



UNIVERSIDAD
NACIONAL
DE COLOMBIA

**TABLAS DE CONTINGENCIA PARA LA ENSEÑANZA DEL
CONCEPTO DE ASOCIACIÓN ENTRE VARIABLES
ALEATORIAS CUALITATIVAS**

Juan Pablo Caro Quintero

Universidad Nacional de Colombia

Facultad de Ciencias

Maestría en Enseñanza de las Ciencias Exactas y Naturales

Bogotá, Colombia

2017

TABLAS DE CONTINGENCIA PARA LA ENSEÑANZA DEL CONCEPTO DE ASOCIACIÓN ENTRE VARIABLES ALEATORIAS CUALITATIVAS

Juan Pablo Caro Quintero

Tesis presentada como requisito parcial para optar al título de:
Magister en Enseñanza de la Ciencias Exactas y Naturales

Director:

Mg. Pedro Nel Pacheco Durán

Docente Departamento de Estadística

Línea de Investigación: Educación Estadística

Universidad Nacional de Colombia

Facultad de Ciencias

Maestría en Enseñanza de las Ciencias Exactas y Naturales

Bogotá, Colombia

2017

Dedicatoria

A mis padres: Ángel Custodio Caro y María Ignacia Quintero, por su esfuerzo, dedicación, apoyo incondicional y amor durante toda mi vida.

A mi hermana Elizabeth por estar siempre cuando emprendo nuevos retos, o la construcción de nuevos sueños.

A mi tía María de Jesús por todo el apoyo cuando mi familia más lo necesitaba.

Agradecimientos

A Dios por permitirme cada día nuevas oportunidades de crecimiento personal y profesional.

Al profesor Pedro Nel por su apoyo incondicional durante el proceso de desarrollo de este trabajo de grado.

Resumen

Este trabajo presenta una propuesta didáctica que pretende facilitar a los estudiantes de grado once del Colegio Saludcoop Sur I.E.D, la construcción del concepto de asociación estadística entre variables cualitativas en tablas de contingencias 2X2; y por otra parte ayudarles en el análisis de las diferentes probabilidades que pueden ser calculadas a partir de frecuencias absolutas, y marginales por filas o por columnas. La propuesta incluye el diseño de objetos virtuales de aprendizaje y un modelo geométrico que permite calcular las probabilidades esperadas de eventos estadísticamente independientes.

Palabras clave: Tablas de contingencia 2X2, asociación de variables cualitativas, probabilidades, condiciones de independencia.

Abstract

This work presents a didactic proposal that aims to facilitate to then students of the eleventh grade at Colegio Saludcoop Sur I.E.D, the construction of the concept of statistical association between qualitative variables in tables of contingencies 2×2 ; on the other hand to help them in the analysis of the different probabilities that can be calculated from absolute frequencies and marginal, frequencies by rows or by columns. The proposal includes the design of virtual objects of learning and a geometric model that allows to calculate the expected probabilities of events statistically independent.

Key words: Contingency tables 2×2 , association of qualitative variables, probabilities, conditions of independence.

Contenido

Resumen	V
Introducción	1
1. Marco Conceptual	4
1.1 Justificación	4
1.2 Situación Problema	4
1.3 Objetivo General.....	5
1.4 Objetivos Específicos	5
2. Marco Teórico	6
2.1 Probabilidades.....	6
2.1.1 Definición clásica.....	6
2.1.2 Definición de la probabilidad como frecuencia relativa o probabilidad empírica	6
2.1.3 Interpretación subjetiva de la probabilidad	6
2.1.4 Elementos básicos de la teoría de conjuntos utilizados para el estudio de la probabilidad.....	7
2.1.5 Probabilidad conjunta, marginal y condicional	8
2.1.5.1 Probabilidad conjunta.....	8
2.1.5.2 Probabilidad marginal.....	9
2.1.5.3 Probabilidad condicional	9
2.2 Eventos estadísticamente independientes	9
2.3 Tablas de Contingencia.....	10
2.3.1 Tablas de Contingencia 2x2.....	12

2.4	Independencia y asociación entre las variables de una Tabla de Contingencias ...	13
2.5	Estudios iniciales sobre coeficientes de asociación de variables categóricas en tablas de contingencia 2X2.....	16
2.6	Signo de la Asociación, Riesgo Relativo y Riesgo de Productos Cruzados en Tablas de Contingencia 2X2	18
2.6.1	Signo de la Asociación.....	18
2.6.2	Riesgo Relativo	19
2.6.3	Riesgo de Productos Cruzados	20
2.7	Asociación en Tablas de Contingencia 2X2, y el coeficiente Φ (Phi) de Pearson ..	21
2.8	Representaciones Semióticas.....	22
2.9	Situaciones Didácticas	24
3.	Propuesta Didáctica.....	26
3.1	Prueba diagnóstica.....	26
3.2	Problema Introdutorio a Tablas de Contingencias 2x2.....	28
3.3	Modelo Geométrico de una Tabla de Contingencias 2X2	32
3.4	Applets de GeoGebra primera fase	46
3.5	Formalización de la Asociación en Tablas de Contingencias 2x2	46
3.6	Applets de GeoGebra segunda fase.....	50
3.7	Tareas enviadas al Moodle del Colegio.....	51
3.8	Solución de un caso hipotético	51
4.	Conclusiones y recomendaciones.....	55
4.1	Conclusiones	55
4.2	Recomendaciones.....	56
A.	Anexo: Cuestionario Evaluación Diagnóstica.....	57
B.	Anexo: Libro de GeoGebra	59
C.	Anexo: Análisis Algebraico del Modelo Geométrico	61
C1.	Primer Procedimiento Propuesto.....	61
C2.	Segundo Procedimiento Propuesto.....	62

C3. Cálculo de Probabilidades con el Modelo Geométrico y su relación con los Cálculos del Método Tradicional	64
D. Fórmula simplificada para calcular el coeficiente de Phi de Pearson.....	67
E. Actividades registradas en el Moodle del colegio	73
Bibliografía	75

Lista de figuras

Figura 1: Tabla de la distribución Chi cuadrada	15
Figura 2: Applet, Prueba Introdutoria Tablas de Contingencia	27
Figura 3: Modelo Geométrico de una tabla de contingencias 2X2	33
Figura 4: Modelo Geométrico aplicado al ejemplo 01	35
Figura 5: Modelo Geométrico aplicado al ejemplo 02	40
Figura 6: Libro de GeoGebra	59
Figura 7: Applet Introdutorio Información de Tablas de Contingencia	59
Figura 8: Modelo Geométrico	60
Figura 9: Applet para el cálculo de porbabilidades condiciones	60
Figura 10: Applet para el cálculo de Riesgos Relativos	60
Figura 11: Applet para calcular probabilidades conjuntas y marginales	60
Figura 12: Tareas en Moodle sobre tablas de contingencia	73

