el ahorro sigue una distribución normal, determinar la probabilidad de que la cuasivarianza muestral subestime la varianza poblacional con un error superior a un 10% de ésta.

# Unidad Temática VI

# Intervalos de Confianza y Contraste de Hipótesis.

# Resumen Teórico

## 6.1 Tipos de estimación

Distinguiremos dos tipos de estimación:

**puntual.-** En la estimación puntual damos un solo punto como valor estimado del parámetro. Este punto será el valor que tome el estimador en la muestra seleccionada.

**por intervalos.-** En la estimación por intervalos no daremos un solo punto, sino un intervalo, pudiendo afirmar que el valor real del parámetro pertenece a dicho intervalo con una determinada probabilidad.

En el tema anterior hemos estudiado la estimación puntual y en este nos dedicaremos a ver la estimación por intervalos. La estimación puntual es más simple, pero también más inexacta. La estimación por intervalos presenta la ventaja de que es posible cuantificar los errores. El intervalo en el que se afirma que se encuentra el parámetro se denomina *intervalo de confianza*. La probabilidad de que el parámetro pertenezca a dicho intervalo se denomina *grado de confianza* y se suele representar como $1-\alpha$.

Para determinar el intervalo de confianza, utilizaremos los estadísticos ya vistos en el tema anterior de los que conocemos su distribución, y a continuación determinaremos una región que contenga al estadístico con probabilidad $1-\alpha$, de modo que deje a cada lado una región con probabilidad $\alpha/2$.

Veamos a continuación un resumen de los estadísticos descritos en el tema anterior:

## 6.2 Resumen de estimadores

Para estimar la media de una población normal, con varianza conocida, usaremos el estimador:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Para estimar la media de una población normal, con varianza desconocida, usaremos el estimador:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

que se distribuirá según una t de Student con n-1 grados de libertad. En el caso de que n sea mayor que 30, podemos usar la aproximación normal.

Para estimar la diferencia de medias de dos poblaciones normales, con varianzas $\sigma_1$ y $\sigma_2$ conocidas, usaremos el estimador:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Para estimar la diferencia de medias de dos poblaciones normales, con varianzas $\sigma_1$ y $\sigma_2$ desconocidas pero iguales, usaremos el estimador:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

que se distribuye según una t de Student con $n_1 + n_2 - 2$ grados de libertad.

Siendo

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Para estimar la diferencia de medias de dos poblaciones normales, con varianzas $\sigma_1$ y $\sigma_2$ desconocidas y distintas, usaremos el estimador:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

que se distribuye según una t de Student con $\nu$ grados de libertad. siendo

$$\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Para estimar la varianza de una población, usaremos el estadístico

$$(n-1)S^2/\sigma^2$$

que se distribuye según una chi-cuadrado con $n-1$ grados de libertad.

Para estimar el cociente entre las varianzas de dos poblaciones, usaremos el estadístico

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

## 6.3 Intervalos de Confianza para la media en poblaciones normales

### 6.3.1 Con varianza conocida

Utilizaremos el estadístico

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Si queremos obtener un intervalo que tenga una probabilidad $1-\alpha$ de contener al parámetro $\mu$, esto equivaldrá a

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$$

Al ser $Z \sim N(0,1)$, el valor de $z_{1-\alpha/2}$ se busca en las tablas considerando que

$$P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$$

Sustituyendo $Z$ se obtiene

$$P\left(-z_{1-\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha$$

y despejando

$$P(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}) = 1 - \alpha$$

de donde el intervalo de confianza para $\mu$ será

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right)$$

### 6.3.2 Con varianza desconocida

Al ser la desviación típica desconocida no podemos utilizar el estadístico anterior. Usaremos en su lugar el siguiente

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

que se distribuirá según una t de Student con n-1 grados de libertad.

Si queremos obtener un intervalo que tenga una probabilidad $1 - \alpha$ de contener al parámetro $\mu$, esto equivaldrá a

$$P(-t_{1-\alpha/2} < T < t_{1-\alpha/2}) = 1 - \alpha$$

Al ser $T$ una t con n-1 grados de libertad, el valor de $t_{1-\alpha/2}$ se busca en las tablas considerando que

$$P(T < t_{1-\alpha/2}) = 1 - \alpha/2$$

Sustituyendo $T$ se obtiene

$$P\left(-t_{1-\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha/2}\right) = 1 - \alpha$$

y despejando

$$P\left(\overline{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2} < \mu < \overline{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}\right) = 1 - \alpha$$

de donde el intervalo de confianza para $\mu$ será

$$\left(\overline{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}, \overline{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2}\right)$$

## 6.4 Intervalos de Confianza para la diferencia de medias en poblaciones normales

Estos estimadores se calculan de modo similar a los anteriores, utilizándose los estadísticos correspondientes. En el caso de que las varianzas sean conocidas o los tamaños muestrales sean grandes se utilizará un estadístico normal. En el caso de que las varianzas sean desconocidas y los tamaños muestrales sean pequeños se utilizará un estadístico t de Student, variando el cálculo de los grados de libertad según sean las varianzas de las dos poblaciones iguales o distintas.

## 6.5 Intervalos de Confianza para la varianza en poblaciones normales

En este caso se utilizará el estadístico

$$(n-1)S^2/\sigma^2$$

que se distribuye según una chi-cuadrado con $n-1$ grados de libertad. Habrá que buscar en las tablas de la chi-cuadrado de esos grados de libertad el punto $\chi^2_{\alpha/2}$ que deje a su izquierda un área de $\alpha/2$ y el punto $\chi^2_{1-\alpha/2}$ que deje a su derecha un área de $\alpha/2$. Tendremos por tanto que

$$P\left(\chi^2_{\alpha/2} < (n-1)S^2/\sigma^2 < \chi^2_{1-\alpha/2}\right) = 1-\alpha$$

y despejando:

$$P\left((n-1)S^2/\chi^2_{1-\alpha/2} < \sigma^2 < (n-1)S^2/\chi^2_{\alpha/2}\right) = 1-\alpha$$

de donde el intervalo de confianza para $\sigma^2$ será

$$\left((n-1)S^2/\chi^2_{1-\alpha/2}, (n-1)S^2/\chi^2_{\alpha/2}\right)$$

Obsérvese que en este caso el valor del estimador $S^2$ está contenido en el intervalo pero no es el punto medio del mismo.

## 6.6 Intervalos de Confianza para el cociente entre varianzas en poblaciones normales

Para estimar el cociente entre las varianzas de dos poblaciones, usaremos el estadístico

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

Buscando en las tablas de la F de Snedecor de los grados de libertad correspondientes el punto $f_{\alpha/2}$ que deje a su izquierda un área de $\alpha/2$ y el punto $f_{1-\alpha/2}$ que deje a su derecha un área de $\alpha/2$, tendremos que:

$$P\left(f_{\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < f_{1-\alpha/2}\right) = 1-\alpha$$

y despejando:

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}}\right) = 1-\alpha$$

de donde el intervalo de confianza para $\sigma^2$ será

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}}, \frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}}\right)$$

## 6.7 Contraste de Hipótesis.

En muchas ocasiones, el problema no es determinar el valor de los parámetros de una población, sino verificar si un conjunto de afirmaciones sobre la población son ciertas o no, es decir, verificar si debe aceptarse o rechazarse una hipótesis.

La veracidad o falsedad de una hipótesis nunca podrá ser conocida con certeza, a menos que analicemos el total de la población. Como este procedimiento es claramente inviable en la mayoría de los casos, lo que haremos será tomar muestras y analizar los datos de ésta, confrontándolos con la hipótesis. Si de acuerdo con ésta es poco probable que se hubieran presentado unos datos como los observados, rechazaríamos la veracidad de la hipótesis. Por ejemplo, si queremos ver si una moneda está trucada o no, podemos realizar el experimento de lanzarla 100 veces. Nuestra hipótesis sería que la moneda no está trucada, es decir, que el valor de $p$ es 0,5. Si obtenemos tan sólo 23 caras, podemos comprobar que este resultado es muy poco probable que ocurra si la moneda no está trucada, con lo cual rechazaríamos la hipótesis. En cambio, si obtenemos 47 caras, esto es bastante probable bajo nuestra hipótesis, con lo cual aceptaríamos que la moneda no está trucada. Obsérvese que este último resultado también sería consistente con la hipótesis $p = 0,45$. Por tanto la moneda podría estar ligeramente trucada y, sin embargo, en virtud de los resultados de nuestro experimento, podríamos concluir que no lo está.

Debemos dejar claro qué significa aceptar y qué rechazar una hipótesis. Rechazar una hipótesis significa concluir que es falsa, mientras que aceptarla significa que no tenemos suficiente información para concluir otra cosa. Evidentemente, tanto si aceptamos como si rechazamos existe la posibilidad de equivocarnos, en nuestro ejemplo anterior es poco probable que al lanzar una moneda no trucada 100 veces nos salgan 23 caras o menos, sin embargo, aunque esta probabilidad sea pequeña existe.

Llamaremos hipótesis nula y denotaremos $H_0$ a la primera hipótesis enunciada. Si rechazamos $H_0$ estamos aceptando la hipótesis alternativa $H_1$. Siempre que tomemos una decisión en un contraste de hipótesis hay dos errores que podemos cometer:

- Podemos rechazar $H_0$ cuando sea verdadera, esto se llama error de tipo I.

- Podemos aceptar $H_0$ cuando sea falsa, esto se llama error de tipo II.

A la probabilidad de cometer un error de tipo I se le llama $\alpha$, también denominado nivel de significación. La probabilidad de cometer un error de tipo II se le llama $\beta$. Al valor $1 - \beta$ se le denomina potencia del contraste. Siempre existirá la posibilidad de cometer alguno de los dos errores. La única manera de hacer 0 una de las dos probabilidades es aceptando o rechazando siempre, lo cual, evidentemente, es un procedimiento no válido.

Lo que se hace usualmente es fijar $\alpha$ y, a partir de ahí, intentar obtener el procedimiento que haga lo más pequeño posible $\beta$, (con lo cual elegiremos el procedimiento que nos haga $1 - \beta$ lo más grande posible, esto es, el de máxima potencia). Los valores más usuales para $\alpha$ son de 0,05 y 0,01, con lo cual la fiabilidad es del 95% y del 99% respectivamente.

Una hipótesis nula referida a un parámetro de la población será siempre enunciada de tal modo que incluya el signo igual, (es decir, admitiremos hipótesis nula del tipo mayor o igual pero no del tipo mayor estricto).

Dentro del contraste de hipótesis podemos hacer la siguiente clasificación:

- Contrastes paramétricos: conocemos qué tipo de distribución sigue la población, y hacemos hipótesis sobre el valor de los parámetros.
- Contrastes no paramétricos: no conocemos el tipo de distribución.

El procedimiento general de operar a la hora de realizar un contraste de hipótesis será la siguiente:

1. Formular las hipótesis.
2. Establecer el nivel de significación.
3. Obtener la muestra.
4. Calcular una función de la muestra cuya distribución bajo $H_0$ sea conocida. A esta función se le denomina estadístico.
5. Si el valor del estadístico está en una determinada región llamada *región de aceptación* se acepta la hipótesis nula. Si no está en esta región, entonces está en la *región crítica* y se rechaza $H_0$. El límite entre ambas regiones se denomina *valor crítico.*

La probabilidad de la región crítica será de $\alpha$, y, por tanto, la de la región de aceptación será de $1 - \alpha$. Habrá que determinar, en primer lugar, si la región crítica es unilateral (caso de $H_1$ del tipo $\mu > 0$ o $\mu < 0$) o bilateral (caso de $H_1$ del tipo $\mu \neq 0$).

## 6.8 Contraste de Hipótesis para la media en una población normal.

Los contrates de hipótesis que estudiaremos a continuación serán los paramétricos sobre el valor de $\mu$ en poblaciones normales. En este caso el estadístico a utilizar será la media muestral estandarizada (dividida por la desviación típica o por un estimador de ésta), que sabemos que sigue una distribución N(0,1) o t de Student.

### 6.8.1 Con varianza conocida

Veamos algunos ejemplos de contraste de hipótesis en poblaciones normales con varianza conocida.

Para estimar una hipótesis del tipo $H_0 : \mu = \mu_0$ contra la alternativa $H_1 : \mu > \mu_0$ en una población normal con varianza $\sigma^2$ conocida, usaremos el estadístico

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

que sabemos que bajo $H_0$ (es decir, si la hipótesis nula es cierta) se distibuye según una N(0,1). Como el contraste es unilateral y la región crítica está por encima de la región de aceptación (ya que la hipótesis alternativa es $H_1 : \mu > \mu_0$, con lo que rechazaremos $H_0$ si el valor obtenido en el estadístico es grande) habrá que buscar en las tablas de la normal un punto $z_{1-\alpha}$ que deje a su derecha un área de $\alpha$. Este es el punto crítico. Si el valor calculado al estadístico es mayor que este punto crítico rechazaremos la hipótesis nula, mientras que si el valor obtenido es inferior a este punto crítico aceptaremos $H_0$.

En el caso de que la hipótesis alternativa fuese $H_1 : \mu < \mu_0$ se procedería del mismo modo, pero ahora la región crítica se obtendrá para valores pequeños del estadístico. El punto crítico será ahora $z_\alpha$ y se rechazará la hipótesis nula si el valor calculado en el estadístico es menor a este punto crítico.

En el caso de que la hipótesis alternativa fuese $H_1 : \mu \neq \mu_0$ el contraste es bilateral. Se procedería del mismo modo, pero ahora la región crítica estará dividida en dos partes, una corresponderá a valores pequeños del estadístico y la otra a valores grandes. Cada una de esas dos partes en las que está dividida la región crítica tendrá un área de $\alpha/2$. Habrá que buscar en las tablas el punto $z_{\alpha/2}$ que deje a su izquierda un área de $\alpha/2$, y el punto $z_{1-\alpha/2}$ que deje a su derecha un área de $\alpha/2$ Se rechazará la hipótesis nula si el valor calculado en el estadístico es menor que $z_{\alpha/2}$ o mayor que $z_{1-\alpha/2}$.

Si el contraste fuese del tipo $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$ se precedería exactamente igual que si fuese del tipo $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$.

### 6.8.2 Con varianza desconocida

En el caso de que la varianza de la población sea desconocida, se utilizará el estadístico

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

que se distribuirá según una t de Student con n-1 grados de libertad. Si el tamaño de la muestra es grande (mayor que 30) se puede realizar una aproximación a la normal, con lo cual estaríamos en el caso anterior.

La realización del contraste sería análogo a lo visto en el apartado anterior, tan solo variaría el hecho de que los puntos críticos habría que buscarlos en las tablas de la t de Student en lugar de en las de la normal.

## 6.9 Contrastes sobre diferencias de medias

Si tenemos dos poblaciones normales, una con media $\mu_1$ y otra con media $\mu_2$, podemos plantearnos contrastes sobre si la diferencia de las dos medias es igual a un valor $d_0$ determinado. El caso más usual será querer contrastar si las dos medias son iguales, es decir, el caso en que $d_0 = 0$.

En estos contrastes habrá que distinguir si las varianzas son conocidas, desconocidas pero iguales o desconocidas y distintas, ya que como se vió en la lección anterior los estadísticos a utilizar serán distintos. Una vez obtenida la distribución del estadístico se calculará la región crítica, y se procederá a comprobar si el valor obtenido en el estadístico está dentro de la región crítica o no.

Puede darse el caso de observaciones apareadas, esto es, cuando tenemos dos poblaciones pero cada elemento de la primera población está relacionada con el correspondiente elemento de la segunda población (por ejemplo, para ver la eficacia de un tratamiento, se hacen mediciones del mismo individuo antes y después de seguir dicho tratamiento). En este caso, construimos una nueva muestra formada por las diferencias entre dos observaciones emparejadas. Como en el caso anterior, usualmente $d_0$ valdrá 0. El estadístico t se construye dividiendo la media de las diferencias entre el estimador de la desviación típica de dichas diferencias.

En la tabla 6.12 se resumen los contrastes de hipótesis más usuales, indicando el estadístico a utilizar en cada caso.

Tabla 6.12: Contrastes de hipótesis más usuales

| $H_0$ | Estadístico |
|---|---|
| $\mu = \mu_0$ | $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ <br> $\sigma$ conocida ó $n \geq 30$ |
| $\mu = \mu_0$ | $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ <br> $\nu = n - 1$ <br> $\sigma$ desconocida y $n < 30$ |
| $\mu_1 - \mu_2 = d_0$ | $z = \dfrac{\overline{x}_1 - \overline{x}_2 - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$ <br> $\sigma_1, \sigma_2$ conocidas |
| $\mu_1 - \mu_2 = d_0$ | $t = \dfrac{\overline{x}_1 - \overline{x}_2 - d_0}{s_p\sqrt{(1/n_1) + (1/n_2)}}$ <br> $\nu = n_1 + n_2 - 2$ <br> $\sigma_1 = \sigma_2$ pero desconocidas <br> $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ |
| $\mu_1 - \mu_2 = d_0$ | $t = \dfrac{\overline{x}_1 - \overline{x}_2 - d_0}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$ <br> $\sigma_1 \neq \sigma_2$ y desconocidas <br> $\nu = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$ |
| $\mu_D = d_0$ | $t = \dfrac{\overline{d} - d_0}{s_d/\sqrt{n}}$ <br> $\nu = n - 1$ <br> observaciones apareadas |

## 6.10 Selección del tamaño de la muestra para el contraste de medias.

El tamaño de la prueba se selecciona, por lo común, para tener una buena potencia con un nivel $\alpha$ fijo y una hipótesis alternativa específica fija.

Supongamos que desea probarse:

$$\begin{array}{lcl} H_0: \mu & = & \mu_0 \\ H_1: \mu & > & \mu_0 \end{array}$$

con un nivel de significación $\alpha$, cuando se conoce $\sigma^2$. Para una alternativa específica, como $\mu = \mu_0 + \delta$, la potencia de la prueba es:

$$1 - \beta = P[\bar{X} > a \quad / \quad \mu = \mu_0 + \delta].$$

Por tanto:

$$\begin{aligned} \beta &= P[\bar{X} < a \,/\, \mu = \mu_0 + \delta] = \\ &= P[\frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} < \frac{a - (\mu_0 + \delta)}{\sigma/\sqrt{n}} \quad \text{cuando} \quad \mu = \mu_0 + \delta] = \\ &= P[Z < \frac{a - \mu_0}{\sigma/\sqrt{n}} - \frac{\delta}{\sigma/\sqrt{n}}] = \\ &= P[Z < z_{1-\alpha} - \frac{\delta}{\sigma/\sqrt{n}}] \end{aligned}$$

o sea $z_\beta = z_{1-\alpha} - \frac{\delta\sqrt{n}}{\sigma}$ , y despejando tenemos:

$$n = \frac{(z_{1-\alpha} - z_\beta)^2 \cdot \sigma^2}{\delta^2} = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \cdot \sigma^2}{\delta^2}.$$

Este resultado también es cierto si $H_1: \mu < \mu_0$.

En el caso en el que el contraste sea de dos colas tendremos:

$$n \approx \frac{(z_{1-\alpha/2} - z_\beta)^2 \cdot \sigma^2}{\delta^2} = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\delta^2}.$$

Para diferencias de medias el tamaño $n = n_1 = n_2$ de la muestra si se conoce $\sigma_1$ y $\sigma_2$ será:

$$\begin{array}{rcl} \mathrm{H}_0: \ \mu_1 - \mu_2 & = & d_0 \\ \mathrm{H}_1: \ \mu_1 - \mu_2 & \neq & d_0 \end{array}$$

el tamaño de la muestra será:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \cdot ({\sigma_1}^2 + {\sigma_2}^2)}{\delta^2}.$$

y para una cola:

$$\begin{array}{rcl} \mathrm{H}_0: \ \mu_1 - \mu_2 & = & d_0 \\ \mathrm{H}_1: \ \mu_1 - \mu_2 & >< & d_0 \end{array}$$

el tamaño de la muestra será:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \cdot ({\sigma_1}^2 + {\sigma_2}^2)}{\delta^2}.$$

Cuando se desconoce la varianza poblacional, la selección del tamaño de la muestra no es tan directa.

## 6.11 Contrastes relacionados con proporciones

Los pasos para probar una hipótesis nula relativa a una proporción en contra de varias alternativas, utilizando las probabilidades binomiales son:

1. $H_0 : p = p_0$.
2. $H_1 : p > p_0$ , $H_1 : p < p_0$ , $H_1 : p \neq p_0$.
3. Escoger el nivel de significación $\alpha$.
4. El estadístico de prueba: la variable binomial $X$ con $p = p_0$.
5. Calcular $x$, la cantidad de éxitos, y calcular el valor apropiado de $p$.

6. Decisión: Tomar las decisiones apropiadas con base al valor $p$.

Cuando $p$ está próximo a 0 ó 1 se puede aproximar por la distribución de Poisson con $\lambda = np_0$.

Si no está cerca y $n$ es grande, haremos la aproximación Normal con $\mu = np_0$ y $\sigma^2 = np_0q_0$.

Para la diferencia de dos proporciones la decisión es: $\hat{p_1} - \hat{p_2}$.

Para muestras grandes ($n \gg$) tendremos

$$\frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} \quad \sim \quad N(0,1).$$

Así cuando $p_1 = p_2 = p$ y $q_1 = q_2 = q$ tendremos:

$$\frac{\hat{p_1} - \hat{p_2}}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}} \quad \sim \quad N(0,1).$$

La estimación combinada de $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ y $\hat{q} = 1 - p$. Y los valores se sustituyen en:

$$\frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} \quad \sim \quad N(0,1).$$

## 6.12 Contrastes sobre varianzas.

Para contrastar la varianza en una población normal:

$$\begin{array}{lrcl} \mathrm{H}_0: & \sigma^2 & = & {\sigma_0}^2 \\ \mathrm{H}_1: & \sigma^2 & \neq & {\sigma_0}^2 \\ & \sigma^2 & < & {\sigma_0}^2 \\ & \sigma^2 & > & {\sigma_0}^2 \end{array}$$

Así, como la distribución del estadístico:

$$\frac{(n-1)S^2}{{\sigma_0}^2} \sim \chi^2(n-1),$$

tendremos que las regiones críticas para las distintas hipótesis alternativas son:

$$\begin{array}{lll}
\text{Si} \quad \sigma^2 = {\sigma_0}^2 & \longrightarrow & \chi^2 > \chi^2_{1-\alpha/2} \quad \text{y} \quad \chi^2 < \chi^2_{\alpha/2} \\
\text{Si} \quad \sigma^2 < {\sigma_0}^2 & \longrightarrow & \chi^2 < \chi^2_{\alpha} \\
\text{Si} \quad \sigma^2 > {\sigma_0}^2 & \longrightarrow & \chi^2 > \chi^2_{1-\alpha}
\end{array}$$

Si los contrastes son relativos a las varianzas de dos poblaciones tendremos:

$$\begin{array}{lccc}
\mathrm{H}_0: & {\sigma_1}^2 & = & {\sigma_2}^2 \\
\mathrm{H}_1: & {\sigma_1}^2 & \neq & {\sigma_2}^2 \\
 & {\sigma_1}^2 & < & {\sigma_2}^2 \\
 & {\sigma_1}^2 & > & {\sigma_2}^2
\end{array}$$

Así, como la distribución del estadístico:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

Buscando en las tablas de la F de Snedecor de los grados de libertad correspondientes el punto $f_{\alpha/2}$ que deje a su izquierda un área de $\alpha/2$ y el punto $f_{1-\alpha/2}$ que deje a su derecha un área de $\alpha/2$, tendremos que las regiones críticas para las distintas hipótesis alternativas son:

$$\begin{array}{lll}
\text{Si} \quad {\sigma_1}^2 = {\sigma_2}^2 & \longrightarrow & f > f_{1-\alpha/2}(v_1, v_2) \quad \text{y} \quad f < f_{\alpha/2}(v_1, v_2) \\
\text{Si} \quad {\sigma_1}^2 < {\sigma_2}^2 & \longrightarrow & f < f_{\alpha}(v_1, v_2) \\
\text{Si} \quad {\sigma_1}^2 > {\sigma_2}^2 & \longrightarrow & f > f_{1-\alpha}(v_1, v_2)
\end{array}$$

siendo $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad.

# Problemas Resueltos

## Problema 6.1

**Una muestra aleatoria de 100 ordenadores de un determinado fabricante mostró que el tiempo que tardaban en realizar una determinada operación era de 11,8 milisegundos, con un estimador de la desviación típica de 8,9 milisegundos. Contrastar, a un nivel de significación $\alpha = 0{,}05$ si estos datos contradicen los informes del fabricante que aseguran que el tiempo tardado en realizar dicha operación es de 10 milisegundos.**

Debemos realizar un contraste de hipótesis. El nivel de significación está fijado en el enunciado como $\alpha = 0{,}05$. Las hipótesis serían:

$H_0$: $\mu = 10$ milisegundos
$H_1$: $\mu > 10$ milisegundos

El estadístico a utilizar será

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

ya que el valor de $n$ es grande.

El contraste es unilateral. La región crítica será

$$z > z_{1-\alpha} = z_{0,95} = 1{,}645.$$

Realizando los cálculos, en nuestro caso el valor del estadístico de contraste es

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{11{,}8 - 10}{8{,}9/10} = 2{,}02.$$

Por tanto, debemos concluir que, al estar en la región crítica, debemos rechazar $H_0$ y admitir, por tanto, que la desviación de los datos respecto a lo anunciado por el fabricante nos lleva a concluir que éste no dice la verdad.

## Problema 6.2

**Repetir el problema anterior con $\alpha = 0{,}01$. Explicar las diferencias de resultados.**

---

Las hipótesis serían las mismas, así como el estadístico a utilizar. Pero en este caso la región crítica sería

$$z > z_{0,99} = 2{,}33.$$

Al no estar ahora en la región crítica, aceptaríamos $H_0$. Al ser $\alpha$ menor, la probabilidad de cometer un error de tipo I es menor, es decir, es más dificil rechazar $H_0$ si es verdadera. En cambio, al mantener fijo el tamaño muestral, al disminuir $\alpha$ aumenta $\beta$, con lo cual ahora es más fácil cometer el error de aceptar $H_0$ siendo ésta falsa.

## Problema 6.3

**Una muestra aleatoria de 16 componentes mostró que la temperatura media a la que empezaban a fallar era de 81 grados celsius. La desviación típica de la temperatura de fallo era conocida por múltiples pruebas anteriores y era igual a 5 grados. Contrastar, a un nivel de significación $\alpha = 0{,}05$ y mediante un contraste bilateral si puede admitirse que la temperatura a la que empiezan a fallar es de 80 grados celsius.**

---

Debemos de realizar un contraste de hipótesis. El nivel de significación está fijado en el enunciado como $\alpha = 0{,}05$. Las hipótesis serían:

$H_0$: $\mu = 80$ grados celsius
$H_1$: $\mu \neq 80$ grados celsius

El estadístico a utilizar será

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

ya que aunque el valor de $n$ no es grande y $\sigma$ es conocida.

El contraste es bilateral. La región crítica será

$$z > z_{1-\alpha/2} \quad \text{ó} \quad z < -z_{1-\alpha/2}$$

que en este caso es

$$z > 1{,}96 \quad \text{ó} \quad z < -1{,}96.$$

Realizando los cálculos, en nuestro caso el valor del estadístico de contraste es

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{81 - 80}{5/4} = 0{,}8.$$

Por tanto, debemos concluir que, al no estar en la región crítica, debemos aceptar $H_0$.

# Problema 6.4

**Repetir el problema anterior si la desviación típica no fuese conocida, siendo 5 grados el valor de la desviación típica muestral.**

---

Las hipótesis serían las mismas, pero el estadístico a utilizar cambia. En efecto, al ser la desviación típica desconocida y la muestra pequeña, no podemos usar el estadístico normal. Debemos usar la $t$ de Student:

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

El contraste es bilateral. La región crítica será

$$t > t_{15;1-\alpha/2} \quad \text{ó} \quad t < -t_{15;1-\alpha/2}$$

que en este caso es, mirando en las tablas de percentiles de la distribución $t$-Student con 15 grados de libertad

$$t > 2{,}131 \quad \text{ó} \quad t < -2{,}131.$$

Realizando los cálculos, en nuestro caso el valor del estadístico de contraste es

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{81 - 80}{5/4} = 0{,}8.$$

Por tanto, debemos concluir que, al no estar en la región crítica, debemos aceptar $H_0$.

# Problema 6.5

**Se probaron los microprocesadores de dos fabricantes para medir**

**su tiempo de vida útil sin sistema de refrigeración a temperatura ambiente de 35 grados centrígados. Se midieron 12 del primer fabricante, obteniéndose una media de 24 minutos y una desviación típica muestral de 4 minutos, y 10 del segundo, obteniéndose una media de 20 minutos y una desviación típica muestral de 5 minutos. Razonar si puede concluirse, para $\alpha = 0{,}05$ que el tiempo de duración de los microprocesadores del primer fabricante superan en más de dos minutos a los del segundo. Supóngase que ambas poblaciones son aproximadamente normales con la misma varianza.**

---

Debemos de realizar un contraste de hipótesis. El nivel de significación está fijado en el enunciado como $\alpha = 0{,}05$. Las hipótesis serían:

$$H_0: \quad \mu_1 - \mu_2 = 2$$
$$H_1: \quad \mu_1 - \mu_2 > 2$$

El estadístico a utilizar será

$$t = \frac{\overline{X}_1 - \overline{X}_2 - d_0}{S_p \sqrt{(1/n_1) + (1/n_2)}}$$

ya que las dos poblaciones tienen la misma varianza pero es desconocida. El estadístico tiene $n_1 + n_2 - 2 = 20$ grados de libertad.

El contraste es unilateral. La región crítica será

$$t > t_{20;1-\alpha} = t_{20;0,95} = 1{,}725.$$

Realizando los cálculos, el valor de $S_p$ es:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(11)(16) + (9)(25)}{12 + 10 - 2} = (4{,}478)^2.$$

En nuestro caso el valor del estadístico de contraste es

$$t = \frac{24 - 20 - 2}{4{,}478\sqrt{(1/12) + (1/10)}} = 1{,}04.$$

Por tanto, debemos concluir que, al no estar en la región crítica, debemos aceptar $H_0$ y admitir, por tanto, que la duración de los microprocesadores del primer fabricante no superan en más de dos minutos a los del segundo.

