RECOPILADO POR:
*MOISÉS MEDINA LÓPEZ*

# ESTADÍSTICA: PROBABILIDAD

BIOESTADÍSTICA

## Tabla de contenido

# PROBABILIDAD:

## Introducción

La **probabilidad** mide la frecuencia con la que aparece un resultado determinado cuando se realiza un experimento.

> **Ejemplo**: tiramos un dado al aire y queremos saber cual es la probabilidad de que salga un 2, o que salga un número par, o que salga un número menor que 4.

**El experimento tiene que ser aleatorio**, es decir, que pueden presentarse diversos resultados, dentro de un conjunto posible de soluciones, y esto aún realizando el experimento en las mismas condiciones. Por lo tanto, a priori no se conoce cual de los resultados se va a presentar:

> **Ejemplos**: lanzamos una moneda al aire: el resultado puede ser cara o cruz, pero no sabemos de antemano cual de ellos va a salir.
>
> En la Lotería de Navidad, el "Gordo" (en España se llama "Gordo" al primer premio) puede ser cualquier número entre el 1 y el 100.000, pero no sabemos a priori cual va a ser (si lo supiéramos no estaríamos aquí escribiendo esta lección).

**Hay experimentos que no son aleatorios** y por lo tanto no se les puede aplicar las reglas de la probabilidad.

> **Ejemplo**: en lugar de tirar la moneda al aire, directamente selccionamos la cara. Aquí no podemos hablar de probabilidades, sino que ha sido un resultado determinado por uno mismo.

Antes de calcular las probabilidades de un experimento aleaotorio hay que definir una serie de conceptos:
**Suceso elemental:** hace referencia a cada una de las posibles soluciones que se pueden presentar.

> **Ejemplo**: al lanzar una moneda al aire, los sucesos elementales son la cara y la cruz. Al lanzar un dado, los sucesos elementales son el 1, el 2, .., hasta el 6.

**Suceso compuesto**: es un subconjunto de sucesos elementales.

**Ejemplo**: lanzamos un dado y queremos que salga un número par. El suceso "numero par" es un suceso compuesto, integrado por 3 sucesos elementales: el 2, el 4 y el 6

O, por ejemplo, jugamos a la ruleta y queremos que salga "menor o igual que 18". Este es un suceso compuesto formado por 18 sucesos elementales (todos los números que van del 1 al 18).

Al conjunto de todos los posibles sucesos elementales lo denominamos **espacio muestral**. Cada experimento aleatorio tiene definido su espacio muestral (es decir, un conjunto con todas las soluciones posibles).

**Ejemplo:** si tiramos una moneda al aíre una sola vez, el espacio muestral será cara o cruz.

Si el experimento consiste en lanzar una moneda al aire dos veces, entonces el espacio muestral estaría formado por (cara-cara), (cara-cruz), (cruz-cara) y (cruz-cruz).

## Probabilidad: Relación entre sucesos

-

Entre los sucesos compuestos se pueden establecer distintas relaciones:

**a) Un suceso puede estar contenido en otro:** las posibles soluciones del primer suceso también lo son del segundo, pero este segundo suceso tiene además otras soluciones suyas propias.

**Ejemplo**: lanzamos un dado y analizamos dos sucesos: a) que salga el número 6, y b) que salga un número par. Vemos que el suceso a) está contenido en el suceso b).

Siempre que se da el suceso a) se da el suceso b), pero no al contrario. Por ejemplo, si el resultado fuera el 2, se cumpliría el suceso b), pero no el el a).

**b) Dos sucesos pueden ser iguales:** esto ocurre cuando siempre que se cumple uno de ellos se cumple obligatoriamente el otro y viceversa.

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que salga múltiplo de 2. Vemos que las soluciones coinciden en ambos casos.

**c) Unión de dos o más sucesos:** la unión será otro suceso formado por todos los elementos de los sucesos que se unen.

**Ejemplo:** lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par y b) que el resultado sea mayor que 3. El suceso unión estaría formado por los siguientes resultados: el 2, el 4, el 5 y el 6

**d) Intersección de sucesos:** es aquel suceso compuesto por los elementos comunes de dos o más sucesos que se intersectan.

**Ejemplo**: lanzamos un dado al aire, y analizamos dos sucesos: a) que salga número par, y b) que sea mayor que 4. La intersección de estos dos sucesos tiene un sólo elemento, el número 6 (es el único resultado común a ambos sucesos: es mayor que 4 y es número par).

**e) Sucesos incompatibles:** son aquellos que no se pueden dar al mismo tiempo ya que no tienen elementos comunes (su interesección es el conjunto vacio).

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga un número menor que 3, y b) que salga el número 6. Es evidente que ambos no se pueden dar al mismo tiempo.

**f) Sucesos complementarios:** son aquellos que si no se da uno, obligatoriamente se tiene que dar el otro.

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga un número par, y b) que salga un número impar. Vemos que si no se da el primero se tiene que dar el segundo (y viceversa).

## Cálculo de probabilidades

### Probabilidad

Como hemos comentado anteriormente, la probabilidad mide la mayor o menor posibilidad de que se dé un determinado resultado (suceso) cuando se realiza un experimento aleatorio.

La probabilidad toma valores entre 0 y 1 (o expresados en tanto por ciento, entre 0% y 100%):

**El valor cero corresponde al suceso imposible**: lanzamos un dado al aire y la probabilidad de que salga el número 7 es cero (al menos, si es un dado certificado por la OMD, "Organización Mundial de Dados").

**El valor uno corresponde al suceso seguro:** lanzamos un dado al aire y la probabilidad de que salga cualquier número del 1 al 6 es igual a uno (100%).

**El resto de sucesos tendrá probabilidades entre cero y uno**: que será tanto mayor cuanto más probable sea que dicho suceso tenga lugar.

**¿Cómo se mide la probabilidad?**
Uno de los métodos más utilizados es aplicando la **Regla de Laplace:** define la probabilidad de un suceso como el cociente entre casos favorables y casos posibles.

**P(A) = Casos favorables / casos posibles**

Veamos algunos **ejemplos**:

**a) Probabilidad de que al lanzar un dado salga el número 2**: el caso favorable es tan sólo uno (que salga el dos), mientras que los casos posibles son seis (puede salir cualquier número del uno al seis). Por lo tanto:

P(A) = 1 / 6 = 0,166 (o lo que es lo mismo, 16,6%)

**b) Probabilidad de que al lanzar un dado salga un número par**: en este caso los casos favorables son tres (que salga el dos, el cuatro o el seis), mientras que los casos posibles siguen siendo seis. Por lo tanto:

P(A) = 3 / 6 = 0,50 (o lo que es lo mismo, 50%)

**c) Probabilidad de que al lanzar un dado salga un número menor que 5**: en este caso tenemos cuatro casos favorables (que salga el uno, el dos, el tres o el cuatro), frente a los seis casos posibles. Por lo tanto:

P(A) = 4 / 6 = 0,666 (o lo que es lo mismo, 66,6%)

**d) Probabilidad de que nos toque el "Gordo" de Navidad**: tan sólo un caso favorable, el número que jugamos (¡qué triste...¡), frente a 100.000 casos posibles. Por lo tanto:

P(A) = 1 / 100.000 = 0,00001 (o lo que es lo mismo, 0,001%)

Merece la pena ...... Por cierto, tiene la misma probabilidad el número 45.264, que el número 00001, pero ¿cuál de los dos comprarías?

Para poder aplicar la **Regla de Laplace** el experimento aleatorio tiene que cumplir **dos requisitos:**

> **a) El número de resultados posibles (sucesos) tiene que ser finito.** Si hubiera infinitos resultados, al aplicar la regla "casos favorables / casos posibles" el cociente siempre sería cero.
>
> **b) Todos los sucesos tienen que tener la misma probabilidad.** Si al lanzar un dado, algunas caras tuvieran mayor probabilidad de salir que otras, no podríamos aplicar esta regla.

A la regla de Laplace también se le denomina **"probabilidad a priori"**, ya que para aplicarla hay que conocer antes de realizar el experimento cuales son los posibles resultados y saber que todos tienen las mismas probabilidades.

**¿Y si el experimento aleatorio no cumple los dos requisitos indicados, qué hacemos?, ¿ponemos una denuncia?**

No, no va a ser necesario denunciar a nadie, ya que en este caso podemos acudir a otro modelo de cálculo de probabilidades que se basa en la experiencia **(modelo frecuentista)**:

> Cuando se realiza un experimento aleatorio un número muy elevado de veces, las probabilidades de los diversos posibles sucesos empiezan a converger hacia valores determinados, que son sus respectivas probabilidades.
>
> **Ejemplo**: si lanzo una vez una moneda al aire y sale "cara", quiere decir que el suceso "cara" ha aparecido el 100% de las veces y el suceso "cruz" el 0%.
>
> Si lanzo diez veces la moneda al aire, es posible que el suceso "cara" salga 7 veces y el suceso "cruz" las 3 restantes. En este caso, la probabilidad del suceso "cara" ya no sería del 100%, sino que se habría reducido al 70%.
>
> Si repito este experimento un número elevado de veces, lo normal es que las probabilidades de los sucesos "cara" y "cruz" se vayan aproximando al 50% cada una. Este 50% será la probabilidad de estos sucesos según el modelo frecuentista.

En este modelo ya no será necesario que el número de soluciones sea finito, ni que todos los sucesos tengan la misma probabilidad.

> **Ejemplo**: si la moneda que utilizamos en el ejemplo anterior fuera defectuosa (o estuviera trucada), es posible que al repetir dicho experimento un número elevado de veces, la "cara" saliera con una frecuencia, por ejemplo, del 65% y la "cruz" del

35%. Estos valores serían las probabilidades de estos dos sucesos según el modelo frecuentista.

A esta definición de la probabilidad se le denomina **probabilidad a posteriori**, ya que tan sólo repitiendo un experimento un número elevado de veces podremos saber cual es la probabilidad de cada suceso.

# Probabilidad de sucesos

### Probabilidad de sucesos

Al definir los sucesos hablamos de las diferentes relaciones que pueden guardar dos sucesos entre sí, así como de las posibles relaciones que se pueden establecer entre los mismos. Vamos a ver ahora cómo se refleja esto en el cálculo de probabilidades.

**a) Un suceso puede estar contenido en otro**: entonces, la probabilidad del primer suceso será menor que la del suceso que lo contiene.

**Ejemplo**: lanzamos un dado y analizamos dos sucesos: a) que salga el número 6, y b) que salga un número par. Dijimos que el suceso a) está contenido en el suceso b).

$$P(A) = 1/6 = 0,166$$

$$P(B) = 3 / 6 = 0,50$$

Por lo tanto, podemos ver que la probabilidad del suceso contenido, suceso a), es menor que la probabilidad del suceso que lo contiene, suceso b).

**b) Dos sucesos pueden ser iguales**: en este caso, las probabilidades de ambos sucesos son las mismas.

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que salga múltiplo de 2. Las soluciones coinciden en ambos casos.

$$P(A) = 3 / 6 = 0,50$$

$$P(B) = 3 / 6 = 0,50$$

**c) Intersección de sucesos**: es aquel suceso compuesto por los elementos comunes de los dos o más sucesos que se intersectan. La probabilidad será igual a la probabilidad de los elemntos comunes.

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que sea mayor que 3. La intersección de estos dos sucesos tiene dos elementos: el 4 y el 6.

Su probabilidad será por tanto:

$$P(A \wedge B) = 2 / 6 = 0,33$$

**d) Unión de dos o más sucesos**: la probabilidad de la unión de dos sucesos es igual a la suma de las probabilidades individuales de los dos sucesos que se unen, menos la probabilidad del suceso intersección

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que el resultado sea mayor que 3. El suceso unión estaría formado por los siguientes resultados: el 2, el 4, el 5 y el 6.

$$P(A) = 3 / 6 = 0,50$$

$$P(B) = 3 / 6 = 0,50$$

$$P (A \wedge B) = 2 / 6 = 0,33$$

Por lo tanto,

$$P (A \text{ u } B) = (0,50 + 0,50) - 0,33 = 0,666$$

**e) Sucesos incompatibles**: la probabilidad de la unión de dos sucesos incompatibles será igual a la suma de las probabilidades de cada uno de los sucesos (ya que su intersección es el conjunto vacio y por lo tanto no hay que restarle nada).

**Ejemplo**: lanzamos un dado al aire y analizamos dos sucesos: a) que salga un número menor que 3, y b) que salga el número 6.

La probabilidad del suceso unión de estos dos sucesos será igual a:

$$P(A) = 2 / 6 = 0,333$$

$$P(B) = 1 / 6 = 0,166$$

Por lo tanto,

$$P(A \text{ u } B) = 0,33 + 0,166 = 0,50$$

**f) Sucesos complementarios:** la probabilidad de un suceso complementario a un suceso (A) es igual a 1 - P(A)

**Ejemplo**: lanzamos un dado al aire. el suceso (A) es que salga un número par, luego su complementario, suceso (B), es que salga un número impar.