Lista de tablas

Tabla 1: Tabla bivariada	8
Tabla 2: Modelo general de Tablas de Contingencia	11
Tabla 3: Modelo 1, Tabla de Contingencias 2X2 para las característica A y B.	12
Tabla 4: Modelo 2 Tabla de Contingencias 2X2 para las característica A y B.	12
Tabla 5. Modelo 3 Tabla de Contingencias 2X2 para las característica A y B.	12
Tabla 6: Signo de la Asociación Tabla de Contingencias 2X2	19
Tabla 7: Interpretaciones de los valores del cálculo de Riesgos Relativos.....	20
Tabla 8: Datos cuestionario introductorio.....	26
Tabla 9: Modelo Tabla de Contingencias Problema Introductorio.....	29
Tabla 10: Tabla para la formalización de cálculos en Tablas de Contingencia	30
Tabla 11: Modelo de tabla de contingencias 2X2 para elaborar la representación gráfica del modelo geométrico propuesto	32
Tabla 12: Tabla Ejemplo 01	34
Tabla 14: Ejercicio propuesto 01	39
Tabla 13: Ejercicio propuesto 02	39
Tabla 15: Información ejemplo 02	39
Tabla 16: Tabla resumen de los errores entre los valores esperados y observados	44
Tabla 17: Resumen de los errores entre los valores esperados y observados ejemplo 2 44	
Tabla 18: Resumen de los errores entre las probabilidades esperadas y las observadas	45
Tabla 19: Resumen de los errores entre las probabilidades esperadas y las observadas del ejemplo 2.....	45
Tabla 20: Tablas de Contingencias Problema Propuesto	47
Tabla 21: Valores para la Interpretación del Phi de Pearson	50

Tabla 22: Problema Evaluación Tablas de Contingencias 2X2.....	52
Tabla 23: Tabla de Contingencias Evaluación Diagnóstica.....	57
Tabla 24: Para el análisis algebraico del modelo geométrico	61
Tabla 25: Comparativo cálculo de probabilidades procedimiento 1	61
Tabla 26: Tabla de contingencias construida a partir de probabilidades	62
Tabla 27: Comparativo cálculo de probabilidades procedimiento 2	63

Lista de Símbolos y Abreviaturas

Símbolo o Abreviatura	Significado o definición
\cap	Intersección
\cup	Unión
\sim	Negación
\emptyset	Vacío
n_{ij}, f_{ij}	Frecuencia absoluta observada, registrada en la fila i y en la columna j
n_{11}, f_{11}	Frecuencia absoluta observada, registrada en la fila 1 y en la columna 1
$f_{\cdot j}$	Total marginal de la columna j
$f_{i \cdot}$	Total marginal de la fila o renglón i
n, T	Total de frecuencias o eventos registrados en la Tabla de Contingencias
a, a_o	Frecuencia observada f_{11} o valores observados de los eventos $(A_1 \cap B_1)$
b, b_o	Frecuencia observada f_{12} o valores observados de los eventos $(A_1 \cap B_2)$
c, c_o	Frecuencia observada f_{21} o valores observados de los eventos $(A_2 \cap B_1)$
d, d_o	Frecuencia observada f_{22} o valores observados de los eventos $(A_2 \cap B_2)$
o	Frecuencias absolutas observadas
A_i, B_j	Eventos de la Tabla de Contingencias
A	Característica A o presencia de la característica A
$\sim A$	Ausencia de la característica A
$P(A_i)$	Probabilidad marginal del evento A_i
$P(B_j)$	Probabilidad marginal del evento B_j

Símbolo o Abreviatura	Significado o definición
$P(A_i \cap B_j)$	Probabilidad conjunta de los eventos A_i y B_j
$P(A_i B_j)$	Probabilidad condicional del evento A_i , dado que ha ocurrido el evento B_j
Pa	Probabilidad observada de a
Pb	Probabilidad observada de b
Pc	Probabilidad observada de c
Pd	Probabilidad observada de d
A_a	Área que representa la ocurrencia de a
A_b	Área que representa la ocurrencia de b
A_c	Área que representa la ocurrencia de c
Ad	Área que representa la ocurrencia de d
T^2	Área que representa la ocurrencia de todos los eventos del espacio muestral
e_{ij}	Frecuencia absoluta esperada para la fila i y la columna j
e	Frecuencias absolutas esperadas
a_e	Frecuencia esperada e_{11} o valores esperados de los eventos ($A_1 \cap B_1$)
b_e	Frecuencia esperada e_{12} o valores esperados de los eventos ($A_1 \cap B_2$)
c_e	Frecuencia esperada e_{21} o valores esperados de los eventos ($A_2 \cap B_1$)
d_e	Frecuencia esperada e_{22} o valores esperados de los eventos ($A_2 \cap B_2$)
$P(a_e)$	Probabilidad calculada con el Modelo Geométrico (Probabilidad esperada) para el evento a
$P(b_e)$	Probabilidad calculada con el Modelo Geométrico (Probabilidad esperada) para el evento b
$P(c_e)$	Probabilidad calculada con el Modelo Geométrico (Probabilidad esperada) para el evento c
$P(d_e)$	Probabilidad calculada con el Modelo Geométrico (Probabilidad esperada) para el evento d
χ^2	Estadístico Chi Cuadrado
!	Factorial
	Valor absoluto
Δ	Error o diferencia entre los valores esperados y los observados

Introducción

El ciudadano actual, ciudadano del siglo XXI, cuenta con diversas fuentes de información estadística para la toma de decisiones relacionadas con su vida profesional, académica, familiar o personal; esta información puede presentarse como indicadores económicos, indicadores de morbilidad, estudios epidemiológicos, horas pico de tráfico, estudios de mercadeo, resultados de encuestas de opinión, etc. Por esta razón corresponde a la formación impartida en los colegios promover en los estudiantes el aprendizaje de la estadística y su implementación en diferentes situaciones hipotéticas o concretas.

De acuerdo con (Artega, Batanero, Contreras, & Cañadas, 2011) quienes citan a (Batanero, 2002) desde hace unas décadas la enseñanza de la estadística se ha incorporado en todos los niveles educativos en forma generalizada, promoviendo así el desarrollo del razonamiento estadístico, basado en la comprensión de ideas fundamentales como los datos, el análisis e interpretación de gráficos o tablas, la variabilidad aleatoria, el análisis y la inferencia de distintas distribuciones de datos, la asociación entre variables categóricas y la correlación para analizar si existe correspondencia o algún tipo de relación de dependencia entre variables cuantitativas; la concepción de probabilidad y el muestro inferencial (Batanero , Contreras, Díaz , & Roa, 2013). Además el razonamiento estadístico presenta los siguientes modos fundamentales según (Wild & Pfannkuch, 1999) quienes son citados por (Batanero et al., 2013)

- Reconocer la necesidad de los datos, en la forma adecuada como se han recogido y en su análisis, pues son ellos la base de la investigación estadística.
- La transnumeración, que ocurre cuando surge una nueva comprensión de los datos, desorganizados en un primer momento, al cambiarlos a otras representaciones que pueden simplificar, en gran medida, la información estadística que se desee comunicar de forma sencilla al público en general. Es posible hablar de transnumeración cuando se emplea una escala o medida para capturar la información de diferentes cualidades y características propias de un evento que se esté analizando; cuando el significado de un estadístico, aplicado a un conjunto de datos, es llevado a un lenguaje comprensible que puede ser entendido por personas no expertas; cuando los datos se representan con gráficas o tablas que permiten descubrir nuevos significados y nuevas relaciones de las variables involucradas.
- La percepción de la variación de eventos aleatorios, que permite identificar sus fuentes, buscar sus explicaciones y sus causas, para proponer ciertas inferencias o predicciones con algún margen de error.
- El razonamiento con modelos estadísticos diseñados para alcázar una mejor comprensión de la realidad. No solo se pueden emplear modelos probabilísticos, también pueden usarse funciones o gráficos para modelar la situación de interés.
- Y por último, en la integración que se debe hacer entre los estudios estadísticos y su contexto, por ejemplo al proponer el modelo y al momento de interpretar sus resultados, contrastándolos con la situación real objeto estudio.

El análisis crítico de la información estadística consignada en tablas de doble entrada o de contingencia, puede ayudar a analizar fenómenos sociales, políticos, económicos, empresariales, institucionales o nacionales.