## Problema 6.6

**Una empresa tiene una serie de factorías en las cuales se producen planchas de acero cuyo peso se distribuye normalmente con media 130 Kg. En una nueva factoría se han empezado a producir planchas, obteniéndose los siguientes pesos:**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **134** | **132** | **129** | **127** | **131** | **139** | **135** | **131** | **129** | **131** |

**(a) ¿Se puede admitir que en esta factoría se están produciendo planchas de mayor peso que en el resto de la empresa?**

**(b) En una segunda factoría se han obtenido los siguientes resultados:**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **128** | **130** | **130** | **132** | **131** | **138** | **135** | **131** | **134** |

**Supuesta igualdad de varianzas, ¿puede admitirse que la segunda factoría tiene mayor media que la primera?**

**Utilizar $\alpha = 0{,}05$.**

---

(a) Para comprobar si puede admitirse que en esta factoría se están produciendo planchas de mayor peso que en el resto de la empresa habrá que realizar un contraste de hipótesis sobre el valor de la media $\mu$ de esta factoría. Seguiremos para ello el siguiente proceso:

1. Formular las hipótesis.
2. Establecer el nivel de significación.
3. Obtener la muestra.
4. Calcular una función de la muestra cuya distribución bajo $H_0$ sea conocida. A esta función se le denomina estadístico.
5. Si el valor del estadístico está en una determinada región llamada *región de aceptación* se acepta la hipótesis nula. Si no está en esta región, entonces está en la *región crítica* y se rechaza $H_0$.

1. Las hipótesis a considerar serían:

$$\begin{aligned} H_0: &\quad \mu = 130 \\ H_1: &\quad \mu > 130 \end{aligned}$$

2. El nivel de significación viene dado en el enunciado ($\alpha = 0{,}05$).
3. Se ha tomado una muestra de 10 elementos, obteniéndose una media muestral de $\bar{X} = 131{,}8$ y una desviación típica muestral de $S = 3{,}45768$. La media obtenida es ligeramente superior a 130, pero debemos comprobar si este valor obtenido puede ser debido al azar o no.
4. Como no conocemos el valor de la varianza de la población y la muestra es de tamaño pequeño, utilizaremos el estadístico

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

que se distribuirá según una $t$ de Student con $\nu = n-1 = 9$ grados de libertad. El valor del estadístico para la muestra considerada es de $t = 1{,}646219054$.
5. La región crítica quedará determinada por aquellos puntos correspondientes a la cola derecha de probabilidad $\alpha$ de una distribución $t$ de Student con 9 grados de libertad. En este caso la región crítica correspondería a valores del estadístico mayores de $t_{9;0,95} = 1{,}833113856$. Como en nuestro caso el valor del estadístico es menor, estamos en la región de aceptación y aceptamos $H_0$. Por tanto no podemos admitir que en esta factoría se produzcan planchas de mayor peso que en el resto.

(b) En este caso debemos realizar la comparación entre la media de la factoría 1, que llamaremos $\mu_1$ y la media de la factoría 2, que llamaremos $\mu_2$. Seguiremos las mismos pasos indicados en el apartado anterior:

1. Las hipótesis a considerar serían:

$$\begin{array}{ll} H_0: & \mu_1 - \mu_2 = 0 \\ H_1: & \mu_1 - \mu_2 < 0 \end{array}$$

2. El nivel de significación viene dado en el enunciado ($\alpha = 0{,}05$).
3. Se ha tomado una muestra de 10 elementos de la primera factoría, obteniéndose una media de $\overline{X}_1 = 131{,}8$ y una desviación típica muestral de $S_1 = 3{,}457680661$. En la segunda factoría se ha tomado una muestra de 9 elementos, obteniéndose una media de $\overline{X}_2 = 132{,}1111111$ y una desviación típica muestral de $S_2 = 3{,}059593292$. La media obtenida en la segunda muestra es

ligeramente superior a la de la primera, con lo que parece que se confirma la hipótesis alternativa, pero debemos comprobar si este valor obtenido puede deberse al azar o no.

4. No conocemos el valor de las varianzas, pero el enunciado nos dice que ambas son iguales. Por tanto, utilizaremos el estadístico

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - d_0}{S_p\sqrt{(1/n_1) + (1/n_2)}}$$

siendo

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

que se distribuirá según una $t$ de Student con $\nu = n_1 + n_2 - 2 = 17$ grados de libertad. El valor del estadístico para la muestra considerada es $t = -0{,}206665$ pues:

$$S_p = \sqrt{\frac{(9 \cdot 11{,}9556 + 8 \cdot 9{,}36111}{17}} = \sqrt{10{,}73466} = 3{,}2763796$$

$$t = \frac{(131{,}8 - 132{,}111111) - 0}{3{,}2763796\sqrt{(1/10) + (1/9)}} = \frac{-0{,}311111}{1{,}50539} = -0{,}206665.$$

5. La región crítica quedará determinada por aquellos puntos correspondientes a la cola izquierda de probabilidad $\alpha$ de la distribución $t$ de Student con 17 grados de libertad. En este caso la región crítica correspondería a valores del estadístico $t$ menores de $t_{17;0,05} = -t_{17;0,95} = -1{,}739606432$. Como en nuestro caso el valor del estadístico $t$ es mayor, estamos en la región de aceptación y aceptamos $H_0$. Por tanto no podemos admitir que exista diferencia en el peso de las planchas producidas en la segunda factoría respecto a la primera.

# Problema 6.7

**La pérdida de peso de un determinado producto dietético en 16 individuos después de un mes fue (en kg):**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **3,2** | **2** | **2,5** | **3,3** | **5** | **4,3** | **2,9** | **4,1** |
| **3,6** | **2,7** | **3,5** | **4,2** | **2,8** | **4,4** | **3,3** | **3,1** |

**Determinar un intervalo de confianza para la varianza con nivel de confianza del 99%, si la pérdida de peso es aproximadamente normal.**

---

Un intervalo de confianza para la varianza vendrá dado por

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\ ;\ \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$$

donde los valores $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ corresponden al valor de la función de distribución Chi-cuadrado con $n-1$ grados de libertad.

En nuestro caso como $n = 16$, y $\alpha = 0{,}01$ tenemos que mirando en las tablas de percentiles de la distribución Chi-cuadrado con 15 grados de libertad $\chi^2_{0{,}005}(15) = 4{,}60087$ y $\chi^2_{0{,}995}(15) = 32{,}80149$

Si calculamos la varianza muestral, tenemos que $S^2 = 0{,}636958$; y por tanto sustituyendo los valores en los del intervalo de confianza para la varianza tenemos

$$\left[\frac{15\cdot 0{,}636958}{32{,}80149}\ ;\ \frac{15\cdot 0{,}636958}{4{,}60087}\right] = \left[\frac{9{,}55437}{32{,}80149}\ ;\ \frac{9{,}55437}{4{,}60087}\right].$$

Y un intervalo de confianza del 99% para la varianza de la pérdida de peso es

$$\sigma^2 \in (0{,}291279 \quad ; \quad 2{,}07664).$$

# Problema 6.8

**Se consideran los siguientes tiempos de reaccción de un producto químico, en segundos:**

**1,4 1,2 1,2 1,3 1,5 1,3 2,2 1,4 1,1**

**Obtener un intervalo de confianza del 90% para el tiempo de reacción. Suponer la variable normal con desviación típica poblacional conocida $\sigma = 0{,}4$.**

---

Para calcular un intervalo de confianza para la media de una población normal de la que conocemos la varianza poblacional, tenemos que usar el estadístico:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Por tanto, necesitamos calcular la media de la muestra. Que en nuestro caso es:

$$\bar{X} = \frac{1{,}4 + 1{,}2 + 1{,}2 + 1{,}3 + 1{,}5 + 1{,}3 + 2{,}2 + 1{,}4 + 1{,}1}{9} = \frac{12{,}6}{9} = 1{,}4.$$

Como queremos obtener un intervalo con nivel de confianza del 90%, tenemos que $\alpha = 0{,}1$. Así un intervalo de confianza del 90% para la media se obtendrá:

$$(\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \; ; \; \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}).$$

En nuestro caso, si miramos en la tabla de la función de distribución de la N(0;1), tenemos que $z_{1-\alpha/2} = z_{0,95} = 1{,}64$ es el valor de la normal estándar que correspondiente al percentil 95.

Un intervalo de confianza del 90% para la media $\mu$ será:

$$(1{,}4 - 1{,}64 \cdot \frac{0{,}4}{3} \; ; \; 1{,}4 + 1{,}64 \cdot \frac{0{,}4}{3}) = (1{,}4 - 0{,}2186 \; ; \; 1{,}4 - 0{,}2186) =$$

$$= (1{,}1813 \; ; \; 1{,}6186).$$

# Problema 6.9

**Hace tiempo, una máquina producía arandelas de 0,05 pulgadas de espesor. Para determinar si sigue en buen estado, se toma una muestra de 10 arandelas, que dan un espesor medio de 0,053 pulgadas y desviación típica de 0,003 pulgadas. Contrastar la hipótesis de que la máquina sigue funcionando bien con niveles de significación $\alpha = 0{,}05$ y $\alpha = 0{,}01$.**

(a) Para comprobar si puede admitirse que la máquina sigue funcionando bien habrá que realizar un contraste de hipótesis sobre el valor de la media $\mu$. Seguiremos para ello el siguiente proceso:

1. Formular las hipótesis.
2. Establecer el nivel de significación.
3. Obtener la muestra.

4. Calcular un estadístico de la muestra cuya distribución bajo $H_0$ sea conocida.
5. Si el valor del estadístico está en una determinada región llamada *región de aceptación* se acepta la hipótesis nula. Si no está en esta región, entonces está en la *región crítica* y se rechaza $H_0$.

1. Las hipótesis a considerar serían:

$$\begin{array}{ll} H_0: & \mu = 0{,}05 \\ H_1: & \mu \neq 0{,}05 \end{array}$$

2. El nivel de significación viene dado en el enunciado ($\alpha = 0{,}05$).
3. Se ha tomado una muestra de 10 elementos, obteniéndose una media de $\bar{X} = 0{,}053$ pulgadas y una desviación típica muestral de $S = 0{,}003$. La media obtenida es ligeramente superior a 0,05 pero debemos comprobar si este valor se debe al azar o no.
4. Como no conocemos el valor de la varianza de la población y la muestra es de tamaño pequeño, utilizaremos el estadístico

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

que se distribuirá según una $t$ de Student con $\nu = n-1 = 9$ grados de libertad. El valor del estadístico para la muestra considerada es de $t = 3{,}162$.
5. Como tenemos un test de dos colas, la región crítica quedará determinada por aquellos puntos correspondientes a la cola derecha de probabilidad $\alpha/2$ de una distribución $t$ de Student con 9 grados de libertad, y a la cola izquierda de probabilidad $\alpha/2$ de una distribución $t$ de Student con 9 grados de libertad.

   En este caso la región crítica correspondería a valores del estadístico mayores de $t_{9;0{,}925} = 2{,}26$ o menores que $-t_{9;0{,}925} = -2{,}26$.

   Como en nuestro caso el valor del estadístico es mayor, estamos en la región crítica y rechazamos $H_0$. Por tanto no podemos admitir que la máquina sigue en buen estado.

(b) Para comprobar si se acepta la hipótesis anterior con nivel de significación $\alpha = 0{,}01$ procedemos de forma análoga al apartado anterior, con lo que tenemos que la región crítica quedará determinada por aquellos puntos correspondientes a la cola derecha de probabilidad $\alpha/2$ de una distribución $t$ de Student con 9 grados de libertad, y a la cola izquierda de probabilidad $\alpha/2$ de una distribución $t$ de Student con 9 grados de libertad.

En este caso la región crítica correspondería a valores del estadístico mayores de $t_{9;0,995} = 3{,}25$ o menores que $-t_{9;0,995} = -3{,}25$.

Como en nuestro caso el valor del estadístico se encuentra comprendido entre los valores $-3{,}25 < t < 3{,}25$; estamos en la región de aceptación del test y aceptamos $H_0$. Por tanto podemos admitir que la máquina sigue en buen estado con nivel de confianza del 99%.

Nota: Como los resultados que hemos obtenido son de alguna forma contradictorios para dos niveles de significación muy similares, sería recomendable examinar otra muestra o revisar la máquina.

## Problema 6.10

**El número diario de piezas fabricadas en 5 días por una máquina $A$ han sido: 50, 42, 53, 60, 37. Mientras que otra máquina $B$ en esos mismos días ha hecho: 40, 51, 62, 55, 64. Si suponemos que la fabricación de piezas se distribuye normalmente en ambas máquinas. Se pide:**

**(a) Contrastar la hipótesis de que las máquinas tienen varianzas iguales con nivel de significación $\alpha = 0{,}05$.**

**(b) Usando el resultado anterior, construir un intervalo de confianza del 95% para la diferencia de medias.**

---

(a) Para comprobar si puede admitirse que las dos máquinas tiene varianzas iguales, habrá que realizar un contraste de hipótesis de igualdad de varianzas, para lo cual seguiremos el proceso del ejemplo anterior.

1. Las hipótesis a considerar serán:

$$\begin{array}{ll} H_0: & \sigma_A^2 = \sigma_B^2 \\ H_1: & \sigma_A^2 < \sigma_B^2 \end{array}$$

2. El nivel de significación es $\alpha = 0{,}05$.

3. El estadístico de decisión del test de hipótesis anterior es

$$F = \frac{S_A^2}{S_B^2}.$$

Con lo que en nuestro caso, tenemos que $n_1 = n_2 = 5$, siendo para las dos máquinas la media y la desviación típica muestral:

$$\bar{X}_A = 48{,}4 \qquad S_A = 9{,}0719347 \qquad \bar{X}_B = 54{,}4 \qquad S_B = 9{,}6072889$$

4. El estadístico de decisión para el test de igualdad de varianzas es

$$F = \frac{S_A^2}{S_B^2}$$

y en nuestro caso, se obtiene que $F = \dfrac{82{,}299}{92{,}3} = 0{,}8916576.$

5. La región crítica de este test de una cola viene dada por aquellos valores de $F$ que son menores que $F_\alpha(n_1 - 1; n_2 - 1)$ ; siendo $F$ el valor de la distribución Fisher-Snedecor con área a la izquierda $\alpha$ , que en nuestro caso usando la propiedad de la distribución $F$ que indica que $F_\alpha(n; m) = 1/F_{1-\alpha}(m; n)$ tenemos

$$F_{0,05}(4,4) = \frac{1}{F_{0,95}(4,4)} = \frac{1}{6{,}39} = 0{,}1564.$$

con lo que tenemos que $F > F_\alpha(n_1 - 1; n_2 - 1)$; y por tanto está en la región de aceptación del test, y concluimos que aceptamos la hipótesis nula de igualdad de varianzas.

(b) Ahora tenemos que calcular un intervalo de confianza para la diferencia de medias con un nivel de significación $\alpha = 0{,}05$. Si utilizamos el resultado del apartado anterior se puede suponer que las varianzas son iguales pero desconocidas, con lo que el intervalo de confianza para la diferencia de medias queda:

$$\left[(\bar{X}_B - \bar{X}_A) - t_{1-\alpha/2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; (\bar{X}_B - \bar{X}_A) + t_{1-\alpha/2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

siendo $S_p^2 = \dfrac{(n_1 - 1)S_A^2 + (n_2 - 1)S_B^2}{n_1 + n_2 - 2}.$

Si calculamos en nuestro caso $S_p^2$ obtenemos

$$S_p^2 = \frac{4(82{,}299 + 92{,}3}{8} = 87{,}2995. \qquad \Longrightarrow \qquad S_p = 9{,}3434.$$

Como el valor del percentil de la distribución $t$-Student con $n_1 + n_2 - 2$ grados de libertad para $\alpha = 0,05$ es $t_{8;0{,}975} = 2{,}306$; sustituyendo los valores en el intervalo de confianza tenemos:

$$\left[(54{,}4 - 48{,}4) - 2{,}306 \cdot 9{,}3434 \cdot \sqrt{\frac{2}{5}}\ ;\ (54{,}4 - 48{,}4) + 2{,}306 \cdot 9{,}3434 \ \sqrt{\frac{2}{5}}\right]$$

$$[6 - 13{,}626841\ ;\ 6 + 13{,}626841]\,.$$

Y un intervalo de confianza para la diferencia de medias con nivel de confianza del 95% es

$$-7{,}6268406 \ < \ \mu_B - \mu_A \ < \ 19{,}626841\,.$$

# Problema 6.11

**Una muestra aleatoria de 36 refrescos de una máquina automática tiene un contenido promedio de 21,9 decilitros, con una desviación típica muestral de 1,42 decilitros. Con una hipótesis de que $\mu = 22{,}2$ decilitros frente a la alternativa $\mu < 22{,}2$ y un nivel de significación de 0,05; decidir si se acepta el test.**

---

Si seguimos el esquema que hemos mantenido en los problemas anteriores, tenemos que el nivel de significación es $\alpha = 0{,}05$; las hipótesis a considerar son:

$$\begin{array}{ll} H_0: & \mu = 22{,}2 \\ H_1: & \mu < 22{,}2 \end{array}$$

Puesto que el tamaño de la muestra es $n = 36 > 30$, y no tenemos ningún dato que nos indique que la población se distribuye normalmente, el test que nos permitirá aceptar o rechazar $H_0$ es:

$$\text{Se rechaza } H_0 \text{ si } Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -z_{1-\alpha},$$

donde $z_{1-\alpha}$ es el valor de la variable aleatoria normal N(0,1) tal que deja a su derecha un área igual a $\alpha$.

Luego $-z_{1-\alpha} = -z_{0,95} = -1{,}645$, y el estadístico $Z$ en nuestro caso vale

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{21{,}9 - 22{,}2}{1{,}42/6} = -1{,}27.$$

Así, tenemos que $Z = -1{,}27 > -1{,}645$, por lo que no está dentro de la región crítica. Luego no teníamos evidencia para rechazar $H_0$, por lo que se acepta como valor de $\mu_0 = 22{,}2$.

# Problema 6.12

**Un fabricante de baterías de automóviles quiere dar para sus baterías el promedio de vida en meses. Si elegida una muestra de 6 de éstas se tienen las siguientes duraciones (en meses)**

**33 44 38 30 39 40,**

**determinar un intervalo de confianza del 90% para dicho promedio de vida en meses, e indicar si es válida la afirmación del fabricante: "Mis baterías duran 42 meses". (Suponer que la duración de estas baterías se distribuye en forma normal).**

---

Un intervalo de confianza del 90% para la media de una población normal vendrá dado por

$$\left[\bar{X} - t_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\,; \bar{X} + t_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right]$$

donde $t_{1-\frac{\alpha}{2}}$ es el valor de la $t$ de Student con $n-1$ grados de libertad que deja un área a la izquierda $1 - \frac{\alpha}{2}$.

En nuestro caso, tenemos que $n = 6$, el nivel de significación es $\alpha = 0{,}1$; el valor de la media muestral es $\bar{X} = 37{,}3333$ meses, y la desviación típica de la muestra es $S = 5{,}04645$ meses.

Si miramos en la tabla de percentiles de la distribución $t$-Student tenemos que $t_{5;0,95} = 2{,}015$; y sustituyendo en los límites del intervalo obtenemos

$$\left[37{,}3333 - 2{,}015 \cdot \frac{5{,}04645}{\sqrt{6}}\,; 37{,}3333 + 2{,}015 \cdot \frac{5{,}04645}{\sqrt{6}}\right].$$

Y un intervalo de confianza del 90% para el promedio de vida de las baterías será:

$$(33{,}18 \text{ meses}\,; 41{,}486 \text{ meses}).$$

Luego la afirmación del fabricante con nivel de confianza del 90% no es válida.

# Problema 6.13

**La pérdida de peso de un determinado producto dietético en 16 individuos después de un mes fué ( en kg):**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **3,2** | **2** | **2,5** | **3,3** | **5** | **4,3** | **2,9** | **4,1** |
| **3,6** | **2,7** | **3,5** | **4,2** | **2,8** | **4,4** | **3,3** | **3,1** |

**Construir un intervalo de confianza del 99% para la pérdida promedio real de peso de dicho producto dietético. (Suponer normalidad)**

Como los datos provienen de una normal de la que desconocemos su varianza, un intervalo de confianza para la media vendra dado por

$$\left[\bar{X} - t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}} \; ; \; \bar{X} + t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]$$

donde $t_{1-\alpha/2}$ es el valor de la función de distribución $t$-Student con $n-1$ grados de libertad.

En nuestro caso tenemos que $\alpha = 0{,}01$ y por tanto $t_{15;0{,}995} = 2{,}947$.

Si calculamos la media muestral y la desviación típica muestral tenemos:

$$\begin{array}{lll} \sum x_i = 54{,}9 & \bar{X} = 3{,}43125 & n = 16 \\ \sum x_i^2 = 197{,}93 & S_x = 0{,}79809669 & \end{array}$$

Con lo que sustituyendo, obtenemos:

$$\left[3{,}43125 - 2{,}947 \cdot \frac{0{,}79809669}{\sqrt{16}} \; ; \; 3{,}43125 + 2{,}947 \cdot \frac{0{,}79809669}{\sqrt{16}}\right].$$

Y un intervalo de confianza del 99% para la pérdida promedio de peso es:

$$(2{,}84325226 \text{ kg} \; ; \; 4{,}019247739 \text{ kg}).$$

## Problema 6.14

**Cinco mediciones del contenido de alquitrán de cierta clase de cigarrillos producen los resultados siguientes (miligramos por cigarrillo):**

**14,5 14,2 14,4 14,3 14,6.**

**Obtener un intervalo de confianza del 99% para la desviación típica $\sigma$, de la población muestreada, suponiendo normalidad.**

---

Un intervalo de confianza para la varianza en poblaciones normales vendrá dado por

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\ ;\ \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$$

donde los valores $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ corresponden al valor de la función de distribución Chi-cuadrado con $n-1$ grados de libertad.

En nuestro caso como $n = 5$, y $\alpha = 0{,}01$; si miramos en las tablas de percentiles de la distribución Chi-cuadrado tenemos que $\chi^2_{0,005}(4) = 0{,}20698$ y $\chi^2_{0,995}(4) = 14{,}86017$.

Si calculamos la varianza muestral, tenemos que

$$\sum x_i = 72 \qquad \sum x_i^2 = 1036{,}9 \qquad \bar{X} = 14{,}4$$

$$S^2 = \frac{\sum x_i^2 - n\bar{X}^2}{n-1} = \frac{1036{,}9 - 5\cdot(14{,}4)^2}{4} = \frac{0{,}1}{4} = 0{,}025.$$

Y por tanto sustituyendo los valores en los del intervalo de confianza para la varianza tenemos

$$\left[\frac{4\cdot 0{,}025}{14{,}86017}\ ;\ \frac{4\cdot 0{,}025}{0{,}20698}\right] = \left[\frac{0{,}1}{14{,}86017}\ ;\ \frac{0{,}1}{0{,}20698}\right].$$

Y un intervalo de confianza del 99% para la varianza del contenido de alquitrán es

$$0{,}00672936 \quad < \sigma^2 < \quad 0{,}483512.$$

Y por tanto un intervalo de confianza del 99% para la desviación típica es

$$\sqrt{0{,}00672936} \quad < \sigma < \quad \sqrt{0{,}483512}$$

$$0{,}0820327 \quad < \sigma < \quad 0{,}69535.$$

# Problema 6.15

**Un agente de ventas tiene un contrato con la empresa en el que se estipula que sus beneficios son de 15000 pesetas fijas más un porcentaje de las ventas que realice. Dicho agente, para examinar si el trabajo que tiene le compensa económicamente, controla los beneficios obtenidos en los últimos 10 meses, obteniendo los siguientes resultados:**

| | | | | |
|---|---|---|---|---|
| **27500** | **26500** | **30560** | **25340** | **22300** |
| **24850** | **22330** | **26420** | **28650** | **29250** |

**Si el beneficio es una variable aleatoria normal, calcular**

**(a) Estimación puntual de la desviación típica.**

**(b) Intervalo de confianza del 90% para el beneficio medio mensual.**

**(c) Intervalo de confianza del 90% para la desviación típica.**

---

(a) Como estamos suponiendo que la variable aleatoria es normal, entonces un estimador puntual de la desviación típica de la población $\sigma$ es la desviación típica muestral $S$.

Así, si calculamos $S^2$ obtenemos:

$$\sum x_i = 263700 \qquad \sum x_i^2 = 7022372000 \qquad n = 10$$

$$\bar{X} = 26370 \qquad S^2 = 7622555{,}555 \qquad S = 2760{,}8976.$$

(b) Como los datos provienen de una normal de la que desconocemos su varianza, un intervalo de confianza para la media vendrá dado por

$$\left[\bar{X} - t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\ ;\ \bar{X} + t_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]$$

donde $t_{1-\alpha/2}$ es el valor de la función de distribución $t$-Student con $n-1$ grados de libertad.

En nuestro caso tenemos que $\alpha = 0{,}1$ y por tanto $t_{9;0{,}95} = 1{,}833$.

Y como tenemos calculadas la media muestral y la desviación típica muestral, en el apartado anterior, sustituyendo obtenemos:

$$\left[26370 - 1{,}833 \cdot \frac{2760{,}8976}{\sqrt{10}} \; ; \; 26370 + 1{,}833 \cdot \frac{2760{,}8976}{\sqrt{10}}\right]$$

Y un intervalo de confianza del 90% para el beneficio medio mensual es:

$$(24769{,}2 \text{ pts} \; ; \; 27970{,}8 \text{ pts}).$$

(c) Un intervalo de confianza para la varianza en poblaciones normales vendrá dado por

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \; ; \; \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$$

donde los valores $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ corresponden al valor de la función de distribución Chi-cuadrado con $n-1$ grados de libertad.

En nuestro caso como $n = 10$, y $\alpha = 0{,}1$ tenemos que $\chi^2_{0{,}05}(9) = 3{,}325$ y $\chi^2_{0{,}95}(9) = 16{,}916$. Y sustituyendo los valores en los del intervalo de confianza para la varianza tenemos

$$\left[\frac{9 \cdot 7622555{,}555}{16{,}916} \; ; \; \frac{9 \cdot 7622555{,}555}{3{,}325}\right] = \left[\frac{68603000}{16{,}916} \; ; \; \frac{68603000}{3{,}325}\right]$$

Y un intervalo de confianza del 90% para la varianza es

$$4054790{,}4 \quad < \sigma^2 < \quad 20632481.$$

Y por tanto un intervalo de confianza para la desviación típica es

$$2013{,}65 \quad < \sigma < \quad 4542{,}3.$$

## Problema 6.16

**Una fábrica de fundas de discos viene utilizando un proceso $A$ en su elaboración. Se acaba de descubrir un nuevo proceso $B$ de fabricación del mismo producto que parece que requiere menos cantidad de materia prima. Para decidir si ésto es cierto se selecciona una muestra de 15 artículos fabricados por el proceso $A$ y otra de 17 fabricados por el $B$. La cantidad de materia prima**

**usada para cada muestra, en gramos, dio como resultados para el proceso $A$ una media de 400 g. y una desviación típica de 9 g.; y para el proceso $B$ una media de 385 g. y desviación típica de 10,5 g. Suponer que la cantidad de materia utilizada en ambos casos sigue una distribución normal.**

(a) **Determinar un intervalo de confianza para el cociente de varianzas con nivel de confianza del 98%.**

(b) **Usar el apartado anterior para decidir si es cierto que el proceso $B$ requiere menos materia prima que el $A$. (Usar $\alpha = 0{,}05$).**

---

(a) Como queremos determinar un intervalo de confianza para el cociente de varianzas de dos poblaciones normales, tenemos

$$\frac{S_A^2}{S_B^2} \cdot \frac{1}{F_{1-\alpha/2}(v_1, v_2)} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{S_A^2}{S_B^2} \cdot F_{1-\alpha/2}(v_2, v_1)$$

donde $F_{1-\alpha/2}(v_2, v_1)$ y $F_{1-\alpha/2}(v_1, v_2)$ corresponden a los valores de los percentiles de la distribución F-Snedecor con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad.

Como en nuestro caso, tenemos que $\alpha = 0{,}02$ si miramos en la tabla de la F-Snedecor obtenemos $F_{0{,}99}(14, 16) = 3{,}450628$ y $F_{0{,}99}(16, 14) = 3{,}618682$.

Como sabemos que

$$\begin{array}{lll} n_1 = 15 & \bar{X}_A = 400 & S_A = 9 \\ n_2 = 17 & \bar{X}_B = 385 & S_B = 10{,}5. \end{array}$$

Si sustituimos, tenemos que un intervalo de confianza del 98% para el cociente de varianzas es

$$\left[\frac{81}{110{,}25} \cdot \frac{1}{3{,}450628} \quad ; \quad \frac{81}{110{,}25} \cdot 3{,}618682\right] = [0{,}212916 \quad ; \quad 2{,}65862].$$

Si ambas varianzas fuesen iguales el cociente debería ser aproximadamente 1. Como en este caso el valor 1 pertenece al intervalo dado, concluimos que las varianzas son iguales.

(b) Como en el intervalo de confianza para el cociente entre varianzas está el valor 1, se puede suponer que las varianzas son iguales pero desconocidas, y para probar la hipótesis de igualdad de medias:

Las hipótesis a considerar son:

$$\begin{aligned} H_0 &: \quad \mu_A = \mu_B \\ H_1 &: \quad \mu_A > \mu_B \end{aligned}$$

El nivel de significación es 0,05.

El estadístico de decisión es:

$$t = \frac{(\bar{X}_A - \bar{X}_B) - d_0}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde $S_p = \sqrt{\frac{(n_1 - 1)S_A^2 + (n_2 - 1)S_B^2}{n_1 + n_2 - 2}}$

Si calculamos $S_p$ obtenemos $S_p = 9{,}828529$ y sustituyendo en el valor del estadístico

$$t = \frac{(400 - 385) - 0}{9{,}828529 \cdot 0{,}354245} = \frac{15}{3{,}4817} = 4{,}30821.$$

Como la región crítica del test viene dada por $t > t_{n_1+n_2-2;1-\alpha}$ siendo $t_{n_1+n_2-2;1-\alpha}$ el valor de la función de distribución de $t$-Student con $n_1+n_2-2$ grados de libertad.