La probabilidad del suceso (A) es igual a :

$$P(A) = 3 / 6 = 0,50$$

Luego, la probabilidad del suceso (B) es igual a:

$$P(B) = 1 - P(A) = 1 - 0,50 = 0,50$$

Se puede comprobar aplicando la regla de "casos favorables / casos posibles":

$$P(B) = 3 / 6 = 0,50$$

**g) Unión de sucesos complementarios:** la probabilidad de la unión de dos sucesos complementarios es igual a 1.

> **Ejemplo**: seguimos con el ejemplo anterior: a) que salga un número par, y b) que salga un número impar. La probabilidad del suceso unión de estos dos sucesos será igual a:

$$P(A) = 3 / 6 = 0,50$$
$$P(B) = 3 / 6 = 0,50$$

Por lo tanto,

$$P(A \cup B) = 0,50 + 0,50 = 1$$

## Combinaciones, Variaciones y Permutaciones (I)

Para aplicar la **Regla de Laplace**, el cálculo de los sucesos favorables y de los sucesos posibles a veces no plantea ningún problema, ya que son un número reducido y se pueden calcular con facilidad:

> **Por ejemplo**: Probabilidad de que al lanzar un dado salga el número 2. Tan sólo hay un caso favorable, mientras que los casos posibles son seis.
>
> Probabilidad de acertar al primer intento el horóscopo de una persona. Hay un caso favorable y 12 casos posibles.

Sin embargo, a veces calcular el número de casos favorables y casos posibles es complejo y hay que aplicar reglas matemáticas:

> **Por ejemplo**: 5 matrimonios se sientan aleatoriamente a cenar y queremos calcular la probabilidad de que al menos los miembros de un matrimonio se sienten junto. En este caso, determinar el número de casos favorables y de casos posibles es complejo.

Las reglas matemáticas que nos pueden ayudar son el cálculo de **combinaciones**, el cálculo de **variaciones** y el cálculo de **permutaciones**.

**a) Combinaciones:**

Determina el número de subgrupos de 1, 2, 3, etc. elementos que se pueden formar con los "n" elementos de una nuestra. Cada subgrupo se diferencia del resto en los elementos que lo componen, sin que influya el orden.

> **Por ejemplo**, calcular las posibles combinaciones de 2 elementos que se pueden formar con los números 1, 2 y 3.
>
> Se pueden establecer 3 parejas diferentes: (1,2), (1,3) y (2,3). En el cálculo de combinaciones las parejas (1,2) y (2,1) se consideran idénticas, por lo que sólo se cuentan una vez.

**b) Variaciones**:

Calcula el número de subgrupos de 1, 2, 3, etc.elementos que se pueden establecer con los "n" elementos de una muestra. Cada subgrupo se diferencia del resto en los elementos que lo componen o en el orden de dichos elementos (es lo que le diferencia de las combinaciones).

> **Por ejemplo**, calcular las posibles variaciones de 2 elementos que se pueden establecer con los número 1, 2 y 3.
>
> Ahora tendríamos 6 posibles parejas: (1,2), (1,3), (2,1), (2,3), (3,1) y (3,3). En este caso los subgrupos (1,2) y (2,1) se consideran distintos.

**c) Permutaciones:**

Cálcula las posibles agrupaciones que se pueden establecer con todos los elementos de un grupo, por lo tanto, lo que diferencia a cada subgrupo del resto es el orden de los elementos.

> **Por ejemplo**, calcular las posibles formas en que se pueden ordenar los número 1, 2 y 3.
>
> Hay 6 posibles agrupaciones: (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2) y (3, 2, 1)

## Combinaciones, Variaciones y Permutaciones (II)

### ¿Cómo se calculan?

**a) Combinaciones:**

Para calcular el número de combinaciones se aplica la siguiente fórmula:

$$C_{m,n} = \frac{m\,!}{n\,! * (m - n)\,!}$$

El termino **" n ! "** se denomina "factorial de n" y es la multiplicación de todos los números que van desde "n" hasta 1.

**Por ejemplo**: 4 ! = 4 * 3 * 2 * 1 = 24

La expresión **"$C_{m,n}$"** representa las combinaciones de "m" elementos, formando subgrupos de "n" elementos.

**Ejemplo**: $C_{10,4}$ son las combinaciones de 10 elementos agrupándolos en subgrupos de 4 elementos:

$$C_{10,4} = \frac{10\,!}{4\,! * (10 - 4)\,!} = \frac{10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1}{(4 * 3 * 2 * 1) \; * (6 * 5 * 4 * 3 * 2 * 1)} = 210$$

Es decir, podríamos formar 210 subgrupos diferentes de 4 elementos, a partir de los 10 elementos.

**b) Variaciones:**

Para calcular el número de variaciones se aplica la siguiente fórmula:

$$V_{m,n} = \frac{m\,!}{(m - n)\,!}$$

La expresión **"$V_{m,n}$"** representa las variaciones de "m" elementos, formando subgrupos de "n" elementos. En este caso, como vimos en la lección anterior, un subgrupo se diferenciará del resto, bien por los elementos que lo forman, o bien por el orden de dichos elementos.

**Ejemplo**: $V_{10,4}$ son las variaciones de 10 elementos agrupándolos en subgrupos de 4 elementos:

$$V_{10,4} = \frac{10\,!}{(10 - 4)\,!} = \frac{10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1}{(6 * 5 * 4 * 3 * 2 * 1)} = 5.040$$

Es decir, podríamos formar 5.040 subgrupos diferentes de 4 elementos, a partir de los 10 elementos.

**c) Permutaciones:**

Para calcular el número de permutaciones se aplica la siguiente fórmula:

$$P_m = m\,!$$

La expresión **"$P_m$"** representa las permutaciones de "m" elementos, tomando todos los elementos. Los subgrupos se diferenciaran únicamente por el orden de los elementos.

**Ejemplo**: $P_{10}$ son las permutaciones de 10 elementos:

$$P_{10} = 10! = 10*9*8*7*6*5*4*3*2*1 = 3.628.800$$

Es decir, tendríamos 3.628.800 formas diferentes de agrupar 10 elementos.

## Combinaciones, Variaciones y Permutaciones (III)

Vamos a analizar ahora que ocurriría con el cálculo de las combinaciones, de las variaciones o de las permutaciones en el **supuesto** de que al formar los subgrupos **los elementos pudieran repetirse**.

**Por ejemplo**: tenemos bolas de 6 colores diferentes y queremos formar subgrupos en los que pudiera darse el caso de que 2, 3, 4 o todas las bolas del subgrupo tuvieran el mismo color. En este caso no podríamos utilizar las fórmulas que vimos en la lección anterior.

**a) Combinaciones con repetición:**
Para calcular el número de combinaciones con repetición se aplica la siguiente fórmula:

$$C'_{m,n} = \frac{(m+n-1)!}{n!*(m-1)!}$$

**Ejemplo**: $C'_{10,4}$ son las combinaciones de 10 elementos con repetición, agrupándolos en subgrupos de 4, en los que 2, 3 o los 4 elementos podrían estar repetidos:

$$C'_{10,4} = \frac{13!}{4!*9!} = \frac{13*12*11*10*9*8*7*6*5*4*3*2*1}{(4*3*2*1)*(9*8*7*6*5*4*3*2*1)} = 715$$

Es decir, podríamos formar 715 subgrupos diferentes de 4 elementos.

**b) Variaciones con repetición:**
Para calcular el número de variaciones con repetición se aplica la siguiente fórmula:

$$V'_{m,n} = m^n$$

**Ejemplo**: $V'_{10,4}$ son las variaciones de 10 elementos con repetición, agrupándolos en subgrupos de 4 elementos:

$$V'_{10,4} = 10^{4} = 10.000$$

Es decir, podríamos formar 10.000 subgrupos diferentes de 4 elementos.

**c) Permutaciones con repetición:**

Para calcular el número de permutaciones con repetición se aplica la siguiente fórmula:

$$P'^{\,x1,\,x2,\,\ldots xk}_{m} = \frac{m\,!}{x_1\,!\,*\,x_2\,!\,*\,\ldots\,*\,x_k\,!}$$

Son permutaciones de "m" elementos, en los que uno de ellos se repite " $x_1$ " veces, otro " $x_2$ " veces y así ... hasta uno que se repite " $x_k$ " veces.

**Ejemplo**: Calcular las permutaciones de 10 elementos, en los que uno de ellos se repite en 2 ocasiones y otro se repite en 3 ocasiones:

$$P'^{\,2,\,3}_{10} = \frac{10\,!}{2\,!\,*\,3\,!} = 302.400$$

Es decir, tendríamos 302,400 formas diferentes de agrupar estos 10 elementos.

# Ejercicios

### 1.- Ejercicio

Calcular la probabilidad de acertar los 14 signos de la quiniela:

**Solución:**

Se aplica la **Regla de Laplace** (casos favorables / casos posibles). El **caso favorable** es tan sólo uno (acertar los 14 signos). Los casos posibles se calculan como variaciones con repetición de 3 elementos (1, X y 2), tomados de 14 en 14 (los signos que hay que rellenar).

Son variaciones y no combinaciones ya que el orden influye: no es lo mismo (1,1,X) que (1, X, 1). Y son con repetición, ya que cualquiera de los signos (1, X y 2) se puede repetir hasta 14 veces.

Por lo tanto, los **casos posibles** son:

$$V'_{3,14} = 3^{\wedge 14} = 4.782.969$$

Y la probabilidad de acertar los 14 resultados es:

$$P(A) = \frac{1}{4.782.969} = 0,000000209$$

No demasiado elevada....pero el que la sigue la consigue.

**2.- Ejercicio**

Y la probabilidad de acertar 12 signos de la quiniela:

**Solución:**

Aplicamos nuevamente la **Regla de Laplace**. En este caso los **casos favorables** se calculan como combinaciones de 14 elementos tomados de 2 en 2, de esta manera obtenemos todas las posibles alternativas de fallar 2 resultados de 14 (lo que equivale a acertar 12 resultados). Utilizamos combinaciones y no variaciones ya que el orden no importa (da lo mismo fallar el 3º y el 6º, que el 6º y el 3º)

$$C_{14,2} = \frac{14\,!}{2\,! * (14 - 2)\,!} = 91$$

Los **casos posibles** siguen siendo los mismos:

$$V'_{3,14} = 3^{\wedge 14} = 4.782.969$$

Por lo que la probabilidad de acertar 12 resultados es:

$$P(A) = \frac{91}{4.782.969} = 0,0000190$$

Por lo tanto, tenemos más probabilidades de acertar 12 resultados que 14 (¿será por eso por lo que pagan menos?).

**3.- Ejercicio**

Calcular la probabilidad de, en una carrera de 12 caballos, acertar los 3 que quedan primeros (sin importar cual de ellos queda primero, cual segundo y cual tercero).

**Solución:**

Se aplica la **Regla de Laplace**. El **caso favorable** es tan sólo uno: los 3 caballos que entran en primer lugar. Los **casos posibles** se calculan como combinaciones de 12 elementos tomados de 3 en 3 (es decir, determinamos todos las posibles alternativas de 3 caballos que pueden entrar en las 3 primeras posiciones). Como el orden de estos 3 primeros caballos no importa, utilizamos combinaciones en lugar de variaciones.

Por lo tanto, los casos posibles son:

$$C_{12,3} = \frac{12\,!}{3\,!\,*\,(12-3)\,!} = 220$$

Por lo que la probabilidad de acertar los 3 caballos ganadores es:

$$P(A) = \frac{1}{220} = 0{,}00455$$

Algo mayor que en las quinielas.... Eso sí, se paga menos.

**4.- Ejercicio**

Y si hubiera que acertar, no sólo los 3 caballos que ganan, sino el orden de su entrada en meta.

**Solución:**

El **caso favorable** sigue siendo uno: los 3 caballos que entran en primer lugar, colocados en su orden correspondiente.

Los **casos posibles** se calculan ahora como variaciones (ya que el orden influye) de 12 elementos tomados de 3 en 3 (calculamos todas las posibles maneras en que los 12 caballos podrían ocupar las 3 primeras posiciones.

$$V_{12,3} = \frac{12\,!}{(12-3)\,!} = 1.320$$

Por lo que la probabilidad de acertar los 3 caballos ganadores es:

$$P(A) = \frac{1}{1.320} = 0{,}00076$$

Menor que en el ejemplo 3º. Ya no vale acertar que 3 caballos entran en primer lugar, sino que tenemos que acertar el orden de su entrada.

## Probabilidad condicionada

Las **probabilidades condicionadas** se calculan una vez que se ha incorporado información adicional a la situación de partida:

> **Ejemplo**: se tira un dado y sabemos que la probabilidad de que salga un 2 es 1/6 (probabilidad a priori). Si incorporamos nueva información (por ejemplo, alguien nos dice que el resultado ha sido un número par) entonces la probabilidad de que el resultado sea el 2 ya no es 1/6.

Las probabilidades condicionadas se calculan aplicando la siguiente **fórmula**:

$$P(B/A) = \frac{P(A \wedge B)}{P(A)}$$

Donde:

**P (B/A)** es la probabilidad de que se de el suceso B condicionada a que se haya dado el suceso A.

**P (B $\wedge$ A)** es la probabilidad del suceso simultáneo de A y de B

**P (A)** es la probabilidad a priori del suceso A

En el **ejemplo** que hemos visto:

**P (B/A)** es la probabilidad de que salga el número 2 (suceso B) condicionada a que haya salido un número par (suceso A).

**P (B $\wedge$ A)** es la probabilidad de que salga el dos y número par.

**P (A)** es la probabilidad a priori de que salga un número par.

Por lo tanto:

$$P(B \wedge A) = 1/6$$
$$P(A) = 1/2$$
$$P(B/A) = (1/6) / (1/2) = 1/3$$

Luego, la probabilidad de que salga el número 2, si ya sabemos que ha salido un número par, es de 1/3 (mayor que su probabilidad a priori de 1/6).

**2º ejemplo:**

En un estudio sanitario se ha llegado a la conclusión de que la probabilidad de que una persona sufra problemas coronarios (suceso B) es el 0,10 (probabilidad a priori).