En el primer capítulo, Marco Conceptual, se describen la justificación, la situación problema, el objetivo general y los objetivos específicos del presente trabajo.

En el segundo capítulo, Marco Teórico, se muestra un breve recorrido por la teoría de probabilidades, tipos de probabilidades que se pueden obtener a partir de las frecuencias de las tablas de contingencia, independencia estadística en tablas de contingencia, y asociación entre variables categóricas. También se mencionan los referentes que, desde la educación matemática, darán soporte al diseño de la propuesta didáctica.

En el tercer capítulo se presenta la propuesta didáctica que contempla: una prueba diagnóstica; el diseño de un modelo geométrico para representar tablas de contingencias 2×2 , cuando sus variables categóricas son estadísticamente independientes; formalización del concepto de asociación y de algunos coeficientes o indicadores del grado de asociación entre variables cualitativas, por último se proponen problemas de aplicación y de Applets en GeoGebra, como complemento a la propuesta didáctica.

En el cuarto y último capítulo se mencionan las principales conclusiones y recomendaciones para ser tenidas en cuenta en caso de desarrollar esta propuesta con otro u otros grupos de estudiantes de grado once o de nivel superior.

En los anexos se expone un análisis algebraico que permite establecer la estrecha relación entre el modelo tradicional, tomando directamente la información de la tabla de contingencias, y el modelo geométrico, cuando se presenta independencia entre sus variables categóricas, al momento de calcular los diferentes tipos de probabilidades de los eventos estudiados con tablas de contingencia 2×2 ; posteriormente, y para finalizar, se muestran imágenes o capturas de pantalla de los Applets propuestos y de actividades complementarias desarrolladas utilizando el Moodle del colegio Saludcoop Sur IED.

1.Marco Conceptual

1.1 Justificación

En la enseñanza estadística, el formato de las tablas de contingencia ayuda a visualizar conceptos y relaciones abstractas difíciles de comprender, pero no es un tema tan sencillo incluso cuando solo se trata de tablas de dos filas y dos columnas (Contreras, Cañadas, & Artea, 2013), por ello es pertinente diseñar propuestas didácticas que permitan a los estudiantes acercarse a los conceptos de probabilidad conjunta, marginal, condicional, condiciones de independencia y grado de asociación, estrechamente relacionados con el estudio de tablas de contingencia y su aplicación a partir de situaciones problema.

1.2 Situación Problema

Partiendo de las posibilidades del análisis de la información en tablas de contingencia, se plantea la siguiente pregunta problema:

¿Qué características debe tener una propuesta didáctica que permita a estudiantes de grado undécimo, del colegio Saludcoop Sur IED jornada tarde, construir tablas de contingencia que los acerquen a los conceptos de asociación entre variables aleatorias cualitativas, probabilidad marginal, conjunta y condicional, a partir del análisis de frecuencias absolutas, relativas y marginales?

1.3 Objetivo General

Diseñar una secuencia didáctica para estudiantes de grado undécimo, del colegio Saludcoop Sur IED jornada tarde, que les permita examinar la información estadística consignada en tablas de contingencia 2X2, y llegar a una primera conceptualización de asociación entre variables cualitativas, probabilidad marginal, conjunta y condicional, a partir del análisis de frecuencias absolutas, relativas y marginales.

1.4 Objetivos Específicos

- Diseñar actividades y guías que permitan a los estudiantes establecer la importancia de la construcción de tablas de contingencia para el análisis de la información de estudios estadísticos de variables cualitativas.
- Diseñar un objeto virtual de aprendizaje que ayude a los estudiantes en la adquisición de habilidades para elaborar y analizar la información consignada en tablas de contingencia, y además facilite la construcción de representaciones gráficas asociadas como los diagramas de barras apilados y adosados.
- Crear actividades que permitan a los estudiantes la obtención de indicadores para evaluar si existe o no asociación entre dos variables cualitativas objeto de estudio.
- Diseñar un conjunto de pruebas que permita evaluar el grado de comprensión de diferentes conceptos relacionados con la asociación entre variables categóricas y de probabilidad vistos a lo largo del desarrollo de la propuesta.

2. Marco Teórico

2.1 Probabilidades

2.1.1 Definición clásica

Si un experimento, que está sujeto al azar, resulta de n formas igualmente probables y mutuamente excluyentes, y si n_A de estos resultados tienen un atributo A , la probabilidad de A es la proporción de n_A con respecto a n . (Canavos, 1988, pág. 29)

2.1.2 Definición de la probabilidad como frecuencia relativa o probabilidad empírica

Si después de n repeticiones de un experimento bajo las mismas condiciones, se observa que una cantidad n_A de los resultados corresponden a la ocurrencia de un atributo A , entonces la probabilidad de A se define como el límite de $\frac{n_A}{n}$ conforme n toma valores cada vez más grandes (Canavos, 1988, pág. 31)

2.1.3 Interpretación subjetiva de la probabilidad

En este caso la probabilidad se interpreta como el grado de creencia, o de convicción personal, respecto a la ocurrencia de una afirmación (Canavos, 1988, pág. 31)

2.1.4 Elementos básicos de la teoría de conjuntos utilizados para el estudio de la probabilidad

1. Espacio muestral (S): Es el conjunto de todos los posibles resultados de un experimento aleatorio; puede ser discreto o continuo según sea la naturaleza de los datos que se estén analizando; su probabilidad de ocurrencia es igual a uno, $P(S) = 1$ (Canavos, 1988, pág. 32)
2. Evento (A): Un evento es un grupo o subconjunto del espacio muestral que se define de acuerdo al cumplimiento de uno o más atributos. Su probabilidad de ocurrencia puede calcularse dentro de intervalo $[0, 1]$
3. Evento ($A \cap B$): Es el evento conformado por todos los resultados comunes entre A y B , es decir son todos los resultados que cumplen simultáneamente con las condiciones de los atributos A y B . La probabilidad de ocurrencia de este evento se expresa $P(A \cap B)$.
4. Evento ($A \cup B$): Es el evento conformado por todos los posibles resultados de A o de B , es decir son todos los resultados que cumplen con las condiciones del atributo A o del atributo B . La probabilidad de ocurrencia de este evento puede calcularse con la siguiente formula:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
5. Evento (A'): El evento A complemento está conformado por todos los posibles resultados de S que no se encuentran en A . Su probabilidad de ocurrencia puede expresarse así $P(A') = 1 - P(A)$
6. Evento vacío o nulo (\emptyset): Es el evento que no contiene ningún resultado posible del espacio muestral, recibe el nombre de evento nulo o vacío y su probabilidad es igual a cero. $P(\emptyset) = 0$.

7. Evento $(A \cap B) = \emptyset$: Si dos eventos son mutuamente excluyentes o disjuntos, su intersección es vacía y se dice que no tienen resultados en común.
8. $(A \subset B)$: El evento A está contenido en el evento B si todo resultado de A también es un resultado de B .

2.1.5 Probabilidad conjunta, marginal y condicional

Para definir las probabilidades conjuntas, marginales y condicionales considérese que se ha realizado una encuesta que busca estudiar dos características, A y B , de un grupo de personas, y que la información se resume en una tabla bivarida de dos filas y dos columnas como la que se muestra a continuación.