Y como en nuestro caso $t_{30;0,95} = 1{,}69726$ tenemos que

$$t = 4{,}30821 > 1{,}69726$$

que está dentro de la región crítica y por tanto rechazo la hipótesis nula. Efectivamente, el proceso $B$ requiere menos materia prima que el $A$.

## Problema 6.17

**Un tratamiento para adelgazar garantiza la pérdida de 4,5 kg. en dos semanas. Para comprobar si esto es estadísticamente aceptable se estudió una muestra de 7 personas, obteniéndose los siguientes resultados:**

| **Antes del tratamiento** | **74** | **60,3** | **64** | **82** | **91** | **56,7** | **62,1** |
|---|---|---|---|---|---|---|---|
| **Después del tratamiento** | **70** | **54,9** | **58,5** | **75** | **84** | **54,4** | **69** |

**¿Se acepta la hipótesis con un nivel de confianza del 95% ?**

Como tenemos que efectuar un test con los datos apareados, y observar las diferencias entre las observaciones, las diferencias que obtenemos son

$$4 \qquad 5{,}4 \qquad 5{,}5 \qquad 7 \qquad 7 \qquad 2{,}3 \qquad -6{,}9.$$

Si seguimos el esquema de los problemas anteriores tenemos:

Las hipótesis a considerar son:

$$\begin{array}{ll} H_0: & d = 4{,}5 \text{ kg.} \\ H_1: & d \neq 4{,}5 \text{ kg.} \end{array}$$

El nivel de significación es 0,05.

El estadístico de decisión es: $t = \dfrac{(\bar{d} - d_0)}{S_d \,/\sqrt{n}}$ siendo

$$\bar{d} = \frac{24{,}3}{7} = 3{,}47142857 \quad \text{y} \quad S_d^2 = \frac{7 \cdot 226{,}31 - (24{,}3)^2}{7 \cdot 6} = 23{,}659048$$

$$S_d = \sqrt{23{,}659048} = 4{,}8640567.$$

Y sustituyendo en el valor del estadístico tenemos que

$$t = \frac{3{,}47142857 - 4{,}5}{4{,}8640567 \,/\sqrt{7}} = -0{,}55948.$$

Como la región crítica del test viene dada por $|t| > t_{n-1;1-\alpha/2}$ siendo $t_{n-1;1-\alpha/2}$ el valor de la función de distribución de $t$-Student con $n-1$ grados de libertad.

En nuestro caso, si miramos en la tabla de percentiles de la $t$-Student tenemos $t_{6;0,925} = 2{,}44691$ y por tanto $|t| = 0{,}55948 < 2{,}44691$ y por tanto aceptamos la hipótesis nula.

El tratamiento para adelgazar es por tanto estadísticamente aceptable para una pérdida de peso de 4,5 kg. en dos semanas.

## Problema 6.18

**Al realizar un estudio sobre la resistencia de ciertos materiales se obtienen los siguientes resultados:**

| | | | | | |
|---|---|---|---|---|---|
| **33,72** | **24,77** | **31,05** | **32,80** | **25,78** | **35,45** |
| **25,26** | **27,99** | **29,43** | **27,00** | **23,22** | |

**Se pide:**

**(a) Realizar un constraste para ver si puede aceptarse que la varianza poblacional vale 25.**

**(b) Si consideramos que la varianza de la población es conocida e igual a 25, realizar contrastes unilaterales y bilaterales para constrastar si puede aceptarse que la media de la población vale 30.**

**(c) Repetir el apartado anterior pero considerando que la varianza de la población es desconocida.**

**Nota: Usar $\alpha = 0{,}05$ y considerar que la población es normal.**

---

(a) Si seguimos el esquema general usado en problemas anteriores, tenemos que el nivel de significación es $\alpha = 0{,}05$; las hipótesis a considerar son:

$$\begin{array}{ll} H_0: & \sigma^2 = 25 \\ H_1: & \sigma^2 \neq 25 \end{array}$$

Puesto que el tamaño de la muestra es $n = 11$, y nos indican que la población se distribuye normalmente, el test que nos permitirá aceptar o rechazar $H_0$ es:

Se acepta $H_0$ si

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \in \left[\chi^2_{\alpha/2}(n-1); \chi^2_{1-\alpha/2}(n-1)\right]$$

donde $\chi^2_{\alpha/2}(n-1)$ y $\chi^2_{1-\alpha/2}(n-1)$ son los valores de la función de distribución de la variable aleatoria Chi-cuadrado con $n-1$ grados de libertad.

Luego $\chi^2_{0{,}025}(10) = 3{,}247$ y $\chi^2_{0{,}975}(10) = 20{,}48$ y el estadístico $\chi^2$ en nuestro caso vale

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{10 \cdot 16{,}28}{25} = 6{,}512.$$

Así tenemos $\chi^2 = 6{,}512$ que está dentro de la región de aceptación del test. Luego no tenemos evidencia suficiente para rechazar $H_0$, por lo que se acepta como valor de $\sigma^2 = 25$.

(b) Si se considera que la varianza de la población es conocida, para efectuar contrastes sobre la media, el estadístico de decisión será

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

En el caso del test bilateral

$$\begin{array}{ll} H_0: & \mu = 30 \\ H_1: & \mu \neq 30 \end{array}$$

La región crítica del test bilateral es $Z > z_{1-\alpha/2}$ ó $Z < -z_{1-\alpha/2}$, siendo $z_{1-\alpha/2}$ el valor de la función de distribución normal estándar. Como $\alpha = 0{,}05$ entonces $z_{1-\alpha/2} = z_{0{,}975} = 1,96$.

En nuestro caso $Z = \dfrac{28{,}77 - 30}{5/\sqrt{11}} = -0{,}82$ y por tanto está dentro de la región de aceptación.

Si repetimos el contraste para el test unilateral

$$\begin{array}{ll} H_0: & \mu = 30 \\ H_1: & \mu < 30 \end{array}$$

La región crítica del test unilateral es $Z < -z_{1-\alpha}$, siendo $z_{1-\alpha}$ el valor de la función de distribución normal estándar. Como $\alpha = 0{,}05$ entonces $-z_{1-\alpha} = -z_{0{,}95} = -1{,}645$.

Y como $Z = -0{,}82$ está dentro de la región de aceptación.

(c) Si se considera que la varianza de la población es desconocida, para efectuar contrastes sobre la media, el estadístico de decisión será

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

En el caso del test bilateral

$$\begin{array}{ll} H_0: & \mu = 30 \\ H_1: & \mu \neq 30 \end{array}$$

La región crítica del test bilateral es $t > t_{n-1;1-\alpha/2}$ ó $t < -t_{n-1;1-\alpha/2}$, siendo $t_{n-1;1-\alpha/2}$ el valor de la función de distribución $t$-Student con $n-1$ grados de libertad. Como $\alpha = 0{,}05$ entonces $t_{10;0{,}975} = 2{,}228$.

En nuestro caso $t = \dfrac{28{,}77 - 30}{4{,}03491/\sqrt{11}} = -1{,}01$ y por tanto está dentro de la región de aceptación.

Si repetimos el contraste para el test unilateral

$$\begin{array}{ll} H_0: & \mu = 30 \\ H_1: & \mu < 30 \end{array}$$

La región crítica del test unilateral es $t < -t_{n-1;1-\alpha}$, siendo $t_{n-1;1-\alpha}$ el valor de la función de distribución $t$-Student con $n-1$ grados de libertad. Como $\alpha = 0{,}05$ entonces $-t_{10;0{,}95} = -1{,}812$.

Al ser $t = -1{,}01 > -t_{n-1;1-\alpha} = -1{,}812$ aceptamos la hipótesis nula.

# Problema 6.19

**Una empresa manda reparar su ordenadores a dos servicios de asistencia técnica. Tras realizar un seguimiento de los tiempos que tardan en la reparación se obtienen los siguientes datos de tiempos de reparación (en minutos):**

| **Servicio 1** | **102** | **86** | **98** | **109** | **92** | | |
|---|---|---|---|---|---|---|---|
| **Servicio 2** | **81** | **165** | **97** | **134** | **92** | **87** | **114** |

**(a) Contrastar la hipótesis de que las dos compañías tardan lo mismo en realizar las reparaciones. (Usar $\alpha = 0{,}05$). Suponer que ambas distribuciones son normales con la misma varianza.**

**(b) Determinar un intervalo de confianza para el cociente de varianzas del 90%. ¿Es adecuada la suposición que se hizo en el apartado anterior de que las varianzas eran iguales?**

(a) Queremos realizar un contraste de igualdad de medias para dos poblaciones normales con varianzas desconocidas pero iguales. El test sería:

$$\begin{array}{ll} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{array}$$

El estadístico de decisión del test será:

$$t = \frac{\bar{X}_2 - \bar{X}_1}{S_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

siendo $S_p = \sqrt{\dfrac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-1}}$ donde el estadístico $t$ se distribuye según una $t$-Student con $n_1 + n_2 - 2$ grados de libertad.

Si calculamos las medias y desviaciones típicas muestrales tenemos:

$$\begin{array}{lll} n_1 = 5 & \bar{X}_1 = 97{,}4 & S_1 = 8{,}8769 \\ n_2 = 7 & \bar{X}_2 = 110 & S_2 = 30{,}221 \end{array}$$

La desviación típica común vale $S_p = 24{,}073221$.

Así el valor del estadístico $t$ es:

$$t = \frac{12{,}6}{24{,}073221 \cdot 0{,}58554} = 0{,}893881.$$

La región crítica del test bilateral es $t > t_{1-\alpha/2}$ ó $t < -t_{1-\alpha/2}$, siendo $t_{1-\alpha/2}$ el valor de la función de distribución $t$-Student con $n_1+n_2-2$ grados de libertad. Como $\alpha = 0{,}05$ entonces $t_{10;0{,}975} = 2{,}228$.

Al ser $t = 0{,}893881$ que no está dentro de la región de rechazo del test, y por tanto aceptamos la hipótesis de igualdad de medias.

(b) Como queremos determinar un intervalo de confianza para el cociente de varianzas de dos poblaciones normales, tenemos

$$\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \cdot F_{1-\alpha/2}(v_2, v_1)$$

donde $F_{1-\alpha/2}(v_2, v_1)$ y $F_{1-\alpha/2}(v_1, v_2)$ corresponden a los valores de los percentiles de la distribución $F$-Snedecor con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad.

Como en nuestro caso, tenemos que $\alpha = 0{,}1$ si miramos en la tabla de percentiles de la $F$-Snedecor obtenemos $F_{0,95}(4,6) = 4{,}533677$ y $F_{0,95}(6,4) = 6{,}163132$.

Como ya tenemos calculadas las desviaciones típicas muestrales, si sustituimos, tenemos que un intervalo de confianza del 90% para el cociente de varianzas es

$$\left[\frac{78{,}8}{913{,}3333}\cdot\frac{1}{4{,}533677} \quad ; \quad \frac{78{,}8}{913{,}3333}\cdot 6{,}163132\right] = [0{,}01903 \quad ; \quad 0{,}5317].$$

Si ambas varianzas fuesen iguales el cociente debería ser aproximadamente 1. Como en este caso el valor 1 no pertenece al intervalo dado, concluimos que las varianzas no son iguales. Por tanto, la suposición que se hizo en el apartado anterior no es adecuada.

# Problemas Propuestos

1. Obtener un intervalo de confianza para $\sigma_1/\sigma_2$ del 98% de poblaciones normales, si se tiene:

$$n_1 = 15 \quad ; \quad S_1 = 0{,}07 \quad ; \quad n_2 = 12 \quad ; \quad S_2 = 0{,}8.$$

2. Los siguientes son los pesos de 10 paquetes de semillas (en kg):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 46 | 45,2 | 46,4 | 46,1 | 45,8 | 47 | 46,1 | 45,9 | 45,8 | 46,9 |

   Encontrar un intervalo de confianza del 95% para la varianza de todos los paquetes de semillas, suponiendo una población normal.

3. De una población normal se ha obtenido una muestra aleatoria de tamaño 20

| | | | | |
|---|---|---|---|---|
| 6,166 | 8,196 | 12,606 | 0,902 | 0,631 |
| 3,667 | 10,758 | 20,047 | 4,668 | 2,547 |
| 5,886 | 9,407 | 21,265 | 6,117 | 3,849 |
| 6,887 | 24,471 | 22,861 | −3,585 | −6,462 |

   Calcular el intervalo de confianza del 95% para la varianza poblacional.

4. Una muestra aleatoria de 8 cigarrillos de una marca determinada tiene un contenido promedio de nicotina de 2,6 miligramos y una desviación típica $S = 0{,}9$ miligramos. Suponiendo que la distribución de los contenidos de nicotina son aproximadamente normales, determinar un intervalo de confianza del 99% para $\sigma$.

5. En una muestra aleatoria de $n = 500$ familias que tienen televisores en una determinada ciudad, se encontró que 340 habían visto un programa de noticias. Encontrar un intervalo de confianza del 95% para

la proporción actual de familias en esta ciudad que se informan con dicho noticiario televisivo.

6. Los contenidos de 7 recipientes similares de ácido sulfúrico son (en litros):

$$9{,}8 \quad 10{,}2 \quad 10{,}4 \quad 9{,}8 \quad 10{,}0 \quad 10{,}2 \quad 9{,}6.$$

Encontrar un intervalo de confianza del 95% para la media de todos los recipientes, suponiendo una distribución normal.

7. Una muestra aleatoria de 100 propietarios de automóvil indica que, en la comunidad autónoma andaluza, un automóvil recorre un promedio de 23500 km. al año con una desviación típica $S = 3900$ km. Determinar un intervalo de confianza del 99% para el promedio de kilómetros que un automóvil recorre anualmente en dicha comunidad.

8. Un fabricante de baterías para automóvil asegura que sus baterías duran, en promedio, 3 años con una desviación típica de un año. Si 5 de estas baterías tienen duraciones (en años) de

$$1{,}9 \quad 2{,}4 \quad 3{,}0 \quad 3{,}5 \quad 4{,}2.$$

Determinar un intervalo de confianza del 95% para $\sigma^2$ e indicar si es válida la afirmación del fabricante de que $\sigma^2 = 1$. Suponer que la duración de las baterías se distribuye en forma normal.

9. Se obtiene una muestra de 20 estudiantes con una media muestral $\bar{X} = 72$ y una cuasivarianza muestral $S^2 = 16$ en un examen de matemáticas. Supongamos que las calificaciones tienen una distribución normal. Determinar un intervalo de confianza del 98% para $\sigma^2$.

10. El consumo regular de cereales azucarados contribuye a la caída de los dientes, enfermedades de corazón y otros procesos degenerativos. En un estudio realizado sobre una muestra de 20 porciones de un cereal, el contenido promedio de azúcar fue de 11,3 g. con una desviación típica $S = 2{,}45$ g. Suponiendo que los contenidos de azúcar están distribuidos normalmente, determinar un intervalo de confianza del 95% para $\sigma$.

11. Una máquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de piezas cuyos diámetros son (en cm.):

$$0{,}99 \quad 1{,}01 \quad 0{,}97 \quad 1{,}03 \quad 1{,}04 \quad 0{,}99 \quad 0{,}98 \quad 1{,}01 \quad 1{,}03.$$

Si se supone una distribución aproximadamente normal. Determinar un intervalo de confianza del 99% para $\sigma^2$.

12. Se selecciona una muestra aleatoria de 200 votantes y se encuentra que 114 respaldan un convenio. Encontrar el intervalo de confianza del 96% para la fracción de la población de votantes que es favorable a dicho convenio.

13. Se selecciona una muestra aleatoria de 500 fumadores de cigarrillos y se encuentra que 86 de ellos prefieren la marca X. encontrar un intervalo de confianza del 90% para la proporción de la población de fumadores que prefieren la marca X.

14. En una muestra aleatoria de 1000 casas en una determinada ciudad, se encuentra que 228 de ellas tienen calefacción. Encontrar un intervalo de confianza del 99% para la proporción de hogares de esta ciudad que tienen calefacción.

15. Las alturas de una muestra aleatoria de 50 estudiantes mostraron una media de 174,5 cm. y una desviación típica muestral $S = 6{,}9$ cm. Determinar un intervalo de confianza del 98% para la altura promedio de todos los estudiantes de dicho centro.

16. Se toma una muestra aleatoria de 12 agujas de tejer en un estudio de la dureza de la cabeza de las agujas. Se realizan las mediciones de la dureza para cada una de las 12 piezas, de lo que se obtiene un valor promedio de 48,5 con una desviación típica $S = 1{,}5$. Suponiendo que las mediciones están normalmente distribuidas, determinar un intervalo de confianza del 90% para la dureza promedio.

17. Una muestra de 12 alumnos de una academia mecanografiaron un promedio de 79,3 palabras por minuto con una desviación típica $S = 7{,}8$ palabras por minuto. Suponiendo una distribución normal para la cantidad de palabras mecanografiadas por minuto, encontrar un intervalo de confianza del 95% para el número promedio de palabras mecanografiadas por minuto por todos los alumnos de esta academia.

18. Una muestra aleatoria de 25 cigarros de una marca determinada tiene un contenido promedio de nicotina de 1,3 miligramos y una desviación típica $S = 0{,}17$ miligramos. Encontrar un intervalo de confianza del 95% para el promedio de nicotina de esta marca.

19. Se registraron las siguientes mediciones del tiempo de secado, en horas, de una marca de pintura acrílica:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4,8 | 4,0 | 5,2 | 3,4 | 2,5 | 4,8 | 2,9 | 3,6 |
| 2,8 | 3,3 | 5,6 | 3,7 | 2,8 | 4,4 | 3,0. | |

Suponiendo que las mediciones representan una muestra aleatoria de una población normal, encontrar un intervalo de confianza del 99% para el tiempo de secado de esta pintura.

20. Con referencia al problema número 10, determinar un intervalo de confianza del 95% para el contenido promedio de azúcar de dicho cereal.

21. Con referencia al problema número 4, determinar un intervalo de confianza del 99% para el contenido promedio real de nicotina de esta marca de cigarrillos.

22. Con referencia al problema número 11, determinar un intervalo de confianza del 99% para el diámetro promedio de piezas de esta máquina.

23. Un fabricante produce focos que tienen un promedio de vida con distribución aproximadamente normal y una desviación típica $\sigma = 40$ horas. Si una muestra de 30 focos tiene una vida promedio de 780 horas, encontrar un intervalo de confianza del 96% para la media poblacional de todos los focos que produce esta empresa.

24. Una máquina de refrescos está ajustada de tal manera que la cantidad de líquido despachada se distribuye aproximadamente en forma normal con una desviación típica $\sigma = 0.15$ decilitros. Encontrar un intervalo de confianza del 95% para la media de todos los refrescos que sirve esta máquina si una muestra aleatoria de 36 refrescos tiene un contenido promedio de 2,25 decilitros.

25. El tiempo de fallo de un dispositivo es una variable aleatoria exponencial con parámetro $\lambda$. Se ponen a prueba 10 de estos dispositivos y se encuentra que la suma de sus tiempos de fallo es 111. Calcular un intervalo de confianza del 95% para $\lambda$.

26. Otro dispositivo tiene un tiempo de fallo que se distribuye según una exponencial con parámetro $\lambda$. Se prueban 20 dispositivos y se encuentra que la suma de tiempos de fallo es 20169 horas. Calcular un intervalo de confianza del 90% para $\lambda$.

27. Un experto en gestión de calidad desea determinar un intervalo de confianza para el tiempo promedio que se necesita para hacer tres perforaciones en una cierta pieza metálica. Se calcula el tiempo promedio de una muestra aleatoria de 36 trabajadores, resultando 2,6 segundos. Encontrar los intervalos de confianza del 95% y del 99% para dicho tiempo de perforación. Supongamos que la desviación típica de la población es conocida y vale $\sigma = 0{,}3$ segundos.

28. Se registraron los tiempos utilizados en la compra para 64 clientes seleccionados al azar en un supermecado. La media de esos 64 tiempos de compra fue de 33 minutos y la cuasivarianza $S^2 = 256$. Estime el promedio real del tiempo utilizado por los clientes en la compra por medio de un intervalo de confianza del 99%.

29. Un método para resolver la carencia de energía eléctrica requiere de la construcción de plantas eléctricas nucleares flotantes unas millas mar adentro. Se necesita una estimación de la densidad del tráfico naval en el área, porque existe una preocupación respecto a una posible colisión por parte de un barco con la planta flotante (aunque anclada). El número de barcos que pasan dentro de un radio de 10 millas de la ubicación propuesta de la planta eléctrica, registrado durante 60 días en julio y agosto, tuvo una media y una cuasivarianza muestral de: $\bar{X}_1 = 7{,}2$ y $S_1^2 = 8{,}8$.

    (a) Obtener un intervalo de confianza del 95% para el promedio de número de barcos que pasan dentro de un radio de 10 millas de la localización propuesta de la planta eléctrica durante un día.

    (b) Se esperaba que la densidad del tráfico naval disminuyera en los meses de invierno. De una muestra de 90 observaciones de barcos durante el invierno, se obtuvieron los siguientes valores de la media y cuasivarianza: $\bar{X}_2 = 4{,}7$ , $S_2^2 = 4{,}9$. Obtener un intervalo de confianza del 90% para la diferencia en la densidad media del tráfico naval entre los meses de verano e invierno.

30. La experiencia muestra que la desviación típica del ingreso anual de trabajadores de la rama industrial en cierto estado es de 400 dólares. ¿Cuántos trabajadores de dicha rama tendrían que ser seleccionados si se quisiera estimar la media poblacional con un error máximo de 50 dólares, con una probabilidad de 0,95?

31. En una planta que fabrica cemento las producciones diarias para la semana pasada fueron 785, 805, 790, 793 y 802 toneladas.

(a) Estimar, a partir de los datos, la media de la producción diaria con un coeficiente de confianza de 0,90.

(b) ¿Se puede afirmar, al mismo nivel de confianza, que el promedio de producción diaria es inferior a 800 toneladas?

32. Se desea estimar la distancia promedio que recorren los empleados de una empresa de camino a su trabajo. Los estudios anteriores de este tipo indican que la desviación típica de esas distancias debe estar cercana a los 2 km. ¿Cuántos empleados deben muestrearse si la estimación debe quedar a menos de 0,1 km. del promedio verdadero, con un coeficiente de confianza de 0,95?

33. El alambre pretensado para reforzar ciertos tubos, se fabrica en rollos grandes. En una inspección de control de calidad se prueban 5 muestras de un rollo y se mide su resistencia máxima a la tensión. Los resultados de las mediciones, en miles de $1b/pulg^2$, fueron 253, 261, 258, 255 y 256. Con estos datos determinar una estimación del intervalo de confianza del 95% para el promedio verdadero de resistencia a la tensión máxima.

34. Por un procedimiento aleatorio se extraen 16 elementos muestrales de una población normal con $\bar{X} = 41{,}5$ y $S = 2{,}795$. ¿Hay motivos para rechazar la hipótesis de que la media de la población es 43? (Usar $\alpha = 0{,}05$.)

35. El vicepresidente del departamento de ventas de una gran coorporación afirma que los vendedores tienen un promedio no mayor de 15 partes de venta por semana. (Desearía aumentar esta cifra). Se seleccionan 36 vendedores al azar para verificar su afirmación y se registra el número de contactos en una sola semana de forma aleatoria. La muestra tiene una media de 15 prospectos y una cuasivarianza de 9.

    ¿Contradicen los hechos la afirmación del vicepresidente? Utilizar un nivel de significación $\alpha = 0{,}005$.

36. Los salarios diarios en la industria naval presentan una distribución normal con media $\mu = 13{,}20\$$ y desviación típica $\sigma = 2{,}50\$$. Si en esta industria, una compañía que fabrica motores para barcos emplea a 40 trabajadores y les paga un promedio de 12,20$, ¿puede acusarse a esta compañía de pagar salarios inferiores a los del sector naval? Utilizar $\alpha = 0{,}01$.

37. El voltaje de salida de un cierto circuito eléctrico debe ser igual a 130, según se especifica. Una muestra de 40 lecturas independientes para este circuito dió una media muestral de 128,6 y una desviación típica $S = 2{,}1$.

    Probar la hipótesis de que el voltaje de salida por término medio es 130, frente a la hipótesis alternativa de que es inferior a 130. Utilizar un nivel de significación del 5%.

38. El índice de dureza de Rockwell para acero se determina al rayar el acero con un punzón de diamante y medir la profundidad de la penetración. En 50 muestras de cierto tipo de acero el índice de dureza de Rockwell tuvo una media de 62 y una desviación típica $S = 8$. El productor afirma que este acero tiene un índice de dureza promedio de por lo menos 64. ¿Hay suficiente evidencia para refutar la afirmación del productor con nivel de significación del 1%?

39. Una máquina expendedora de café que funciona a base de monedas, se diseñó para servir, en promedio, 7 centilitros de bebida por vaso. Con el objeto de verificar lo anterior se eligieron 10 vasos llenos de bebida y se midieron los contenidos. La media y la desviación típica de las diez mediciones fueron $\bar{x} = 7{,}1$ cl. y $S = 0{,}12$ cl., respectivamente.

    (a) ¿Presentan estos datos suficiente evidencia para afirmar que la descarga media difiere de 7 cl.? Utilizar $\alpha = 0{,}01$.

    (b) ¿Qué puede decirse si $\alpha = 0{,}1$?

40. Una encuesta realizada por la revista *Investigación y Desarrollo industrial* reveló que los sueldos del personal de investigación y desarrollo aumentaron un 9,6% en 1998 hasta alcanzar un promedio de 31221 euros al año. Si una muestra aleatoria de sueldos de 20 expertos en física tuvo una media de 33120 y una desviación típica $S = 2140$, ¿concluiría que se paga más en promedio a los físicos que la media global de 31221 euros por año? Utilizar $\alpha = 0{,}05$.

41. Se enseñó a dos grupos de niños de la escuela primaria a leer por dos métodos diferentes, 50 por cada método. Al terminar el periodo de instrucción, una prueba de lectura dió los siguientes resultados:

$$\bar{Y}_1 = 74 \quad , \quad \bar{Y}_2 = 71 \quad , \quad S_1 = 9 \quad , \quad S_2 = 10$$

    (a) ¿Cuál es el nivel de significación si desea verificarse si hay evidencia de una diferencia real entre las dos medias poblacionales?

(b) ¿Cuál será la conclusión si utilizamos $\alpha = 0{,}05$?

42. Se efectuó un estudio por parte de la comisión de medio ambiente del Patronato de Doñana para estimar las cantidades de residuos químicos encontrados en los tejidos celebrales de patos malvasía. En una prueba sobre DDT, muestras aleatorias de $n_1 = 10$ patos jóvenes y $n_2 = 13$ polluelos recién nacidos dieron como resultados:

$$\bar{Y}_1 = 0{,}041 \quad , \quad \bar{Y}2 = 0{,}026 \quad , \quad S_1 = 0{,}017 \quad , \quad S_2 = 0{,}006$$

Probar la hipótesis de que no existe diferencia en las cantidades promedio de DDT encontradas en ambas muestras, frente a la alternativa de que los patos jóvenes presentan un índice mayor que los polluelos. Utilizar $\alpha = 0{,}05$.

43. La resistencia del "concreto" depende del método que se utiliza para el secado. Dos diferentes métodos de secado mostraron como resultados:

$$\begin{array}{lll} n_1 = 7 \quad , & \bar{Y}_1 = 3250 \quad , & S_1 = 210 \\ n_2 = 10 \quad , & \bar{Y}_2 = 3240 \quad , & S_2 = 190 \end{array}$$

¿Parecen producir los dos métodos "concretos" con una resistencia diferente? Utilizar $\alpha = 0{,}05$.

44. Una muestra aleatoria de tamaño 25 de una población normal con desviación típica conocida $\sigma_1 = 5{,}2$ nos da una media de 81. Otra muestra de tamaño 36 para una población normal de desviación típica $\sigma_2 = 3{,}4$ nos da una media de 76. Probar la hipótesis de que ambas medias sean iguales frente a que sean distintas. Utilizar $\alpha = 0{,}05$.

45. Se quiere comprobar si tres grupos de alumnos de distintos colegios obtienen las mismas notas en selectividad.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| A | 7,3 | 8,9 | 8,2 | 4,3 | 8,0 | 7,3 | 6,6 | 6,0 |
| B | 8,8 | 7,8 | 4,8 | 9,1 | 5,1 | 8,5 | 7,4 | 7,7 |
| C | 6,8 | 7,9 | 5,6 | 9,1 | 7,1 | 7,1 | 8,7 | 4,1 |
| A | 4,5 | 9,3 | 3,6 | 7,7 | | | | |
| B | 3,1 | 7,8 | 6,2 | 7,6 | 9,6 | 8,0 | 5,6 | |
| C | 5,9 | 6,8 | 5,3 | 7,9 | 1,5 | | | |

Probar la hipótesis de que ambas medias sean iguales frente a que sean distintas. Utilizar $\alpha = 0{,}05$.

46. Se afirma que una de las piezas de un motor, producida por una compañía, tiene una varianza del diámetro no mayor que 0,0002 pulgadas$^2$. Una muestra aleatoria de 10 informes reveló una cuasivarianza muestral de 0,0003 pulgadas$^2$.

    ¿Se pude afirmar que la varianza es mayor de 0,002? Contrastar a un nivel de significación de 0,05.

47 Queremos comparar la varianza de la compañía anterior con la de su competidor. Tomando una muestra para éste de tamaño 20, se obtiene una varianza de 0,0001 pulgadas$^2$. ¿Presentan los datos suficiente información para indicar una menor varianza en los diámetros fabricados por el competidor? Utilizar $\alpha = 0{,}05$.

48. Se desea estimar, con una confianza del 99%, la intensidad media que circula por una componente de un circuito en circunstancias diversas. Se supone que la intensidad, en miliamperios, sigue una distribución normal de varianza 144. Llevadas a cabo 25 medidas en instantes elegidos al azar, se obtuvo una media muestral de $\bar{x} = 85$ mA.