Además, la probabilidad de que una persona sufra problemas de obesidad (suceso A) es el 0,25 y la probabilidad de que una persona sufra a la vez problemas de obesidad y coronarios (suceso intersección de A y B) es del 0,05.

Calcular la probabilidad de que una persona sufra problemas coronarios si está obesa (probabilidad condicionada P(B/A)).

$$P(B \wedge A) = 0{,}05$$
$$P(A) = 0{,}25$$
$$P(B/A) = 0{,}05 / 0{,}25 = 0{,}20$$

Por lo tanto, la probabilidad condicionada es superior a la probabilidad a priori. No siempre esto es así, a veces la probabilidad condicionada es igual a la probabilidad a priori o menor.

**Por ejemplo**: probabilidad de que al tirar un dado salga el número 2, condicionada a que haya salido un número impar.

La probabilidad condicionada es en este caso cero, frente a una probabilidad a priori de 1/6.

## Probabilidad compuesta

La **probabilidad compuesta** (o regla de multiplicación de probabilidades) se deriva de la probabilidad condicionada:

> La probabilidad de que se den simultáneamente dos sucesos (suceso intersección de A y B) es igual a la probabilidad a priori del suceso A multiplicada por la probabilidad del suceso B condicionada al cumplimiento del suceso A.

La **fórmula** para calcular esta probabilidad compuesta es:

$$P(A \wedge B) = P(B/A) * P(A)$$

**Ejemplo 1º** : Estudiamos el suceso A (porcentaje de varones mayores de 40 años casados) y el suceso B (varones mayores de 40 años con más de 2 hijos) y obtenemos la siguiente información:

> Un 35% de los varones mayores de 40 años están casados.
>
> De los varones mayores de 40 años y casados, un 30% tienen más de 2 hijos (suceso B condicionado al suceso A).

**Calcular la probabilidad de que un varón mayor de 40 años esté casado y tenga más de 2 hijos** (suceso intersección de A y B).
Por lo tanto:

$$P(A) = 0{,}35$$
$$P(B/A) = 0{,}30$$
$$P(A \wedge B) = 0{,}35 * 0{,}30 = 0{,}105$$

Es decir, un 10,5% de los varones mayores de 40 años están casados y tienen más de 2 hijos.
**2º ejemplo:** Estudiamos el suceso A (alumnos que hablan inglés) y el suceso B (alumnos que hablan alemán) y obtenemos la siguiente información:

> Un 50% de los alumnos hablan inglés.
>
> De los alumnos que hablan inglés, un 20% hablan también alemán (suceso B condicionado al suceso A).

**Calcular la probabilidad de que un alumno hable inglés y alemán** (suceso intersección de A y B).
Por lo tanto:

$$P(A) = 0{,}50$$

$$P\ (B/A) = 0{,}20$$

$$P\ (A \wedge B) = 0{,}50 * 0{,}20 = 0{,}10$$

Es decir, un 10% de los alumnos hablan inglés y alemán.

## Teorema de la probabilidad total

El **Teorema de la probabilidad total** nos permite calcular la probabilidad de un suceso a partir de probabilidades condicionadas:

> **Ejemplo:** supongamos que si llueve la probabilidad de que ocurra un accidentes es x% y si hace buen tiempo dicha probabilidad es y%. Este teorema nos permite deducir cuál es la probabilidad de que ocurra un accidente si conocemos la probabilidad de que llueva y la probabilidad de que haga buen tiempo.

La **fórmula** para calcular esta probabilidad es:

$$P\ (B) = \Sigma\ (A_i) * P\ (B/A_i) \qquad \text{(donde "i" toma valores entre 1 y n)}$$

Es decir, **la probabilidad de que ocurra el suceso B** (en nuestro ejemplo, que ocurra un accidente) **es igual a la suma de multiplicar cada una de las probabilidades condicionadas** de este suceso con los diferentes sucesos A (probabilidad de un accidente cuando llueve y cuando hace buen tiempo) **por la probabilidad de cada suceso A.**

Para que este teorema se pueda aplicar hace falta cumplir **un requisito**:

> **Los sucesos A tienen que formar un sistema completo**, es decir, que contemplen todas las posibilidades (la suma de sus probabilidades debe ser el 100%).
>
> **Ejemplo:** al tirar una moneda, el suceso "salir cara" y el suceso "salir cruz" forman un sistema completo, no hay más alternativas: la suma de sus probabilidades es el 100%
>
> **Ejemplo:** al tirar un dado, que salga el 1, el 2, el 3, o el 4 no forman un sistema completo, ya que no contempla todas las opciones (podría salir el 5 o el 6). En este caso no se podría aplicar el teorema de la probabilidad total.

**Ejercicio 1º:** En un saquito hay papeletas de tres colores, con las siguientes probabilidades de ser elegidas:

**a) Amarilla**: probabilidad del 50%.

**b) Verde**: probabilidad del 30%

**c) Roja**: probabilidad del 20%.

Según el color de la papeleta elegida, podrás participar en diferentes sorteos. Así, si la papeleta elegida es:

**a) Amarilla**: participas en un sorteo con una probabilidad de ganar del 40%.

**b) Verde**: participas en otro sorteo con una probabilidad de ganar del 60%

**c) Roja**: participas en un tercer sorteo con una probabilidad de ganar del 80%.

Con esta información, **¿qué probabilidad tienes de ganar el sorteo en el que participes?:**

1.- Las tres papeletas forman un sistema completo: sus probabilidades suman 100%

2.- Aplicamos la fórmula:

$$P(B) = \Sigma (A_i) * P(B/A_i)$$

Luego,

P (B) = (0,50 * 0,40) + (0,30 * 0,60) + (0,20 * 0,80) = 0,54

Por tanto, la probabilidad de que ganes el sorteo es del 54%.

**Ejercicio 2º:** Van a cambiar a tu jefe y se barajan diversos candidatos:

**a) Carlos**, con una probabilidad del 60%

**b) Juan**, con una probabilidad del 30%

**c) Luis**, con una probabilidad del 10%

En función de quien sea tu próximo jefe, la probabilidad de que te suban el sueldo es la siguiente:

**a) Si sale Carlos:** la probabilidad de que te suban el sueldo es del 5%.

**b) Si sale Juan:** la probabilidad de que te suban el sueldo es del 20%.

**c) Si sale Luis:** la probabilidad de que te suban el sueldo es del 60%.

En definitiva, **¿cual es la probabilidad de que te suban el sueldo?**:

1.- Los tres candidatos forman un sistema completo

2.- Aplicamos la fórmula:

P (B) = (0,60 * 0,05) + (0,30 * 0,20) + (0,10 * 0,60) = 0,15

Por tanto, la probabilidad de que te suban el sueldo es del 15%. Lo llevas claro amigo...

## Teorema de Bayes

El **Teorema de Bayes** viene a seguir el **proceso inverso** al que hemos visto en el **Teorema de la probabilidad total**:

**Teorema de la probabilidad total:** a partir de las probabilidades del suceso A (probabilidad de que llueva o de que haga buen tiempo) deducimos la probabilidad del suceso B (que ocurra un accidente).

**Teorema de Bayes:** a partir de que ha ocurrido el suceso B (ha ocurrido un accidente) deducimos las probabilidades del suceso A (¿estaba lloviendo o hacía buen tiempo?).

La **fórmula** del Teorema de Bayes es:

$$P\,(Ai/B) = \frac{P\,(Ai) * P\,(B/Ai)}{\sum P\,(Ai) * P\,(B/Ai)}$$

Tratar de explicar estar fórmula con palabras es un galimatías, así que vamos a intentar explicarla con un ejemplo. De todos modos, antes de entrar en el ejercicio, recordar que este teorema también exige que el **suceso A forme un sistema completo.**

**Ejercicio 1º:** El parte meteorológico ha anunciado tres posibilidades para el fin de semana:

**a) Que llueva**: probabilidad del 50%.

**b) Que nieve**: probabilidad del 30%

**c) Que haya niebla**: probabilidad del 20%.

Según estos posibles estados meteorológicos, la posibilidad de que ocurra un accidente es la siguiente:

**a) Si llueve**: probabilidad de accidente del 10%.

**b) Si nieva**: probabilidad de accidente del 20%

**c) Si hay niebla**: probabilidad de accidente del 5%.

Resulta que efectivamente ocurre un accidente y como no estabamos en la ciudad no sabemos que tiempo hizo (nevó, llovío o hubo niebla). El teorema de Bayes nos permite calcular estas probabilidades:

Las probabilidades que manejamos antes de conocer que ha ocurrido un accidente se denominan **"probabilidades a priori"** (lluvia con el 60%, nieve con el 30% y niebla con el 10%).

Una vez que incorporamos la información de que ha ocurrido un accidente, las probabilidades del suceso A cambian: son probabilidades condicionadas P (A/B), que se denominan **"probabilidades a posteriori"**.

Vamos a aplicar la fórmula:

$$P(Ai/B) = \frac{P(Ai) * P(B/Ai)}{\Sigma P(Ai) * P(B/Ai)}$$

a) **Probabilidad de que estuviera lloviendo:**

$$P(Ai/B) = \frac{0{,}50 * 0{,}20}{(0{,}50 * 0{,}20) + (0{,}30 * 0{,}10) + (0{,}20 * 0{,}05)} = 0{,}714$$

La probabilidad de que efectivamente estuviera lloviendo el día del accidente (probabilidad a posteriori) es del 71,4%.

b) **Probabilidad de que estuviera nevando:**

$$P(Ai/B) = \frac{0{,}30 * 0{,}10}{(0{,}50 * 0{,}20) + (0{,}30 * 0{,}10) + (0{,}20 * 0{,}05)} = 0{,}214$$

La probabilidad de que estuviera nevando es del 21,4%.

c) **Probabilidad de que hubiera niebla:**

$$P(Ai/B) = \frac{0{,}20 * 0{,}05}{(0{,}50 * 0{,}20) + (0{,}30 * 0{,}10) + (0{,}20 * 0{,}05)} = 0{,}071$$

La probabilidad de que hubiera niebla es del 7,1%.

## Independencia de sucesos

Dos **sucesos son independientes** entre sí, si la ocurrencia de uno de ellos no afecta para nada a la ocurrencia del otro:

> **Ejemplo:** el suceso estatura de los alumnos de una clase y el color del pelo son independientes: el que un alumno sea más o menos alto no va a influir en el color de su cabello, ni viceversa.

Para que dos sucesos sean independientes tienen que verificar al menos una de las siguientes **condiciones**:

> **P (B/A) = P (B)** es decir, que la probabilidad de que se de el suceso B, condicionada a que previamente se haya dado el suceso A, es exactamente igual a la probabilidad de B.
>
> > **Ejemplo:** la probabilidad de que al tirar una moneda salga cara (suceso B), condicionada a que haga buen tiempo (suceso A), es igual a la propia probabilidad del suceso B.
>
> **P (A/B) = P (A)** es decir, que la probabilidad de que se de el suceso A, condicionada a que previamente se haya dado el suceso B, es exactamente igual a la probabilidad de A.
>
> > **Ejemplo:** la probabilidad de que haga buen tiempo (suceso A), condicionada a que al tirar una moneda salga cara (suceso B), es igual a la propia probabilidad del suceso A.
>
> **P (A $\wedge$ B) = P (A) * P (B)** es decir, que la probabilidad de que se de el suceso conjunto A y B es exactamente igual a la probabilidad del suceso A multiplicada por la probabilidsad del suceso B.
>
> > **Ejemplo:** la probabilidad de que haga buen tiempo (suceso A) y salga cara al tirar una moneda (suceso B), es igual a la probabilidad del suceso A multiplicada por la probabilidad del suceso B

Si el **suceso A es independiente del suceso B**, entonces **el suceso B** también es **independiente del suceso A.**

**Ejemplo 1º:** analicemos dos sucesos:

> **Suceso A**: la probabilidad de que haga buen tiempo es del 0,4
>
> **Suceso B**: la probabilidad de tener un accidente es del 0,1
>
> **Suceso intersección**: la probabilidad de que haga buen tiempo y tener un accidente es del 0,08

Veamos si se cumple alguna de las condiciones señaladas:

P (B/A) = P (A ∧ B) / P (A) = 0,08 / 0,4 = 0,2 (que no es igual a P (B))

P (A/B) = P (A ∧ B) / P (B) = 0,08 / 0,6 = 0,133 (que no es igual a P (A))

P (A ∧ B) = 0,08 (que no es igual a P (A) multiplicado por P (B))

Por lo tanto, no se cumple ninguna de las tres condiciones señaladas por lo que **estos dos sucesos no son independientes**, sino que existe algún grado de dependencia entre ellos.

**Ejemplo 2º:** analicemos dos sucesos:

**Suceso A**: la probabilidad de que haga buen tiempo es del 0,4

**Suceso B**: la probabilidad de salir cara al lanzar una moneda es del 0,5

**Suceso intersección**: la probabilidad de que haga buen tiempo y que salga cara es 0,2

Veamos si se cumple alguna de las condiciones señaladas:

P (B/A) = P (A ∧ B) / P (A) = 0,2 / 0,4 = 0,5 (igual que P (B))

P (A/B) = P (A ∧ B) / P (B) = 0,2 / 0,6 = 0,4 (igual que P (A))

P (A ∧ B) = 0,2 (igual a P (A) multiplicado por P (B))

Por lo tanto, **estos dos sucesos sí son independientes**.