Tabla 1: Tabla biviada

	B_1	B_2	Totales
A_1	n_{11}	n_{12}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$n_{2\cdot}$
Totales	$n_{\cdot 1}$	$n_{\cdot 2}$	n

Si S es el espacio muestral que contiene a los eventos disjuntos A_1 , A_2 , B_1 , y B_2 , entonces se tienen las siguientes probabilidades:

2.1.5.1 Probabilidad conjunta

Se llama probabilidad conjunta a la probabilidad de ocurrencia, de forma simultánea, de los eventos A_i y B_j en relación al total de eventos del espacio muestral.

$$P(A_i \cap B_j) = \frac{n_{ij}}{n} \quad (1)$$

En el caso particular de la tabla 1, se tiene que

$$P(A_1 \cap B_1) = \frac{n_{11}}{n} \quad (2)$$

2.1.5.2 Probabilidad marginal

Se llama probabilidad marginal a la probabilidad de ocurrencia de los eventos A_i o B_j , independientemente de cualquier otro evento del espacio muestral. Otra forma de definir la probabilidad marginal es como la suma de sus probabilidades conjuntas.

$$P(A_i) = \sum_{j=1}^2 \frac{n_{ij}}{n} = \sum_{j=1}^2 P(A_i \cap B_j) \quad (3)$$

$$P(B_j) = \sum_{i=1}^2 \frac{n_{ij}}{n} = \sum_{i=1}^2 P(A_i \cap B_j) \quad (4)$$

2.1.5.3 Probabilidad condicional

Se llama probabilidad condicional a la probabilidad de que ocurra un evento A_i cuando se sabe que ha ocurrido evento un B_j , esto se denota así $P(A_i|B_j)$ y se lee "Probabilidad de A_i dado B_j o la probabilidad de que ocurra A_i , dado que ha ocurrido B_j ". Una fórmula que puede utilizarse para calcular esta probabilidad es la siguiente:

$$P(A_i|B_j) = \frac{P(A_i \cap B_j)}{P(B_j)}, \quad \text{con } P(B_j) > 0 \quad (5)$$

2.2 Eventos estadísticamente independientes

Cuando se quiere calcular la probabilidad condicional de ocurrencia de un evento A , dado que ha ocurrido un evento B , se presupone que las probabilidades de ocurrencia de estos dos eventos están relacionadas y que son dependientes entre sí; dicho de otra forma, la probabilidad de que se presente A es afectada por la ocurrencia de B , sin embargo pueden presentarse situaciones en las que la ocurrencia de B no influye sobre las probabilidades de ocurrencia de A . Cuando se presentan estas situaciones se dice que los eventos A y B son estadísticamente independientes.

Sean A y B dos eventos cualesquiera de un espacio muestral S , se dice que el evento A es **estadísticamente independiente** del evento B , con $P(B) > 0$, si

$P(A|B) = P(A)$, igualmente B es **estadísticamente independiente** del evento A , con $P(A) > 0$, si $P(B|A) = P(B)$, esto tiene la siguiente implicación:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ pero como } P(A|B) = P(A) \text{ tenemos que}$$

$$P(A) = \frac{P(A \cap B)}{P(B)}, \text{ por lo tanto } P(A)P(B) = P(A \cap B)$$

En resumen, si A y B son eventos **estadísticamente independientes**, se debe verificar una de las siguientes condiciones de independencia:

i. $P(A|B) = P(A)$ (6)

ii. $P(B|A) = P(B)$ (7)

iii. $P(A \cap B) = P(A)P(B)$ (8)

2.3 Tablas de Contingencia

Las tablas de contingencia, término acuñado por Karl Pearson en 1904 (Sanchez Rivero, 1998), se utilizan en estudios estadísticos interesados en la posible relación entre dos atributos cualitativos de los individuos de una población. Usualmente cada atributo se maneja como una variable cualitativa X o Y , y la variación de cada variable está representada por los valores $x_1, x_2, \dots, x_i, \dots, x_r$ y $y_1, y_2, \dots, y_j, \dots, y_c$ respectivamente. La información se organizan en un tabla de doble entrada de observaciones bivariadas que tienen como primera columna y primera fila o renglón, los valores que asumen X y Y (Cañadas G. , 2010)

En general si se toma una muestra de tamaño n de una población, y se desean estudiar dos características de un mismo individuo o grupo de personas, se tiene que:

Sean estas características X y Y la muestra se divide en:

- Clase x_i para la variable X
- Clase y_j para la variable Y

Dónde:

- f_{ij} es la frecuencia absoluta o frecuencia observada en cada celda
- $f_{i\cdot}$ es la frecuencia acumulada en $x_{i\cdot}$, distribución marginal por filas o renglones
- $f_{\cdot j}$ es la frecuencia acumulada en $y_{j\cdot}$, distribución marginal por columnas
- $h_{ij} = \frac{f_{ij}}{n}$ es la frecuencia relativa
- $h_{ij} \left(\frac{x_i}{y_j} \right) = \frac{f_{ij}}{f_{\cdot j}} = \frac{f_{ij}}{\widehat{f_{y_j}}} = \frac{h_{ij}}{h_{\cdot j}}$ representa las distribuciones condicionales por columna
- $h_{ij} \left(\frac{y_j}{x_i} \right) = \frac{f_{ij}}{f_{i\cdot}} = \frac{f_{ij}}{\widehat{f_{x_i}}} = \frac{h_{ij}}{h_{i\cdot}}$ representa las distribuciones condicionales por fila o renglón

Tabla 2: Modelo general de Tablas de Contingencia

	y_1	y_2	...	y_j	...	y_c	Total
x_1	f_{11}	f_{12}	...	f_{1j}	...	f_{1c}	$f_{1\cdot}$
x_2	f_{21}	f_{22}	...	f_{2j}	...	f_{2c}	$f_{2\cdot}$
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{ic}	$f_{i\cdot}$
⋮	⋮	...	⋮	⋮
x_r	f_{r1}	f_{r2}	...	f_{rj}	...	f_{rc}	$f_{r\cdot}$
Total	$f_{\cdot 1}$	$f_{\cdot 2}$...	$f_{\cdot j}$...	$f_{\cdot c}$	n

2.3.1 Tablas de Contingencia 2x2

Un caso particular son las tablas de contingencia de 2X2 como las que se muestran a continuación:

Tabla 3: Modelo 1, Tabla de Contingencias 2X2 para las característica A y B.

	<i>B</i>	<i>No B</i>	Total
<i>A</i>	f_{11}	f_{12}	$f_{1\cdot}$
<i>No A</i>	f_{21}	f_{22}	$f_{2\cdot}$
Total	$f_{\cdot 1}$	$f_{\cdot 2}$	n

Tabla 4: Modelo 2 Tabla de Contingencias 2X2 para las característica A y B.

	<i>B</i>	<i>No B</i>	Total
<i>A</i>	a	b	$a + b$
<i>No A</i>	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

Tabla 5. Modelo 3 Tabla de Contingencias 2X2 para las característica A y B.

	B_1	B_2	Total
A_1	a	b	$a + b$
A_2	c	d	$c + d$
Total	$a + c$	$b + d$	$T = a + b + c + d$

Dónde:

a, b, c y d : Representan frecuencias absolutas.

De acuerdo a (Contreras J. M., Cañadas, Gea, & Arteaga, 2012), quienes citan a (Díaz & De la Fuente, 2005) por cada celda de la tabla de contingencias pueden calcularse tres frecuencias relativas diferentes, y además cada frecuencia relativa se relaciona con el cálculo de una probabilidad diferente.