49 Los siguientes datos son los volúmenes de ventas diarios de una empresa, expresados en millones de pesetas, durante 35 días elegidos al azar

| | | | | | | |
|---|---|---|---|---|---|---|
| 0,360 | 1,185 | 0,524 | 0,870 | 0,356 | 2,567 | 0,566 |
| 1,789 | 0,578 | 0,578 | 0,892 | 0,345 | 0,256 | 0,987 |
| 0,355 | 0,989 | 0,412 | 0,453 | 1,987 | 0,544 | 0,798 |
| 0,634 | 0,355 | 0,455 | 0,445 | 0,755 | 0,423 | 0,754 |
| 0,452 | 0,452 | 0,450 | 0,511 | 1,234 | 0,543 | 1,501 |

Un histograma de estos datos mostraría claramente una fuerte asimetría a la derecha, lo cual sugiere que los valores de actividad no siguen una distribución normal. Determinar un intervalo de confizanza del 95% para el volumen de ventas.

50. Se quiere determinar un intervalo de confianza del 95% de la proporción de ciudadanos que juegan al bingo, lotería, quinielas, etc. Para ello se seleccionó una muestra simple de 200 personas con la que se estimó que el 15% tienen este tipo de hábito.

51. Admitiendo que el número de erratas por página de cierto libro sigue una distribución de Poisson, se quiere determinar un intervalo de confianza al 95% del número promedio de erratas por página que contiene.

Para ello se eligieron al azar y con reemplazamiento 100 páginas en las que se observó una media muestral de $\bar{x} = 0{,}04$ erratas por página.

52. Se quiere determinar un intervalo de confianza para la varianza de la variable del problema número 49. (a) Dar dicho intervalo con nivel de confianza del 99%, (b) con nivel de confianza del 90%.

53. Con objeto de estudiar la efectividad de un compresor de imágenes, 6 de ellas fueron capturadas en ficheros sin comprimir mientras otras 5 fueron capturadas y comprimidas. El tamaño de los archivos, en Kb, resultó ser

| | | | | | | |
|---|---|---|---|---|---|---|
| Sin compresión: | 20,4 | 62,5 | 61,3 | 44,2 | 11,1 | 23,7 |
| Comprimidas: | 1,2 | 6,9 | 38,7 | 20,4 | 17,2 | |

Supuesto que el tamaño de una imagen, tanto en un caso como en otro, sigue una distribución normal de parámetros desconocidos, dar el intervalo de confianza para el cociente de varianzas poblacionales, con nivel de confianza del 95%. ¿Sugiere el intervalo de confianza admitir que ambas varianzas poblacionales son sensiblemente iguales?

54. Con referencia al ejercicio anterior, determinar el intervalo de confianza para la diferencia de medias poblacionales con nivel de confianza del 95%.

55. Se intenta comparar el efecto de una temperatura elevada sobre dos tipos de procesadores. Para ello, se hicieron trabajar a altas temperaturas 250 equipos con el primer procesador y 200 con el segundo. 183 de los primeros y 90 de los segundos manifestaron un funcionamiento anormal. Determinar el intervalo de confianza del 95% para la diferencia de proporciones de procesadores sensibles al calor.

56. Con objeto de averiguar si el calor disipado por su funcionamiento hace variar significativamente la eficiencia de cierto tipo de procesador, se midió el tiempo empleado en ejecutar una serie de programas, nada más iniciarse el funcionamiento del equipo $X_i$, y tras dos horas de funcionamiento $Y_i$. Los datos obtenidos, en milisegundos, fueron

| | | | | | |
|---|---|---|---|---|---|
| $X_i$ | 169,7 | 168,5 | 165,9 | 177,8 | 179,6 |
| $Y_i$ | 168,2 | 165,5 | 164,4 | 175,7 | 176,6 |
| $X_i$ | 168,9 | 169,2 | 167,9 | 181,8 | 163,3 |
| $Y_i$ | 166,1 | 167,1 | 166,3 | 179,7 | 161,5 |

Determinar un intervalo de confianza del 95% para la disminución media del tiempo de ejecución.

57. Se quiere analizar la precisión de dos *scaner* en la conversión en ficheros ASCII de documentos mecanografiados, determinando un intervalo de confianza para la diferencia del número medio de palabras "mal leídas" por página.

    Para ello se seleccionaron al azar 40 páginas, anotándose el número de palabras "mal leídas" por el primer *scaner*, así como el de palabras "mal leídas" por el segundo, en otras 40 páginas elegidas también al azar. Si el número medio de tales palabras fue, respectivamente, $\bar{x}_1 = 4$ y $\bar{x}_2 = 3$ con la misma dispersión, $S_1 = S_2 = 2$, determinar un intervalo de confianza para la diferencia de medias con nivel de confianza del 95%.

58. Se quiere determinar un intervalo de confianza del 95% para el número medio de accidentes de tráfico en cada fin de semana, en una determinada zona. Se supone que dicho número sigue una distribución de Poisson de parámetro desconocido y se dispone de los datos correspondientes a 30 semanas en las que se han observado 217 accidentes.

59. Se seleccionó una muestra aleatoria de 100 muertes durante un año, dando una vida promedio de 71,8 años. Suponiendo una desviación típica poblacional $\sigma = 8{,}9$ años. ¿Se podría afirmar que la vida promedio hoy en día es mayor de 70 años, con nivel $\alpha = 0{,}05$ ?

60. Un fabricante ha desarrollado un nuevo sedal sintético para pesca que se considera tiene resistencia a la ruptura de 8 kg. con una desviación típica $\sigma = 0{,}5$ kg. Probar la hipótesis de que $\mu = 8$ kg. frente a que $\mu \neq 8$ kg. Se prueban 50 sedales y se encuentra que tienen una resistencia promedio de 7,8 kg. Usar un nivel $\alpha = 0{,}01$.

61. El Instituto de consumo afirma que un aspirador consume en promedio 46 kilovatios-hora al año. Si una muestra de 12 hogares indica que los aspiradores consumen en promedio 42 kilovatios-hora al año con una desviación típica $S = 11{,}9$ kilovatios-hora, ¿sugiere esto que los aspiradores consumen, en promedio, menos de 46 kilovatios-hora al año, con nivel 0,05? Supongamos que la población es normal.

62. Se llevó a cabo un experimento para comparar el deterioro abrasivo de dos materiales laminados diferentes. Se probaron 12 piezas del material 1 y 10 del material 2, se observó la profundidad del deterioro. Las

muestras del material 1 dieron un deterioro promedio de 85 unidades con una desviación típica $S_1 = 4$, mientras para el material 2 dieron un promedio de 81 y desviación típica $S_2 = 5$. ¿Puede decirse que el deterioro abrasivo del material 1 excede al del material 2 en más de 2 unidades con un nivel de significación de 0,05? Suponer poblaciones normales con varianzas desconocidas pero iguales.

63. Se cree que un medicamento en el mercado, que por lo común se prescribe para aliviar la tensión nerviosa es eficaz sólo en el 60% de los casos. Un nuevo medicamento se administra a una muestra de 100 adultos que sufren tensión nerviosa, de los que 70 experimentaron alivio. ¿Es evidencia suficiente para deducir que el nuevo medicamento es mejor que el que se suministraba comunmente? Usar un nivel de significación 0,05.

64. Supóngase que se desea probar la hipótesis

$$\begin{aligned} H_0 &: \quad \mu = 68 \text{ kg.} \\ H_1 &: \quad \mu > 68 \text{ kg.} \end{aligned}$$

para los pesos de un grupo de hombres. Si se sabe que $\sigma = 5$. Con un nivel de significación $\alpha = 0{,}05$ determinar el tamaño de la muestra, si la potencia debe ser de 0,95 cuando la media verdadera es de 69 kg.

65. Se va a construir una planta química en un pueblo. Para determinar si hay diferencia significativa entre la proporción de votantes del pueblo y de las cercanías que están a favor de la construcción, se toma un grupo de ellos. Si 120 de 200 votantes del pueblo y 240 de 500 de las cercanías están a favor de la propuesta, ¿estaríamos de acuerdo en que la proporción de votantes del pueblo es mayor que la proporción de votantes de cercanías que están a favor de la construcción de dicha planta? Usar un nivel de 0,025.

66. Un fabricante de baterías para automóvil asegura que la duración de sus baterías tiene una distribución aproximadamente normal con desviación típica conocida $\sigma = 0{,}9$ años. Si una muestra de 10 de éstas tiene una desviación $S = 1{,}2$ años. ¿Es $\sigma > 0{,}9$ años con un nivel de 0,05?

67. Al probar la diferencia en el desgaste de dos materiales en el problema número 62, se supuso que las varianzas eran iguales pero desconocidas. ¿Es correcto que $\sigma_1 = \sigma_2$ con un nivel de 0,1?

68. Una firma de cigarros distribuye dos marcas. Si se encuentra que 56 de 200 fumadores prefieren la marca A y que 29 de 150 fumadores prefieren la marca B, ¿puede concluirse que la marca A aventaja en ventas a la B con nivel de significación 0,06?

69. En un colegio se estima que a lo sumo el 25% de los estudiantes se traslada a clase en bicilceta. ¿Es válida esta afirmación si, en una muestra de 90 estudiantes, se encuentra que 28 usan este transporte? Usar nivel de significación de 0,05.

70. En un cierto estudio para estimar la proporción de residentes de una ciudad y sus suburbios que está de acuerdo con la construcción de una planta de energía nuclear, se encontró que 63 de 100 residentes urbanos favorecen la construcción, mientras que 59 de 125 residentes suburbanos se oponen. ¿Existe alguna diferencia significativa entre las proporciones de residentes urbanos y suburbanos que favorecen la construcción de la planta nuclear?

71. Se sabe que la capacidad de los recipientes de un determinado lubricante tiene una distribución normal con varianza $\sigma^2 = 0{,}03$.

    (a) Probar la hipótesis $\sigma^2 = 0{,}03$ en contraposición a la alternativa $\sigma^2 \neq 0{,}03$ para la muestra aleatoria de 10 recipientes (en litros): (Usar $\alpha = 0{,}01$).

    10,2 9,7 10,1 10,3 10,1 9,8 9,9 10,4 10,3 9,8

    (b) Probar la hipótesis de que el contenido promedio es de 10 litros con nivel de significación de 0,01.

72. Se sabe que el contenido de nicotina de una marca de cigarros es aproximadamente normal con una varianza $\sigma^2 = 1{,}3$. Probar la hipótesis $\sigma^2 = 1{,}3$ en contraposición a la alternativa $\sigma^2 \neq 1{,}3$ si una muestra aleatoria de 8 de éstos tiene una desviación típica $S = 1{,}8$ mg. Utilizar un nivel $\alpha = 0{,}05$.

73. Se lleva a cabo un estudio para comparar el tiempo que tardan hombres y mujeres en armar un producto determinado. Las experiencias anteriores indican que la distribución de tiempos tanto para hombres como para mujeres es aproximadamente normal. Pero la varianza de los tiempos para las mujeres es menor que la de los hombres. Una muestra aleatoria de tiempos para 11 hombres y 14 mujeres arroja los

siguientes datos

| | | |
|---|---|---|
| Hombres | $n_1 = 11$ | $S_1 = 6{,}1$ |
| Mujeres | $n_2 = 14$ | $S_2 = 5{,}3.$ |

Probar la hipótesis de que $\sigma_1^2 = \sigma_2^2$ en contraposición a la alternativa $\sigma_1^2 > \sigma_2^2$. Usar $\alpha = 0{,}01$.

74. Una gran compañía de automóviles quiere decidir si comprar una marca A o una marca B de neumáticos. Para decidirse se lleva a cabo un experimento utilizando 12 de cada marca. Los resultados son

| | | |
|---|---|---|
| Marca A | $\bar{x}_1 = 37900$ km | $S_1 = 5100$ km |
| Marca B | $\bar{x}_2 = 39800$ km | $S_2 = 5900$ km |

Probar la hipótesis de que no hay diferencia en el promedio recorrido en las dos marcas de neumáticos. Usar nivel 0,05. Supongamos que las dos poblaciones son aproximadamente normales.

75. Con referencia al problema anterior, probar la hipótesis de que $\sigma_1 = \sigma_2$ en contraposición a la alternativa $\sigma_1 < \sigma_2$ donde $\sigma_1$ y $\sigma_2$ son las desviaciones típicas de las distancias recorridas por los neumáticos de las marcas A y B. Usar un nivel de significación de 0,05.

76. Se afirma que una máquina que sirve refrescos está fuera de control si la varianza de los contenidos excede de 1,15 $\mathrm{dl}^2$. Si una muestra aleatoria de 25 refrescos de esta máquina tien una cuasivarianza $S = 2{,}03$ $\mathrm{dl}^2$, ¿indica ésto, con en un nivel 0,05, que la máquina está fuera de control? Supongamos que los contenidos tienen distribución aproximadamente normal.

77 Una muestra de 36 refrescos de una máquina automática tiene un contenido promedio de 21,9 decilitros por refresco, con una desviación típica $S = 1{,}42$ decilitros. Decidir si se acepta el test con una hipótesis nula $H_0 : \mu = 22{,}2$ decilitros frente a $H_1 : \mu < 22{,}2$, y un nivel de significación 0,05.

78. ¿Qué tamaño de la muestra se requiere en el problema anterior si la potencia de la prueba debe ser 0,9 cuando la media verdadera es 21,3 decilitros? Supongamos que la desviación típica es conocida $\sigma = 1{,}42$ decilitros.

79. La altura promedio de un colectivo de mujeres es de 162,5 cm con una desviación típica de 6,9 cm. ¿Hay alguna razón para creer que existe un cambio en la altura promedio si una muestra aleatoria de 50 mujeres del grupo actual tiene una altura promedio de 165,2 cm?

80. ¿Qué tamaño se requiere para una muestra en el problema anterior si la potencia es 0,95 cuando la altura promedio real difiere de 162,5 (altura promedio del colectivo) en 3,1 cm?

81. Una compañía de carburantes asegura que una quinta parte de los hogares en una cierta ciudad se calientan con gasoil. ¿Se tiene alguna razón para dudar de esta afirmación si, en una muestra de 1000 hogares se encuentra que 236 se calientan con gasoil? Utilizar un nivel 0,05.

82. Un fabricante afirma que la resistencia promedio a la tensión de los tornillos A excede la de los tornillos B al menos en 12 kg. Para probar esta afirmación se examinan 50 piezas de cada tipo. El tornillo A tuvo una resistencia promedio $\bar{x}_1 = 86{,}7$ kg con una desviación típica $S_1 = 6{,}28$ kg; mientras para el tornillo tipo B estos mismos parámetros fueron $\bar{x}_2 = 77{,}8$ kg y $S_2 = 5{,}61$ kg. Comprobar la afirmación del fabricante usando un nivel de significación de 0,05.

# Soluciones de los Problemas Propuestos.

# Capítulo 1

## 1.1

Parám. de centralización:

$$\overline{X} = 0,507; Me = 0; Mo = 0$$

Parám. de dispersión:

$Re = 4; Var(x) = 0,502; \sigma = 0,708; Cv = 139,64\%; Dc = 0,5; Dp_{10-90} = 0,5$

Otros:

$$As = 0,715; g_1 = 1,39; g_2 = 1,9 \text{ (Leptocúrtica)}$$

## 1.2

(a) Me=28

(b) $63,\widehat{63}$

## 1.3

Grafico

## 1.4

- Parám. de centralizacion:

  $$\overline{X} = 4,445; Me = 4,5; Mo = 5$$

  Parám. de dispersion:

  $Re = 7; Var(x) = 2,327; \sigma = 1,525; Cv = 34,31\%; Dc = 1,5; Dp_{10-90} = 2$

- $P_{20} = 3$

## 1.5

No, puesto que sería necesario conocer el número de vuelos totales de cada año para poder realizar un estudio comparativo. Un estudio más en profundidad nos llevaría a considerar otros factores como (calidad de los aparatos, nivel de aprendizaje de los pilotos, condiciones atmosféricas, accidentes debidos a causa mayor...)

## 1.6

$$\overline{X} = 325,5; Mo = 312,8\widehat{3}\ y\, 348,071; Me = 327; Var(x) = 2144$$

## 1.7

| $[L_{i-1} - L_i)$ | $n_i$ | $x_i$ | $f_i$ | $N_i \downarrow$ | $N_i \uparrow$ | $F_i \downarrow$ | $F_i \uparrow$ | $h_i$ |
|---|---|---|---|---|---|---|---|---|
| 1-3 | 3 | 2 | 0,03 | 3 | 100 | 0,03 | 1 | 1,5 |
| 3-7 | 29 | 5 | 0,29 | 32 | 97 | 0,32 | 0,97 | 7,25 |
| 7-8 | 35 | 7,5 | 0,35 | 67 | 68 | 0,67 | 0,68 | 35 |
| 8-10 | 26 | 9 | 0,26 | 93 | 33 | 0,93 | 0,33 | 13 |
| 10-13 | 6 | 11,5 | 0,06 | 99 | 7 | 0,99 | 0,07 | 2 |
| 13-20 | 1 | 16,5 | 0,01 | 100 | 1 | 1 | 0,01 | 0,14 |

## 1.8

Grafico

## 1.9

NOTA: consideraremos el último intervalo en 75

- **HOMBRES**

  Parám. de centralización:

$$\overline{X} = 25,368; Me = 23,9; Mo = 23,886$$

  Parám. de dispersion:

$$Re \doteq 75; Var(x) = 55,096; \sigma = 7,423; Cv = 29,3\%; Dc = 2,9; Dp_{10-90} = 5,916$$

- **MUJERES**

Parám. de centralizacion:

$$\overline{X} = 22,318; Me = 22,764; Mo = 23,92$$

Parám. de dispersion:

$$Re = 75; Var(x) = 66,789; \sigma = 8,172; Cv = 36,616\%; Dc = 2,325; Dp_{10-90} = 9,467$$

# 1.10

| $[L_{i-1} - L_i)$ | $x_i$ | $n_i$ | $N_i \downarrow$ | $f_i$ | $F_i \downarrow$ |
|---|---|---|---|---|---|
| 70-75 | 72,5 | 9 | 9 | 0,009 | 0,009 |
| 75-80 | 77,5 | 48 | 57 | 0,048 | 0,057 |
| 80-85 | 82,5 | 147 | 254 | 0,197 | 0,254 |
| 85-90 | 87,5 | 321 | 575 | 0,321 | 0,575 |
| 90-95 | 92,5 | 259 | 834 | 0,259 | 0,834 |
| 95-100 | 97,5 | 121 | 955 | 0,121 | 0,955 |
| 100-105 | 102,5 | 45 | 1000 | 1 | 1 |

- Parám. de centralizacion:

$$\overline{X} = 89,08; Me = 88,831; Mo = 87,839$$

- Parám. de dispersion:

$$Re = 35; Var(x) = 37,953; \sigma = 6,161; Cv = 6,91\%; Dc = 4,24; Dp_{10-90} = 8,317$$

# 1.11

Tipificando ambas variables ($\overline{Z} = 0; \sigma = 1$) para un estudio conjunto obtenemos que $Z_1 = 2$ y $Z_2 = 2,5$. Luego la segunda nota es más meritoria que la primera puesto que $Z_2 > Z_1$.

# 1.12

(a) Parám. de centralizacion:

$$\overline{X} = 20,32; Me = 20; Mo = 19$$

Parám. de dispersion:

$$Re = 9; Var(x) = 3,938; \sigma = 1,984; Cv = 9,76\%; Dc = 1,5; Dp_{10-90} = 2,5$$

(b)

$$\overline{X} = 20,34; Me = 20,522; \sigma = 2,1$$

(c) 91,836%

# 1.13

- Parám. de centralizacion:

$$\overline{X} = 174,5; Me = 174,571; Mo = 175,265$$

- Parám. de dispersion:

$Re = 30; Var(x) = 34; \sigma = 5,831; Cv = 3,34\%; Dc = 3,912; Dp_{10-90} = 8,083$

- El centil correspondiente es el **82**

# 1.14

(a) Gráfico

(b) Parám. de centralizacion:

$$\overline{X} = 10,5; Me = 10,5; Mo = 10\, y\, 11$$

Parám. de dispersion:

$Re = 15; Var(x) = 8,75; \sigma = 2,958; Cv = 28,17\%; Dc = 2,5; Dp_{10-90} = 3,5$

(c) NOTA:Transformamos la distribucion en Tipo II

$$(7,542 - 13,458) \Longrightarrow 66,776\% \simeq 67\%$$

$$(4,584 - 16,416) \Longrightarrow 95,83\% \simeq 96\%$$

# 1.15

(a) Parám. de centralizacion:

$$\overline{X} = 538,5; Me = 541,935; Mo = 545,\widehat{5}$$

Parám. de dispersion:

$Re = 700; Var(x) = 16917,75; \sigma = 130,068; Cv = 24,15\%; Dc = 91,563; Dp_{10_90} =$

(b)

- $P_{10} = 368,75; P_{70} = 609,756; P_{90} = 713,043$

- $225cm. \Longrightarrow \mathbf{P = 2} \qquad 575cm \Longrightarrow \mathbf{P = 60}$

(c)

$$(408,432 - 668,568) \Longrightarrow 67,49\%$$

$$(278,36 - 798,636) \Longrightarrow 96,1\%$$

# 1.16

| Minutos | $n_i$ | $f_i$ | $N_i \downarrow$ | $F_i \uparrow$ |
|---|---|---|---|---|
| 0 | 4 | 0,08 | 4 | 1 |
| 1 | 5 | 0,10 | 9 | 0,92 |
| 2 | 5 | 0,10 | 14 | 0,82 |
| 3 | 6 | 0,12 | 20 | 0,72 |
| 4 | 4 | 0,08 | 24 | 0,60 |
| 5 | 4 | 0,08 | 28 | 0,52 |
| 6 | 6 | 0,12 | 34 | 0,44 |
| 7 | 6 | 0,12 | 40 | 0,32 |
| 8 | 4 | 0,08 | 44 | 0,20 |
| 9 | 3 | 0,06 | 47 | 0,12 |
| 10 | 3 | 0,06 | 50 | 0,06 |

# 1.17

| litros | $n_i$ | $f_i$ | $N_i \downarrow$ |
|---|---|---|---|
| [0 – 15) | 1 | 0,016 | 1 |
| [15 – 30) | 5 | 0,083 | 6 |
| [30 – 45) | 8 | 0,13 | 14 |
| [45 – 60) | 21 | 0,35 | 35 |
| [60 – 75) | 17 | 0,283 | 52 |
| [75 – 90) | 6 | 0,1 | 58 |
| [90 – 105) | 2 | 0,033 | 60 |

# 1.18

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | 10 | 0,25 | | |
| | | 0,375 | 31 | 0,775 |
| | | | | |
| | 3 | 0,075 | 40 | |

# 1.19

•

| Nº de Km. | $n_i$ | $N_i \downarrow$ |
|---|---|---|
| (1 – 15] | 61 | 61 |
| (15 – 30] | 17 | 78 |
| (30 – 45] | 7 | 85 |
| (45 – 60] | 1 | 86 |
| (60 – 75] | 2 | 88 |
| (75 – 90] | 2 | 90 |
| Más de 90 | 1 | 91 |

• 8 operarios

# 1.20

Gráfico

# 1.21

$$\overline{X} = 7,\overset{\frown}{90}; G = 6,283; H = 4,68; C = 9,24$$

# 1.22

$\overline{X} = 28436,\overset{\frown}{2}$

# 1.23

$n_3 = 7$

# 1.24

i) $Me = 3$

ii) Me$=53,58\,\overset{\frown}{3}$

# 1.25

(a) Gráfico

(b) $100\,alumnos \Longrightarrow 66,225\%$

(c) $74,172\%$

(d) $100\%$

(e) $76,821\%$

# 1.26

Gráfico

# 1.27

(a)

| $[L_{i-1} - L_i)$ | $n_i$ | $N_i \downarrow$ | $N_i \uparrow$ | $f_i$ | $F_i \downarrow$ | $F_i \uparrow$ |
|---|---|---|---|---|---|---|
| 32-40 | 8 | 8 | 48 | 0,167 | 0,167 | 1 |
| 40-48 | 10 | 18 | 40 | 0,208 | 0,375 | 0,833 |
| 48-56 | 13 | 31 | 30 | 0,271 | 0,646 | 0,625 |
| 56-64 | 8 | 39 | 17 | 0,167 | 0,812 | 0,354 |
| 64-72 | 9 | 48 | 9 | 0,187 | 1 | 0,187 |

# 1.28

| Intervalo | | | $n_i$ | $N_i$ | $f_i$ | $F_i$ | $x_i$ | $a_i$ | altura |
|---|---|---|---|---|---|---|---|---|---|
| | - | | | | 0,01 | 0,01 | | 30000 | $6,67 \cdot 10^{-5}$ |
| 50000 | - | | 6 | 8 | 0,03 | | 55000 | 10000 | $6 \cdot 10^{-4}$ |
| | - | 70000 | 16 | 24 | 0,08 | | 65000 | | $1,6 \cdot 10^{-3}$ |
| 70000 | - | 80000 | 26 | 50 | | 0,25 | | 10000 | $2,6 \cdot 10^{-3}$ |
| 80000 | - | 100000 | | 94 | 0,22 | 0,47 | 90000 | 20000 | |
| | - | 125000 | 25 | | 0,125 | 0,595 | 112500 | 25000 | |
| 125000 | - | | 22 | 141 | | 0,705 | 137500 | 25000 | $8,8 \cdot 10^{-4}$ |
| | - | 200000 | 30 | | 0,15 | 0,855 | 175000 | | $6 \cdot 10^{-4}$ |
| 200000 | - | | 17 | 188 | 0,085 | 0,94 | 225000 | 50000 | $3,4 \cdot 10^{-4}$ |
| 250000 | - | | | | 0,06 | 1 | | 600000 | $2 \cdot 10^{-5}$ |

# 1.29

(a) 1.

| $x_i$ | $n_i$ |
|---|---|
| 19000 | 3 |
| 20000 | 3 |
| 21000 | 4 |
| 22000 | 4 |
| 23000 | 5 |
| 24000 | 5 |
| 25000 | 6 |
| 26000 | 8 |
| 27000 | 7 |
| 28000 | 6 |
| 29000 | 5 |
| 30000 | 4 |
| | 60 |

2.

| $[L_{i-1} - L_i)$ | $n_i$ | $N_i$ | $x_i$ |
|---|---|---|---|
| 185000-215000 | 10 | 10 | 20000 |
| 215000-245000 | 14 | 24 | 23000 |
| 245000-275000 | 21 | 45 | 26000 |
| 275000-305000 | 15 | 60 | 29000 |

(b) Gráfico

(c) 1. $68,3\%$

2. $63,\widehat{3}\ \%$

(d) $25357,143$

(e) $26051,724$

(f) 1. $25066,667$

2. $25050$

# 1.30

(a) $77,5$

(b) $Q_1 = 62 \quad Q_2 = 73,\widehat{3} \quad Q_3 = 92$

(c) $Mo = 64,286 \quad y \quad 86,\widehat{6}$

# 1.31

(a)

$$\overline{X}_{Fac.1} = 975; \overline{X}_{Fac.2} = 975; \overline{X}_{Fac.3} = 787,5; \overline{X}_{Total} = 2412,5$$

(b)

$$\overline{X}_{Fac.1} = 9,489; \overline{X}_{Fac.2} = 14,233; \overline{X}_{Fac.3} = 11,496; \overline{X}_{Total} = 35,219$$

# 1.32

(a) $205,714$

(b) $205,706$

(c) $218,603$

# 1.33

(a) 1344000

(b) $26,88$

(c) $28,\widehat{4}$

(d) $28,64$

(e) $Var(x) = 49,786 \quad Cv = 26,249\%$

(f) $g_1 = -0,833$

# 1.34

(a) A

(b) A (Según Cv)

# 1.35

96 Km/h (Media armónica)

# 1.36

(a) $D_{\overline{x}} = 0,533$

(b) $\sigma = 0,656$

(c) $\sigma = 0,636$

# 1.37

71,428

# 1.38

(a)

| $[L_{i-1} - L_i)$ | $n_i$ | $x_i$ | $N_i \downarrow$ | $f_i$ | $F_i \downarrow$ |
|---|---|---|---|---|---|
| 30-37 | 2 | 33,5 | 2 | 0,0625 | 0,0625 |
| 37-44 | 2 | 40,5 | 4 | 0,0625 | 0,125 |
| 44-51 | 4 | 47,5 | 8 | 0,125 | 0,25 |
| 51-58 | 11 | 54,5 | 19 | 0,34375 | 0,59375 |
| 58-65 | 8 | 61,5 | 27 | 0,25 | 0,84375 |
| 65-72 | 5 | 68,5 | 32 | 0,15625 | 1 |

(b) Gráfico

(c) $\overline{X} = 55,375$

(d) $Me = 56,\widehat{09}$

(e) $Q_1 = 51 \quad Q_3 = 62,375$

(f) $P_8 = 38,96 \quad P_{90} = 67,52$

(g) $Mo = 55,\widehat{6}$

# 1.39

$$6,976\% \simeq 7\% \text{menos de dispersión relativa}$$

# 1.40

(a) Gráfico

(b)

$$10075000 ptas/mes; \overline{X} = 74629,6 ptas/mes; Mo = 68571,4 ptas/mes$$

(c)

$$D_{c_1} = 19642,8 ptas/mes \quad D_{c_2} = 15222,2 ptas/mes$$

(d)

$$Var(x) = 603196,1 ptas/mes \quad \sigma = 24560 ptas/mes$$

# 1.41

(a)

$$\overline{X} = 4,66; \quad Mo = 5; \quad Me = 5$$

(b)

$$Var(x) = 1,984; \quad \sigma = 1,409$$

(c)

$$g_1 = -0,585; \quad As = -0,724$$

(d)

$$g_2 = -0,157 \text{ (Casi Mesocúrtica)}$$

# 1.42

(a) $\overline{X} = 18,5$

(b) $\sigma = 6,517$

(c) Cv=35,227%

(d) 46,11%

(e) $Q_1 = 13,\overset{\frown}{6}$

(f) Mo=17,227

(g) 5550

# 1.43

(a) 5,167

(b)

$$Cv_A = 0,1\overset{\frown}{6}; Cv_B = 0,188; Cv_C = 0,18; Cv_D = 0,2$$

(c) A

(d) Var(x)=1,91

# 1.44

(a)

$$\overline{X} = 125000 ptas/mes; Cv = 0,248$$

(b) A.