# EJERCICIOS RESUELTOS DE PROBABILIDAD

1. Un dado tiene 3 caras de 1 punto, dos caras con una X y una cara con 2 puntos. Calcular la probabilidad de que salga una X o un 2.

Solución: El evento "una X o un 2" tiene 3 posibilidades de ocurrir (casos favorables), mientras que el espacio muestral consiste de las 6 caras (espacio muestral). Por tanto la probabilidad pedida es 3/6=1/2=0.5.

2. En una urna hay 3 bolas blancas, 2 rojas y 4 azules. Calcula la probabilidad de que al extraer una bola al azar, ésta sea roja.

Respuesta: 2/9

3. Se extrae una carta de una baraja española de 40 cartas, y se consideran los siguientes sucesos: *A*= "obtener una de oros", *B* = "obtener una sota" y *C* = "obtener un tres". Decide si estos tres eventos son compatibles o incompatibles. Justifica tu respuesta.

Nota: Se dice que dos o más eventos son incompatibles (mutuamente excluyentes) si no es posible que ocurran al mismo tiempo. En otras palabras, si la ocurrencia de uno de ellos excluye o impide la ocurrencia de al menos uno de los otros.

*Solución:*

Son incompatibles, porque *B* y *C* no se pueden dar a la vez (*B* y *C* son incompatibles).

4. Un dado está trucado para que el 6 tenga una probabilidad de 0.25. ¿Cuál es la probabilidad de no obtener un 6?

Respuesta: 0.75

5. En el lanzamiento de un dado, consideramos los eventos *A* = {2, 3} y *B* = {2, 4, 6}. Describe el evento unión *AUB* y el evento intersección de *AB*.

6. Se lanza 100 veces un dado y se obtiene:

| Cara | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Frecuencia absoluta | 12 | 17 | 18 | 16 | 18 | 19 |

Calcula la frecuencia relativa del suceso "obtener múltiplo de 3".

Resúesta: 0.37

7. Se lanza dos veces un dado. Representamos el espacio muestral de la siguiente forma: {(1, 1), (1, 2), (1, 3), ..., (2, 1), (2, 2), (2, 3), ..., (6, 6)} donde en cada pareja el primer número representa lo que se obtiene en la primera tirada y el segundo en la segunda. Si *A* = "la suma de las dos tiradas es 7" y *B* = "el primer número es par". ¿Cuál es la probabilidad de *AUB?*

Respuesta: 21/36

8. Se lanza una moneda dos veces. Si *A* = "obtener lo mismo en las dos tiradas", *B* = "la primera vez sale águila" y *C* = "obtener al menos un sello", describe los eventos AB, AUB, BC, BUC, CUA, CA.

Respuesta: AB={aa}, AUB={aa,as,ss}, BC={as}, BUC={aa,as,sa,ss}, CUA={ss,sa,aa,as}, CA={ss}

9. Se lanza 100 veces un dado y se obtiene:

| Cara | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Frecuencia absoluta | 12 | 17 | 18 | 16 | 18 | 19 |

Calcula la frecuencia relativa de los siguientes sucesos:

a) *A* = par.

b) *B* = No par.

Respuesta:

a) *fr* (*A*) = 0.52
b) *fr* (*B*) = 0.48

10. En el lanzamiento de un dado, consideramos los sucesos *A* = {2, 3} y *B* = {2, 4, 6}. Calcular la probabilidad del suceso AUB (unión de *A* y *B)*.

Respuesta: 4/6=2/3

11. Calcular la probabilidad de aprobar un examen de matemáticas si la probabilidad de no aprobar es 0.4.

Respuesta: 0.6.

12. Se lanza dos veces un dado. Representamos el espacio muestral de la siguiente forma: {(1, 1), (1, 2), (1, 3), ..., (2, 1), (2, 2), (2, 3), ..., (6, 6)} donde en cada pareja el primer número representa lo que se obtiene en la primera tirada y el segundo en la segunda. Sean los sucesos: *A* = "obtener primero un 4 y después un 3" = (4, 3), *B* = "la suma de las dos tiradas es 7", *C* = "el primer número es par" y *D* = "obtener el mismo número en las dos tiradas". Describir los siguientes eventos: AUB,BC,AUD,CD,BD.

Respuesta

$A \cup B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$

$B \cap C = \{(2,5), (4,3), (6,1)\}$

$A \cup D = \{(4, 3), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$

$C \cap D = \{(2, 2), (4, 4), (6, 6)\}$

$B \cap D = \varnothing =$ evento imposible.

13. Una urna contiene 3 bolas blancas (*B*), 2 rojas (*R*) y 1 amarilla (*A*). Se extrae una bola al azar. Indica cuáles son los eventos elementales, el evento seguro y el evento imposible.

Respuesta:

**Eventos elementales**: *{B}, {R},{A}*.

**Evento seguro**: {A,B,R}.

**Evento imposible**: extraer una bola azul.

14. En el lanzamiento de un dado, consideramos los eventos *A* = {2, 3} y *B* = {2, 4, 6}. Describir el evento AUB, unión de *A* y *B,* y el evento AB, intersección de *A* y *B*.

Respuesta: AUB={2,3,4,6}, AB={2}.

15. ¿Cuál es la probabilidad de obtener el mismo número de puntos en cada dado al lanzar dos dados?

Respuesta: 6/36=1/2.

# MODELOS DE DISTRIBUCIÓN DE PROBABILIDAD

## Variables aleatorias discretas

### <u>Modelo de Bernouilli</u>

- **Corresponde a experimentos como el lanzamiento de una moneda. Sirve de modelo para muchas situaciones en las que <u>sólo puede haber dos posibles resultados complementarios</u> (*A* y *no A*): uno de ellos con probabilidad *p* y el otro con probabilidad (1-*p*).**

**<u>Ejemplos</u>:**

- **Inspeccionar un objeto para ver si es o no es defectuosos.**
- **Preguntar a una persona si tiene o no tiene trabajo**

- **Comprobar si una empresa está o no está en quiebra**
- **Ver si un alumno apruebe o no aprueba un examen**

- **Normalmente se denomina éxito (*x*=1) al suceso con probabilidad *p* y fracaso (*x*=0) al suceso con probabilidad 1-*p*. Por tanto, diremos que una variable aleatoria *x* tiene una distribución de Bernouilli si:**

$$P\{exito\} = P\{x = 1\} = p$$

$$P\{fracaso\} = P\{x = 0\} = 1 - p$$

- **Si *x* es una variable aleatoria con distribución de Bernouilli su media será:**

$$m_x = \sum_{i=1}^{k} x_i p_i = 1 \times p + 0 \times (1 - p) = p$$

**y su desviación típica:**

$$\sigma_x = \sqrt{\sum_{i=1}^{k} x_i^2 p_i - m_x^2} = \sqrt{1^2 \times p + 0^2 \times (1 - p) - p^2} = \sqrt{p - p^2} =$$

$$= \sqrt{p(1 - p)}$$

**<u>Ejemplo</u>:**

**Sabemos que una máquina produce un 3% de piezas defectuosas.**

**La variable es *x*=1 si la pieza no es defectuosa y *x*=0 si la pieza es defectuosa, es decir:**

$$P\{x = 0\} = 1 - p = 0{,}03$$

$$P\{x = 1\} = p = 0{,}97$$

**La variable *x* sigue una distribución de Bernouilli con *p*=0,97, luego:**

$$m_x = p = 0{,}97$$

$$\sigma_x = \sqrt{p(1-p)} = \sqrt{0{,}97 \times 0{,}03} = 0{,}1706$$

## La distribución Binomial

- **Se repite *n* veces de forma independiente un experimento de Bernouilli con probabilidad de éxito igual a *p*. La variable aleatoria *x* que expresa el número de "éxitos" obtenidos en este proceso sigue una distribución binomial con parámetros *n* y *p*: *B(n,p).***

  **Ejemplo:**

**En un país en el que está en paro el 25% de la población activa, se realiza una encuesta sobre distintos temas a 12 personas.**

**La variable aleatoria *x* que expresa el nº de encuestados que están en paro sigue una binomial con parámetros *n*=12 y *p*=0,25, es decir, una *B*(12,0,25).**

- **La distribución de Bernouilli es *B(1,p)*, un caso particular de la binomial en que el experimento se realiza una sóla vez.**

- **Para conocer la distribución de una variable binomial *x* tendremos que especificar la probabilidad de que tome cualquier valor *k* entre 0 y *n*. La Tabla 1 del Apéndice B de Peña y Romo, presenta la probabilidad de *k* éxitos en una *B*(*n*,*p*), para distintos valores de *n* y de *p.***

**<u>Ejemplo</u>: (continuación del *B*(12,025))**

**La probabilidad de que al realizar la encuesta se pregunte a 4 personas en paro (es decir, P($x$=4)) se puede encontrar en la Tabla 1 (con $n$=12, $k$=4 y $p$=0,25) y es igual a 0,1936.**

- **Si $x$ es una variable *B*($n$, $p$) su media, varianza y desviación típica serán:**

$$m_x = np$$

$$\sigma_x^2 = np(1-p)$$

$$\sigma_x = \sqrt{np(1-p)}$$

**Para un valor de $n$, la dispersión es máxima cuando $p$=0,5.**

**Ejemplo: (continuación del *B*(12,025))**

$$m_x = np = 12 \times 0{,}25 = 3$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{12 \times 0{,}25 \times (1-0{,}25)} = 1{,}5$$

- **La forma de la distribución depende del valor de *p*: si es menor que 0,5 es asimétrica a la derecha, si es mayor que 0,5 asimétrica a la izquierda y si es igual a 0,5 es simétrica.**

**EJERCICIOS**

**Ejercicio 16.2 (Peña y Romo)**

**Se sabe que el 40% de los habitantes de una ciudad consumen**

**café diariamente:**

a) **Se pregunta a una persona si toma café a diario. La variable aleatoria $x_1$ vale 1 si la respuesta es afirmativa y 0 en caso contrario. Hallar la media y la desviación típica de $x_1$.**

**La distribución de $x_1$ es una Bernoulli con *p*=0,4:**

**Si toma café:** $x_1 = 1$ **con probabilidad *p*=0,4**

**Si no toma café:** $x_1 = 0$ **con probabilidad 1-*p*=0,6**

**Luego:**

$$m_{x_1} = p = 0{,}4$$

$$\sigma_{x_1} = \sqrt{p(1-p)} = \sqrt{0{,}4 \times 0{,}6} = 0{,}4899$$

**b)Se encuesta a 20 personas sobre su consumo diario de café. Sea *x* el nº de personas encuestadas que consume café a diario. Calcular la probabilidad de que *x* sea igual a 12. Hallar la media y desviación típica de *x*. Obtener la probabilidad de que nadie tome café a diario y de que lo hagan al menos tres personas.**

**La variable aleatoria *x* sigue una distribución *B*(20,0,4) (*n*=20 y *p*=0,4).**

**Mirando en la Tabla 1 para *k*=12, *n*=20 y *p*=0,4 tenemos que:**

$$P(x = 12) = 0{,}0355$$

**La media y desviación típica de *x* son:**

$$m_x = np = 20 \times 0{,}4 = 8$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{20 \times 0{,}4 \times (1-0{,}4)} = 2{,}1908$$

**La probabilidad de que ninguna persona tome café diariamente, es decir,** $P(x = 0)$ **podemos encontrarla en la Tabla 1 para *k*=0, *n*=20 y *p*=0,4:**

$$P(x = 0) = 0$$

**La probabilidad de que al menos tres personas tomen café a diario será:**

$$P(x \geq 3) = 1 - P(x < 3) = 1 - \left[ P(x = 0) + P(x = 1) + P(x = 2) \right]$$

**Mirando en la Tabla 1 los valores para *k*=0, *k*=1 y *k*=2 (para *n*=20 y *p*=0,4) tenemos:**

$$P(x \geq 3) = 1 - (0 + 0{,}0005 + 0{,}0031) = 0{,}9964$$

<u>Ejercicio 16.3</u> **(Peña y Romo)**

**Un partido político consigue el 20% de los votos en unas elecciones. Se realiza una encuesta a 15 personas.**

**a) ¿Cuál es la probabilidad de que no haya entre ellas ningún votante del partido?**

**Sea *x* el nº de votantes del partido entre los encuestados, que sigue una distribución *B*(15,0,2) (*n*=15 y *p*=0,2), entonces nos piden** $P(x = 0)$**. Si miramos en la Tabla 1 para *k*=0, *n*=15 y *p*=0,2, tendremos:**

$$P(x = 0) = 0{,}0352$$

**b) Hallar la probabilidad de que no haya más de 3 votantes de ese partido.**

$$P(x \leq 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$$

**Si miramos los valores en la Tabla 1 para *k*=0,1,2 y 3 (para *n*=15 y *p*=0,2) obtenemos:**

$$P(x \leq 3) = 0{,}0352 + 0{,}1319 + 0{,}2309 + 0{,}2501 = 0{,}6481$$

**c) Obtener la probabilidad de que al menos tres personas voten a dicho partido**

$$P(x \geq 3) = 1 - P(x < 3) = 1 - \left[P(x = 0) + P(x = 1) + P(x = 2)\right] =$$
$$= 1 - (0{,}0352 + 0{,}1319 + 0{,}2309) = 0{,}602$$

**d) Calcular la media y la desviación típica del nº de votantes entre los 15 encuestados.**

$$m_x = np = 15 \times 0{,}2 = 3$$

$$\sigma_x = \sqrt{np(1 - p)} = \sqrt{15 \times 0{,}2 \times (1 - 0{,}2)} = 1{,}549$$