Por ejemplo en relación a la celda ***a*** o celda (1,1) tenemos:

- Frecuencia relativa doble y el cálculo de la probabilidad conjunta de los eventos A_1 y B_1

$$\frac{f_{11}}{n} = \frac{a}{a+b+c+d} = P(A_1 \cap B_1) \quad (9)$$

- Frecuencia relativa con respecto a su fila o renglón, y el cálculo de la probabilidad condicional B_1 dado que ha ocurrido A_1 :

$$\frac{f_{11}}{f_{1\cdot}} = \frac{a}{a+b} = P(B_1|A_1) \quad (10)$$

- Frecuencia relativa con respecto a su columna, y el cálculo de la probabilidad condicional A_1 dado que ha ocurrido B_1

$$\frac{f_{11}}{f_{\cdot 1}} = \frac{a}{a+c} = P(A_1|B_1) \quad (11)$$

- Frecuencias relativas marginales y el cálculo de probabilidades marginales por filas o columnas es el siguiente:

$$\text{Por fila } \frac{f_{1\cdot}}{n} = \frac{a+b}{a+b+c+d} = P(A_1) \quad (12)$$

$$\text{Por columna } \frac{f_{\cdot 1}}{n} = \frac{a+c}{a+b+c+d} = P(B_1) \quad (13)$$

2.4 Independencia y asociación entre las variables de una Tabla de Contingencias

Cuando se estudia la posible asociación o relación de dependencia entre las variables categóricas de una tabla de contingencias, se parte de la hipótesis de que

no existe relación alguna y por lo tanto cada frecuencia observada (f_{ij}) es equivalente a una frecuencia esperada o teórica (e_{ij})

Antes de realizar los cálculos de independencia se plantean las siguientes dos hipótesis:

Hipótesis nula (H_0): Las variables categóricas de la tabla de contingencias son independientes entre sí.

Hipótesis alternativa (H_1): Las variables categóricas de la tabla de contingencias son dependientes entre sí.

Para determinar las frecuencias esperadas se tienen en cuenta los valores marginales de la tabla de contingencias; se asume igual la probabilidad para cada categoría y que no existe asociación entre las variables estudiadas.

$$e_{ij} = \frac{(Total\ de\ la\ columna\ j)(Total\ de\ renglón\ i)}{Tamaño\ de\ la\ muestra} \quad (14)$$

El grado de concordancia entre las frecuencias esperadas y las frecuencias observadas se calcula utiliza el estadísticos χ^2 (Chi cuadrado)

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad (15)$$

Una versión simplificada para calcular este estadístico es:

$$\chi^2 = \sum \frac{(o-e)^2}{e} \quad (16)$$

Este estadístico permite identificar relaciones de dependencia entre variables cualitativas, y afirmar con cierto nivel de confianza si los valores alcanzados por una de las variables influyen sobre otra. (Otero & Medina Mora, 2005)

Cuando se calcula el valor de χ^2 , y resulta mayor que algún valor crítico, tal como el de $\chi^2_{0,95}$, se puede decir que las frecuencias observadas difieren significativamente de las frecuencias esperadas, entonces se rechaza la hipótesis inicial H_0 . En caso contrario, se acepta o al menos no se rechaza. Este procedimiento se conoce como ensayo o prueba de Chi-cuadrado de hipótesis. (Spiegel & Stephens, 2009)

En el caso particular de tablas de contingencia de 2X2, si decide trabajar con un nivel de significancia de 0.05 y un valor crítico de $\chi^2_{0,95}$ se tiene como regla de decisión la siguiente:

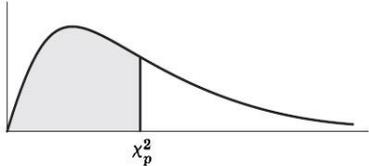
Si $\chi^2 \geq 3.84$, entonces H_0 es rechazada, lo que significa la existencia de dependencia entre las variables

Si $\chi^2 < 3.84$ No se rechaza la hipótesis H_0 , lo que significa que existe independencia entre las variables.

Estos valores se encuentran en tablas de la distribución Chi cuadrada como la que se muestra a continuación, que pertenece al del libro de Estadística de Spiegel y Stephens, 2009.

Apéndice IV

Valores percentiles (χ^2_p) correspondientes a la distribución ji cuadrada con ν grados de libertad (área sombreada = p)



ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989

Figura 1: Tabla de la distribución Chi cuadrada

2.5 Estudios iniciales sobre coeficientes de asociación de variables categóricas en tablas de contingencia 2X2

Según Fienberg (1978), citado por (Sanchez Rivero, 1998), los estudios iniciales que se ocuparon del análisis de las variables categóricas en tablas de contingencias fueron los desarrollados por Pearson y Yule a inicios del siglo XX.

Pearson definió tres coeficientes:

1) Coeficiente de correlación Tetracórico

$$r_T = \frac{f_{12}f_{21}}{f_{11}f_{22}} \quad (17)$$

Si $\frac{f_{12}f_{21}}{f_{11}f_{22}} > 1$ entonces r_T es positivo, la relación entre las variables es positiva

Si $\frac{f_{12}f_{21}}{f_{11}f_{22}} < 1$ entonces r_T es negativo, la relación entre las variables es negativa

2) Coeficiente Phi de Pearson:

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad (18)$$

El tamaño de la muestra no influye en el valor que pueda tomar este coeficiente, por lo que es posible comparar los resultados de diferentes tablas. Este coeficiente únicamente toma valores mayores o iguales a cero, a medida que se aleje de cero el grado de asociación entre las variables será mayor, si su valor es cero se presume que las variables muestran independencia entre sí.

3) Coeficiente de contingencia:

Una de las principales dificultades del coeficiente Phi de Pearson es que puede tomar valores superiores a 1, por ello Pearson propuso en 1904 (Sanchez Rivero,

1998) el coeficiente de contingencia, el cual tomará siempre valores en el intervalo 0 a 1. Si $C = 0$ entonces las variables muestran independencia, pero si se acerca a 1 entonces se dice que presentan cierto grado de asociación.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (19)$$

Entre 1954 y 1972

Goodman y Kruskal proponen un conjunto de medidas para evaluar la independencia y la asociación entre variables categóricas. Entre estas medidas pueden mencionarse las siguientes (Sanchez Rivero, 1998):

1) Test de corrección de continuidad de Yates:

$$\chi^2 = \frac{n(|f_{11}f_{22} - f_{12}f_{21}| - 0.5n)^2}{f_{1\cdot} \cdot f_{2\cdot} \cdot f_{\cdot 1} \cdot f_{\cdot 2}} \quad (20)$$

Este test se comporta como una distribución χ^2 con un grado de libertad y puede variar desde cero hasta infinito. Si el resultado obtenido es mayor a 3,84 puede decirse que las variables de la tabla de contingencias muestran que son independientes.

2) Test exacto de Fisher

Si las frecuencias observadas en la tabla de contingencias son menores a cinco, no es apropiado utilizar el test de χ^2 , lo recomendado es utilizar el siguiente test

$$P = \frac{f_{1\cdot}! \cdot f_{2\cdot}! \cdot f_{\cdot 1}! \cdot f_{\cdot 2}!}{n! \cdot f_{11}! \cdot f_{12}! \cdot f_{21}! \cdot f_{22}!} \quad (21)$$

3) Test Q de Yule:

$$Q = \frac{f_{11}f_{22} - f_{12}f_{21}}{f_{11}f_{22} + f_{12}f_{21}} \quad (22)$$

Posibles resultados

$Q = 0$ Independencia

$Q < 0$ Asociación negativa

$Q > 0$ Asociación positiva

$Q = 1$ Asociación perfecta estricta positiva

$Q = -1$ Asociación perfecta estricta negativa

4) Coeficiente Y de Yule:

$$\hat{Y} = \frac{\sqrt{\hat{\alpha}} - 1}{\sqrt{\hat{\alpha}} + 1} \quad (23)$$

En donde $\sqrt{\hat{\alpha}} = \sqrt{\frac{f_{11}f_{22}}{f_{12}f_{21}}}$

El coeficiente de Yule puede tomar valores entre 1 y -1 . Si se da el caso de que $\hat{Y} = 0$ entonces se puede decir que las variables de la tabla de contingencias son independientes.