1.

$$\overline{X} = 131250 ptas/mes; Cv = 0,248 \text{ (Permanece igual)}$$

2.

$$\overline{X} = 130500 ptas/mes; Cv = 0,237$$

3.

$$\overline{X} = 131125 ptas/mes; Cv = 0,243$$

B. El segundo caso (ii)

## 1.45

Parám. de centralización:

$$\overline{X} = 28,275; Me = 24; Mo = 16,522$$

Parám. de dispersión:

$Re = 120; Var(x) = 400,212; \sigma = 20,005; Cv = 70,75\%; Dc = 10'\widehat{2}; Dp_{10-90} = 22'083$

Otros:

$$As = 0,588; g_1 = 1,808; g_2 = 3,671 \quad \text{(Muy Leptocurtica)}$$

## 1.46

(a)

$$\overline{X} = 4,371; Var(x) = 7,483; \sigma = 2,736; Cv = 62,62\%; Me = 3,75; Mo = 2,425$$

(b) 2248400 ptas

## 1.47

(a)

- 1986 $\Longrightarrow$ 5130 millones de ptas.
- 1987 $\Longrightarrow$ 6050 millones de ptas.

(b)

- $\overline{X}_{1986} = 31280487,8$ millones de ptas.
- $\overline{X}_{1987} = 34571428,6$ millones de ptas.

(c) $13,398\%$

# 1.48

(a) (Atendiendo exclusivamente al rendimiento global),**Sí**,puesto que ponen más ladrillos que el segundo grupo.

$$\text{grupo1} \Longrightarrow 1240\, ladrillos \qquad \text{grupo2} \Longrightarrow 798\, ladrillos$$

(b) El grupo 2 es más homogéneo, el trabajo se reparte más equilibradamente. En el grupo 1, sin embargo, mientras unos ponen pocos ladrillos los otros trabajan en exceso.

# 1.49

$\overline{X} = 52114,551$

# 1.50

$\overline{X} = 730$

# 1.51

(a)

$$Z_{9l.} = 2,857; \quad Z_{6l.} = -1,428; \quad Z_{5,5l.} = -2,143$$

(b) **Sí**, tendrá un consumo relativo mayor que la primera marca.

**Marca 1**$\Longrightarrow$ Z=1,428
**Marca 2**$\Longrightarrow$ Z=1,875

# 1.52

El **A** es el adecuado.

$$Cv_A = 0,9121\% \qquad Cv_B = 1,842\%$$

## 1.53

$X_{Barc.} = 611,25 ptas.; X_{Gran.} = 593,25 ptas; X_{Bilb.} = 608 ptas; X_{Cac.} = 587,5 ptas;$

## 1.54

La mejor posición se produce durante su permanencia en el hogar.

$$Z_{Hogar} = 0,921 \qquad Z_{Vac.} = 0,4$$

## 1.55

## 1.56

(a)

$$Peones = 74,4747\%; \quad Oficiales = 12,4124\%; \quad Especialistas = 13,1129\%$$

(b) Aumentando en 4655, 6 ptas. para el total

## 1.57

(a) $\overline{X} = 72,727$ ptas/Kg

(b) $Var(x) = 110,744\ (ptas/Kg)^2$

## 1.58

Son más estable las *importaciones* que las *exportaciones*.

$$Cv_{Imp.} = 7,506\% \qquad Cv_{Exp.} = 12,2511\%$$

## 1.59

## 1.60

(a) $\overline{X} = 23,5$

(b) $\overline{X}_T = 8,225$

(c) Mo$=15,55$

(d) Me$=20$

# Capítulo 2

## 2.1

$\bar{X} = 35,9 \quad \bar{Y} = 35,75 \quad \mathrm{Var}(X) = 106,19 \quad \mathrm{Var}(Y) = 122,44$

## 2.2

$$\begin{aligned} x &= -0,22y + 34 \\ y &= -3,78x + 150 \\ r &= -0,9115 \end{aligned}$$

## 2.3

$$\begin{aligned} x &= 1,16y + 16,876 \\ y &= 0,346x + 1,636 \\ r &= 0,63 \end{aligned}$$

## 2.4

$$\begin{aligned} x &= 1,039x - 0,795 \\ y &= 0,817x + 1,6 \\ r &= 0,92 \\ r_{y/x}(8,3) \Longrightarrow y &= 8,38 \end{aligned}$$

## 2.5

$\mathrm{Cov}(X,Y) = 3,95$

## 2.6

$\bar{X} = 1,713 \quad \bar{Y} = 70 \quad \mathrm{Var}(Y) = 4,868 \quad \mathrm{Var}(X) = 0,0355$

## 2.7

$\bar{X} = 9,21 \quad \bar{Y} = 0,94 \quad \mathrm{Var}(X) = 8,87 \quad \mathrm{Var}(Y) = 0,46 \quad \mathrm{Cov}(X,Y) = 1,78$

**2.8** $r_{y/x} \Longrightarrow y = -3,79x + 150,01$

**2.9** $y = 0,7717 + 0,1421x + 0,0567x^2$

**2.10** No hay dependencia

**2.11** (b) $\bar{X} = 69$ $\text{Var}(X) = 8,5$ $\bar{Y} = 1,71$ $\text{Var}(Y) = 0,14$ (c) $\text{Cov}(X,Y) = 0,77$

**2.12** (b) $\bar{X} = 4,6$ $Mo = 5$ $\bar{Y} = 1,08$ $Mo = 1$

(c) $\sigma_x = 1,027$ $A_s = -0,38$ $\sigma_x = 0,75$ $A_s = 0,1$

**2.13** $r = 0,539$ (hay poca dependencia lineal).

**2.14** $r_{(x,y)} = 0,974$ $r_{(y,z)} = 0,105$ $r_{(z,v)} = -1$

**2.15** $r = 0,9764$

**2.16** $r = 0,9356$

**2.17** (b) $Me(Y) = 35,94$ $Q_3 = 0,11$

**2.18** En efecto, son independientes.

**2.19** $r = 0,13$ (No están correlacionados).

**2.20**

(a) $n_{1.} = 60$ $n_{2.} = 80$ $n_{.2} = 180$ $n_{.3} = 184$

(b) $f_{12} = 0,025$ $f_{23} = 0,025$ $f_{34} = 0,042$ $f_{42} = 0,083$ $f_{2.} = 0,16$ $f_{3.} = 0,41$

(c) $f(X_1/Y = 350000) = 0,066$ $\quad f(Y_2/X = 200000) = 0,4$

(d) $\bar{X} = 222916,67$ $\quad \bar{Y} = 1224583,33$

## 2.21

(b) $\bar{X} = 3,5$ $\quad \bar{Y} = 3,21,$ $\quad \text{Cov}(X,Y) = -0,9,$ $\quad Porc = 66,67\%$

## 2.22

(a) $\bar{X} = 290,$ $\quad \bar{X}(x/y = 1) = 290,$ $\quad \bar{X}(x/y = 2) = 290,$ $\quad \bar{X}(x/y = 3) = 290$

$\bar{X}(x/y = 4) = 290,$ $\quad \bar{X}(x/y = 5) = 2904,$ $\quad \bar{Y} = 3,6,$ $\quad \bar{Y}(y/x = 1) = 3,6$

$\bar{Y}(y/x = 2) = 3,6,$ $\quad \bar{Y}(y/x = 3) = 3,6,$ $\quad \bar{Y}(y/x = 4) = 3,6$

$\text{Cov}(X,Y) = 0,$ $\quad \text{Var}(X) = 12900,$ $\quad \text{Var}(Y)(y = 4) = 12900,$ $\quad \text{Var}(Y) = 1,44$ $\quad \text{Var}(Y)(x = 200) = 1,44$

(b) Son independientes.

## 2.23

$rango = 4$

## 2.24

(a)

$$\begin{aligned} x &= 3,75y - 7,32 \\ y &= 0,26x + 1,98 \end{aligned}$$

(b) Centro de gravedad $= (\bar{X}, \bar{Y}) = (3,97, 3,01)$

(c) $y = 4,58$

(d) $x = 9,9$

(e) Las dos rectas coinciden.

(f) Como las dos rectas son las mismas, las pendientes de las rectas son inversas entre sí.

(g) Cuanto más cercano esté $(1 - R^2)$ de 0, el ajuste será mejor.

# 2.25

(a) Ajuste parabólico $y = 3,15 + 3877,5x + 1,96x^2$

(b) $R^2 = 0,93$ (Ajuste prácticamente perfecto).

(c) $y = 27242$

# 2.26

(a) $y = 2,096\, x^{1,47}$

(b) $R^2 = 0,99$

(c) Ajuste bastante bueno.

# 2.27

(a) $y = 1,0044\, x^{1,393}$

(b) $R^2 = 0,99$

(c) Ajuste bastante bueno.

# 2.28

(a) $y = (3,098)(2,098)^{-0,2x}$

(b) $R^2 = 0,99$

(c)

# 2.29

(a) $y = (2,2436615)(0,730284244)^{-0,2x}$

(b) $R^2 = 0,98$

(c) Ajuste bastante bueno.

# 2.30

(a)

| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $\phi(x)$ | 3 | 2,875 | 3,579 | 4 | 4,577 | 4,962 | 5,844 | 6,357 | 6,76 |

(b) 0,54

(c) $y = 1,982732872 + 0,53258851 \cdot x$

(d) $R^2 = 0,453$

# 2.31

(a)

| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\phi(x)$ | 5,256 | 5,111 | 5,794 | 6,615 | 6,786 | 6,675 | 8,125 |

(b) 0,9

(c) $y = 4,19680203 + 0,56639037 \cdot x$

(d) $R^2 = 0,143$

# 2.32

(a)

| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\phi(x)$ | 4,349 | 5,919 | 6,571 | 7,607 | 8,5 | 9,4 | 10,3 | 10,85 | 11,3 | 11,67 |

El valor de la razón de correlación es 0,4

(b)

$$\begin{aligned} y &= 0,68x - 0,91 \\ y &= 2,6626 - 0,7805x + 0,1086x^2 \\ y &= (0,7182)(1,2101)^x \end{aligned}$$

(c) $R^2 = 0,61$, $R^2 = 0,75$, $R^2 = 0,58$

# 2.33

(b)

$$\begin{aligned} y &= 0,1011x + 2,6963 \\ y &= 2,4119 + 0,1451x - 0,0011x^2 \end{aligned}$$

# 2.34

(a) Si llamamos $X$=años, $Y$=importación y $Z$=exportación,

$$\begin{aligned} z &= 623,77\,x - 1232904,3 \\ y &= 398,24\,x - 786900,78 \end{aligned}$$

(b) $y(1991) = 9027,14$ $y(1992) = 9650,92$ $y(1993) = 10274,72$, $r = 0,98$

$z(1991) = 6005,93$ $z(1992) = 6404,17$ $z(1993) = 6802,42$, $r = 0,98$

(c) $w = 225,53\,x - 446003,58$ $R^2 = 0,695$ (No es un buen ajuste).

(d) Se importará más de lo que se exportará.

# 2.35

(a) $y_1 = 0,073\,x - 142,967$

(b) $y_2 = 0,1005\,x - 196,896$

(c) Sin esos últimos años, el ajuste es mejor, ya que entre esos 4 años la diferencia entre los porcentajes es mayor de lo normal.

(d) $y_1(1990) = 2,69$ $y_1(1991) = 2,76$ $y_2(1990) = 3,19$ $y_2(1991) = 3,29$

## 2.36

(a) $y_1 = 5,97\,x - 11474,087$

(b) $y_2 = 7,65\,x - 14742,025$

(c) $R_1^2 = 0,83 \quad R_2^2 = 0,86$ (Ambos ajustes son buenos, pero el segundo es mejor).

(d) $y_1(1951) = 174,759 \quad y_2(1951) = 183,127$

## 2.37

(a) $r = -0,036$ (Es un mal ajuste).

## 2.38

(a) $r = 0,917$ (Sí, concuerdan)

## 2.39

(a) $r = -0,231$ (Mal ajuste).

## 2.40

$C = 0,012$ (No existe apenas relación).

# Soluciones Unidad Temática 3

## 3.1

(a) 1/15

(b) 1/3

## 3.2

(a) 0,165

(b) 0,081

## 3.3

$\frac{365\cdot 364\cdot 363\cdots(365-N+1)}{365^{N}}$

## 3.4

0,243

## 3.5

1/50 , 1/4950

## 3.6

56/1024

## 3.7

0,011

## 3.8

1/20 , 6/20 , 3/4

**3.9** 1/16

**3.10** 1/5 , 1/15 , 0

**3.11** 1/3 , 1/10 , 1/5 , 9/17

**3.12** 1/11 , 1/3

**3.13** 2/5

**3.14** 0,161

**3.15** 0,087

**3.16** 0,199 , 1/199

**3.17** 37/256 , 29/128 , 22/64

**3.18** 600/840

**3.19** 220/9880

**3.20** 1/6

**3.21**

(a) $0,7 \cdot 0,8 \cdot 0,9 = 0,504$

(b) $0,7 \cdot 0,8 \cdot 0,1 + 0,7 \cdot 0,2 \cdot 0,9 + 0,3 \cdot 0,8 \cdot 0,9 + 0,7 \cdot 0,8 \cdot 0,9 = 0,902$

**3.22** $P(9) = 25/216$ , $P(10) = 27/216$

**3.23** $(1/2)^n$

**3.24** $1-(5/6)^n$

**3.25** $\frac{\binom{a}{x}\cdot\binom{b}{k-x}}{\binom{a+b}{k}}$

**3.26** $p' = 1-(\frac{35}{36})^n \quad \approx \frac{1}{2} \quad \Rightarrow \quad n \approx 25$

**3.27** $P(0) = 1/270725$ , $P(1) = 13/270725$ , $P(2) = 78/270725$ , $P(3) = 286/270725$ , $P(4) = 715/270725$

**3.28** $1/4$

**3.29** $P(\bar{B}) = \sum_{i=1}^{n} \binom{100}{i} \cdot (0,1)^i \cdot (0,9)^{100-i} \Rightarrow \quad P(B) = 1 - P(\bar{B})$

**3.30** $\frac{\binom{n}{h}\cdot(m-1)^{n-h}}{m^n}$

**3.31** $1-0,8^3$

**3.32** 0,2 , 0,5 , 0,4

**3.33**

(a) 4/7

(b) 2/3

(c) 1

(d) 2/3

(e) 5/6

**3.34** $P(B/A) = \frac{P(A\cap B)}{P(A)}$ $\quad y \quad$ $P(C/A) = \frac{P(A\cap C)}{P(A)}$

Como B y C son subconjuntos de A $\Rightarrow A \cap B = B$ , $A \cap C = C$
Por lo tanto $\Rightarrow P(A \cap B) = P(B)$ , $P(A \cap C) = P(C) \Rightarrow P(B/A) = \frac{P(B)}{P(A)}$ , $P(C/A) = \frac{P(C)}{P(A)} \Rightarrow \frac{P(B/A)}{P(C/A)} = \frac{P(B)P(A)}{P(C)P(A)} = \frac{P(B)}{P(C)}$

## 3.35

Hay que probar que $P[(A \cup B) \cap C] = P(A \cup B)P(C)$
$P[(A \cup B) \cap C] = P[A \cap C) \cup (B \cap C)] = P(A \cap C) + P(B \cap C) - P[(A \cap C) \cap (B \cap C)] = P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)$
Por condiciones de independencia, $P(A \cap C) = P(A)P(C) y P(B \cap C) = P(B)P(C)$
Por otra parte, $P(A \cap B \cap C) = P(A)P(B)P(C) = [P(A)P(B)]P(C) = P(A \cap B)P(C)$
Entonces, $P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) = P(A)P(C) + P(B)P(C) - P(A \cap B)P(C) = [P(A) + P(B) - P(A \cap B)]P(C) = P(A \cup B)P(C)$
Por tanto son independientes

## 3.36

(a) $A \cap \bar{B} = A - B y (A - B) \cup (A \cap B) = A$
$P(A - B) + P(A \cap B) = P(A)$
$P(A \cap \bar{B}) = P(A - B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(\bar{B})$
Por tanto son independientes

(b) $E = (\bar{A} \cap \bar{B}) \cup (A \cup B)$
$P(E) = 1 = P(\bar{A} \cap \bar{B}) + P(A \cup B)$
$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(A \cap B)] = 1 - [P(A) + P(B) - P(A)P(B) = 1 - P(A) - P(B)[1 - P(A)] = 1 - P(A) - P(B)P(\bar{A}) = P(\bar{A}) - P(B)P(\bar{A}) = P(\bar{A}) - P(\bar{A})P(B) = P(\bar{A})[1 - P(B)] = P(\bar{A})P(\bar{B})$
Por tanto son independientes.

## 3.37

$P(A/A \cup B) = \frac{P(A \cap (A \cup B))}{P(A \cup B)}$
A es un subconjunto de $A \cup B$ , por tanto $A \cap (A \cup B) = A \Rightarrow P[A \cap (A \cup B)] = P(A)$.
Al ser A y B disjuntos, $A \cap B = \emptyset$ $y$ $P(A \cap B) = 0$, entonces $P(A \cup B) = P(A) + P(B)$
Por tanto, $P(A/A \cup B) = \frac{P[A \cap (A \cup B)]}{P(A \cup B)} = \frac{P(A)}{P(A) + P(B)}$

# 3.38

(a) 0,000975

(b) 0,0791

# 3.39

0,02 , 0,26

# 3.40

0,7995

# 3.41

(a) 0,00004032

(b) 0,001701

# 3.42

(a) 0,1681

(b) 0,3087

(c) 0,8370

(d) 0,0309

(e) 0,441

# 3.43

(a) 0,19

(b) 0,2

(c) 0,25

(d) 0,1

(e) 0,55

## 3.44
0,3 , 0,6

## 3.45
0,966

## 3.46
0,3625

## 3.47
0,903

## 3.48

(a) $\frac{1}{40}$

(b) $\frac{1}{40} \cdot \frac{1}{39} \cdot \frac{1}{38} \cdot \frac{1}{37}$

## 3.49
0,4 , Hipótesis 1 y 2 imposible, Hipótesis 3 posible

## 3.50
No son disjuntos, son independientes

## 3.51

(a) 0,000547

(b) 0,003283

(c) 0,000383

## 3.52
Sí

## 3.53

(a) 101

(b) 252

(c) 153

(d) 310

# 3.54

(a) 34

(b) 196

(c) 38

(d) 254,36,196,34

# 3.55

# 3.56

(a) {(V,V), (V,H), (H,V), (H,H)}

(b) {(X,X,X), (X,X,C), (X,C,X), (X,C,C), (C,X,X), (C,X,C), (C,C,X), (C,C,C)}

(c) {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

# 3.57

(a) $E \cap F \cap \bar{G}$

(b) $F \cap \bar{E} \cap \bar{G}$

(c) $G \cap \bar{E} \cap \bar{G}$

(d) $E \cup F \cup G$

(e) $(E \cap \bar{F} \cap \bar{G}) \cup (\bar{E} \cap F \cap \bar{G}) \cup (\bar{E} \cap \bar{F} \cap G)$

(f) $E \cap F \cap G$

(g) $(E \cap F \cap \bar{G}) \cup (E \cap G \cap \bar{F}) \cup (F \cap G \cap \bar{E}) \cup (F \cap G \cap E)$

(h) $(E \cap F \cap \bar{G}) \cup (E \cap G \cap \bar{F}) \cup (F \cap G \cap \bar{E})$

(i) $(E \cap \bar{F} \cap \bar{G}) \cup (\bar{E} \cap F \cap \bar{G}) \cup (\bar{E} \cap \bar{F} \cap G) \cup (\bar{E} \cap \bar{F} \cap \bar{G}) \cup H$

(j) $(\bar{E} \cap \bar{F} \cap \bar{G}) \cup e$

(k) $(E \cup F \cap \bar{G}) - (E \cap F)$

**3.58** $P(\bar{E}) = 12/20$ , $P(\bar{F}) = 8/20$, $P(\bar{E} \cap \bar{F}) = 6/20$ , $P(E \cup F) = 14/20$, $P(\overline{(E \cup F)}) = 6/20$ , $P(E - F) = 2/20$ , $P(F/E) = 6/8$ , $P(E/\bar{F}) = 1/6$ , $P((E \cup F)/E) = 1/4$ , $P((E \bar{\cup} F)/E) = 0$.

**3.59** $(S \cup E \cup C) = 495 \neq 500$

**3.60** 11/36

**3.61** 1/4

**3.62** 1/3

**3.63** $n = 25$

**3.64**

(a) 11/188

(b) 83/188

(c) 6/47

**3.65** 0,044

**3.66** 0,31

**3.67** 5/16

**3.68** 1/36

**3.69** 0,4914

**3.70** 0 ó 1

**3.71** En A

**3.72** 1/2

**3.73** M/N

**3.74** $(n/(n+b)) * (n+r/(n+r+b)) * ((n+2r)/(n+b+2r))$

**3.75**

(a) 3/7

(b) 5/140

**3.76** 49/90

**3.77** 2/3

**3.78** $B_1$

**3.79** $N = 4$

**3.80** 4/7

**3.81** 108/343

**3.82** 0,0967

**3.83** $n = 23$

**3.84** (a) $5{,}13 \cdot 10^{-5}$ (b) 0,5973 (c) 0,98898

**3.85** 30/61

**3.86** 0,26

**3.87** 1000/1009

**3.88** $P[X = 0] =0{,}504$ $P[X = 1] = 0{,}398$ $P[X = 2] = 0{,}092$ $P[X = 3] = 0{,}006$.

**3.89** 0,8742

**3.90** 4/7 , 2/7 , 1/7 respectivamente

**3.91** (a) 0,22 (b) 0,7.

**3.92**

(a) $\frac{4}{7} \cdot 0{,}3$

(b) 3/7

(c) 19/70.

**3.93** (a) 0,0012 (b) 4/19.

**3.94** 14/57

**3.95** 97,9%

**3.96** 19/118

**3.97** 0,8864 ; 3/1108

**3.98** (a) 0,046 (b) 9/23.

**3.99** (a) 1/2 (b) 0,3.

**3.100** (a) 0,028 (b) 45/56.

**3.101** (a) 0,26 (b) 0,1 (c) 0,02 (d) 0,076923 (e) 2,3 Mb.

**3.102** 0,00141734 ; 0,15174 ; $E[X] = 1{,}5$ ; $\mathrm{Var}[X] = 0{,}9643$.

**3.103** Si $p =$P[cara] entonces P[lanzar $n$ veces]$= (n-1)p^2 \cdot q^{n-2}$.

**3.104** (a) 1/25 (b) 262/300.

**3.105** $n = 71$ temas.

# 3.106

$$
\begin{array}{llll}
P_x: & \mathrm{R} & \longrightarrow & \mathrm{R} \\
& \text{llave1} & \longrightarrow & \frac{1}{4} \qquad\qquad P(x=1)=\frac{1}{4} \\
& \text{llave2} & \longrightarrow & \frac{1}{4} \qquad\qquad P(x=2)=\frac{1}{4} \\
& \text{llave3} & \longrightarrow & \frac{1}{4} \qquad\qquad P(x=3)=\frac{1}{4} \\
& \text{llave4} & \longrightarrow & \frac{1}{4} \qquad\qquad P(x=4)=\frac{1}{4}
\end{array}
$$

# 3.107

$$
P(Y=y)=\begin{cases} \binom{5}{y}\cdot(0,5)^y\cdot(0,5)^{5-y} & si \quad y=0,1,2,3,4,5 \\ 0 & si \quad y>5 \end{cases}
$$

$P(Rojos)=0,5;\quad P(negros)=0,5$

# 3.108

NOTA:Consideraremos como máximo una entrada por día.Con la misma probabilidad de visita como no

$$
P(Z=z)=\begin{cases} \binom{7}{z}\cdot(0,5)^z\cdot(0,5)^{7-z} & si \quad z=0,1,2,3,4,5,6,7 \\ 0 & si \quad z>7 \end{cases}
$$

# 3.109

$$
P(x)=\begin{cases} \frac{1}{2} & si \quad x=0,-2 \\ 0 & \text{en otro caso} \end{cases}
$$

- $P(-1\leq x\leq 1)=\frac{1}{2}$

# 3.110

$$
P(W)=\begin{cases} \frac{1}{3} & si \quad x=3,6 \\ \frac{1}{6} & si \quad x=4,5 \\ 0 & \text{en otro caso} \end{cases}
$$

- $P(3<W\leq 5)=\frac{1}{3}$

# 3.111

$$P(Y) = \begin{cases} \frac{1}{4} & si \quad x = 0 \\ \frac{1}{12} & si \quad x = 5 \\ \frac{1}{6} & si \quad x = 7,102 \\ \frac{1}{3} & si \quad x = 100 \\ 0 & \text{en otro caso} \end{cases}$$

- $P(Y \leq 100) = \frac{5}{6}$

# 3.112

$$F_z(x) = \begin{cases} 0 & si \quad x < 0 \\ \frac{1}{3} & si \quad 0 \leq x < 1 \\ \frac{2}{3} & si \quad 1 \leq x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

# 3.113

$$F_U(x) = \begin{cases} 0 & si \quad x < -3 \\ \frac{1}{3} & si \quad -3 \leq x < 0 \\ \frac{1}{2} & si \quad 0 \leq x < 4 \\ \frac{5}{6} & si \quad 4 \leq x \end{cases}$$

No es función de probabilidad, pues no se obtiene una probabilidad final 1.

# 3.114

$$f_x(x) = \begin{cases} \frac{1}{2} & \text{si} \quad x = -1, 1 \\ 0 & \text{en otro caso} \end{cases}$$

- $P(-\frac{1}{2} < x < \frac{1}{2}) = \frac{1}{2}$

# 3.115

$$f_y(y) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{si} \quad 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

$\bullet P(-\frac{1}{4} < x < \frac{1}{4}) = \frac{1}{2}$

# 3.116

$$f_z(z) = \begin{cases} 2z & \text{si} \quad 0 \leq z \leq \frac{1}{2} \\ 6 - 6z & \text{si} \quad \frac{1}{2} \leq z \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

# 3.117

No es función de distribución:

-Decrece en el intervalo [0,1/2].

-No está acotada superiormente por 1.

# 3.118

-Continua con discontinuidad de salto finito.

-Acotada entre 0 y 1.

-Monótona creciente.

# 3.119

$$F_x(x) = \begin{cases} 0 & si \quad x \leq 99 \\ x - 99 & si \quad 99 < x < 100 \\ 1 & si \quad x \geq 100 \end{cases}$$

# 3.120

$$F_y(Y) = \begin{cases} 0 & si \quad y < 0 \\ 2y - y^2 & si \quad 0 \leq y \leq 1 \\ 1 & si \quad y > 1 \end{cases}$$

# 3.121

$$F(z) = \begin{cases} 0 & si \quad z \leq 0 \\ 1 - e^{-10 \quad z} & si \quad z > 0 \end{cases}$$

# 3.122

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 0,336 & si \quad 0 \leq x < 1 \\ 0,788 & si \quad 1 \leq x < 2 \\ 0,976 & si \quad 2 \leq x < 3 \\ 1 & si \quad x \geq 3 \end{cases}$$

# 3.123

$$F(x) = \begin{cases} 0 & si \quad x < 3 & \frac{5}{8} & si \quad 11 \leq x < 12 \\ \frac{1}{216} & si \quad 3 \leq x < 4 & \frac{20}{27} & si \quad 12 \leq x < 13 \\ \frac{1}{54} & si \quad 4 \leq x < 5 & \frac{181}{216} & si \quad 13 \leq x < 14 \\ \frac{5}{108} & si \quad 5 \leq x < 6 & \frac{49}{54} & si \quad 14 \leq x < 15 \\ \frac{5}{54} & si \quad 6 \leq x < 7 & \frac{103}{108} & si \quad 15 \leq x < 16 \\ \frac{35}{216} & si \quad 7 \leq x < 8 & \frac{53}{54} & si \quad 16 \leq x < 17 \\ \frac{7}{27} & si \quad 8 \leq x < 9 & \frac{215}{216} & si \quad 17 \leq x < 18 \\ \frac{3}{8} & si \quad 9 \leq x < 10 & 1 & si \quad x \geq 13 \\ \frac{1}{2} & si \quad 10 \leq x < 11 & & \end{cases}$$

# 3.124

$$f(x) = \begin{cases} \frac{3 \cdot (x-2)^2}{8} & 2 < x \leq 4 \\ 0 & \text{en otro caso} \end{cases}$$

$P(1 < x < 3) = \frac{1}{8}$; $\quad P(x \geq 3) = \frac{7}{8}$

$P(x < 3) = 0$; $\quad P(x \leq 3) = \frac{1}{8}$

$P(x \geq 4) = 0$; $\quad P(x \geq 2) = 1$

# 3.125

$$F(x) = \begin{cases} 0 & si \quad x < 2 & \frac{21}{36} & si \quad 7 \leq x < 8 \\ \frac{1}{36} & si \quad 2 \leq x < 3 & \frac{26}{36} & si \quad 8 \leq x < 9 \\ \frac{3}{36} & si \quad 3 \leq x < 4 & \frac{30}{36} & si \quad 9 \leq x < 10 \\ \frac{6}{36} & si \quad 4 \leq x < 5 & \frac{33}{36} & si \quad 10 \leq x < 11 \\ \frac{10}{36} & si \quad 5 \leq x < 6 & \frac{35}{36} & si \quad 11 \leq x < 12 \\ \frac{15}{36} & si \quad 6 \leq x < 7 & 1 & si \quad x \geq 12 \end{cases}$$

# 3.126

(a)

$$P(x) = \begin{cases} \frac{1}{36} & si \quad x = 2, 12 \\ \frac{2}{36} & si \quad x = 3, 11 \\ \frac{3}{36} & si \quad x = 4, 10 \\ \frac{4}{36} & si \quad x = 5, 9 \\ \frac{5}{36} & si \quad x = 6, 8 \\ \frac{6}{36} & si \quad x = 7 \end{cases}$$

(b) Igual al del ejercicio anterior.