**<u>Ejercicio 16.5</u> (Peña y Romo)**

**Un examen consta de 15 preguntas cada una de las cuales tiene 4 posibles respuestas. Una persona sin**

**conocimientos del tema responde las preguntas al azar.**

**a) ¿Cuál es la probabilidad de que acierte la respuesta si contesta sólo una pregunta?**

$P(acertar) = p = 1/4 = 0{,}25$

**b) Hallar la probabilidad de que dicha persona no conteste bien a ninguna de las 15 preguntas. Calcular la probabilidad de que acierte alguna.**

**Sea la variable aleatoria *x* el nº de aciertos en las 15 preguntas que sigue una distribución *B*(15,0,25) (*n*=15 y *p*=0,25).**

**Nos piden la probabilidad de que no acierte ninguna, es decir,** $P(x = 0)$ **que, mirando en la Tabla 1 para *k*=0, *n*=15 y *p*=0,25, será:**

$P(x = 0) = 0{,}0134$

**La probabilidad de que acierte alguna será:**

$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x = 0) = 1 - 0{,}0134 = 0{,}9866$

**c) Obtener la probabilidad de que responda bien a todas las preguntas.**

**Mirando en la Tabla 1 para *k*=15, *n*=15 y *p*=0,25, tendremos que:**

$$P(x = 15) = 0$$

**d) ¿Cuál es la probabilidad de que conteste acertadamente a más de la mitad de las cuestiones?**

**Mirando en la Tabla 1 para *k*=8,9,..,15, (para *n*=15 y *p*=0,25), tendremos que:**

$$P(x \geq 8) = P(x = 8) + P(x = 9) + \ldots + P(x = 15) =$$
$$= 0{,}0131 + 0{,}0034 + 0{,}0007 + 0{,}0001 = 0{,}0173$$

**e) ¿Cuál es la probabilidad de que el nº de preguntas acertado sea distinto de tres?**

$$P(x \neq 3) = 1 - P(x = 3)$$

**Mirando en la Tabla 1 para *k*=3 (para *n*=15 y *p*=0,25), tendremos que:**

$$P(x \neq 3) = 1 - 0{,}2252 = 0{,}7748$$

**<u>Ejercicio 16.9</u> (Peña y Romo)**

**El 25% de las personas con tarjeta de crédito liquidan sus pagos cada mes. Se pregunta a 15 personas con tarjeta.**

**a) Hallar el nº esperado entre ellos que liquidan sus deudas cada mes. ¿Cuál es la desviación típica de esta variable?**

**Sea *x* el nº de personas (de entre los 15) que liquidan sus deudas cada mes que se distribuye como una *B*(15,0,25) (*n*=15 y *p*=0,25). Por tanto:**

$$m_x = np = 15 \times 0{,}25 = 3{,}75 \approx 4$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{15 \times 0{,}25 \times (1 - 0{,}25)} = 1{,}677$$

**b) Obtener la probabilidad de que ninguno pague todas sus deudas mensualmente.**

**Nos piden** $P(x=0)$ **que, mirando en la Tabla 1 para *k*=0, *n*=15 y *p*=0,25, será:**

$$P(x=0)=0{,}0134$$

**c) Calcular la probabilidad de que el nº de personas encuestadas que liquida mensualmente las deudas supere en más de una desviación típica al valor esperado.**

$$m_x+\sigma_x=3{,}75+1{,}677=5{,}427$$

$$\begin{aligned}P(x\geq 6) &= P(x=6)+P(x=7)+\ldots+P(x=15)=\\ &=0{,}0917+0{,}0393+0{,}0131+0{,}0034+0{,}0007+0{,}0001=\\ &=0{,}1483\end{aligned}$$

## La distribución geométrica

- **Supongamos que un experimento Bernouilli se repite hasta que aparece el primer éxito. Se llama <u>distribución geométrica</u> a la de una variable *x* que expresa el instante en que ocurre el primer "éxito". Por ejemplo, el nº de personas a las que hay que entrevistar hasta encontrar una que esté en paro.**

- **Esta distribución depende sólo de la probabilidad *p* de obtener "éxito" en cada ensayo y se representa como *G(p)*. La probabilidad de que el primer éxito surja en el instante *k* es:**

$$P(x = k) = p(1 - p)^{k-1} \qquad k = 1,2,\dots$$

**<u>Ejemplo</u>:**

**Si la probabilidad de que una persona esté en paro es *p*=0,25, la probabilidad de que haya que realizar**

**6 entrevistas hasta encontrar una persona en paro es:**

$$P(x=6)=0,25(1-0,25)^{6-1}=0,05$$

- **Si *x* es una variable *G*(*p*) su media y desviación típica serán:**

$$m_x=\frac{1}{p}$$

$$\sigma_x=\sqrt{\frac{1-p}{p^2}}$$

**<u>Ejemplo</u>: (continuación)**

**El número medio de entrevistas que habrá que realizar para encontrar una persona en paro será:**

$$m_x=\frac{1}{p}=\frac{1}{0,25}=4$$

**y la desviación típica es:**

$$\sigma_x=\sqrt{\frac{1-p}{p^2}}=\sqrt{\frac{1-0,25}{0,25^2}}=3,46$$

**<u>Ejercicio 16.4</u> (Peña y Romo)**

**En base al Ejercicio 16.3 (un partido político obtiene el 20% de los votos), se pide:**

**a) Obtener la probabilidad de que el primer votante del partido al que se pregunta sea la tercera persona entrevistada.**

**Sea *x* la variable aleatoria que representa el número de entrevistas que habrá que realizar hasta encontrar al primer votante del partido que sigue una distribución *G(*0,20*)*. Entonces:**

$$P(x = 3) = p(1 - p)^{k-1} = 0{,}2(1 - 0{,}2)^{3-1} = 0{,}128$$

**b) ¿Cuál es el número medio de personas que hay que entrevistar hasta llegar al primer votante del partido?**

$m_x = \frac{1}{p} = \frac{1}{0{,}2} = 5$ **personas**

c) **Hallar la probabilidad de que sean necesarias más de 6 entrevistas para llegar al primer votante del partido.**

$$P(x > 6) = 1 - P(x \leq 6) = 1 - [P(x = 1) + P(x = 2) + \ldots + P(x = 6)] =$$
$$= 1 - [0{,}2 \times 0{,}8^0 + 0{,}2 \times 0{,}8^1 + \ldots + 0{,}2 \times 0{,}8^5] = 0{,}2621$$

**<u>Ejercicio 16.6</u> (Peña y Romo)**

**En base al Ejercicio 16.5 (un examen en el que cada pregunta tiene 4 posibles respuestas), y suponiendo que el examinado contesta tantas preguntas como sea necesario, se pide:**

a) **Calcular la probabilidad de que la primera pregunta acertada sea la quinta.**

**Sea *x* la variable aleatoria que representa el número de preguntas que habrá que contestar hasta acertar la primera respuesta que sigue una distribución *G(*0,25*)*. Entonces:**

$$P(x = 5) = p(1 - p)^{k-1} = 0{,}25(1 - 0{,}25)^{5-1} = 0{,}0791$$

**b) ¿Cuál es el número esperado de preguntas que debe responder hasta contestar una correctamente?**

$$m_x = \frac{1}{p} = \frac{1}{0{,}25} = 4$$ **preguntas**

## La distribución de Poisson

- **Una variable aleatoria *x* con distribución de Poisson expresa el nº de "sucesos raros" que ocurren en una proporción fija de espacio o de tiempo. Por ejemplo, el nº de accidentes en un día, el nº de llamadas a un teléfono en una hora, nº de erratas por página en un libro…**

- **La intensidad con que aparecen dichos sucesos se representa mediante el parámetro positivo *λ*.**

- **La probabilidad de que *x* tome el valor *k* en una distribución de Poisson con parámetro de intensidad *λ* puede verse en la Tabla 2 del Apéndice B (de Peña y Romo).**

**<u>Ejemplo 16.3 de Peña y Romo</u>:**

**La variable aleatoria *x*, con distribución de Poisson con parámetro $\lambda$=0,8, representa el número de accidentes diarios en una ciudad. La probabilidad de que hoy ocurran exactamente 3 accidentes será (ver Tabla 2 con *k*=3 y $\lambda$=0,8):**

$$P(x = 3) = 0{,}0383$$

- **Si *x* es una variable con distribución Poisson su media y varianza serán:**

$$m_x = \sigma_x^2 = \lambda$$

**y su desviación típica**

$$\sigma_x = \sqrt{\lambda}$$

**<u>Ejemplo 16.4 de Peña y Romo</u>:**

**El número medio de errores que comete una persona al mecanografiar una página es 2. Si suponemos que la distribución de *x* (errores al mecanografiar una página) es de**

**Poisson entonces, la probabilidad de que en una página no haya ningún error será (ver Tabla 2 con *k*=0 y $\lambda$=2):**

$$P(x = 0) = 0{,}1353$$

## Ejercicio 16.7 (Peña y Romo)

**El número de clientes *x* que llegan a la caja de un supermercado en un cuarto de hora sigue una distribución de Poisson con media 5.**

a) **Hallar la probabilidad de que lleguen al menos 4 personas en un cuarto de hora.**

**La variable *x* sigue una distribución de Poisson con** $m_x = \lambda = 5$**. Se pide (mirar las probabilidades en la Tabla 2 para** $\lambda = 5$ **y *k*=0,1,2 y 3):**

$$P(x \geq 4) = 1 - P(x < 4) = 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)] =$$
$$= 1 - (0{,}0067 + 0{,}0337 + 0{,}0842 + 0{,}104) = 0{,}735$$

**b) Obtener la probabilidad de que no llegue nadie a la caja en un cuarto de hora.**

$$P(x = 0) = 0{,}0067$$

**c) ¿Cuál es la probabilidad de que lleguen entre 2 y 4 clientes a la caja en un cuarto de hora?**

$$P(2 \leq x \leq 4) = P(x = 2) + P(x = 3) + P(x = 4) =$$
$$= 0{,}0842 + 0{,}1404 + 0{,}1755 = 0{,}4001$$

## Variables aleatorias continuas

### La distribución normal o gaussiana

- **Se utiliza como modelo para variables como el peso, la altura, la calificación en un examen..., es decir, en variables cuya <u>distribución es simétrica</u> con <u>respecto a un valor central</u> (alrededor del cual toma valores con gran probabilidad) y apenas aparecen valores extremos.**

- **Si una variable aleatoria $x$ tiene distribución normal suele representarse como $N(m,\sigma)$ donde $m=m_x$ es la media o valor esperado de la variable y $\sigma= \sigma_x$ es la desviación típica de la variable, que son los dos parámetros que caracterizan la distribución normal.**

**Ver en Figuras 18.2 y 18.3 la densidad de una variable normal.**

- **En la distribución normal, la mayoría de la probabilidad se concentra en la zona central.**

**Ver Figura 18.5**

**<u>Ejercicio 18.8</u> (Peña y Romo)**

**El peso de los atletas de pruebas de medio fondo sigue una distribución normal con media 64,3 kilos y desviación típica 2,3 kilos. Hallar un intervalo centrado alrededor de la media que contenga:**

**a) El 68,3% de la población.**

**En la distribución normal se concentra el 68,3% de la probabilidad en el intervalo de valores comprendido entre** $m_x + \sigma_x$ **y** $m_x - \sigma_x$**. En este caso como** $m_x = 64{,}3 \quad \sigma_x = 2{,}3$ **será:**

$$m_x + \sigma_x = 64{,}3 + 2{,}3 = 66{,}6$$

$$m_x - \sigma_x = 64{,}3 - 2{,}3 = 62$$

**b) El 95,5% de la población.**

**En la distribución normal se concentra el 95,5% de la probabilidad en el intervalo de valores comprendido entre** $m_x + 2\sigma_x$ **y** $m_x - 2\sigma_x$**. Luego será:**

$$m_x + 2\sigma_x = 64{,}3 + 2 \times 2{,}3 = 68{,}9$$

$$m_x - 2\sigma_x = 64{,}3 - 2 \times 2{,}3 = 59{,}7$$

**c) El 99,7% de la población**

**En la distribución normal se concentra el 99,7% de la probabilidad en el intervalo de valores comprendido entre** $m_x + 3\sigma_x$ **y** $m_x - 3\sigma_x$**. Luego será:**

$$m_x + 3\sigma_x = 64,3 + 3 \times 2,3 = 71,2$$

$$m_x - 3\sigma_x = 64,3 - 3 \times 2,3 = 57,4$$

- **Las <u>transformaciones lineales</u> de una variable que sigue una distribución normal, también tendrán una distribución normal. Es decir:**

  ***x*** **es** ***N***$(m_x, \sigma_x)$ **entonces** ***y=ax+b*** **será** ***N***$(am_x + b, |a|\sigma_x)$

  **<u>Ejemplo</u>:**

  **La variable** ***x*** **que expresa el tiempo en horas que tarda un empleado en hacer una tarea sigue una distribución normal con** $m_x = 1,5$ **y** $\sigma_x = 0,1$**.**

**¿Cómo será la distribución de la variable *y* que expresa lo mismo pero en minutos?**

$y = 60x$ **luego seguirá una distribución normal con:**

$m_y = 60m_x = 60 \times 1{,}5 = 90$

$\sigma_y = 60\sigma_x = 60 \times 0{,}1 = 6$

- **Si tipificamos una variable *x* con una distribución *N*$(m_x, \sigma_x)$ la nueva variable $z = \dfrac{x - m_x}{\sigma_x}$ tiene una distribución *N*(0,1) o <u>normal estándar</u>.**