2.6 Signo de la Asociación, Riesgo Relativo y Riesgo de Productos Cruzados en Tablas de Contingencia 2X2

2.6.1 Signo de la Asociación.

En una tabla de contingencias de dos filas y dos columnas, las diagonales principales a, d y b, c aportan información sobre el tipo de dependencia y

asociación entre sus variables. Valores en las celdas **a** (presencia de **A** y de **B**) y en la celda **d** (ausencia de **A** y de **B**) indican asociación directa o positiva. Valores en las celdas **b** (presencia de **A** y ausencia de **B**) y en la celda **c** (ausencia de **A** y presencia de **B**) indican asociación inversa o negativa. (Cañadas G. , 2013)

Tabla 6: Signo de la Asociación Tabla de Contingencias 2X2

	<i>B</i>	<i>No B</i>
<i>A</i>	<i>a</i> <i>Asociación Directa</i>	<i>b</i> <i>Asociación Inversa</i>
<i>No A</i>	<i>c</i> <i>Asociación Inversa</i>	<i>d</i> <i>Asociación Directa</i>

Para determinar el grado de asociación entre las variables cualitativas de una tabla de contingencias se pueden utilizar los siguientes coeficientes de asociación.

2.6.2 Riesgo Relativo

El Riesgo Relativo puede calcularse tanto por filas como por columnas. Si es por columnas permite establecer que tan probable es que se dé **A** dado que se ha dado **B** en relación a los casos en los que no se ha presentado **B**. El riesgo relativo por filas muestra que tanto más probable es la presencia de **B** dado que ha ocurrido **A** en relación a aquellos casos en los que no se presenta **A** (Cañadas G. , Contreras, Arteaga, & Gea, 2013)

$$RR_C = \frac{P(A/B)}{P(A/No B)} = \frac{a/(a+c)}{b/(b+d)} = \frac{a(b+d)}{b(a+c)} \quad (24)$$

$$RR_F = \frac{P(B/A)}{P(B/No A)} = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)} \quad (25)$$

Para este indicador se establecen las siguientes interpretaciones

Tabla 7: Interpretaciones de los valores del cálculo de Riesgos Relativos

$RR < 1$	$RR = 1$	$RR > 1$
Existe asociación negativa entre las variables	No existe asociación entre las variables	Existe asociación positiva entre las variables

2.6.3 Riesgo de Productos Cruzados

Es el cociente que se calcula entre el producto de las celdas favorables a la asociación positiva y el de las celdas que son favorables a la asociación negativa

Para calcular el riesgo de productos cruzado se procede así

- 1) Se calcula la razón entre los caso favorables a que se dé **B** dado que ha ocurrido **A** y los casos favorables a que **no** se de **B** dado que ha ocurrido **A**

$$R_1 = \frac{P(B|A)}{P(No B|A)} = \frac{\frac{a}{a+b}}{\frac{b}{a+b}} = \frac{a(a+b)}{b(a+b)} = \frac{a}{b}$$

- 2) Se calcula la razón entre los caso favorables a que se dé **B** dado que no ha ocurrido **A** y los casos favorables a que **no** se de **B** dado que **no** ha ocurrido **A**

$$R_2 = \frac{P(B|No A)}{P(No B|No A)} = \frac{\frac{c}{c+d}}{\frac{d}{c+d}} = \frac{c(c+d)}{d(c+d)} = \frac{c}{d}$$

- 3) Finalmente se calcula la razón entre R_1 y R_2

$$RC = \frac{R_1}{R_2} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

En resumen

$$RC = \frac{ad}{bc} \quad (26)$$

Para este indicador se establecen las siguientes interpretaciones

Si $RC < 1$: Se dice que las variables de la tabla de contingencias presentan asociación inversa.

Si $RC = 1$: Se dice que las variables de la tabla de contingencias presentan independencia entre sí.

Si $RC > 1$: Se dice que las variables de la tabla de contingencias presentan asociación directa.

2.7 Asociación en Tablas de Contingencia 2X2, y el coeficiente Φ (Phi) de Pearson

El coeficiente Phi de Pearson utiliza el valor de Chi-cuadrado y toma valores entre -1 y 1 . Puede calcularse utilizando una de las siguientes fórmulas (Cañadas G., et al. 2013)

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad (27.1)$$

$$\Phi = \frac{f_{11}f_{22} - f_{12}f_{21}}{\sqrt{f_{1\cdot} \cdot f_{\cdot 1} f_{2\cdot} \cdot f_{\cdot 2}}} \quad (27.2)$$

$$\Phi = \frac{ad - bc}{\sqrt{(a+c)(a+b)(b+d)(c+d)}} \quad (27.3)$$

La equivalencia entre estas expresiones puede verificarse en el anexo D

Este coeficiente se calcula a partir de las distancias entre las frecuencias observadas y las esperadas en caso de independencia. Si las variables son

independientes entonces no existe diferencia entre las frecuencias observadas y las esperadas, por lo tanto, el resultado de este coeficiente es cero.

Para interpretar el resultado obtenido con este coeficiente se deben tener en cuenta las siguientes observaciones:

- 1) Cuando la dependencia es directa y perfecta, todos los caso se concentran en las celdas a y d , se obtiene el valor máximo de 1 . Cuanto más se acerque a 1 la dependencia es más alta.
- 2) Si la dependencia es inversa y perfecta, todos los caso se concentran en las celdas b y c , se obtiene el valor mínimo de -1 . Cuanto más se acerque a -1 la dependencia inversa y más alta.
- 3) Si existe total independencia, el valor del coeficiente es cero
- 4) Este coeficiente no depende las frecuencias marginales.
- 5) Si se dividen o multiplican todas las frecuencias de la tabla de contingencias por el mismo número, el valor de este coeficiente no varía.

2.8 Representaciones Semióticas

A continuación se describen los elementos sugeridos por la teoría de las representaciones semióticas para que los estudiantes logren la comprensión o conceptualización de un objeto matemático. En este caso en particular se utilizarán en la enseñanza de conceptos de probabilidades conjuntas, marginales, condicionales, condiciones de independencia y asociación, relacionados con tablas de contingencias 2×2 .

El aprendizaje de las matemáticas está mediado por la utilización de diferentes representaciones semióticas que permiten la adquisición, construcción y manipulación de diversos conceptos, por ejemplo, si se piensa en un punto, una recta o un cuadrado resulta casi imposible actuar sobre estos objetos matemáticos

sin recurrir a una imagen mental, una representación tabular o gráfica que de ellos se tenga.