# 3.127

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 0,024 & si \quad 0 \leq x < 1 \\ 0,212 & si \quad 1 \leq x < 1 \\ 0,664 & si \quad 1 \leq x < 2 \\ 1 & si \quad x \geq 3 \end{cases}$$

# 3.128

$\mu = 0; \quad \sigma^2 = \frac{35}{6}$

# 3.129

(a)

$$\begin{array}{llcl} G: & E & \longrightarrow & R \\ & \text{2 rojas} & \longrightarrow & \text{2puntos} \\ & \text{1 roja 1 verde} & \longrightarrow & \text{3 puntos} \\ & \text{1 roja 1 negra} & \longrightarrow & \text{4 puntos} \\ & \text{2 verdes} & \longrightarrow & \text{4 puntos} \\ & \text{1 verde 1 negra} & \longrightarrow & \text{5 puntos} \end{array}$$

(b)

$$F(x) = \begin{cases} 0 & para \quad x < 2 \\ \frac{3}{15} & para \quad 2 \leq x < 3 \\ \frac{9}{15} & para \quad 3 \leq x < 4 \\ \frac{13}{15} & para \quad 4 \leq x < 5 \\ 1 & para \quad x \geq 5 \end{cases}$$

# 3.130 $K = 4$

# 3.131

(a) 4/210; (*b*) 52/210; (*c*) 55/210;

# 3.132

(a)

$$f(x) = \begin{cases} 0 & \text{en otro caso} \\ nx^{n-1} & \text{si} \quad 0 < x \leq 1 \end{cases}$$

(b) $\ln M_e = \frac{1}{n} \cdot \ln 0'5$

(c) $E(x) = \frac{n}{n+1}$; $Var(x) = \frac{n}{(n+2)\cdot(n+1)^2}$

# 3.133

(a)

$$F(x) = \begin{cases} 0 & 0 \leq x \\ x^3 & 0 < x < 1 \\ 0 & x \geq 0 \end{cases}$$

(b) $F(2/3) = (2/3)^3$; $F(9/10) = (9/10)^3$; $P(1/3 < x \leq 1/2) = \frac{19}{216}$

(c) $a = \sqrt[3]{\frac{1}{4}}$

(d) $E[x] = 0,75$; $Var(x) = 3/80$

## 3.134

$2,7\%$

## 3.135

$5.000 ptas$.

## 3.136

(b)

$$f(x) = \begin{cases} 0 & \text{en otro caso} \\ \frac{1}{2} & 0 \leq x \leq 1 \\ \frac{1}{4} & 2 \leq x \leq 4 \end{cases}$$

(c) $P(x > 3,5/x > 1) = 0,25$

## 3.137

(a)

$$P(Y = 0) = 0,4$$

$$P(Y = 1) = 0,4$$

$$P(Y = 4) = 0,2$$

(c) $E(x^2) = 1,2$

## 3.138

$\gamma_1 = \frac{K^2 - 3K + 2}{(1-K)^{3\ 2}}$

# 3.139

$$K = \frac{2}{\pi}; \quad F(x) = \frac{4}{\pi} \cdot \arctan e^x - \frac{\pi}{4} \text{ para } x \geq 0$$

# 3.140

$[-1, 0] \cup [1, 2]$ aunque no cumple la condición:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

# 3.141

$P(1,5 < X \leq 2) = 0,125; \quad P(X \leq 5) = 0,6; \quad P(4,5 < X \leq 5,5) = 0,4$

# 3.142

(a)

$$F(x) = \begin{cases} 0,15x & 0 < 2 \leq 2 \\ \frac{x^2}{32} - \frac{x}{4} + \frac{4}{5} & 4 \leq x \leq 8 \\ -\frac{x^2}{5} + \frac{22}{5} \cdot x - 23,2 & 10 < x \leq 11 \end{cases}$$

(b) $P(X \leq 9) = 0,8$

(c) $P(1 < X \leq 10,5) = 0,8$

(d) $P(X \geq 6) = 0,575$

# 3.143

(a) $P(-2/a < X < 2/a) = 0,8646$; $(b)$ $a = 0,0744$

# 3.144

$E(x) = 1,7; \quad Var(x) = 2,41$

# 3.145

(a) $Me = 0,693$; (b) $Q_3 = 1,386$; (c) $P_{32} = 0,386$

## 3.146

En los intervalos que se verifique:

$$f(x) \leq 0 \qquad y \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

## 3.147

$K = \frac{1}{8}$; $a = 4$

## 3.148

(a) $K = 100$; (b) $\frac{1}{3}$; (c) $\frac{2}{3}$

## 3.149

(a) $g(y) = 2y$ para $0 < y \leq 1$

(b) $g(y) = \frac{1}{\sqrt{y}}$ para $0 < y \leq 1$

(c) $g(y) = \frac{1}{y^2}$ para $0 < y \leq 1$

## 3.150

$0,8032$

## 3.151

$0,049$

## 3.152

La probabilidad exacta 0,952 es mayor que 0,75(Chebychev).

## 3.153

(a) $E(g_1(x)) = 0,8\widehat{6}$; (b) $E(g_2(x)) = 5,767$

## 3.154

(a) $f(x_p) = \frac{3\cdot(40-x)^2}{8000}$

(b) $E(x_p) = 25$

(c) $f(x_v) = \frac{3\cdot(40-\sqrt{1600-8x})^2}{1600\cdot\sqrt{1600-8x}}$

(d) $E(x_v) = 180$

## 3.155

$f(y) = \frac{3}{2}\ \sqrt{y}$ en el intervalo $(0,1)$

## 3.156

(a)

$$F(X_{costes}) = \frac{1}{11}\cdot(2X-\frac{4}{3});\ \ F(X_{ventas}) = \frac{1}{33}\cdot(-6X+38);\ \ F(X_{benef}) = \frac{1}{132}\cdot(-6x+34)$$

(b) $P(X<0) = \frac{37}{132}$

(c) $E(X) = \frac{18}{11}$

## 3.157

$\varphi_x(t) = \frac{2}{it}e^{it} - \frac{2}{i^2\cdot t^2}\cdot(e^{it}+1)$

## 3.158

$f(x) = \frac{2y}{\sqrt{2\pi}}\cdot e^{-\frac{y^4}{2}}$

## 3.159

$E(x) = 200\ \ ptas.$

## 3.160

(a) $\varphi_x(t) = \frac{0{,}1}{it-0{,}1}$

(b) $\alpha_1 = 10;\ \ \alpha_2 = 200$

(c) $E(x) = 10;\ \ X_c = 14,14;\ \ Var(x) = 100$

## 3.161

$\mu = M_o = M_e = 0,5$

## 3.162

(a) es preferible la edición de la obra.

(b) Para $E[x] < 300$ sería no editar, si E[300] editar.

# 3.163

(a) La tercera alternativa.

(b) El coste de oportunidad será 3,02%.

# 3.164 3

# 3.165

$$P_{x,y}(X,Y) = \begin{cases} P(0,0) = \frac{1260}{2652} & P(1,2) = \frac{24}{2704} \\ P(0,1) = \frac{864}{2652} & P(2,0) = \frac{6}{2652} \\ P(0,2) = \frac{132}{2652} & P(2,1) = \frac{6}{2652} \\ P(1,0) = \frac{216}{2652} & P(2,2) = 0 \\ P(1,1) = \frac{144}{2652} & \end{cases}$$

- $P(X > Y) = \frac{228}{2652}$

# 3.166

| $Y \backslash X$ | 0 | 1 | 2 | $f_1(x)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | $\frac{6}{90}$ | $\frac{6}{90}$ |
| 1 | 0 | $\frac{42}{90}$ | 0 | $\frac{42}{90}$ |
| 2 | $\frac{42}{90}$ | 0 | 0 | $\frac{42}{90}$ |
| $f_2(y)$ | $\frac{42}{90}$ | $\frac{42}{90}$ | $\frac{6}{90}$ | |

- $P(X \leq Y) = \frac{84}{90}$

NOTA: Comprobamos dos formas diferentes de representación de la función de probabilidad en los ejercicios 1 y 2.

# 3.167 $f_{x,y}(x,y) = \frac{1}{\pi} \quad si \quad x^2 - y^2 \leq 1$

$f_x(x) = f_y(y) = \frac{1-t^2}{\pi}$

## 3.168 $A = \frac{2}{Ln2}$

## 3.169 $\frac{9}{40}$

## 3.170 $f_x(x) = 1 \quad si\ 0 < x < 1$

$f_y(y) = -Lny \quad si\ 0 < y < 1$

## 3.171

(a) $a = \frac{1}{\sqrt{2}}$

(b) $f_x(x) = f_y(y) = \frac{1}{\sqrt{2}}$

## 3.172

(a) $a = 1$

(b) $f_x(x) = x + 1 \quad si \ -1 < x < 0$

$f_y(y) = 1 - y \quad si\ 0 < x < 1$

## 3.173

(a) $k = 1$

(b)

$$F_{x,y}(x,y) = \begin{cases} 0 & si \quad x < 0 \ , \ y < 0 \\ \text{xy} & si \quad |x| \leq y \ \ 0 \leq y \leq 1 \\ 1 & si \quad x > y \ , \ y > 1 \end{cases}$$

## 3.174

(a)

$$F_{x,y}(x,y) = \begin{cases} 0 & si \quad x \le 0, y \le 0 \\ e^{-x-y} - e^{-x} - e^{-y} + 1 & si \quad 0 < x, 0 < y \end{cases}$$

(b) $P(x \le y) = 0,5, \quad P(x = y) = 0$

## 3.175

$P_x(x = i) = \frac{1}{2^i} \quad si \quad i = 1, 2, ...$

$P_y(y = j) = \frac{1}{2^j} \quad si \quad j = 1, 2, ...$

## 3.176

$F_x(x) = \frac{x^2}{4}; \qquad F_y(y) = y^2$

$F(x/y) = \frac{x^2}{4}; \qquad F(y/x) = 16$

## 3.177

(a)

$$F_{x,y}(x,y) = \begin{cases} 0 & si \quad x < 0, y > 3 \\ \frac{1}{8} & si \quad 0 \le < 1, 1 < y \le 3 \\ \frac{4}{8} & si \quad 0 \le < 1, 1 < y \le 3 \\ \frac{7}{8} & si \quad 2 \le < 3, 0 < y \le 1 \\ 1 & si \quad x \ge 3, y \le 0 \end{cases}$$

(b) No son independientes.

## 3.178

(a) $f_x(x) = 2; \qquad f_y(y) = 2 \; f(x/y) = 1 \quad si \; 0 < x < y; \qquad f(y/x) = 1 \quad si \; x < y < 1$

(b) $P(y > 1/2/x = 1/2) = \frac{1}{2}; \qquad P[x > 1/3/y = 2/3] = \frac{1}{27}$

(c) No son independientes.

## 3.179

$P[Y > 1/X < 1/2] = \frac{4}{7}$

# 3.180

$$P[X \leq 4] = 0,7; \;\; P[X \geq 2] = 0,83; \;\; P[Y = 3/x \geq 4] = 0,3871$$

$$P[Y > 2/x \geq 3] = 0,6071; \;\; P[2 < x \leq 4/y = 1] = 0,\widehat{3}; \;\; P[x \geq 3/y \geq 2] = 0,7263$$

# 3.181

(a) $f_1(x) = 1 \quad para \quad -1 \leq x \leq 1; \qquad f_2(x) = 1 \quad para \quad -1 \leq x \leq 1$

(b) No son independientes.

# 3.182 $E[X + Y^2] = \frac{83}{8}$

# 3.183 $E[X + Y^2] = \frac{5}{6}$

# 3.184

(a)

$$P_x(x) = \begin{cases} \frac{4}{14} & si \quad x = 1 \\ \frac{5}{14} & si \quad x = 2 \\ \frac{5}{14} & si \quad x = 3 \end{cases} \qquad P_y(y) = \begin{cases} \frac{3}{14} & si \quad y = 0 \\ \frac{2}{14} & si \quad y = 1 \\ \frac{5}{14} & si \quad y = 2 \\ \frac{4}{14} & si \quad y = 3 \end{cases}$$

(b)

$$E[X] = \frac{29}{14}; \;\; E[Y] = \frac{24}{14}; \;\; E[XY] = \frac{48}{14}$$
$$E[X^2] = \frac{69}{14}; \;\; E[Y^2] = \frac{58}{14}; \;\; \sigma_{x,y} = -\frac{6}{49}$$

# 3.185

(a) $F_x(x) = x; \qquad F_y(y) = y$

(b) $E[X] = 0$; $E[Y] = \frac{2}{3}$; $E[XY] = 0$; $\sigma_{x,y} = 0$

# 3.186

(a) $F_x(x) = \frac{x^2}{4}$; $F_y(y) = y$

(b)

$$E[X] = \frac{4}{3};\ E[Y] = \frac{2}{3};\ E[XY] = 1$$

$$Var(x) = \frac{2}{9};\ Var(y) = \frac{2}{9};\ \sigma_{x,y} = \frac{1}{9}$$

# 3.187

(a) $P(X \leq 4) = 0,78$; $P(Y \geq 2) = 0,91$

(b) $P(Y = 3/X \leq 4) = 0,346$; $P(Y > 2/X \geq 3) = 0,645$

(c) $P(2 < X \leq 4/Y = 1) = 0,\overset{\frown}{6}$; $P(X \geq 3/Y \geq 2) = 0,857$

# 3.188

(a) $K = \frac{500}{3}$

(b) $f_x(x) = \frac{125}{2} \cdot x^2 e^{-5x}$; $f_y(y) = \frac{8}{3} \cdot y^3 e^{-2y}$

(c) Las variables aleatorias son independientes.

# 3.189

(a) $K = \frac{1}{2}$

(b) $F(xy) = \frac{1}{2} \cdot xy(x + y^2)$

(c) $F_x(x) = \frac{1}{2} \cdot x(x + 1)$; $F_y(y) = \frac{1}{2} \cdot y(1 + y^2)$

(d) $F(y/x) = \frac{2xy+y^3}{1+2x}$; $F(x/y) = \frac{x^2+3xy^2}{1+3y^2}$

(e) $P(X \leq 0,1; Y \leq 0,2) = 0,0014$

(f) $P(X \geq 0,1; Y \leq 0,8) = 0,6264$

(g) $P(0,1 < X \leq 0,5; 0,3 < Y \leq 0,8) = 0,157$

(h) $P(X \leq 0,3) = 0,805$

(i) $a = 0,69$

(j) $P(X \leq 0,5/Y \leq 0,2) = 0,2596$

(k) $P(0,8 < Y \leq 0,9/0,5 < X \leq 0,7) = 0,1532$

(l) $P(X \leq 0,4/Y = 0,6) = 0,2846$

(m) $a = 0,145$

## 3.190

(a) $P(X \leq \frac{1}{2};\ \ Y \geq 0,2) = 0,27$

(b) $P(0,6 < X \leq 0,8\ /\ 0 < Y \leq 0,2;\ \ Z \geq 0,7) = 0,2276$

(c) $P(0,8 \leq X;\ \ Z \leq 0,5\ /\ Y \leq 0,9) = 0,1103$

## 3.191

$P(0 \leq x \leq 1; 0 \leq y \leq 1) = 0,4$

## 3.192

$a \geq 57,25$

## 3.193

(a) $E[x] = 4$

(b) $E[z] = 2$

(c) $\rho_{x,y} = 0,71$

# 3.194

| $X$ $Y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | $8 \cdot 10^{-}$[illegible] | $3,2 \cdot 10-4$ | 0,0128 | 0,512 |
| 1 | $7,2 \cdot 10^{-}$[illegible] | $2,8 \cdot 10^{-}$[illegible] | 0,1152 | 0 |
| 2 | $6,48 \cdot 10^{-}$[illegible] | 0,026 | 0 | 0 |
| 3 | $5,832 \cdot 10^{-}$[illegible] | 0 | 0 | 0 |

# 3.195

$E[x] = 245$ horas ; $\quad Var(x) = 9$

# 3.196

(a)

(b) $f_x(x) = 2x; \quad f_y(y) = \frac{2+2y}{3}$ $F_x(x) = x^2; \quad F_y(y) = \frac{2y+y^2}{3}$

(c) $f(x/y) = 2x; \quad f(y/x) = \frac{1+y}{3}$ $F(x/y) = x^2; \quad F(y/x) = \frac{2y+y^2}{3}$

(d) Son independientes.

# 3.197

$$f_x(x) = 3x^2 \text{ para } 0 < x < 1; \quad f_y(y) = \frac{3}{2} \cdot (1 - y^2) \text{ para } 0 < y < x$$

$f(x/y) = \frac{2x}{1-y^2}$ para $y < x < 1$

# 3.198

(a) No son independientes.

(b) $E[X_2 - X_1] = 5,5$

# 3.199

$\alpha = 0,4161$

# Soluciones Unidad Temática 4

## 4.1

(a)

$$P_x(x=k)=\begin{cases}\binom{4}{k}\cdot(0.8)^k\cdot(0.2)^{4-k} & 0\le k\le 4\\ 0 & \text{fuera}\end{cases}$$

(b)

$$F_x(x=k)=\begin{cases}\binom{4}{k}\cdot(0.8)^k\cdot(0.2)^{4-k} & 0\le k\le 4\\ 1 & \text{fuera}\end{cases}$$

## 4.2

0,65576

## 4.3

0,5138

## 4.4

$k=\ln 3/2$

## 4.5

(a) 0,0788 ; (b) 0,469 ; (c) 0,7981

## 4.6

0,28

## 4.7

(a) 0,1653 ; (b) 0,2678 ; (c) 0,5372 ; (d) 0,2717

## 4.8

(a) $F_x = \begin{cases} \dfrac{i}{20} & 0 \leq i \leq 20 \end{cases}$

(b) $P(x \leq 7) = 7/20$

(c) $E(x) = 10,5$ ; $Var(x) = 100/3$

(d) $P(x = 12) = 0$

**4.9** 0,0034

**4.10** $\mu - \sigma \cdot \sqrt{3}$ ; $\mu + \sigma \cdot \sqrt{3}$

**4.11** (a) 0,9995 ; (b) 0,0002 ; (c) 100 ; (d) 33 ;, (e) 1576

**4.12** (a) 0,0956 ; (b) 0,2487 ; (c) 0,4069

**4.13** Se ajusta

**4.14** $6,5685 \cdot 10^{-8}$

**4.15** 0,084

**4.16** (a) 0,8647 ; (b) 0,1954 ; (c) 0,0733 ; (d) 0,835

**4.17** (a) 0,002 ; (b) 0,0337

**4.18** 0,1606

**4.19** (a) 0,1404 ; (b) $3 \leq k \leq 4$

**4.20** (a) 0,4714 ; (b)(469;531)

**4.21** 7

**4.22** (a) 0,7867 ; (b) $\approx$ 6000000 ptas.

**4.23** (a) 0,1525 ; (b) 0,9925 ; (c) 0.0011

**4.24** (a) 0,2853 ; (b) 0,2796 ; (c) 0,7687 ; (d) -2,02

**4.25** 0,25

**4.26** $\mu = 74,32$ ; $\sigma = 3,22$

**4.27** (a) 0,8790 ; (b) 80,37 ; (c)5 Kg. 90 g.

**4.28** (a) 0,3678 ; (b) 11520 ; (c) 72,5

**4.29** (a) $\pm 3,92$ ; (b) +1,23 ; (c) 16 pesadas.

**4.30** (a) (i) 202 (ii) 127 ; (b) 0,8602 ; (c) 0,19025

**4.31** 3/4 ; 0,8783

**4.32** (a) $\alpha = 468,75$ ; $\sigma = 625$ (b) 727,2

**4.33** (a) 5/8, 73/960 , -1/64 ; (b) -0,205

**4.34** 0 , 0

**4.35** N(28; $\sqrt{101}$)

**4.36** No es normal bivariante.

**4.37** (a) 28,86 ; 72,97 ; (b) 2,08 ; 2,014

**4.38** (a) 2,695 ; (b) 2,135 ; (c) 0,162

**4.39** 2,028 ; 81,03

**4.40** $Var(W) = a^2 + 2b^2 \quad Var(T) = 2c^2 + d^2 \quad r = 0$

**4.41** Segundo

**4.42** 0,6642

**4.43** 0,5654

**4.44** $\frac{5 - 5e^{-\frac{x}{5}} - xe^{-\frac{x}{5}}}{x^2}$ con $x \geq 0$

**4.45** (a) 0,8264 ; (b) 14828

**4.46** 11621,6

**4.47** (a) 0,732 ; (b) 48,075 ; (c) 0,9 ; (d) 1,25

**4.48** 140

**4.49** (a) 0,0228 ; (b) 0,5403

**4.50** (a) 0,392 ; (b)0,108 ; (c) 0,634 ; (d) 0,980 ; (e) 0,554 ; (f) 0,613 ; (g) 0,115 ; (h) 0,276 ; (i) 0,799 ; (j) 0,642 ; (k) 0,667 ; (l) 0,0114

**4.51** 12,138

**4.52** 58

**4.53** (a) 0,8867 ; (b) 0,8914 ; (c) 0,5116 ; (d) 25,369 ; (e) 23,05 ; (f) 1,236

**4.54** (a) 0,9018 ; (b) 0,2563 ; (c) 0,6174 ; (d) 108,9628 ; (e) 154,8734

**4.55** (a) 0,7556 ; (b) 0,9390 ; (c) 5,276

**4.56** 3,2686 ó 0,7314

**4.57** (a) 0,863 ; (b) 0,081 ; (c) 0,034 ; (d) 0,654 ; (e) 0,014 ; (f) 0,098 ; (g) 0,799 ; (h) -0,978 ; (i) 0,308 ; (j) 0,285 ; (k) 0,852 ; (l) 3,392 ; (m) 0,728 ; (n) 0,774 ; (o) 1,531

**4.58** (a) 0,0124 ; (b) 0,4938 ; (c) 0,9415

**4.59** 51,71

**4.60** 2736,7

**4.61** 0,1685

**4.62** (a) 0,036 ; (b) 217 ó más

**4.63** N(25 ; 2,5)

**4.64** 0,994

**4.65** 198 personas

**4.66** (a) 0,0338 (b) 0,8778 (c) 0,1859

**4.67** $\mu = 6 \quad \sigma^2 = 3,6$

**4.68** 3/10

**4.69** $f(x) = 1/25 \quad si \quad x = 1, 2, ..., 25$

**4.70** $\mu = 5,5 \quad \sigma^2 = 8,25$

**4.71** (a) 16/81 (b) 64/81

**4.72** (a) 0,7073 (b) 0,4613 (c) 0,1484

**4.73** 0,1240

**4.74** 0,0006

**4.75** 0,0096

**4.76** 0,041

**4.77** 27/256

**4.78** 63/64

**4.79** (a) 0,1429 (b) 0,1353

**4.80** 0,0515

**4.81** (a) 0,0458 (b) 0,0060

**4.82** (a) 0,3840 (b) 0,1395 (c) 0,0552

**4.83** $P(x = 0) = 1/56$ $P(x = 1) = 15/56$ $P(x = 2) = 30/56$ $P(x = 3) = 10/56$

**4.84** 0,3011

**4.85** 0,3113

**4.86** 53/65

**4.87** 5/14

**4.88** 0,9517

**4.89** 3,25 ; de 0,52 a 5,98

**4.90** (a) 1/6 (b) 29/30

**4.91** 0,1042

**4.92** 0,0487

**4.93** 0,3134

**4.94** (a) 0,8635 (b) 0,2486 (c) 0

**4.95** (a) 0,00056 (b) Sería una buena compra

**4.96** (a) 1/50 (b) 0,0890

**4.97** 0,6288

**4.98** 0,2657

**4.99** (a) 0,5506 (b) 0,188 (c) 0,00586

**4.100** (a) 0,992451 (b) 0,2297 (c) 0,07353 (d) 0,991802 (e) 0,054001 (f) 0,72769

**4.101** (a) 2,23 (b) -0,72 (c) 1,03

**4.102** 0,28575

**4.103** (a) 0,8577 (b) 0,1423 (c) 0,1423

**4.104** (a) 0,152490 (b) 0,967580 (c) 0,00486

**4.105** $(\mu - 1,96\sigma; \mu + 1,96\sigma)$

**4.106** $\mu = 5 \quad \sigma = 2$

**4.107** $8,208\%$

**4.108** $\mu = 5 \quad \sigma = 1$

**4.109** $(\mu - 0,68\sigma; \mu + 0,68\sigma)$

**4.110** (a) 0,0071 (b) 0,2151 (c) 14,39 (d) 17,18

**4.111** (a) (2,56 ; 8,44) (b) (4,48 ; 6.52) (c) a partir de 7,42

# Soluciones Unidad Temática 5

## 5.1

## 5.2

(a) $\bar{X} = 6,35$ cm ; (b) $S^2 = 0,00055$ cm$^2$

## 5.3

(a) $\bar{X} = 67,45$ kg ; (b) $S^2 = 8,6136$ kg$^2$

## 5.4

La mediana es insesgada y no eficiente. Así Me= 6,36 cm

## 5.5

(a) 9,5 kg ; (b) 0,74 kg$^2$ ; (c) 0,78 y 0,86 kg.

## 5.6

(a) $\widehat{\theta} = (2x_1 + x_2)/2n$ ; (b) $\widehat{\theta} = 0,55$ y frecuencias 0,3025; 0,495 y 0,2025.

## 5.7

(a) 1200 horas ; (b) 105,4 horas.

## 5.8

(a) Las estimaciones de las desviaciones típicas de la población para los tamaños de muestra 30, 50 y 100 tubos son, respectivamente, 101,7; 101,0 y 100,5 horas. Las estimaciones de la media de la población son 1200 horas en todos los casos.

## 5.9

$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$

**5.10** $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

**5.11** $\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

**5.12** $\widehat{k} = -1 - \dfrac{n}{\ln(x_1 x_2 \cdots x_n)}$

**5.13** $\widehat{\alpha} = \frac{3n}{2(x_1^2+\cdots+x_n^2)}$

**5.14** Usar la descomposición factorial de Fisher-Neyman.

**5.15** $\widehat{\theta} \equiv \text{máx}\{X_{(n)}, -X_{(1)}\}$.

**5.16** El EMV para $\theta$ es $\widehat{\theta} = 28,8$ dias.

**5.17** $\widehat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} x_i$.

**5.18** Usar la descomposición factorial de Fisher-Neyman.

**5.19** $\widehat{\theta} = 0,5$, y las frecuencias 100, 50, 25, 25.

**5.20** $\widehat{c} = X_{(1)}$ ; $\widehat{\lambda} = \left(\bar{X} - X_{(1)}\right)^{-1}$.

**5.21** 0,2714

**5.22** 0,2894

**5.23** 0,0456

**5.24** $\widehat{p} = \frac{25}{150} = \frac{1}{6}$ ; 0,0505

**5.25** 0,215

**5.26** 0,926

**5.27** 0,506

**5.28** 0,5145

**5.29** 28 segundos.

**5.30** $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$

**5.31** $\widehat{\theta} = x_{(n)} = \max\{x_1, x_2, \cdots, x_n\} = 7,4$

**5.32** 0,008

**5.33** 0,377

# Soluciones Unidad Temática 6

**6.1** $0,04453 < \frac{\sigma_1}{\sigma_2} < 0,18144$

**6.2** $0,135 < \sigma^2 < 0,953$

**6.3** $42,053 < \sigma^2 < 155,116$

**6.4** $0,528788 < \sigma < 2,394079$

**6.5** $p \in (0,599112; 0,680888)$

**6.6** $\mu \in (9,74; 10,26)$

**6.7** $\mu \in (22475,9; 24524,1)$

**6.8** $0,2923 < \sigma^2 < 6,73554$ Afirmación válida.

**6.9** $8,3978 < \sigma^2 < 39,8427$

**6.10** $1,8619 < \sigma < 3,5777$

**6.11** $2,1981 \cdot 10^{-4} < \sigma^2 < 3,596 \ \ 10^{-3}$

**6.12** $0,498 < p < 0,642$

**6.13** $0,1405795 < p < 0,19542$

**6.14** $0,194 < p < 0,262$

**6.15** $172,15 < \mu < 176,85$

**6.16** $\mu \in (47,722; 49,278)$

**6.17** $\mu \in (74,344; 84,256)$

**6.18** $\mu \in (1,23; 1,37)$

**6.19** $\mu \in (3,04; 4,53)$

**6.20** $\mu \in (10,15; 12,45)$

**6.21** $\mu \in (1,49; 3,71)$

**6.22** $\mu \in (0,978; 1,033)$

**6.23** $\mu \in (765; 795)$

**6.24** $\mu \in (2,20; 2,30)$

**6.25** $\lambda \in (0,043; 0,154)$

**6.26** $\lambda \in (6,512 \cdot 10^{-4}; 1,374 \cdot 10^{-3})$

**6.27**

(a) $\mu \in (2,5; 2,7)$

(b) $\mu \in (2,47; 2,73)$

**6.28** $\mu \in (27,7; 38,3)$

**6.29**

(a) $\mu \in (6,449; 7,951)$

(b) $\mu_1 - \mu_2 \in (1,762; 3,238)$

**6.30** $n = 246$

**6.31**

(a) $\mu \in (787,05; 802,95)$

(b) No, porque $-t_\alpha < t$.

**6.32** $n = 1537$

**6.33** $\mu \in (252,81; 260,39)$

**6.34** No los hay porque $t < -t_\alpha$

**6.35** Afirmación del vicepresidente incorrecta,porque $Z > -Z_\alpha$.

**6.36** Si,porque $Z < -Z_\alpha$

**6.37** Inferior a 130.

**6.38** Acepto $\mu = 64$

## 6.39

(a) Acepto media 7.

(b) Rechazo,porque $t > t_\alpha$.

## 6.40

Si,porque $t > t_\alpha$.

## 6.41

(a) $\alpha \geq 0,06$

(b) No hay diferencia real.

## 6.42

Hay diferencia,porque $t_\alpha < t$.

## 6.43

Las resistencias son las mismas,porque $|t| < t_{\alpha/2}$

## 6.44

Las medias no son iguales,porque $|Z| > Z_{\alpha/2}$.

## 6.45

Si,las medias son iguales porque $|t| < t_{\alpha/2}$.

## 6.46

No.

## 6.47

Si,existen suficientes datos porque $F > F_\alpha$.