- **La <u>Tabla 3 del Apéndice B</u> (Peña y Romo) nos proporciona la probabilidad de que una variable *z* con una distribución *N*(0,1) tome un valor menor que una cierta cantidad, es decir $P(z<b)$. A partir de ellas también podremos obtener:**

$P(z>a)=1-P(z<a)$

$P(a<z<b)=P(z<b)-P(z<a)$

**Ver Figura 18.7**

**<u>Ejemplo</u>: Sea *z* una variable normal estándar**

**-** $P(z<1{,}35)=0{,}9115$ **(mirar en la fila de 1,3 y en la columna de 0,05)**

**-** $P(z>0{,}86)=1-P(z<0{,}86)=1-0{,}8051=0{,}1949$

**(mirar en la fila de 0,8 y columna de 0,06)**

**-** $P(-0{,}3<z<0{,}83)=P(z<0{,}83)-P(z<-0{,}3)=$

$$=0{,}7967-0{,}3821=0{,}4105$$

**(mirar en la fila de 0,8 y columna de 0,03 y en la fila de –0,3 y columna de 0,00)**

## **<u>Ejercicio 18.2</u> (Peña y Romo)**

**Si *z* es una variable normal estándar hallar:**

**a)** $P(z < 2{,}23) = 0{,}9871$

**b)** $P(z < 3{,}48) = 0{,}9997$

**c)** $P(z < -1{,}76) = 0{,}0392$

**d)** $P(z > 2{,}45) = 1 - P(z < 2{,}45) = 1 - 0{,}9929 = 0{,}0071$

**e)** $P(z > 3{,}23) = 1 - P(z < 3{,}23) = 1 - 0{,}9994 = 0{,}0006$

**f)** $P(z > -3{,}07) = 1 - P(z < -3{,}07) = 1 - 0{,}0011 = 0{,}9989$

**g)** $P(1{,}13 < z < 2{,}69) = P(z < 2{,}69) - P(z < 1{,}13) =$

$$= 0{,}9964 - 0{,}8708 = 0{,}1256$$

**h)** $P(-0{,}86 < z < 1{,}28) = P(z < 1{,}28) - P(z < -0{,}86) =$

$$= 0{,}8997 - 0{,}1949 = 0{,}7048$$

**i)** $P(-2{,}98 < z < -1{,}32) = P(z < -1{,}32) - P(z < -2{,}98) =$

$$= 0{,}0934 - 0{,}0014 = 0{,}092$$

**<u>Ejercicio 18.6</u> (Peña y Romo)**

**Hallar el valor *a* de la variable *z* normal estándar tal que:**

**a)** $P(z < a) = 0{,}2033$

$a = -0{,}83$

**b)** $P(z < a) = 0{,}7734$

$a = 0{,}75$

**c)** $P(z > a) = 0{,}922$

**Sabemos que** $P(z > a) = 1 - P(z < a)$ **luego:**

$P(z < a) = 1 - 0{,}922 = 0{,}0778$

$a = -1{,}42$

**d)** $P(z > a) = 0{,}0314$

**Sabemos que** $P(z > a) = 1 - P(z < a)$ **luego:**

$P(z < a) = 1 - 0{,}0314 = 0{,}9686$

$a = 1{,}86$

- **Como al tipificar cualquier variable *x* con una distribución *N*$(m_x, \sigma_x)$, obtenemos una variable *z* con una distribución normal estándar, podremos calcular probabilidades para cualquier variable normal usando la Tabla 3.**

- **Si *x* tiene una distribución *N*$(m_x, \sigma_x)$, podremos calcular $P(x < b)$ de la siguiente manera:**

$$P(x < b) = P\left(\frac{x - m_x}{\sigma_x} < \frac{b - m_x}{\sigma_x}\right) = P\left(z < \frac{b - m_x}{\sigma_x}\right)$$

**y calcular esta probabilidad con la Tabla 3.**

**<u>Ejemplo</u>:**

**Si *x* tiene una distribución *N*$(3{,}2, 1{,}22)$, entonces:**

- $P(x < 3{,}45) = P\left(\frac{x - 3{,}2}{1{,}22} < \frac{3{,}45 - 3{,}2}{1{,}22}\right) = P(z < 0{,}20) = 0{,}5793$

**El valor de $P(z < 0{,}20)$ se busca en la Tabla 3**

- $P(2{,}92 < x < 3{,}43) = P\left(\frac{2{,}92 - 3{,}2}{1{,}22} < \frac{x - 3{,}2}{1{,}22} < \frac{3{,}43 - 3{,}2}{1{,}22}\right) =$

$$= P\left(-0,22 < z < 0,19\right) =$$

$$= P(z < 0,19) - P(z < -0,22) =$$

$$= 0,5753 - 0,4129 = 0,1624$$

- **Si *x* tiene una distribución *N*$(m_x, \sigma_x)$ e *y* una distribución *N*$(m_y, \sigma_y)$ y son independientes, entonces *x+y* tendrá una distribución *N*$(m_x + m_y, \sqrt{\sigma_x^2 + \sigma_y^2})$**

**<u>Ejercicio 18.4</u> (Peña y Romo)**

**Si *x* es una variable *N*$(8,47;1,15)$, hallar:**

**a)** $P(x < 9,12)$

$$P(x < 9,12) = P\left(\frac{x - 8,47}{1,15} < \frac{9,12 - 8,47}{1,15}\right) = P\left(z < 0,56\right) = 0,7123$$

**b)** $P(x < 12,34)$

$$P(x < 12,34) = P\left(\frac{x - 8,47}{1,15} < \frac{12,34 - 8,47}{1,15}\right) = P\left(z < 3,36\right) = 0,9996$$

**c)** $P(x < 6,42)$

$$P(x < 6,42) = P\left(z < \frac{6,42 - 8,47}{1,15}\right) = P(z < -1,78) = 0,0375$$

**d)** $P(x > 10,53)$

$$P(x > 10,53) = P\left(z > \frac{10,53 - 8,47}{1,15}\right) = P(z > 1,79) =$$

$$= 1 - P(z < 1,79) = 1 - 0,9633 = 0,0367$$

**e)** $P(x > 12,62)$

$$P(x > 12,62) = P\left(z > \frac{12,62 - 8,47}{1,15}\right) = P(z > 3,61) =$$

$$= 1 - P(z < 3,61) \approx 1 - 1 \approx 0$$

**f)** $P(x > 4,01)$

$$P(x > 4,01) = P\left(z > \frac{4,01 - 8,47}{1,15}\right) = P(z > -3,88) =$$

$$= 1 - P(z < -3,88) \approx 1 - 0 \approx 1$$

**g)** $P(6,12 < x < 11,92)$

$$P(6,12 < x < 11,92) = P\left(\frac{6,12 - 8,47}{1,15} < z < \frac{11,92 - 8,47}{1,15}\right) =$$

$$= P(-2,04 < z < 3) =$$

$$= P(z < 3) - P(z < -2{,}04) =$$

$$= 0{,}9987 - 0{,}0207 = 0{,}978$$

**h)** $P(7{,}52 < x < 10{,}32)$

$$P(7{,}52 < x < 10{,}32) = P\left(\frac{7{,}52 - 8{,}47}{1{,}15} < z < \frac{10{,}32 - 8{,}47}{1{,}15}\right) =$$

$$= P\left(-0{,}83 < z < 1{,}61\right) =$$

$$= P(z < 1{,}61) - P(z < -0{,}83) =$$

$$= 0{,}9463 - 0{,}2033 = 0{,}743$$

**i)** $P(5{,}06 < x < 6{,}84)$

$$P(5{,}06 < x < 6{,}84) = P\left(\frac{5{,}06 - 8{,}47}{1{,}15} < z < \frac{6{,}84 - 8{,}47}{1{,}15}\right) =$$

$$= P\left(-2{,}96 < z < -1{,}42\right) =$$

$$= P(z < -1{,}42) - P(z < -2{,}96) =$$

$$= 0{,}0778 - 0{,}0015 = 0{,}0763$$

**<u>Ejercicio 18.7</u> (Peña y Romo)**

**Si *x* es una variable con distribución *N*(4,3;1,2) hallar el valor de *a* tal que:**

**c)** $P(x < a) = 0{,}7389$

$$P(x < a) = P\left(\frac{x - 4{,}3}{1{,}2} < \frac{a - 4{,}3}{1{,}2}\right) = P\left(z < \frac{a - 4{,}3}{1{,}2}\right) = 0{,}7389$$

**luego:**

$$\frac{a - 4{,}3}{1{,}2} = 0{,}64 \rightarrow a = 0{,}64 \times 1{,}2 + 4{,}3 = 5{,}068$$

**d)** $P(x < a) = 0{,}6179$

$$P(x < a) = P\left(\frac{x - 4{,}3}{1{,}2} < \frac{a - 4{,}3}{1{,}2}\right) = P\left(z < \frac{a - 4{,}3}{1{,}2}\right) = 0{,}6179$$

**luego:**

$$\frac{a - 4{,}3}{1{,}2} = 0{,}3 \rightarrow a = 0{,}3 \times 1{,}2 + 4{,}3 = 4{,}66$$

**a)** $P(x > a) = 0{,}2981$

$$P(x > a) = P\left(z > \frac{a - 4{,}3}{1{,}2}\right) = 1 - P\left(z < \frac{a - 4{,}3}{1{,}2}\right) = 0{,}2981$$

**luego:**

$$P\left(z < \frac{a-4{,}3}{1{,}2}\right) = 1 - 0{,}2981 = 0{,}7019$$

$$\frac{a-4{,}3}{1{,}2} = 0{,}53 \rightarrow a = 0{,}53 \times 1{,}2 + 4{,}3 = 4{,}936$$

**b)** $P(x > a) = 0{,}5871$

$$P(x > a) = P\left(z > \frac{a-4{,}3}{1{,}2}\right) = 1 - P\left(z < \frac{a-4{,}3}{1{,}2}\right) = 0{,}5871$$

**luego:**

$$P\left(z < \frac{a-4{,}3}{1{,}2}\right) = 1 - 0{,}5871 = 0{,}4129$$

$$\frac{a-4{,}3}{1{,}2} = -0{,}22 \rightarrow a = -0{,}22 \times 1{,}2 + 4{,}3 = 4{,}036$$

**<u>Ejercicio 18.3</u> (Peña y Romo)**

**La variable *x* que expresa la altura en metros de los jugadores de baloncesto tiene una distribución normal con media $m_x$=1,89 y desviación típica $\sigma_x$=0,07. Si la variable *y* es la altura en centímetros:**

**a) Escribir la relación entre *x* e *y*.**

$y = 100x$

**b) Obtener la distribución de *y*.**

**Como *y* es una transformación lineal de una variable con distribución normal tendrá una distribución *N*($m_y, \sigma_y$) con:**

$$m_y = 100m_x = 100 \times 1{,}89 = 189$$

$$\sigma_y = 100\sigma_x = 100 \times 0{,}07 = 7$$

**c) Hallar la probabilidad de que un jugador de baloncesto elegido al azar mida más de 180 centímetros.**

$$P(y > 180) = P\left(z > \frac{180 - 189}{7}\right) = 1 - P(z < -1{,}28) = 1 - 0{,}1003 = 0{,}8997$$

**Aproximación mediante la normal**

- **Si *x* tiene una distribución *B*$(n, p)$ entonces la variable tipificada $\frac{x-np}{\sqrt{np(1-p)}}$ tiene una distribución próxima a la normal estándar. La aproximación será buena si *n*>30 y tanto *np* como *n(1-p)* son mayores o iguales que 5.**

- **Podremos entonces calcular probabilidades de *x* aproximándolas mediante la normal estándar (usando los valores de la Tabla 3).**

- **Si *x* es una variable discreta (la binomial) e *y* una variable continua (la normal) si queremos aproximar probabilidades de *x* mediante *y* funcionará mejor si hacemos la corrección por continuidad:**

$$P(a < x < b) \rightarrow P(a - 0{,}5 < y < b + 0{,}5)$$

$$P(x < b) \rightarrow P(y < b + 0{,}5)$$

$$P(x > a) \rightarrow P(y > a - 0{,}5)$$

**Ejemplo:**

**El 35% de los habitantes de una ciudad votan a un partido. Se hace una encuesta a 200 personas. La variable *x* que representa el nº de personas encuestadas que vota al partido sigue una distribución *B*(*n*=200;*p*=0,35).**

**Por tanto:**

$$m_x = np = 200 \times 0{,}35 = 70$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{200 \times 0{,}35 \times (1 - 0{,}35)} = 6{,}74$$

**Queremos saber la probabilidad de que haya entre 82 y 106 votantes del partido:** $P(82 < x < 106)$**. Podemos aproximarla por la normal estándar (*n*=200>30;**

**np=70>5 y n(1-p)=130>5) para lo que haremos la corrección por continuidad y tipificaremos la variable:**

$$P(82 < x < 106) \approx P\left(\frac{82 - 0{,}5 - 70}{6{,}74} < z < \frac{106 + 0{,}5 - 70}{6{,}74}\right) =$$

$$= P(1{,}70 < z < 5{,}41) = P(z < 5{,}41) - P(z < 1{,}70) =$$

$$= 1 - 0{,}9554 = 0{,}0446$$

**<u>Ejercicio 18.9</u> (Peña y Romo)**

**El 40% de los relojes que se venden en una tienda son digitales y el resto analógicos. Se consideran las ventas de los 100 próximos relojes.**