Cuando se busca la comprensión de un objeto matemático, desde el enfoque teórico que se quiere manejar, se deben tener en cuenta diversos aspectos dentro de los cuales se resaltan:

- Los grados de libertad, es decir, “los registros de representación semiótica de los que dispone un sujeto [en este caso el estudiante], para objetivarse él mismo [un concepto matemático] o una idea aún confusa” (Duval, 1999, pág. 29)
- “Las representaciones semióticas son a la vez externas y conscientes, y cumplen con las funciones cognitivas de transformación intencional, expresión y objetivación.” (Duval, 1999, pág. 34). En este sentido son externas y conscientes toda vez que posibilitan al individuo comunicar su comprensión acerca de un objeto.
- Se debe lograr la congruencia entre representaciones, posibilitando una correspondencia semántica de los elementos significantes [de cada representación.] (Duval, 1999, pág. 50) es decir que los elementos característicos del objeto permanezcan invariantes ante el cambio de registro de representación.
- Cuando se comparan dos representaciones, por ejemplo, la tabular y la gráfica “Las organizaciones respectivas de las unidades significante de las dos representaciones comparadas, [conducen] a que las unidades en correspondencia semántica sean *aprehendidas* en el mismo orden en las dos representaciones. (Duval, 1999, págs. 50-51)
- Las transformaciones al interior de los registros, los tratamientos, las transformaciones externas, y conversiones de un objeto matemático evidencian el grado de conceptualización que ha alcanzado el estudiante.

- Por último según (Duval, 1999, págs. 13,14) “...no puede haber comprensión en matemáticas si no se distingue un objeto de su representación. Desde esta perspectiva es esencial no confundir jamás los objetos matemáticos, es decir, los números, las funciones, las rectas, etc., con sus representaciones, es decir, las escrituras decimales o fraccionarias, los símbolos, los gráficos, los trazados de figuras..., pues un mismo objeto matemático puede darse a través de representaciones muy diferentes...es el objeto matemático representado lo que importa y no sus diversas representaciones semióticas posibles (Deledicq al, 1979). Toda confusión entre el objeto y su representación provoca, en un plazo más o menos amplio, una pérdida de comprensión: los conocimientos adquiridos se hacen rápidamente inutilizables por fuera de su contexto de aprendizaje, sea por no recordarlos o porque permanecen como representaciones “inertes” que no sugieren ninguna transformación productora. En virtud de su pluralidad potencial, las diversas representaciones semióticas de los objetos matemáticas sería, pues, secundarias y extrínsecas a la aprehensión conceptual de los objetos...”

2.9 Situaciones Didácticas

Según (Godino, 2003) citando a Brousseau

“...el conocimiento existe y tiene sentido para el sujeto cognoscente sólo porque representa una solución óptima en un sistema de restricciones (Brousseau, 1986, pág. 368)...”

De acuerdo a esta teoría, el profesor puede diseñar artificialmente diferentes situaciones con ciertas restricciones, que guíen a los estudiantes en la construcción de nuevos conocimientos o en otros casos a modificar sus conocimientos previos.

Brousseau, propone varios tipos de situaciones didácticas, cuya secuenciación puede provocar la génesis artificial de un concepto matemático. Dichos tipos de situaciones son las siguientes:

Tipos de situaciones didácticas propuestas por Brousseau (Godino, 2003)

1) Situaciones centradas sobre “la acción”

Son las situaciones en las que los estudiantes intentan, por primera vez, resolver por sus propios medios, un problema planteado por el profesor.

2) Situaciones centradas en la “comunicación”

En este tipo de situaciones los estudiantes deben comunicar a sus compañeros y su profesor los resultados parciales o finales alcanzados al intentar encontrar la solución al problema propuesto.

3) Situaciones centradas sobre la “validación”

En estas situaciones los estudiantes deben argumentar sus hallazgos, utilizando formalmente los conocimientos matemáticos de los que disponen, dejando de lado las simples conjeturas basadas en experiencia empíricas.

4) Situaciones de institucionalización

En este punto los estudiantes, después de un proceso de diálogo con sus compañeros y con su profesor, resumen los procedimientos, las ideas y los conceptos matemáticos de acuerdo a la terminología utilizada de forma oficial.

En cada uno de los momentos de la clase ya mencionados se da un componente que Brousseau designa como “adidáctico”, en el que son los estudiantes los responsables de gestionar autónomamente la situación en un tiempo y un espacio determinado por el profesor.

3. Propuesta Didáctica

3.1 Prueba diagnóstica

Antes de iniciar la implementación de la propuesta se realiza una prueba diagnóstica en dos fases para determinar si los estudiantes cuentan con conocimientos sobre el cálculo de probabilidades en una tabla de contingencias. En la primera fase se propone que solucionen un cuestionario adaptado de un ejercicio del libro “Estadística Elemental de Freund John E. y Simón, G. A. (1992)”. (Ver Anexo A)

Se realiza una encuesta a 260 solteros, 400 casados y 340 viudos o divorciados, sobre los aspectos que más contribuyen a su felicidad. Los resultados se muestran en la siguiente tabla.

Tabla 8: Datos cuestionario introductorio

	Solteros	Casados	Divorciados o viudos
Amigos y vida social	140	200	160
Trabajo/Estudio	40	120	120
Salud y condición física	80	80	60

Teniendo en cuenta la información de la tabla de contingencias, los estudiantes deben contestar preguntas como las siguientes:

- 1) Si se desea cuantificar la posibilidad de seleccionar una persona que asegure que se siente más feliz cuando esta con sus amigos departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?
- 2) Si se desea cuantificar la posibilidad de seleccionar una persona soltera que asegure que se siente más feliz cuando esta con sus amigos departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?
- 3) Si se desea cuantificar la posibilidad de seleccionar una persona divorciada o viuda que asegure que se siente más feliz cuando esta con sus amigos departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?

En la segunda fase se presenta a los estudiantes un Applet en GeoGebra (ver anexo B) para familiarizarlos con las ventajas de organizar información en tablas de contingencia y adicionalmente mostrar que a partir de la información registrada es posible calcular:

- 1) Frecuencias por filas y columnas
- 2) Porcentajes
- 3) Razones que posteriormente se utilizarán para el cálculo de probabilidades

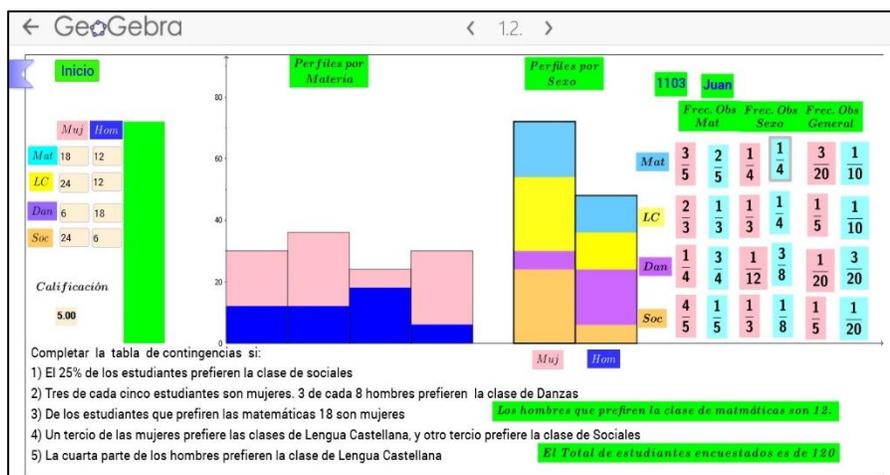


Figura 2: Applet, Prueba Introdutoria Tablas de Contingencia

El Applet también permite mostrar a los estudiantes las representaciones gráficas de las tablas de contingencias utilizando diagramas de barras apilados y adosados.

Una vez los estudiantes han desarrollado el Applet, se pide que en clase contesten preguntas como las siguientes:

- 1) ¿Qué frecuencias absolutas permiten calcular las siguientes probabilidades?
 - a. Probabilidad de seleccionar un estudiante que sea mujer y prefiera la clase de Danzas.
 - b. Probabilidad de seleccionar un estudiante que sea hombre y prefiera la clase de Sociales.