## 6.48

$\mu \in (78,4; 91,59)$

## 6.49

$\mu \in (0,5813; 0,9247)$

## 6.50

$p \in (0,1005; 0,1995)$

## 6.51

$\lambda \in (0,0008; 0,0792)$

**6.52** (a) $\sigma^2 \in (73,22; 337,45)$

(b) $\sigma^2 \in (91,6; 240,9)$

**6.53** (a) $\frac{\sigma_1^2}{\sigma_2^2} \in (0,247; 17,117)$ , (b) Si.

**6.54** $\mu_1 - \mu_2 \in (-5,822; 46,462)$

**6.55** $p_1 - p_2 \in (0,194; 0,370)$

**6.56** $d \in (1,727; 2,573)$

**6.57** $\mu_1 - \mu_2 \in (0,1235; 1,8765)$

**6.58** $\lambda \in (6,27; 8,19)$

**6.59** La vida promedio es mayor de 70 años.

**6.60** Se rechaza $H_0$. La resistencia es menor de 8 kg.

**6.61** No, el consumo no es inferior a 46 kw/h.

**6.62** No,porque $t < t_\alpha$

**6.63** El nuevo medicamento es mejor.

**6.64** $n = 271$.

**6.65** Si.

**6.66** No, pero con nivel 0,07 se acepta que $\sigma > 0,9$.

**6.67** Si,porque $F \in (0,354609929; 2,82)$.

**6.68** Si,porque $Z > Z_\alpha$.

**6.69** Estimación válida porque $Z < Z_\alpha$.

**6.70** Si, para un nivel 0.01.

**6.71** (a) No se rechaza $H_0$, (b) Se acepta $\mu = 10$.

**6.72** Se rechaza $H_0$, $\sigma^2 > 1,3$.

**6.73** Se acepta la hipótesis nula porque $F < F_{1-\alpha}$.

**6.74** Se acepta que no hay diferencia porque $|t| < t_{\alpha/2}$.

**6.75** Se acepta la hipótesis nula porque $F > F_\alpha$.

**6.76** La máquina estááaveriada.

**6.77** Si,se acepta que $\mu = 22.2$ porque $Z > Z_\alpha$.

**6.78** $n = 22$.

**6.79** Si,la altura promedio $\mu > 162,5$ cm porque $|Z| > Z_{\alpha/2}$.

**6.80** $n = 64$.

**6.81** Se rechaza $H_0$, así $p > \frac{1}{5}$.

**6.82** Se rechaza $H_0$, y por tanto $\mu_1 - \mu_2 < 12$ kg.

# Anexo I: Tablas Estadísticas.

Tabla de la Normal

| | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,00 | 0,50000 | 0,50399 | 0,50798 | 0,51197 | 0,51595 | 0,51994 | 0,52392 | 0,52790 | 0,53188 | 0,53586 |
| 0,10 | 0,53983 | 0,54380 | 0,54776 | 0,55172 | 0,55567 | 0,55962 | 0,56356 | 0,56749 | 0,57142 | 0,57535 |
| 0,20 | 0,57926 | 0,58317 | 0,58706 | 0,59095 | 0,59483 | 0,59871 | 0,60257 | 0,60642 | 0,61026 | 0,61409 |
| 0,30 | 0,61791 | 0,62172 | 0,62552 | 0,62930 | 0,63307 | 0,63683 | 0,64058 | 0,64431 | 0,64803 | 0,65173 |
| 0,40 | 0,65542 | 0,65910 | 0,66276 | 0,66640 | 0,67003 | 0,67364 | 0,67724 | 0,68082 | 0,68439 | 0,68793 |
| 0,50 | 0,69146 | 0,69497 | 0,69847 | 0,70194 | 0,70540 | 0,70884 | 0,71226 | 0,71566 | 0,71904 | 0,72240 |
| 0,60 | 0,72575 | 0,72907 | 0,73237 | 0,73565 | 0,73891 | 0,74215 | 0,74537 | 0,74857 | 0,75175 | 0,75490 |
| 0,70 | 0,75804 | 0,76115 | 0,76424 | 0,76730 | 0,77035 | 0,77337 | 0,77637 | 0,77935 | 0,78230 | 0,78524 |
| 0,80 | 0,78814 | 0,79103 | 0,79389 | 0,79673 | 0,79955 | 0,80234 | 0,80511 | 0,80785 | 0,81057 | 0,81327 |
| 0,90 | 0,81594 | 0,81859 | 0,82121 | 0,82381 | 0,82639 | 0,82894 | 0,83147 | 0,83398 | 0,83646 | 0,83891 |
| 1,00 | 0,84134 | 0,84375 | 0,84614 | 0,84849 | 0,85083 | 0,85314 | 0,85543 | 0,85769 | 0,85993 | 0,86214 |
| 1,10 | 0,86433 | 0,86650 | 0,86864 | 0,87076 | 0,87286 | 0,87493 | 0,87698 | 0,87900 | 0,88100 | 0,88298 |
| 1,20 | 0,88493 | 0,88686 | 0,88877 | 0,89065 | 0,89251 | 0,89435 | 0,89617 | 0,89796 | 0,89973 | 0,90147 |
| 1,30 | 0,90320 | 0,90490 | 0,90658 | 0,90824 | 0,90988 | 0,91149 | 0,91308 | 0,91466 | 0,91621 | 0,91774 |
| 1,40 | 0,91924 | 0,92073 | 0,92220 | 0,92364 | 0,92507 | 0,92647 | 0,92785 | 0,92922 | 0,93056 | 0,93189 |
| 1,50 | 0,93319 | 0,93448 | 0,93574 | 0,93699 | 0,93822 | 0,93943 | 0,94062 | 0,94179 | 0,94295 | 0,94408 |
| 1,60 | 0,94520 | 0,94630 | 0,94738 | 0,94845 | 0,94950 | 0,95053 | 0,95154 | 0,95254 | 0,95352 | 0,95449 |
| 1,70 | 0,95543 | 0,95637 | 0,95728 | 0,95818 | 0,95907 | 0,95994 | 0,96080 | 0,96164 | 0,96246 | 0,96327 |
| 1,80 | 0,96407 | 0,96485 | 0,96562 | 0,96638 | 0,96712 | 0,96784 | 0,96856 | 0,96926 | 0,96995 | 0,97062 |
| 1,90 | 0,97128 | 0,97193 | 0,97257 | 0,97320 | 0,97381 | 0,97441 | 0,97500 | 0,97558 | 0,97615 | 0,97670 |
| 2,00 | 0,97725 | 0,97778 | 0,97831 | 0,97882 | 0,97932 | 0,97982 | 0,98030 | 0,98077 | 0,98124 | 0,98169 |
| 2,10 | 0,98214 | 0,98257 | 0,98300 | 0,98341 | 0,98382 | 0,98422 | 0,98461 | 0,98500 | 0,98537 | 0,98574 |
| 2,20 | 0,98610 | 0,98645 | 0,98679 | 0,98713 | 0,98745 | 0,98778 | 0,98809 | 0,98840 | 0,98870 | 0,98899 |
| 2,30 | 0,98928 | 0,98956 | 0,98983 | 0,99010 | 0,99036 | 0,99061 | 0,99086 | 0,99111 | 0,99134 | 0,99158 |
| 2,40 | 0,99180 | 0,99202 | 0,99224 | 0,99245 | 0,99266 | 0,99286 | 0,99305 | 0,99324 | 0,99343 | 0,99361 |
| 2,50 | 0,99379 | 0,99396 | 0,99413 | 0,99430 | 0,99446 | 0,99461 | 0,99477 | 0,99492 | 0,99506 | 0,99520 |
| 2,60 | 0,99534 | 0,99547 | 0,99560 | 0,99573 | 0,99585 | 0,99598 | 0,99609 | 0,99621 | 0,99632 | 0,99643 |
| 2,70 | 0,99653 | 0,99664 | 0,99674 | 0,99683 | 0,99693 | 0,99702 | 0,99711 | 0,99720 | 0,99728 | 0,99736 |
| 2,80 | 0,99744 | 0,99752 | 0,99760 | 0,99767 | 0,99774 | 0,99781 | 0,99788 | 0,99795 | 0,99801 | 0,99807 |
| 2,90 | 0,99813 | 0,99819 | 0,99825 | 0,99831 | 0,99836 | 0,99841 | 0,99846 | 0,99851 | 0,99856 | 0,99861 |
| 3,00 | 0,99865 | 0,99869 | 0,99874 | 0,99878 | 0,99882 | 0,99886 | 0,99889 | 0,99893 | 0,99896 | 0,99900 |
| 3,10 | 0,99903 | 0,99906 | 0,99910 | 0,99913 | 0,99916 | 0,99918 | 0,99921 | 0,99924 | 0,99926 | 0,99929 |
| 3,20 | 0,99931 | 0,99934 | 0,99936 | 0,99938 | 0,99940 | 0,99942 | 0,99944 | 0,99946 | 0,99948 | 0,99950 |
| 3,30 | 0,99952 | 0,99953 | 0,99955 | 0,99957 | 0,99958 | 0,99960 | 0,99961 | 0,99962 | 0,99964 | 0,99965 |
| 3,40 | 0,99966 | 0,99968 | 0,99969 | 0,99970 | 0,99971 | 0,99972 | 0,99973 | 0,99974 | 0,99975 | 0,99976 |
| 3,50 | 0,99977 | 0,99978 | 0,99978 | 0,99979 | 0,99980 | 0,99981 | 0,99981 | 0,99982 | 0,99983 | 0,99983 |
| 3,60 | 0,99984 | 0,99985 | 0,99985 | 0,99986 | 0,99986 | 0,99987 | 0,99987 | 0,99988 | 0,99988 | 0,99989 |
| 3,70 | 0,99989 | 0,99990 | 0,99990 | 0,99990 | 0,99991 | 0,99991 | 0,99992 | 0,99992 | 0,99992 | 0,99992 |
| 3,80 | 0,99993 | 0,99993 | 0,99993 | 0,99994 | 0,99994 | 0,99994 | 0,99994 | 0,99995 | 0,99995 | 0,99995 |
| 3,90 | 0,99995 | 0,99995 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99997 | 0,99997 |
| 4,00 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99998 | 0,99998 | 0,99998 | 0,99998 |

# Tabla de la distribución chi-cuadrado 1/2

| n | 0,001 | 0,005 | 0,010 | 0,025 | 0,050 | 0,100 | 0,200 | 0,300 | 0,400 | 0,500 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,000002 | 0,00004 | 0,00016 | 0,00098 | 0,00393 | 0,01579 | 0,06418 | 0,14847 | 0,27500 | 0,45494 |
| 2 | 0,00200 | 0,01002 | 0,02010 | 0,05064 | 0,10259 | 0,21072 | 0,44629 | 0,71335 | 1,02165 | 1,38629 |
| 3 | 0,02430 | 0,07172 | 0,11483 | 0,21579 | 0,35185 | 0,58438 | 1,00517 | 1,42365 | 1,86917 | 2,36597 |
| 4 | 0,09080 | 0,20698 | 0,29711 | 0,48442 | 0,71072 | 1,06362 | 1,64878 | 2,19470 | 2,75284 | 3,35669 |
| 5 | 0,21022 | 0,41175 | 0,55430 | 0,83121 | 1,14548 | 1,61031 | 2,34253 | 2,99991 | 3,65550 | 4,35146 |
| 6 | 0,38104 | 0,67573 | 0,87208 | 1,23734 | 1,63538 | 2,20413 | 3,07009 | 3,82755 | 4,57015 | 5,34812 |
| 7 | 0,59850 | 0,98925 | 1,23903 | 1,68986 | 2,16735 | 2,83311 | 3,82232 | 4,67133 | 5,49324 | 6,34581 |
| 8 | 0,85715 | 1,34440 | 1,64651 | 2,17972 | 2,73263 | 3,48954 | 4,59357 | 5,52742 | 6,42264 | 7,34412 |
| 9 | 1,15191 | 1,73491 | 2,08789 | 2,70039 | 3,32512 | 4,16816 | 5,38006 | 6,39330 | 7,35703 | 8,34283 |
| 10 | 1,47865 | 2,15585 | 2,55820 | 3,24696 | 3,94030 | 4,86518 | 6,17908 | 7,26722 | 8,29547 | 9,34182 |
| 11 | 1,83375 | 2,60320 | 3,05350 | 3,81574 | 4,57481 | 5,57779 | 6,98867 | 8,14787 | 9,23729 | 10,34100 |
| 12 | 2,21413 | 3,07379 | 3,57055 | 4,40378 | 5,22603 | 6,30380 | 7,80733 | 9,03428 | 10,18197 | 11,34032 |
| 13 | 2,61720 | 3,56504 | 4,10690 | 5,00874 | 5,89186 | 7,04150 | 8,63386 | 9,92568 | 11,12914 | 12,33975 |
| 14 | 3,04072 | 4,07466 | 4,66042 | 5,62872 | 6,57063 | 7,78954 | 9,46733 | 10,82148 | 12,07848 | 13,33927 |
| 15 | 3,48251 | 4,60087 | 5,22936 | 6,26212 | 7,26093 | 8,54675 | 10,30696 | 11,72117 | 13,02975 | 14,33886 |
| 16 | 3,94171 | 5,14216 | 5,81220 | 6,90766 | 7,96164 | 9,31224 | 11,15212 | 12,62435 | 13,98273 | 15,33850 |
| 17 | 4,41624 | 5,69727 | 6,40774 | 7,56418 | 8,67175 | 10,08518 | 12,00226 | 13,53068 | 14,93727 | 16,33818 |
| 18 | 4,90480 | 6,26477 | 7,01490 | 8,23074 | 9,39045 | 10,86494 | 12,85695 | 14,43986 | 15,89321 | 17,33790 |
| 19 | 5,40666 | 6,84392 | 7,63270 | 8,90651 | 10,11701 | 11,65091 | 13,71579 | 15,35166 | 16,85044 | 18,33765 |
| 20 | 5,92101 | 7,43381 | 8,26037 | 9,59077 | 10,85080 | 12,44260 | 14,57844 | 16,26585 | 17,80883 | 19,33743 |
| 21 | 6,44669 | 8,03360 | 8,89717 | 10,28291 | 11,59132 | 13,23960 | 15,44461 | 17,18227 | 18,76831 | 20,33723 |
| 22 | 6,98287 | 8,64268 | 9,54249 | 10,98233 | 12,33801 | 14,04149 | 16,31404 | 18,10072 | 19,72880 | 21,33704 |
| 23 | 7,52912 | 9,26038 | 10,19569 | 11,68853 | 13,09051 | 14,84795 | 17,18650 | 19,02109 | 20,69020 | 22,33688 |
| 24 | 8,08466 | 9,88620 | 10,85635 | 12,40115 | 13,84842 | 15,65868 | 18,06180 | 19,94323 | 21,65249 | 23,33673 |
| 25 | 8,64943 | 10,51965 | 11,52395 | 13,11971 | 14,61140 | 16,47341 | 18,93975 | 20,86704 | 22,61558 | 24,33658 |
| 26 | 9,22224 | 11,16022 | 12,19818 | 13,84388 | 15,37916 | 17,29188 | 19,82019 | 21,79240 | 23,57943 | 25,33646 |
| 27 | 9,80288 | 11,80765 | 12,87847 | 14,57337 | 16,15139 | 18,11389 | 20,70298 | 22,71923 | 24,54400 | 26,33634 |
| 28 | 10,39071 | 12,46128 | 13,56467 | 15,30785 | 16,92788 | 18,93924 | 21,58797 | 23,64746 | 25,50925 | 27,33623 |
| 29 | 10,98614 | 13,12107 | 14,25641 | 16,04705 | 17,70838 | 19,76774 | 22,47505 | 24,57698 | 26,47514 | 28,33613 |
| 30 | 11,58763 | 13,78668 | 14,95346 | 16,79076 | 18,49267 | 20,59924 | 23,36411 | 25,50776 | 27,44162 | 29,33603 |
| 31 | 12,19612 | 14,45774 | 15,65547 | 17,53872 | 19,28056 | 21,43357 | 24,25506 | 26,43970 | 28,40868 | 30,33594 |
| 32 | 12,81042 | 15,13402 | 16,36220 | 18,29079 | 20,07191 | 22,27059 | 25,14778 | 27,37277 | 29,37629 | 31,33586 |
| 33 | 13,43120 | 15,81518 | 17,07348 | 19,04666 | 20,86652 | 23,11019 | 26,04221 | 28,30691 | 30,34441 | 32,33579 |
| 34 | 14,05677 | 16,50130 | 17,78910 | 19,80624 | 21,66428 | 23,95225 | 26,93827 | 29,24205 | 31,31302 | 33,33570 |
| 35 | 14,68808 | 17,19173 | 18,50887 | 20,56938 | 22,46501 | 24,79665 | 27,83588 | 30,17817 | 32,28211 | 34,33564 |
| 36 | 15,32433 | 17,88675 | 19,23263 | 21,33587 | 23,26862 | 25,64329 | 28,73496 | 31,11521 | 33,25166 | 35,33557 |
| 37 | 15,96520 | 18,58588 | 19,96027 | 22,10562 | 24,07494 | 26,49209 | 29,63547 | 32,05316 | 34,22163 | 36,33552 |
| 38 | 16,61090 | 19,28882 | 20,69141 | 22,87849 | 24,88389 | 27,34296 | 30,53734 | 32,99194 | 35,19201 | 37,33545 |
| 39 | 17,26123 | 19,99583 | 21,42614 | 23,65430 | 25,69538 | 28,19579 | 31,44051 | 33,93155 | 36,16279 | 38,33540 |
| 40 | 17,91664 | 20,70658 | 22,16420 | 24,43306 | 26,50930 | 29,05052 | 32,34495 | 34,87195 | 37,13397 | 39,33534 |
| 41 | 18,57586 | 21,42075 | 22,90556 | 25,21452 | 27,32556 | 29,90708 | 33,25060 | 35,81308 | 38,10550 | 40,33530 |
| 42 | 19,23838 | 22,13838 | 23,65014 | 25,99866 | 28,14405 | 30,76542 | 34,15740 | 36,75496 | 39,07738 | 41,33525 |
| 43 | 19,90535 | 22,85957 | 24,39757 | 26,78537 | 28,96471 | 31,62546 | 35,06533 | 37,69753 | 40,04960 | 42,33520 |
| 44 | 20,57640 | 23,58362 | 25,14801 | 27,57454 | 29,78750 | 32,48713 | 35,97435 | 38,64079 | 41,02216 | 43,33516 |
| 45 | 21,25092 | 24,31098 | 25,90120 | 28,36618 | 30,61226 | 33,35038 | 36,88441 | 39,58470 | 41,99503 | 44,33512 |
| 46 | 21,92888 | 25,04130 | 26,65719 | 29,16002 | 31,43900 | 34,21517 | 37,79548 | 40,52924 | 42,96821 | 45,33508 |
| 47 | 22,60974 | 25,77450 | 27,41582 | 29,95616 | 32,26761 | 35,08142 | 38,70752 | 41,47438 | 43,94168 | 46,33504 |
| 48 | 23,29441 | 26,51067 | 28,17697 | 30,75450 | 33,09807 | 35,94914 | 39,62051 | 42,42013 | 44,91543 | 47,33501 |
| 49 | 23,98257 | 27,24937 | 28,94059 | 31,55493 | 33,93029 | 36,81823 | 40,53441 | 43,36644 | 45,88947 | 48,33497 |
| 50 | 24,67356 | 27,99082 | 29,70673 | 32,35738 | 34,76424 | 37,68864 | 41,44921 | 44,31331 | 46,86378 | 49,33494 |

# Tabla de la distribución chi-cuadrado 2/2

| n | 0,6 | 0,7 | 0,8 | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 | 0,999 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,7083 | 1,0742 | 1,6424 | 2,7055 | 3,8415 | 5,0239 | 6,6349 | 7,8794 | 10,8274 |
| 2 | 1,8326 | 2,4079 | 3,2189 | 4,6052 | 5,9915 | 7,3778 | 9,2104 | 10,5965 | 13,8150 |
| 3 | 2,9462 | 3,6649 | 4,6416 | 6,2514 | 7,8147 | 9,3484 | 11,3449 | 12,8381 | 16,2660 |
| 4 | 4,0446 | 4,8784 | 5,9886 | 7,7794 | 9,4877 | 11,1433 | 13,2767 | 14,8602 | 18,4662 |
| 5 | 5,1319 | 6,0644 | 7,2893 | 9,2363 | 11,0705 | 12,8325 | 15,0863 | 16,7496 | 20,5147 |
| 6 | 6,2108 | 7,2311 | 8,5581 | 10,6446 | 12,5916 | 14,4494 | 16,8119 | 18,5475 | 22,4575 |
| 7 | 7,2832 | 8,3834 | 9,8032 | 12,0170 | 14,0671 | 16,0128 | 18,4753 | 20,2777 | 24,3213 |
| 8 | 8,3505 | 9,5245 | 11,0301 | 13,3616 | 15,5073 | 17,5345 | 20,0902 | 21,9549 | 26,1239 |
| 9 | 9,4136 | 10,6564 | 12,2421 | 14,6837 | 16,9190 | 19,0228 | 21,6660 | 23,5893 | 27,8767 |
| 10 | 10,4732 | 11,7807 | 13,4420 | 15,9872 | 18,3070 | 20,4832 | 23,2093 | 25,1881 | 29,5879 |
| 11 | 11,5298 | 12,8987 | 14,6314 | 17,2750 | 19,6752 | 21,9200 | 24,7250 | 26,7569 | 31,2635 |
| 12 | 12,5838 | 14,0111 | 15,8120 | 18,5493 | 21,0261 | 23,3367 | 26,2170 | 28,2997 | 32,9092 |
| 13 | 13,6356 | 15,1187 | 16,9848 | 19,8119 | 22,3620 | 24,7356 | 27,6882 | 29,8193 | 34,5274 |
| 14 | 14,6853 | 16,2221 | 18,1508 | 21,0641 | 23,6848 | 26,1189 | 29,1412 | 31,3194 | 36,1239 |
| 15 | 15,7332 | 17,3217 | 19,3107 | 22,3071 | 24,9958 | 27,4884 | 30,5780 | 32,8015 | 37,6978 |
| 16 | 16,7795 | 18,4179 | 20,4651 | 23,5418 | 26,2962 | 28,8453 | 31,9999 | 34,2671 | 39,2518 |
| 17 | 17,8244 | 19,5110 | 21,6146 | 24,7690 | 27,5871 | 30,1910 | 33,4087 | 35,7184 | 40,7911 |
| 18 | 18,8679 | 20,6014 | 22,7595 | 25,9894 | 28,8693 | 31,5264 | 34,8052 | 37,1564 | 42,3119 |
| 19 | 19,9102 | 21,6891 | 23,9004 | 27,2036 | 30,1435 | 32,8523 | 36,1908 | 38,5821 | 43,8194 |
| 20 | 20,9514 | 22,7745 | 25,0375 | 28,4120 | 31,4104 | 34,1696 | 37,5663 | 39,9969 | 45,3142 |
| 21 | 21,9915 | 23,8578 | 26,1711 | 29,6151 | 32,6706 | 35,4789 | 38,9322 | 41,4009 | 46,7963 |
| 22 | 23,0307 | 24,9390 | 27,3015 | 30,8133 | 33,9245 | 36,7807 | 40,2894 | 42,7957 | 48,2676 |
| 23 | 24,0689 | 26,0184 | 28,4288 | 32,0069 | 35,1725 | 38,0756 | 41,6383 | 44,1814 | 49,7276 |
| 24 | 25,1064 | 27,0960 | 29,5533 | 33,1962 | 36,4150 | 39,3641 | 42,9798 | 45,5584 | 51,1790 |
| 25 | 26,1430 | 28,1719 | 30,6752 | 34,3816 | 37,6525 | 40,6465 | 44,3140 | 46,9280 | 52,6187 |
| 26 | 27,1789 | 29,2463 | 31,7946 | 35,5632 | 38,8851 | 41,9231 | 45,6416 | 48,2898 | 54,0511 |
| 27 | 28,2141 | 30,3193 | 32,9117 | 36,7412 | 40,1133 | 43,1945 | 46,9628 | 49,6450 | 55,4751 |
| 28 | 29,2486 | 31,3909 | 34,0266 | 37,9159 | 41,3372 | 44,4608 | 48,2782 | 50,9936 | 56,8918 |
| 29 | 30,2825 | 32,4612 | 35,1394 | 39,0875 | 42,5569 | 45,7223 | 49,5878 | 52,3355 | 58,3006 |
| 30 | 31,3159 | 33,5302 | 36,2502 | 40,2560 | 43,7730 | 46,9792 | 50,8922 | 53,6719 | 59,7022 |
| 31 | 32,3486 | 34,5981 | 37,3591 | 41,4217 | 44,9853 | 48,2319 | 52,1914 | 55,0025 | 61,0980 |
| 32 | 33,3809 | 35,6649 | 38,4663 | 42,5847 | 46,1942 | 49,4804 | 53,4857 | 56,3280 | 62,4873 |
| 33 | 34,4126 | 36,7307 | 39,5718 | 43,7452 | 47,3999 | 50,7251 | 54,7754 | 57,6483 | 63,8694 |
| 34 | 35,4438 | 37,7954 | 40,6756 | 44,9032 | 48,6024 | 51,9660 | 56,0609 | 58,9637 | 65,2471 |
| 35 | 36,4746 | 38,8591 | 41,7780 | 46,0588 | 49,8018 | 53,2033 | 57,3420 | 60,2746 | 66,6192 |
| 36 | 37,5049 | 39,9220 | 42,8788 | 47,2122 | 50,9985 | 54,4373 | 58,6192 | 61,5811 | 67,9850 |
| 37 | 38,5348 | 40,9839 | 43,9782 | 48,3634 | 52,1923 | 55,6680 | 59,8926 | 62,8832 | 69,3476 |
| 38 | 39,5643 | 42,0450 | 45,0763 | 49,5126 | 53,3835 | 56,8955 | 61,1620 | 64,1812 | 70,7039 |
| 39 | 40,5935 | 43,1053 | 46,1730 | 50,6598 | 54,5722 | 58,1201 | 62,4281 | 65,4753 | 72,0550 |
| 40 | 41,6222 | 44,1649 | 47,2685 | 51,8050 | 55,7585 | 59,3417 | 63,6908 | 66,7660 | 73,4029 |
| 41 | 42,6506 | 45,2236 | 48,3628 | 52,9485 | 56,9424 | 60,5606 | 64,9500 | 68,0526 | 74,7441 |
| 42 | 43,6786 | 46,2817 | 49,4560 | 54,0902 | 58,1240 | 61,7767 | 66,2063 | 69,3360 | 76,0842 |
| 43 | 44,7063 | 47,3390 | 50,5480 | 55,2302 | 59,3035 | 62,9903 | 67,4593 | 70,6157 | 77,4184 |
| 44 | 45,7336 | 48,3957 | 51,6389 | 56,3685 | 60,4809 | 64,2014 | 68,7096 | 71,8923 | 78,7487 |
| 45 | 46,7607 | 49,4517 | 52,7288 | 57,5053 | 61,6562 | 65,4101 | 69,9569 | 73,1660 | 80,0776 |
| 46 | 47,7874 | 50,5071 | 53,8177 | 58,6405 | 62,8296 | 66,6165 | 71,2015 | 74,4367 | 81,3998 |
| 47 | 48,8139 | 51,5619 | 54,9056 | 59,7743 | 64,0011 | 67,8206 | 72,4432 | 75,7039 | 82,7198 |
| 48 | 49,8401 | 52,6161 | 55,9926 | 60,9066 | 65,1708 | 69,0226 | 73,6826 | 76,9689 | 84,0368 |
| 49 | 50,8660 | 53,6697 | 57,0786 | 62,0375 | 66,3387 | 70,2224 | 74,9194 | 78,2306 | 85,3499 |
| 50 | 51,8916 | 54,7228 | 58,1638 | 63,1671 | 67,5048 | 71,4202 | 76,1538 | 79,4898 | 86,6603 |