**a) Hallar el nº esperado de relojes digitales que se venderán entre los 100. ¿Cuál es la desviación típica?**

**La variable aleatoria *x* representa el nº de relojes digitales vendidos entre los 100 próximos. Sigue una distribución *B*(100,0,4).**

$$m_x = np = 100 \times 0,4 = 40$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{100 \times 0,4 \times (1-0,4)} = 4,9$$

**b) Obtener la probabilidad de que se vendan entre 30 y 50 digitales.**

$$P(30 < x < 50) \approx P\left(\frac{30-0,5-40}{4,9} < z < \frac{50+0,5-40}{4,9}\right) =$$

$$= P(-2,14 < z < 1,94) = P(z < 1,94) - P(z < -2,14) =$$

$$= 0,9738 - 0,016 = 0,9578$$

**c) Calcular la probabilidad de que se vendan al menos 15 relojes digitales.**

$$P(x > 15) \approx P\left(z > \frac{15-0,5-40}{4,9}\right) = P(z > -5,20) =$$

$$= 1 - P(z < -5,20) \approx 1 - 0 \approx 1$$

**d) Hallar la probabilidad de que no se vendan más de 60 relojes digitales.**

$$P(x < 60) \approx P\left(z < \frac{60 + 0{,}5 - 40}{4{,}9}\right) = P(z < 4{,}18) \approx 1$$

**<u>Ejercicio 18.10</u> (Peña y Romo)**

**Una empresa tiene 2000 trabajadoras de las que el 60% son mujeres. Se encuesta a 200 personas de la empresa tomadas al azar.**

**a) Hallar la probabilidad aproximada de que al menos 40 de los encuestados sean mujeres.**

**La variable aleatoria *x* representa el nº de mujeres entre los 200 entrevistados. Sigue una distribución *B*(200;0,6).**

$$m_x = np = 200 \times 0{,}6 = 120$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{200 \times 0{,}6 \times (1 - 0{,}6)} = 6{,}9$$

$$P(x > 40) \approx P\left(z > \frac{40 - 0{,}5 - 120}{6{,}9}\right) = P(z > -11{,}67) =$$

$$= 1 - P(z < -11{,}67) \approx 1 - 0 \approx 1$$

**b) Obtener la probabilidad de que el nº de encuestadas esté entre 80 y 150.**

$$P(80 < x < 150) \approx P\left(\frac{80 - 0{,}5 - 120}{6{,}9} < z < \frac{150 + 0{,}5 - 120}{6{,}9}\right) =$$

$$= P(-5{,}87 < z < 4{,}42) = P(z < 4{,}42) - P(z < -5{,}87) =$$

$$\approx 1 - 0 \approx 1$$

**c) Calcular la probabilidad de que no se pregunte a más de 150 mujeres.**

$$P(x < 150) \approx P\left(z < \frac{150 + 0{,}5 - 120}{6{,}9}\right) = P(z < 4{,}42) \approx 1$$

## Ejercicios de Modelos Probabilísticos.

**1.** En un examen de 15 preguntas de falso y verdadero a) ¿Cuál es la probabilidad de que un estudiante no preparado obtenga las 15 preguntas contestadas correctamente por casualidad? b) Si el contestar 10 o más preguntas constituyen el grado de aprobación del examen, cuál es la probabilidad de que un estudiante lo apruebe?. d) Cuál de que no apruebe. e) Encuentre en el número esperado de respuestas correctas.

**2.** Suponga que en la universidad el 20% de los estudiantes están en ultimo año. Se elige un grupo de 10 de estudiantes de todos los semestres, encontrar la probabilidad de que el grupo éste formado por a) 5 estudiantes de último año. b) Más de 7 estén en último año o menos de 4. c) A lo sumo 2 estudiantes estén en último año. d) Entre 2 y 6 inclusive.

**3.** Se sabe que el 30% de los pacientes que han recibido cierto medicamento sufren efectos secundarios, como fiebre alta. Hallar la probabilidad de que de 20 pacientes que hayan recibido esta medicina: a) 3 pacientes a lo sumo sufrieron tal efecto secundario. b) 4 pacientes por lo menos sufrieron la fiebre alta. c) Sufrieron el efecto menos de 10 o más de 15 pacientes. d) Cuál es el número esperado de personas que sufren el efecto, interprete. e) Encuentre la desviación e interprete.

4. El 20% de cierta población tiene sangre tipo A. Para un grupo de 20 personas, ¿cuál es la probabilidad de que: a) Se encuentren exactamente 5 personas con grupo sanguíneo A. b) Se encuentren 3 o más. c) Menos de 3. d) Menos de la mitad tengan tipo de sangre tipo A.

5. En el Municipio de Andes, el 20% tiene SISBEN. Si se eligen 10 personas, cuál es la probabilidad de que: a) ninguno de ellos lo tenga. b) todos lo tengan. c) El 50% de la muestra lo tenga. d) Calcule el valor esperado y la desviación. Interprételas.

6. Una EPS detecta que sólo 4 de cada 10 usuarios que llegan a solicitar información se afilian a ella. Cuál es la probabilidad de que de los siguientes 20 pacientes: a) la mitad o más se afilien. b) 5 o 6 lo hagan. c) ninguno lo haga. d) Encuentre el valor esperado de persona que no se afilian.

7. El 60% de los trabajadores de una IPS se oponen a incrementar la jornada laboral. Halle la probabilidad de que de 20 trabajadores, el número de los que se oponen sea: a) 12. b) menos de 4 c) menos de 8 pero más de 4. d) siquiera 3.

8. Si el número medio de accidentes graves por año en una fabrica grande es de 4, encuentre la probabilidad de que en el próximo año, a) Se tenga exactamente accidentes. b) ningún accidente. c) a lo sumo 6 accidentes.

9. De acuerdo con las estimaciones de una compañía de seguros, la probabilidad de que se registre un incendio en una casa en un año es de 0.5%. La compañía asegura 600 casas. Cuál es la probabilidad de que se registre: a) al menos un incendio?, b) ninguna se incendie. c) 3 incendios. d) Cuál es el número esperado de casas incendiadas y no incendiadas?.

10. La probabilidad de vender un seguro de vida a personas que solicitan información se estima que es de 0.0025. Sobre esta base, si 1000 personas solicitan información, ¿Cuál es la probabilidad de que: a) nadie compre el

seguro de vida. b) Por lo menos 2 lo compren. c) Más de 10 lo compren. d) Siquiera el 40% lo compren. e) Todos menos 5 lo compren.

11. El número medio de computadoras que vende un almacén por día es de 1.5. Calcule la probabilidad de que el almacén venda por lo menos 3 computadoras durante un período de a) 2 días b) 3 días c) Calcule la probabilidad de que venda entre 4 y 8 cada dos días.

12. Las llamadas promedio de emergencia registradas en el conmutador de un Centro de Salud es de 5 por día en una semana normal. a) En un lapso de 1 día, ¿cuántas llamadas de emergencia se espera recibir? b) ¿Cuál es la probabilidad de que no se registren llamadas en un lapso de dos días?. c) En un lapso de medio día, ¿cuál es la probabilidad de que se registren por lo menos 4 llamadas?

13. Se cree que de cada 100 colombianos que inician la educación primaria solo 5 terminan carrera universitaria. Si en una escuela inician 100 alumnos. Cuál es la probabilidad de que: a) Terminen carrera universitaria 5 o más de ellos. b) Más de las tres cuartas partes. c) Entre 3 y 6 inclusive.

14. Supóngase que en un período de varios años, el número promedio de muertes por cierta enfermedad no contagiosa es de 5. Si el número de muertes por esa enfermedad sigue la distribución de Poisson, ¿cuál es la probabilidad de que durante el año en curso: a) Exactamente 7 personas mueran por esa enfermedad? b) 6 o más personas mueran por esa enfermedad. c) No haya muertes por esa enfermedad.

15. Supóngase que las edades de inicio de cierta enfermedad tienen una distribución aproximadamente normal, con una media de 11.2 años y una desviación típica de 3 años. Un niño contrae la enfermedad. ¿Cuál es la probabilidad de que el niño tenga
    a. Edad entre 8 ½ y 14? b. Más de 10 años. ?
    c. Menos de 12? d. Cuál es la edad de tal manera que el 50% de los niños supera dicho valor?

    d. Supóngase que el tiempo medio de permanencia hospitalaria por enfermedad crónica para un tipo de pacientes es de 60 días, con una varianza de 225. Si es razonable suponer que se tiene una distribución normal para el tiempo de hospitalización, calcular la probabilidad de que un paciente elegido al azar de entre ese grupo, tenga una hospitalización:
        a. Mayor que 50 días. b. Entre 1 ó 2 meses.
        c. Más de 90 días.
        d. Encuentre los valores simétricamente distribuidos alrededor de la media de tal manera que el 80% de los días de hospitalización se encuentren ahí..

e. Si el nivel total de colesterol en cierta población tiene una distribución normal con media de 200 mg/100 ml y una varianza de 400, calcular la probabilidad de que un individuo, elegido aleatoriamente de entre esa población, tenga un nivel de colesterol:
    a. Entre 180 y 200 mg/100 ml
    b. Mayor que 225 mg/100 ml
    c. Entre cuales dos valores simétricamente distribuidos alrededor de la media recaerán el 96% de los niveles de colesterol.

f. Los pesos de una población de mujeres jóvenes con la mayoría de edad tienen una distribución aproximadamente normal con una media de 132 libras y una varianza de 225. Calcular la probabilidad de que una joven elegida al azar de entre esa población, pese:
    a. Más de 155 b. 100 libras o menos.
    c. Entre 105 y 143 libras.

16. Las puntuaciones, sobre 100, en un examen se distribuyen en forma normal con media de 64 puntos y desviación típica de 10 puntos, ¿Qué porcentaje de estudiantes
    a. Ganaran el examen (obtener 60 o más puntos)?
    b. Obtuvieron menos de 40 puntos?
    c. Entre 45 y 75 puntos?

17. La probabilidad de curarse que tiene un individuo con cierto síndrome es de 0.01. Se seleccionan 100 individuos con éste síndrome, cuál es la probabilidad de que: a) No se cure más de uno de ellos?, b) se curen 5 o menos?, c) se curen entre 3 y 8 inclusive. Si se escogen 500 personas con el síndrome y la probabilidad de curarse es de 0.10, d) halle la probabilidad de que se curen 45 o menos, e) más de 60, f) entre 55 y 65.

18. Sólo el 10% de los insectos expuestos a un insecticida en condiciones de laboratorio pudieron sobrevivir. Si se exponen 600 insectos, ¿cuál es la probabilidad (aproximada) de que:
    a. Sobrevivan exactamente 75 insectos?
    b. No sobreviva ningún insecto?
    c. No sobrevivan más de 50 insectos?

19. En el Programa de Gerencia de Sistemas de información en Salud, las calificaciones promedio de sus 220 estudiantes tienen una distribución aproximadamente normal con media 3.5 y varianza de 0.1444.

    a. Cuál es la probabilidad de que un estudiante seleccionado al azar no pueda habilitar?
    b. Cuantos estudiantes se espera tenga una calificación de como mínimo 3.5?
    c. Que índice de calificación alcanzará el 88% del cuerpo estudiantil?

20. Una máquina de cerveza está regulada para que descargue un promedio de 207 ml. Por vaso. Si la cantidad del líquido está distribuida aproximadamente normalmente con desviación estándar igual a 15 ml?.

    a). Que fracción de los frascos contendrá más de 231 ml.?
    b). Cuantos vasos contienen entre 198 y 210 mililitros?
    c). Bajo que valor se obtendrá el 20% de los vasos con menor cantidad.

21. Por experiencia pasada se sabe que el 80% de los estudiantes que comienzan GESIS logran obtener su titulo. Si en el segundo semestre del 2002 ingresaron 20 estudiantes calcular.

    a). La probabilidad de que por lo menos se gradué la mitad.
    c). La probabilidad de que se gradúen entre 10 y 15 inclusive.
    d). Si la carrera la empiezan 120 estudiantes, cuantos se espera que se gradúen.

# TIPOS DE MUESTREO

Los autores proponen diferentes criterios de clasificación de los diferentes tipos de muestreo, aunque en general pueden dividirse en dos grandes grupos: métodos de muestreo probabilísticos y métodos de muestreo no probabilísticos.

## Muestreo probabilístico

Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño n tienen la misma probabilidad de ser elegidas. Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.

Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos:

El método otorga una probabilidad conocida de integrar la muestra a cada elemento de la población, y dicha probabilidad no es nula para ningún elemento.

Los métodos de muestreo no probabilísticos no garantizan la representatividad de la muestra y por lo tanto no permiten realizar estimaciones inferenciales sobre la población.

(En algunas circunstancias los métodos estadísticos y epidemiológicos permiten resolver los problemas de representatividad aun en situaciones de muestreo no probabilistico, por ejemplo los estudios de caso−control, donde los casos no son seleccionados aleatoriamente de la población.)

Entre los métodos de muestreo probabilísticos más utilizados en investigación encontramos:

- **Muestreo aleatorio simple:**

  El procedimiento empleado es el siguiente:

  1. Se asigna un número a cada individuo de la población
  2. A través de algún medio mecánico (bolas dentro de una bolsa, tablas de números aleatorios, números aleatorios

     generados con una calculadora u ordenador, etc.) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

     Este procedimiento, atractivo por su simpleza, tiene poca o nula utilidad práctica cuando la población que estamos manejando es muy grande.

     Ejemplo: formar el equipo de fútbol de la universidad seleccionando 11 boletas de una urna con el nombre de todos los alumnos de la universidad.

## Muestreo aleatorio sistemático:

Este procedimiento exige, como el anterior, numerar todos los elementos de la población, pero en lugar de extraer n números aleatorios sólo se extrae uno. Se parte de ese número aleatorio i, que es un número elegido

al azar, y los elementos que integran la muestra son los que ocupa los lugares i, i+k, i+2k, i+3k,...,i+(n−1)k, es

decir se toman los individuos de k en k, siendo k el resultado de dividir el tamaño de la población entre el tamaño de la muestra: k= N/n. El número i que empleamos como punto de partida será un número al azar entre 1 y k.

El riesgo este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante (k) podemos introducir una homogeneidad que no se da en la población.

Imaginemos que estamos seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son varones y los 5 últimos mujeres, si empleamos un muestreo aleatorio sistemático con k=10 siempre seleccionaríamos o sólo hombres o sólo mujeres, no podría haber una representación de los

dos sexos.

## Muestreo aleatorio estratificado:

Trata de obviar las dificultades que presentan los anteriores ya que simplifican los procesos y suelen reducir el error muestral para un tamaño dado de la muestra. Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según la profesión, el municipio de residencia, el sexo, el estado civil, etc.).

Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población.

(Tamaño geográfico, sexos, edades,...).

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

Afijación Simple: A cada estrato le corresponde igual número de elementos muéstrales.

Afijación Proporcional: La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

Afijación Optima: Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

## Muestreo aleatorio por conglomerados:

Los métodos presentados hasta ahora están pensados para seleccionar directamente los elementos de la población, es decir, que las unidades muéstrales son los elementos de la población.

En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, los departamentos universitarios, una caja de determinado producto, etc., son conglomerados naturales.

En otras ocasiones se pueden utilizar conglomerados no naturales como, por ejemplo, las urnas electorales. Cuando los conglomerados son áreas geográficas suele hablarse de "muestreo por áreas".

El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

## Métodos de muestreo no probabilísticos

A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aun siendo conscientes de que no sirven para realizar generalizaciones, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de se elegidos.

En general se seleccionan a los sujetos siguiendo determinados criterios procurando que la muestra sea representativa.

## Muestreo por cuotas:

También denominado en ocasiones "accidental". Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más "representativos" "adecuados" para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél.

En este tipo de muestreo se fijan unas "cuotas" que consisten en un número de individuos que reúnen unas determinadas condiciones, por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en Gijón. Una vez determinada la cuota se eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

## Muestreo opinático o intencional:

Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.

## Muestreo casual o incidental:

Se trata de un proceso en el que el investigador selecciona directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento el utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos).

**Bola de nieve:**

Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones

**Conceptos básicos**

Para introducir los conceptos básicos consideremos el siguiente ejemplo:

Supongamos que estamos interesados en determinar el número medio de televisores por hogar en una ciudad.

Para ello consideraremos primeramente:

**Población: Conjunto de personas u objetos de interés en una Investigación.**

Ej: Todos los hogares de la ciudad

**Muestra**

Es una porción representativa de elementos de una población, elegida para su examen o medición directa.

Note que generalmente es costoso el análisis de todos los datos, así que se hace necesario realizar las mediciones de interés sólo en una porción representativa de la población e inferir de ella resultados que corresponden a la población entera.

Ej: Medir la cantidad de televisores en un grupo de hogares de varias localidades, municipios de la ciudad, escogidos aleatoriamente de manera conveniente.

**Parámetro**

Es cualquier característica de una población, como la media de la población, la desviación de la población, etc. Ej: Número promedio de televisores por hogar en toda la ciudad.

**Estadístico**

Es cualquier característica de una muestra, como la media de la muestra, la desviación de la muestra, etc. Ej: Número promedio de televisores calculado sólo a partir de los hogares que fueron seleccionados en la muestra.

**Muestreo**

Proceso de selección de muestras, se utiliza cuando no es posible contar o medir todos los elementos de la población objeto de estudio.

## Tipos de Muestreo

Existen dos métodos para seleccionar muestras de poblaciones:

a) Muestreo no aleatorio o de juicio: Se emplea el conocimiento y la opinión personal para identificar aquellos elementos de la población que deben incluirse en la muestra.

b) Muestreo aleatorio o de probabilidad: En el cual todos los elementos de la población tienen la oportunidad de ser escogidos para la muestra. Dentro de este tipo de muestreo se encuentran:

b.1) Muestreo aleatorio simple: el cual es un método de selección de muestras que permite que cada muestra posible pueda ser elegida con la misma probabilidad. Por su parte cada elemento de la población tiene la misma oportunidad igual de ser incluido en la muestra.

b.2) Muestreo sistemático: método en el cual los elementos que se muestrearán se seleccionan de la población en un intervalo uniforme que se mide con respecto al tiempo, al orden o al espacio.

b.3) Muestreo estratificado: método en el que la población se divide en grupos homogéneos, o estratos, y después se toma una muestra aleatoria simple de cada estrato.

Aquí la variabilidad dentro de cada grupo es pequeña y entre los grupos es grande.

b.4) Muestreo de racimo: método en el que la población se divide en grupos o racimos de elementos, y luego se selecciona una muestra aleatoria de estos racimos. La variabilidad dentro de cada grupo es grande y entre los grupos es pequeña; es como si cada racimo fuese un pequeña representación de la población en si mima.

El seleccionar uno u otro tipo de muestreo depende del problema en cuestión.

Analicemos nuestro ejemplo.

**Error Muestral**

Es la diferencia entre el parámetro de la población y el estadístico de la muestra utilizado para estimar el parámetro.

**Distribución muestral**

Es una lista de todos los valores posibles de un estadístico y la probabilidad asociada a cada valor. Se considerarán la distribución muestral de medias y la de proporciones.

*Teorema del límite central:*

> *Para muestras grandes, se puede obtener una aproximación cercana de la distribución muestral de la media con una distribución normal.*

Si $\mu_{\bar{x}} = \mu$ y ; $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ para muestras aleatorias de poblaciones infinitas, encontramos que si X es la media de una muestra aleatoria de tamaño **n** de una población infinita con la media □ y la desviación estándar □□ y n es grande, entonces

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

es un valor de una variable aleatoria que tiene aproximadamente la distribución normal estándar.

El teorema del límite central es de importancia fundamental para la estadística porque justifica el uso de métodos de curva normal en una gran variedad de problemas; se aplica a poblaciones infinitas y también a poblaciones finitas cuando *n*, a pesar de ser grande, no constituye más que una pequeña porción de la población. Es difícil señalar con precisión qué tan grande debe ser n de modo que se pueda aplicar el teorema del límite central, pero a menos de que la distribución de la población tenga una forma muy inusual, por lo regular se considera que *n* = 30 es lo suficientemente alto. Nótese que cuando en realidad estamos tomando una muestra de una población, la distribución del muestreo de la media es una distribución normal, no obstante el tamaño de *n*.

*Ejemplo* base en el teorema del límite central, ¿cuál es la probabilidad de que el error sea menor que 5, cuando se usa la media de una muestra aleatoria de tamaño *n* = 64 para estimar la media de una población infinita con □□ = 20?

*Solución* La probabilidad se obtiene por medio del área de la zona blanca bajo la curva de la figura 1, específicamente, por medio del área de curva normal estándar entre

$$z = \frac{-5}{20/\sqrt{64}} = -2 \quad \text{y} \quad z = \frac{5}{20/\sqrt{64}} = 2$$

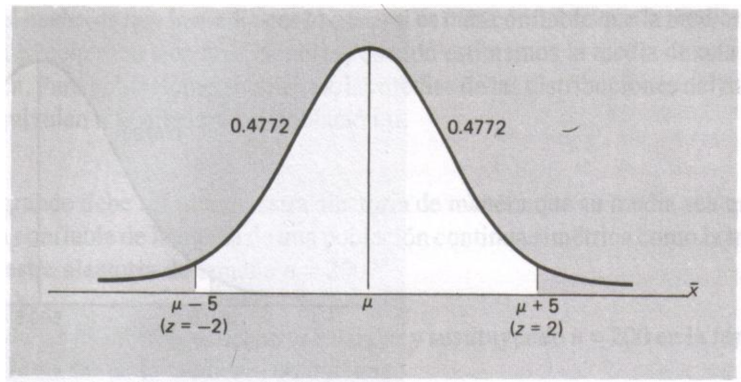



FIGURA 1 Distribución muestral de la media.

Dado que la entrada de la tabla correspondiente a z = 2.00 es 0.4772, la probabilidad que se pide es 0.4772 + 0.4772 = 0.9544. Así, sustituimos la afirmación de que la probabilidad es "como mínimo de 0.75" por una aseveración más firme de que la probabilidad es aproximadamente de 0.95 (de que la media de una muestra aleatoria de tamaño n = 64 de la población de referencia difiera de la población por menos de 5).

También se puede usar el teorema del límite central para poblaciones finitas, pero una descripción precisa de las situaciones en que se puede hacer esto sería más bien complicada. El uso apropiado más común es en el caso en que n es grande mientras que *n* / N es pequeña. Este es el caso en la mayoría de las encuestas políticas.

Teorema Central del Límite.

Para muestras grandes, se puede obtener una aproximación cercana de la distribución muestral de la media con una distribución normal.

Teniendo en cuenta que ya sabemos la media y desviación típica de la distribución muestral, podemos decir que:

————————————x =———————————— y $\sigma_x = \dfrac{\sigma}{\sqrt{n}}$

para muestras aleatorias infinitas con media ———————————— y desviación típica ———————————— y n grande, entonces:

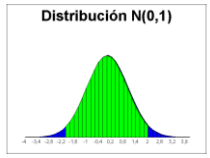
$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

es un valor de una variable N(0,1)

Este teorema es muy importante, puesto que justifica el uso de los métodos de la curva normal en una gran cantidad de problemas se utiliza para poblaciones infinitas y para poblaciones finitas cuando n a pesar de ser grande representa una porción muy pequeña de la población.

Es difícil señalar con precisión qué tan grande debe ser n de modo que podamos aplicar el Teorema Central del límite, pero a no ser que la distribución sea muy Inusual, por lo general se considera que n =30 es lo suficientemente alto.

Veamos el mismo ejemplo anterior aplicando el Teorema Central del Límite.




La probabilidad se obtiene por medio del área marcada de la zona gris, específicamente por medio del área de la N(0,1) entre:

$$z = \frac{-5}{20/\sqrt{64}} = -2 \text{ y } z = \frac{5}{20/\sqrt{64}} = 2$$

lo que consultando en las tablas da una probabilidad de 0,9544. Así sustituimos la afirmación de que la probabilidad es “como mínimo 0,75” por una aseveración más firme de que la probabilidad es aproximadamente de 0,95 ( de que la muestra aleatoria de tamaño n=64 de la población de referencia difiera de la de la población menos de 5 unidades)

También se puede usar el teorema Central del límite para poblaciones finitas, pero una descripción precisa de las situaciones en que se puede hacer esto, sería más bien complicada. El uso apropiado más común es en el caso en que n es grande y n/N es pequeña. Este es el caso de la mayoría de las encuestas políticas.

Problema 1. Extenso Tenemos una población formada por tres datos de valores 1, 2 y 6. a) Obtén todas las posibles muestras aleatorias simples de n=2 y calcula para cada una de ellas la media aritmética y la proporción de datos con valores inferiores a 5; b) A partir de la distribución muestral de medias, calcula el valor esperado y el error tipo; c) haz lo mismo con la distribución muestral de proporciones; d) Calcula los dos valores esperados y los dos errores tipo pero esta vez a partir de los datos de la población.

2. En una población de gran tamaño, el porcentaje de personas que leen un periódico al menos cinco días a la semana es del 45%. a) ¿Cuál es la desviación tipo poblacional? b) Si extraemos muestras de 49 personas, ¿Cuál es el error tipo de la proporción?

3. Según el texto AR4, la ansiedad-rasgo se distribuye en la población con una desviación tipo de valor 16. ¿Cuál es la desviación tipo de la distribución muestral de medias obtenidas a partir de muestras con n=64?

4. ¿En qué situación la desviación tipo poblacional es el doble que el error tipo de medias o de proporciones?

5. ¿Cuál es el valor de la desviación tipo poblacional si el error tipo de la media aritmética con muestras de tamaño n=49 tiene el valor 0,4?

6. ¿Qué valor de proporción hace que la varianza de proporciones sea máxima?

7. En una clase hemos preguntado por el grado en que consideran que la población estudiantil tiene poder para cambiar el mundo. La respuesta se concreta en el intervalo (0,10). Los resultados son: 4, 7, 5, 6, 3, 5, 3, 0, 10, 5, 0, 4, 6, 5, 4, 5, 6, 3, 10, 2, 6, 7, 4, 10, 5, 0, 5, 7, 6, 10, 7, 6, 4, 0, 7, 3, 5, 6, 5, 4 a) Calcula la media y la desviación tipo de la distribución muestral de medias con muestras de 9 personas. b) Obtén la media y la desviación tipo de la distribución muestral de proporciones donde las respuestas son mayores a 5 y n=25.

8. En un caso de medias aritméticas, la desviación tipo poblacional y el error tipo tienen alguno de estos valores: 0,5 y 5, aunque no se especifica cuál de estos valores se corresponde con la desviación tipo poblacional y cuál con el error tipo. Calcula el tamaño de las muestras utilizadas en la construcción de la distribución muestral, sabiendo que la población se ha considerado como infinita.

9. ¿Cuál es el valor del error tipo de proporciones si estamos utilizando muestras de tamaño 36 provenientes de una población infinita y el valor esperado es 0,4?