# Tabla de la Distribución t de Student

| n | 0,65 | 0,7 | 0,75 | 0,8 | 0,85 | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,50953 | 0,72654 | 1,00000 | 1,37638 | 1,96261 | 3,07768 | 6,31375 | 12,706 | 31,82 | 63,66 |
| 2 | 0,44475 | 0,61721 | 0,81650 | 1,06066 | 1,38621 | 1,88562 | 2,91999 | 4,30266 | 6,96455 | 9,92499 |
| 3 | 0,42420 | 0,58439 | 0,76489 | 0,97847 | 1,24978 | 1,63775 | 2,35336 | 3,18245 | 4,54071 | 5,84085 |
| 4 | 0,41416 | 0,56865 | 0,74070 | 0,94096 | 1,18957 | 1,53321 | 2,13185 | 2,77645 | 3,74694 | 4,60408 |
| 5 | 0,40823 | 0,55943 | 0,72669 | 0,91954 | 1,15577 | 1,47588 | 2,01505 | 2,57058 | 3,36493 | 4,03212 |
| 6 | 0,40431 | 0,55338 | 0,71756 | 0,90570 | 1,13416 | 1,43976 | 1,94318 | 2,44691 | 3,14267 | 3,70743 |
| 7 | 0,40154 | 0,54911 | 0,71114 | 0,89603 | 1,11916 | 1,41492 | 1,89458 | 2,36462 | 2,99795 | 3,49948 |
| 8 | 0,39947 | 0,54593 | 0,70639 | 0,88889 | 1,10815 | 1,39682 | 1,85955 | 2,30601 | 2,89647 | 3,35538 |
| 9 | 0,39787 | 0,54348 | 0,70272 | 0,88340 | 1,09972 | 1,38303 | 1,83311 | 2,26216 | 2,82143 | 3,24984 |
| 10 | 0,39659 | 0,54153 | 0,69981 | 0,87906 | 1,09306 | 1,37218 | 1,81246 | 2,22814 | 2,76377 | 3,16926 |
| 11 | 0,39555 | 0,53994 | 0,69744 | 0,87553 | 1,08767 | 1,36343 | 1,79588 | 2,20099 | 2,71808 | 3,10582 |
| 12 | 0,39469 | 0,53862 | 0,69548 | 0,87261 | 1,08321 | 1,35622 | 1,78229 | 2,17881 | 2,68099 | 3,05454 |
| 13 | 0,39396 | 0,53750 | 0,69383 | 0,87015 | 1,07947 | 1,35017 | 1,77093 | 2,16037 | 2,65030 | 3,01228 |
| 14 | 0,39333 | 0,53655 | 0,69242 | 0,86805 | 1,07628 | 1,34503 | 1,76131 | 2,14479 | 2,62449 | 2,97685 |
| 15 | 0,39279 | 0,53573 | 0,69120 | 0,86624 | 1,07353 | 1,34061 | 1,75305 | 2,13145 | 2,60248 | 2,94673 |
| 16 | 0,39232 | 0,53501 | 0,69013 | 0,86467 | 1,07114 | 1,33676 | 1,74588 | 2,11990 | 2,58349 | 2,92079 |
| 17 | 0,39190 | 0,53438 | 0,68919 | 0,86328 | 1,06903 | 1,33338 | 1,73961 | 2,10982 | 2,56694 | 2,89823 |
| 18 | 0,39153 | 0,53382 | 0,68836 | 0,86205 | 1,06717 | 1,33039 | 1,73406 | 2,10092 | 2,55238 | 2,87844 |
| 19 | 0,39120 | 0,53331 | 0,68762 | 0,86095 | 1,06551 | 1,32773 | 1,72913 | 2,09302 | 2,53948 | 2,86094 |
| 20 | 0,39091 | 0,53286 | 0,68695 | 0,85996 | 1,06402 | 1,32534 | 1,72472 | 2,08596 | 2,52798 | 2,84534 |
| 21 | 0,39064 | 0,53246 | 0,68635 | 0,85907 | 1,06267 | 1,32319 | 1,72074 | 2,07961 | 2,51765 | 2,83137 |
| 22 | 0,39039 | 0,53209 | 0,68581 | 0,85827 | 1,06145 | 1,32124 | 1,71714 | 2,07388 | 2,50832 | 2,81876 |
| 23 | 0,39017 | 0,53175 | 0,68531 | 0,85753 | 1,06034 | 1,31946 | 1,71387 | 2,06865 | 2,49987 | 2,80734 |
| 24 | 0,38997 | 0,53144 | 0,68485 | 0,85686 | 1,05932 | 1,31784 | 1,71088 | 2,06390 | 2,49216 | 2,79695 |
| 25 | 0,38978 | 0,53115 | 0,68443 | 0,85624 | 1,05838 | 1,31635 | 1,70814 | 2,05954 | 2,48510 | 2,78744 |
| 26 | 0,38961 | 0,53089 | 0,68404 | 0,85567 | 1,05752 | 1,31497 | 1,70562 | 2,05553 | 2,47863 | 2,77872 |
| 27 | 0,38945 | 0,53065 | 0,68369 | 0,85514 | 1,05673 | 1,31370 | 1,70329 | 2,05183 | 2,47266 | 2,77068 |
| 28 | 0,38930 | 0,53042 | 0,68335 | 0,85465 | 1,05599 | 1,31253 | 1,70113 | 2,04841 | 2,46714 | 2,76326 |
| 29 | 0,38916 | 0,53021 | 0,68304 | 0,85419 | 1,05530 | 1,31143 | 1,69913 | 2,04523 | 2,46202 | 2,75639 |
| 30 | 0,38903 | 0,53002 | 0,68276 | 0,85377 | 1,05466 | 1,31042 | 1,69726 | 2,04227 | 2,45726 | 2,74998 |
| 35 | 0,38850 | 0,52921 | 0,68156 | 0,85201 | 1,05202 | 1,30621 | 1,68957 | 2,03011 | 2,43772 | 2,72381 |
| 40 | 0,38810 | 0,52861 | 0,68067 | 0,85070 | 1,05005 | 1,30308 | 1,68385 | 2,02107 | 2,42326 | 2,70446 |
| 45 | 0,38779 | 0,52814 | 0,67998 | 0,84968 | 1,04852 | 1,30065 | 1,67943 | 2,01410 | 2,41212 | 2,68959 |
| 50 | 0,38754 | 0,52776 | 0,67943 | 0,84887 | 1,04729 | 1,29871 | 1,67591 | 2,00856 | 2,40327 | 2,67779 |
| 55 | 0,38734 | 0,52745 | 0,67898 | 0,84820 | 1,04630 | 1,29713 | 1,67303 | 2,00404 | 2,39608 | 2,66822 |
| 60 | 0,38717 | 0,52720 | 0,67860 | 0,84765 | 1,04547 | 1,29582 | 1,67065 | 2,00030 | 2,39012 | 2,66027 |
| 70 | 0,38691 | 0,52680 | 0,67801 | 0,84679 | 1,04417 | 1,29376 | 1,66692 | 1,99444 | 2,38080 | 2,64790 |
| 80 | 0,38671 | 0,52650 | 0,67757 | 0,84614 | 1,04319 | 1,29222 | 1,66413 | 1,99007 | 2,37387 | 2,63870 |
| 90 | 0,38655 | 0,52626 | 0,67723 | 0,84563 | 1,04244 | 1,29103 | 1,66196 | 1,98667 | 2,36850 | 2,63157 |
| 100 | 0,38643 | 0,52608 | 0,67695 | 0,84523 | 1,04184 | 1,29008 | 1,66023 | 1,98397 | 2,36421 | 2,62589 |
| 110 | 0,38633 | 0,52592 | 0,67673 | 0,84490 | 1,04134 | 1,28930 | 1,65882 | 1,98177 | 2,36072 | 2,62127 |
| 120 | 0,38624 | 0,52580 | 0,67654 | 0,84463 | 1,04093 | 1,28865 | 1,65765 | 1,97993 | 2,35783 | 2,61742 |
| 200 | 0,38587 | 0,52524 | 0,67572 | 0,84342 | 1,03913 | 1,28580 | 1,65251 | 1,97189 | 2,34513 | 2,60063 |
| 300 | 0,38569 | 0,52496 | 0,67531 | 0,84282 | 1,03823 | 1,28438 | 1,64995 | 1,96790 | 2,33884 | 2,59231 |
| 400 | 0,38560 | 0,52482 | 0,67510 | 0,84252 | 1,03778 | 1,28367 | 1,64867 | 1,96591 | 2,33571 | 2,58817 |
| 500 | 0,38554 | 0,52474 | 0,67498 | 0,84234 | 1,03751 | 1,28325 | 1,64791 | 1,96472 | 2,33383 | 2,58569 |
| 1000 | 0,38543 | 0,52457 | 0,67473 | 0,84198 | 1,03697 | 1,28240 | 1,64638 | 1,96234 | 2,33008 | 2,58075 |
| $\infty$ | 0,38532 | 0,52440 | 0,67449 | 0,84162 | 1,03643 | 1,28155 | 1,64485 | 1,95996 | 2,32634 | 2,57583 |

# Tabla de la distribución Fn,m de Snedecor $\alpha = 0,9$

| | n | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 25 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 40 | 50 | 54 | 56 | 57 | 58 | 59 | 59 | 60 | 60 | 60 | 61 | 61 | 62 | 62 | 62 | 63 | 63 | 63 | 63 |
| 2 | 8,5 | 9,0 | 9,2 | 9,2 | 9,3 | 9,3 | 9,3 | 9,4 | 9,4 | 9,4 | 9,4 | 9,4 | 9,4 | 9,4 | 9,5 | 9,5 | 9,5 | 9,5 | 9,5 | 9,5 |
| 3 | 5,54 | 5,46 | 5,39 | 5,34 | 5,31 | 5,28 | 5,27 | 5,25 | 5,24 | 5,23 | 5,22 | 5,22 | 5,20 | 5,18 | 5,17 | 5,17 | 5,16 | 5,15 | 5,14 | 5,13 |
| 4 | 4,54 | 4,32 | 4,19 | 4,11 | 4,05 | 4,01 | 3,98 | 3,95 | 3,94 | 3,92 | 3,91 | 3,90 | 3,87 | 3,84 | 3,83 | 3,82 | 3,80 | 3,79 | 3,78 | 3,76 |
| 5 | 4,06 | 3,78 | 3,62 | 3,52 | 3,45 | 3,40 | 3,37 | 3,34 | 3,32 | 3,30 | 3,28 | 3,27 | 3,24 | 3,21 | 3,19 | 3,17 | 3,16 | 3,14 | 3,12 | 3,10 |
| 6 | 3,78 | 3,46 | 3,29 | 3,18 | 3,11 | 3,05 | 3,01 | 2,98 | 2,96 | 2,94 | 2,92 | 2,90 | 2,87 | 2,84 | 2,81 | 2,80 | 2,78 | 2,76 | 2,74 | 2,72 |
| 7 | 3,59 | 3,26 | 3,07 | 2,96 | 2,88 | 2,83 | 2,78 | 2,75 | 2,72 | 2,70 | 2,68 | 2,67 | 2,63 | 2,59 | 2,57 | 2,56 | 2,54 | 2,51 | 2,49 | 2,47 |
| 8 | 3,46 | 3,11 | 2,92 | 2,81 | 2,73 | 2,67 | 2,62 | 2,59 | 2,56 | 2,54 | 2,52 | 2,50 | 2,46 | 2,42 | 2,40 | 2,38 | 2,36 | 2,34 | 2,32 | 2,29 |
| 9 | 3,36 | 3,01 | 2,81 | 2,69 | 2,61 | 2,55 | 2,51 | 2,47 | 2,44 | 2,42 | 2,40 | 2,38 | 2,34 | 2,30 | 2,27 | 2,25 | 2,23 | 2,21 | 2,18 | 2,16 |
| 10 | 3,29 | 2,92 | 2,73 | 2,61 | 2,52 | 2,46 | 2,41 | 2,38 | 2,35 | 2,32 | 2,30 | 2,28 | 2,24 | 2,20 | 2,17 | 2,16 | 2,13 | 2,11 | 2,08 | 2,06 |
| 11 | 3,23 | 2,86 | 2,66 | 2,54 | 2,45 | 2,39 | 2,34 | 2,30 | 2,27 | 2,25 | 2,23 | 2,21 | 2,17 | 2,12 | 2,10 | 2,08 | 2,05 | 2,03 | 2,00 | 1,97 |
| 12 | 3,18 | 2,81 | 2,61 | 2,48 | 2,39 | 2,33 | 2,28 | 2,24 | 2,21 | 2,19 | 2,17 | 2,15 | 2,10 | 2,06 | 2,03 | 2,01 | 1,99 | 1,96 | 1,93 | 1,90 |
| 13 | 3,14 | 2,76 | 2,56 | 2,43 | 2,35 | 2,28 | 2,23 | 2,20 | 2,16 | 2,14 | 2,12 | 2,10 | 2,05 | 2,01 | 1,98 | 1,96 | 1,93 | 1,90 | 1,88 | 1,85 |
| 14 | 3,10 | 2,73 | 2,52 | 2,39 | 2,31 | 2,24 | 2,19 | 2,15 | 2,12 | 2,10 | 2,07 | 2,05 | 2,01 | 1,96 | 1,93 | 1,91 | 1,89 | 1,86 | 1,83 | 1,80 |
| 15 | 3,07 | 2,70 | 2,49 | 2,36 | 2,27 | 2,21 | 2,16 | 2,12 | 2,09 | 2,06 | 2,04 | 2,02 | 1,97 | 1,92 | 1,89 | 1,87 | 1,85 | 1,82 | 1,79 | 1,76 |
| 16 | 3,05 | 2,67 | 2,46 | 2,33 | 2,24 | 2,18 | 2,13 | 2,09 | 2,06 | 2,03 | 2,01 | 1,99 | 1,94 | 1,89 | 1,86 | 1,84 | 1,81 | 1,78 | 1,75 | 1,72 |
| 17 | 3,03 | 2,64 | 2,44 | 2,31 | 2,22 | 2,15 | 2,10 | 2,06 | 2,03 | 2,00 | 1,98 | 1,96 | 1,91 | 1,86 | 1,83 | 1,81 | 1,78 | 1,75 | 1,72 | 1,69 |
| 18 | 3,01 | 2,62 | 2,42 | 2,29 | 2,20 | 2,13 | 2,08 | 2,04 | 2,00 | 1,98 | 1,95 | 1,93 | 1,89 | 1,84 | 1,80 | 1,78 | 1,75 | 1,72 | 1,69 | 1,66 |
| 19 | 2,99 | 2,61 | 2,40 | 2,27 | 2,18 | 2,11 | 2,06 | 2,02 | 1,98 | 1,96 | 1,93 | 1,91 | 1,86 | 1,81 | 1,78 | 1,76 | 1,73 | 1,70 | 1,67 | 1,63 |
| 20 | 2,97 | 2,59 | 2,38 | 2,25 | 2,16 | 2,09 | 2,04 | 2,00 | 1,96 | 1,94 | 1,91 | 1,89 | 1,84 | 1,79 | 1,76 | 1,74 | 1,71 | 1,68 | 1,64 | 1,61 |
| 21 | 2,96 | 2,57 | 2,36 | 2,23 | 2,14 | 2,08 | 2,02 | 1,98 | 1,95 | 1,92 | 1,90 | 1,87 | 1,83 | 1,78 | 1,74 | 1,72 | 1,69 | 1,66 | 1,62 | 1,59 |
| 22 | 2,95 | 2,56 | 2,35 | 2,22 | 2,13 | 2,06 | 2,01 | 1,97 | 1,93 | 1,90 | 1,88 | 1,86 | 1,81 | 1,76 | 1,73 | 1,70 | 1,67 | 1,64 | 1,60 | 1,57 |
| 23 | 2,94 | 2,55 | 2,34 | 2,21 | 2,11 | 2,05 | 1,99 | 1,95 | 1,92 | 1,89 | 1,87 | 1,84 | 1,80 | 1,74 | 1,71 | 1,69 | 1,66 | 1,62 | 1,59 | 1,55 |
| 24 | 2,93 | 2,54 | 2,33 | 2,19 | 2,10 | 2,04 | 1,98 | 1,94 | 1,91 | 1,88 | 1,85 | 1,83 | 1,78 | 1,73 | 1,70 | 1,67 | 1,64 | 1,61 | 1,57 | 1,53 |
| 25 | 2,92 | 2,53 | 2,32 | 2,18 | 2,09 | 2,02 | 1,97 | 1,93 | 1,89 | 1,87 | 1,84 | 1,82 | 1,77 | 1,72 | 1,68 | 1,66 | 1,63 | 1,59 | 1,56 | 1,52 |
| 30 | 2,88 | 2,49 | 2,28 | 2,14 | 2,05 | 1,98 | 1,93 | 1,88 | 1,85 | 1,82 | 1,79 | 1,77 | 1,72 | 1,67 | 1,63 | 1,61 | 1,57 | 1,54 | 1,50 | 1,46 |
| 40 | 2,84 | 2,44 | 2,23 | 2,09 | 2,00 | 1,93 | 1,87 | 1,83 | 1,79 | 1,76 | 1,74 | 1,71 | 1,66 | 1,61 | 1,57 | 1,54 | 1,51 | 1,47 | 1,42 | 1,38 |
| 60 | 2,79 | 2,39 | 2,18 | 2,04 | 1,95 | 1,87 | 1,82 | 1,77 | 1,74 | 1,71 | 1,68 | 1,66 | 1,60 | 1,54 | 1,50 | 1,48 | 1,44 | 1,40 | 1,35 | 1,29 |
| 120 | 2,75 | 2,35 | 2,13 | 1,99 | 1,90 | 1,82 | 1,77 | 1,72 | 1,68 | 1,65 | 1,63 | 1,60 | 1,55 | 1,48 | 1,44 | 1,41 | 1,37 | 1,32 | 1,26 | 1,19 |
| $\infty$ | 2,71 | 2,30 | 2,08 | 1,94 | 1,85 | 1,77 | 1,72 | 1,67 | 1,63 | 1,60 | 1,57 | 1,55 | 1,49 | 1,42 | 1,38 | 1,34 | 1,30 | 1,24 | 1,17 | 1,00 |

## Tabla de la distribución Fn,m de Snedecor $\alpha = 0,95$

| m \ n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 25 | 30 | 40 | 60 | 120 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161 | 199 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 243 | 244 | 246 | 248 | 249 | 250 | 251 | 252 | 253 | 254 |
| 2 | 18,5 | 19,0 | 19,2 | 19,2 | 19,3 | 19,3 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,5 | 19,5 | 19,5 | 19,5 | 19,5 | 19,5 |
| 3 | 10,1 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,89 | 8,85 | 8,81 | 8,79 | 8,76 | 8,74 | 8,70 | 8,66 | 8,63 | 8,62 | 8,59 | 8,57 | 8,55 | 8,53 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 | 5,94 | 5,91 | 5,86 | 5,80 | 5,77 | 5,75 | 5,72 | 5,69 | 5,66 | 5,63 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,77 | 4,74 | 4,70 | 4,68 | 4,62 | 4,56 | 4,52 | 4,50 | 4,46 | 4,43 | 4,40 | 4,37 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 | 4,03 | 4,00 | 3,94 | 3,87 | 3,83 | 3,81 | 3,77 | 3,74 | 3,70 | 3,67 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,64 | 3,60 | 3,57 | 3,51 | 3,44 | 3,40 | 3,38 | 3,34 | 3,30 | 3,27 | 3,23 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,35 | 3,31 | 3,28 | 3,22 | 3,15 | 3,11 | 3,08 | 3,04 | 3,01 | 2,97 | 2,93 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,14 | 3,10 | 3,07 | 3,01 | 2,94 | 2,89 | 2,86 | 2,83 | 2,79 | 2,75 | 2,71 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 | 2,94 | 2,91 | 2,85 | 2,77 | 2,73 | 2,70 | 2,66 | 2,62 | 2,58 | 2,54 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 | 2,82 | 2,79 | 2,72 | 2,65 | 2,60 | 2,57 | 2,53 | 2,49 | 2,45 | 2,40 |
| 12 | 4,75 | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 | 2,72 | 2,69 | 2,62 | 2,54 | 2,50 | 2,47 | 2,43 | 2,38 | 2,34 | 2,30 |
| 13 | 4,67 | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 | 2,63 | 2,60 | 2,53 | 2,46 | 2,41 | 2,38 | 2,34 | 2,30 | 2,25 | 2,21 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 | 2,57 | 2,53 | 2,46 | 2,39 | 2,34 | 2,31 | 2,27 | 2,22 | 2,18 | 2,13 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 | 2,51 | 2,48 | 2,40 | 2,33 | 2,28 | 2,25 | 2,20 | 2,16 | 2,11 | 2,07 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,46 | 2,42 | 2,35 | 2,28 | 2,23 | 2,19 | 2,15 | 2,11 | 2,06 | 2,01 |
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 | 2,41 | 2,38 | 2,31 | 2,23 | 2,18 | 2,15 | 2,10 | 2,06 | 2,01 | 1,96 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,37 | 2,34 | 2,27 | 2,19 | 2,14 | 2,11 | 2,06 | 2,02 | 1,97 | 1,92 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 | 2,34 | 2,31 | 2,23 | 2,16 | 2,11 | 2,07 | 2,03 | 1,98 | 1,93 | 1,88 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 | 2,31 | 2,28 | 2,20 | 2,12 | 2,07 | 2,04 | 1,99 | 1,95 | 1,90 | 1,84 |
| 21 | 4,32 | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,49 | 2,42 | 2,37 | 2,32 | 2,28 | 2,25 | 2,18 | 2,10 | 2,05 | 2,01 | 1,96 | 1,92 | 1,87 | 1,81 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 | 2,26 | 2,23 | 2,15 | 2,07 | 2,02 | 1,98 | 1,94 | 1,89 | 1,84 | 1,78 |
| 23 | 4,28 | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,44 | 2,37 | 2,32 | 2,27 | 2,24 | 2,20 | 2,13 | 2,05 | 2,00 | 1,96 | 1,91 | 1,86 | 1,81 | 1,76 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 | 2,22 | 2,18 | 2,11 | 2,03 | 1,97 | 1,94 | 1,89 | 1,84 | 1,79 | 1,73 |
| 25 | 4,24 | 3,39 | 2,99 | 2,76 | 2,60 | 2,49 | 2,40 | 2,34 | 2,28 | 2,24 | 2,20 | 2,16 | 2,09 | 2,01 | 1,96 | 1,92 | 1,87 | 1,82 | 1,77 | 1,71 |
| 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 | 2,13 | 2,09 | 2,01 | 1,93 | 1,88 | 1,84 | 1,79 | 1,74 | 1,68 | 1,62 |
| 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 | 2,04 | 2,00 | 1,92 | 1,84 | 1,78 | 1,74 | 1,69 | 1,64 | 1,58 | 1,51 |
| 60 | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 | 1,95 | 1,92 | 1,84 | 1,75 | 1,69 | 1,65 | 1,59 | 1,53 | 1,47 | 1,39 |
| 120 | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,18 | 2,09 | 2,02 | 1,96 | 1,91 | 1,87 | 1,83 | 1,75 | 1,66 | 1,60 | 1,55 | 1,50 | 1,43 | 1,35 | 1,25 |
| $\infty$ | 3,84 | 3,00 | 2,60 | 2,37 | 2,21 | 2,10 | 2,01 | 1,94 | 1,88 | 1,83 | 1,79 | 1,75 | 1,67 | 1,57 | 1,51 | 1,46 | 1,39 | 1,32 | 1,22 | 1,00 |

## Tabla de la distribución Fn,m de Snedecor $\alpha = 0,95$

| | n | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 25 | 30 | 40 | 60 | 120 | $\infty$ |
| 1 | 4052 | 4999 | 5404 | 5624 | 5764 | 5859 | 5928 | 5981 | 6022 | 6056 | 6083 | 6107 | 6157 | 6209 | 6240 | 6260 | 6286 | 6313 | 6340 | 6366 |
| 2 | 98,5 | 99,0 | 99,2 | 99,3 | 99,3 | 99,3 | 99,4 | 99,4 | 99,4 | 99,4 | 99,4 | 99,4 | 99,4 | 99,4 | 99,5 | 99,5 | 99,5 | 99,5 | 99,5 | 99,5 |
| 3 | 34,1 | 30,8 | 29,5 | 28,7 | 28,2 | 27,9 | 27,7 | 27,5 | 27,3 | 27,2 | 27,1 | 27,1 | 26,9 | 26,7 | 26,6 | 26,5 | 26,4 | 26,3 | 26,2 | 26,1 |
| 4 | 21,2 | 18,0 | 16,7 | 16,0 | 15,5 | 15,2 | 15,0 | 14,8 | 14,7 | 14,5 | 14,5 | 14,4 | 14,2 | 14,0 | 13,9 | 13,8 | 13,7 | 13,7 | 13,6 | 13,5 |
| 5 | 16,3 | 13,3 | 12,1 | 11,4 | 11,0 | 10,7 | 10,5 | 10,3 | 10,2 | 10,1 | 9,96 | 9,89 | 9,72 | 9,55 | 9,45 | 9,38 | 9,29 | 9,20 | 9,11 | 9,02 |
| 6 | 13,7 | 10,9 | 9,78 | 9,15 | 8,75 | 8,47 | 8,26 | 8,10 | 7,98 | 7,87 | 7,79 | 7,72 | 7,56 | 7,40 | 7,30 | 7,23 | 7,14 | 7,06 | 6,97 | 6,88 |
| 7 | 12,2 | 9,55 | 8,45 | 7,85 | 7,46 | 7,19 | 6,99 | 6,84 | 6,72 | 6,62 | 6,54 | 6,47 | 6,31 | 6,16 | 6,06 | 5,99 | 5,91 | 5,82 | 5,74 | 5,65 |
| 8 | 11,3 | 8,65 | 7,59 | 7,01 | 6,63 | 6,37 | 6,18 | 6,03 | 5,91 | 5,81 | 5,73 | 5,67 | 5,52 | 5,36 | 5,26 | 5,20 | 5,12 | 5,03 | 4,95 | 4,86 |
| 9 | 10,6 | 8,02 | 6,99 | 6,42 | 6,06 | 5,80 | 5,61 | 5,47 | 5,35 | 5,26 | 5,18 | 5,11 | 4,96 | 4,81 | 4,71 | 4,65 | 4,57 | 4,48 | 4,40 | 4,31 |
| 10 | 10,0 | 7,56 | 6,55 | 5,99 | 5,64 | 5,39 | 5,20 | 5,06 | 4,94 | 4,85 | 4,77 | 4,71 | 4,56 | 4,41 | 4,31 | 4,25 | 4,17 | 4,08 | 4,00 | 3,91 |
| 11 | 9,65 | 7,21 | 6,22 | 5,67 | 5,32 | 5,07 | 4,89 | 4,74 | 4,63 | 4,54 | 4,46 | 4,40 | 4,25 | 4,10 | 4,01 | 3,94 | 3,86 | 3,78 | 3,69 | 3,60 |
| 12 | 9,33 | 6,93 | 5,95 | 5,41 | 5,06 | 4,82 | 4,64 | 4,50 | 4,39 | 4,30 | 4,22 | 4,16 | 4,01 | 3,86 | 3,76 | 3,70 | 3,62 | 3,54 | 3,45 | 3,36 |
| 13 | 9,07 | 6,70 | 5,74 | 5,21 | 4,86 | 4,62 | 4,44 | 4,30 | 4,19 | 4,10 | 4,02 | 3,96 | 3,82 | 3,66 | 3,57 | 3,51 | 3,43 | 3,34 | 3,25 | 3,17 |
| 14 | 8,86 | 6,51 | 5,56 | 5,04 | 4,69 | 4,46 | 4,28 | 4,14 | 4,03 | 3,94 | 3,86 | 3,80 | 3,66 | 3,51 | 3,41 | 3,35 | 3,27 | 3,18 | 3,09 | 3,00 |
| 15 | 8,68 | 6,36 | 5,42 | 4,89 | 4,56 | 4,32 | 4,14 | 4,00 | 3,89 | 3,80 | 3,73 | 3,67 | 3,52 | 3,37 | 3,28 | 3,21 | 3,13 | 3,05 | 2,96 | 2,87 |
| 16 | 8,53 | 6,23 | 5,29 | 4,77 | 4,44 | 4,20 | 4,03 | 3,89 | 3,78 | 3,69 | 3,62 | 3,55 | 3,41 | 3,26 | 3,16 | 3,10 | 3,02 | 2,93 | 2,84 | 2,75 |
| 17 | 8,40 | 6,11 | 5,19 | 4,67 | 4,34 | 4,10 | 3,93 | 3,79 | 3,68 | 3,59 | 3,52 | 3,46 | 3,31 | 3,16 | 3,07 | 3,00 | 2,92 | 2,83 | 2,75 | 2,65 |
| 18 | 8,29 | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,84 | 3,71 | 3,60 | 3,51 | 3,43 | 3,37 | 3,23 | 3,08 | 2,98 | 2,92 | 2,84 | 2,75 | 2,66 | 2,57 |
| 19 | 8,18 | 5,93 | 5,01 | 4,50 | 4,17 | 3,94 | 3,77 | 3,63 | 3,52 | 3,43 | 3,36 | 3,30 | 3,15 | 3,00 | 2,91 | 2,84 | 2,76 | 2,67 | 2,58 | 2,49 |
| 20 | 8,10 | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,70 | 3,56 | 3,46 | 3,37 | 3,29 | 3,23 | 3,09 | 2,94 | 2,84 | 2,78 | 2,69 | 2,61 | 2,52 | 2,42 |
| 21 | 8,02 | 5,78 | 4,87 | 4,37 | 4,04 | 3,81 | 3,64 | 3,51 | 3,40 | 3,31 | 3,24 | 3,17 | 3,03 | 2,88 | 2,79 | 2,72 | 2,64 | 2,55 | 2,46 | 2,36 |
| 22 | 7,95 | 5,72 | 4,82 | 4,31 | 3,99 | 3,76 | 3,59 | 3,45 | 3,35 | 3,26 | 3,18 | 3,12 | 2,98 | 2,83 | 2,73 | 2,67 | 2,58 | 2,50 | 2,40 | 2,31 |
| 23 | 7,88 | 5,66 | 4,76 | 4,26 | 3,94 | 3,71 | 3,54 | 3,41 | 3,30 | 3,21 | 3,14 | 3,07 | 2,93 | 2,78 | 2,69 | 2,62 | 2,54 | 2,45 | 2,35 | 2,26 |
| 24 | 7,82 | 5,61 | 4,72 | 4,22 | 3,90 | 3,67 | 3,50 | 3,36 | 3,26 | 3,17 | 3,09 | 3,03 | 2,89 | 2,74 | 2,64 | 2,58 | 2,49 | 2,40 | 2,31 | 2,21 |
| 25 | 7,77 | 5,57 | 4,68 | 4,18 | 3,85 | 3,63 | 3,46 | 3,32 | 3,22 | 3,13 | 3,06 | 2,99 | 2,85 | 2,70 | 2,60 | 2,54 | 2,45 | 2,36 | 2,27 | 2,17 |
| 30 | 7,56 | 5,39 | 4,51 | 4,02 | 3,70 | 3,47 | 3,30 | 3,17 | 3,07 | 2,98 | 2,91 | 2,84 | 2,70 | 2,55 | 2,45 | 2,39 | 2,30 | 2,21 | 2,11 | 2,01 |
| 40 | 7,31 | 5,18 | 4,31 | 3,83 | 3,51 | 3,29 | 3,12 | 2,99 | 2,89 | 2,80 | 2,73 | 2,66 | 2,52 | 2,37 | 2,27 | 2,20 | 2,11 | 2,02 | 1,92 | 1,80 |
| 60 | 7,08 | 4,98 | 4,13 | 3,65 | 3,34 | 3,12 | 2,95 | 2,82 | 2,72 | 2,63 | 2,56 | 2,50 | 2,35 | 2,20 | 2,10 | 2,03 | 1,94 | 1,84 | 1,73 | 1,60 |
| 120 | 6,85 | 4,79 | 3,95 | 3,48 | 3,17 | 2,96 | 2,79 | 2,66 | 2,56 | 2,47 | 2,40 | 2,34 | 2,19 | 2,03 | 1,93 | 1,86 | 1,76 | 1,66 | 1,53 | 1,38 |
| $\infty$ | 6,63 | 4,61 | 3,78 | 3,32 | 3,02 | 2,80 | 2,64 | 2,51 | 2,41 | 2,32 | 2,25 | 2,18 | 2,04 | 1,88 | 1,77 | 1,70 | 1,59 | 1,47 | 1,32 | 1,00 |

Se terminó de componer este libro
el 16 de octubre fiesta de
Santa Eduvigis, patrona de Polonia,
a la que la alta condición social no impidió
ni la sobriedad del atuendo ni la sinceridad de la vida

SERVICIO·DE·PUBLICACIONES
UNIVERSIDAD·DE·CADIZ

ISBN 84-7786-685-6
9 788477 866855