- 2) ¿Qué frecuencias relativas permiten calcular las siguientes probabilidades?
 - a. La probabilidad de que sea hombre si se sabe que prefiere las matemáticas.
 - b. La probabilidad de que prefiera a clase de sociales si se sabe que es mujer.

- 3) ¿Cuál es la probabilidad de seleccionar un estudiante que sea mujer?

- 4) ¿Cuál es la probabilidad de seleccionar un estudiante que prefiera la clase de danzas?

3.2 Problema Introdutorio a Tablas de Contingencias 2x2

Se propone a los estudiantes la siguiente situación:

Se ha realizado un estudio a 90 bachilleres entre los 19 y 22 años, para determinar quiénes tienen mayor posibilidad de seguir sus estudios en instituciones de educación superior.

El diseño de la encuesta es el siguiente:

1) Indique el estrato socioeconómico al cual pertenece

1	2	3	4	5	6
0	0	0	0	0	0

2) ¿Actualmente está matriculado en una institución de educación superior?

Si__ No__

Los resultados se registraron en la siguiente tabla de contingencias 2X2:

Tabla 9: Modelo Tabla de Contingencias Problema Introdutorio

TC	B_1	B_2	Totales
A_1	24	16	40
A_2	30	20	50
Totales	54	36	90

En donde las variables categóricas se han definido así:

A_1 : Pertenece a los estratos socioeconómicos 4, 5 o 6

A_2 : Pertenece a los estratos socioeconómicos 1, 2 o 3

B_1 : Esta matriculado en una institución de educación superior

B_2 : No está matriculado en ninguna institución de educación superior

Los estudiantes deben contestar preguntas como las siguientes:

- 1) ¿La probabilidad de seleccionar un estudiante que este matriculado en la universidad y no pertenezca al estratos socioeconómicos 4, 5 o 6, es?

- 2) ¿La probabilidad de seleccionar un estudiante que pertenezca a los estratos socioeconómicos 4, 5 o 6 y no este matriculado en alguna institución de educación superior, es?

- 3) ¿La probabilidad de escoger aleatoriamente una persona que en la actualidad no esté estudiando es de?
- 4) ¿La probabilidad de seleccionar al azar una persona que pertenezca al estrato socioeconómico 1, 2 o 3 es de?
- 5) Si se sabe que la persona seleccionada está matriculada en una institución de educación superior ¿La probabilidad de que pertenezca a estrato socioeconómico 4, 5 o 6 es de?
- 6) Si se sabe que la persona seleccionada pertenece a los estratos 1, 2 o 3 ¿La probabilidad de que no esté matriculada en una institución de educación superior es de?

Al finalizar esta actividad y después de discutir los diferentes resultados obtenidos por los estudiantes, se procede a formalizar los procedimientos que se deben utilizar para analizar la información registrada en tablas de contingencias 2X2.

Tabla 10: Tabla para la formalización de cálculos en Tablas de Contingencia

TC	B_1	B_2	Totales
A_1	a	b	$a + b$
A_2	c	d	$d + d$
Totales	$a + c$	$b + d$	$T = a + b + c + d$

Cálculo de Probabilidades Conjuntas

$$P(A_1 \cap B_1) = Pa = \frac{a}{a+b+c+d} = \frac{a}{T} \quad (28) \quad P(A_1 \cap B_2) = Pb = \frac{b}{a+b+c+d} = \frac{b}{T} \quad (29)$$

$$P(A_2 \cap B_1) = Pc = \frac{c}{a+b+c+d} = \frac{c}{T} \quad (30) \quad P(A_2 \cap B_2) = Pd = \frac{d}{a+b+c+d} = \frac{d}{T} \quad (31)$$

Cálculo de Probabilidades Marginales

$$P(B_1) = \frac{a+c}{a+b+c+d} = \frac{a+c}{T} \quad (32)$$

$$P(B_2) = \frac{b+d}{a+b+c+d} = \frac{b+d}{T} \quad (33)$$

$$P(A_1) = \frac{a+b}{a+b+c+d} = \frac{a+b}{T} \quad (34)$$

$$P(A_2) = \frac{c+d}{a+b+c+d} = \frac{c+d}{T}; \quad (35)$$

Probabilidades Condicionales

$$P(B_1|A_1) = \frac{a}{a+b} \quad (36)$$

$$P(B_2|A_1) = \frac{b}{a+b} \quad (37)$$

$$P(A_1|B_1) = \frac{a}{a+c} \quad (38)$$

$$P(A_2|B_1) = \frac{c}{a+c} \quad (39)$$

Cuando se analiza la información de una tabla de contingencias, no solo interesan los cálculos de las probabilidades que ya se mencionaron, es muy importante evaluar la independencia entre sus variables, por ello es indispensable verificar si las probabilidades de los eventos A_1, A_2, B_1 y B_2 cumplen con las siguientes condiciones de independencia.

Condiciones para verificar independencia:

- 1) La probabilidad de ocurrencia de (A_1) es independiente de la ocurrencia o no ocurrencia de (B_1)

$$P(A_1) = P(A_1|B_1)$$

- 2) La probabilidad de ocurrencia de (B_1) es independiente de la ocurrencia o no ocurrencia de (A_1)

$$P(B_1) = P(B_1|A_1)$$

- 3) La probabilidad conjunta de los eventos (A_1) y (B_1) es igual al producto de sus probabilidades marginales

$$P(A_1 \cap B_1) = P(A_1)P(B_1)$$

Por último se orienta a los estudiantes en la construcción de diagramas de barras apilados y adosados para resumir gráficamente la información de la tabla de contingencias.

3.3 Modelo Geométrico de una Tabla de Contingencias 2X2

Después de utilizar diagramas de barras apilados y adosados se hacen las siguientes preguntas a los estudiantes:

¿Es posible diseñar una representación gráfica que de razón de las probabilidades que se pueden calcular a partir de tablas de contingencia de dos filas y dos columnas?

¿Qué características debe tener este modelo gráfico?

¿Cómo se puede verificar que realmente representa la información registrada en la tabla de contingencias?

Después de discutir algunas alternativas propuestas se sugiere el siguiente procedimiento:

Dada una tabla de contingencias como la que se muestra a continuación, dibujar un cuadrado que permita representar la participación de la probabilidad de cada uno de los eventos: a , b , c y d ; con relación a la probabilidad total.

Tabla 11: Modelo de tabla de contingencias 2X2 para elaborar la representación gráfica del modelo geométrico propuesto

TC 01	B_1	B_2	Totales
A_1	a	b	$a + b$
A_2	c	d	$c + d$
Totales	$a + c$	$b + d$	$a + b + c + d = T$

Para establecer las razones que permitirán calcular diferentes probabilidades de la tabla de contingencias, utilizando el modelo geométrico, se deben tener en cuenta las siguientes consideraciones:

- 1) El área total del cuadrado representará la ocurrencia de todos los eventos registrados en la tabla de contingencias.
- 2) Horizontalmente el cuadrado se dividirá en dos regiones $(a + c)$ y $(b + d)$
- 3) Verticalmente el cuadrado se dividirá en dos regiones $(c + d)$ y $(a + b)$
- 4) El área total del cuadrado será igual a T^2 , en donde $T^2 = (a + b + c + d)^2$
- 5) El cuadrado quedará dividido en cuatro áreas más pequeñas

$$Aa = (a + c)(a + b) \quad (40)$$

$$Ab = (b + d)(a + b) \quad (41)$$

$$Ac = (a + c)(c + d) \quad (42)$$

$$Ad = (b + d)(c + d) \quad (43)$$

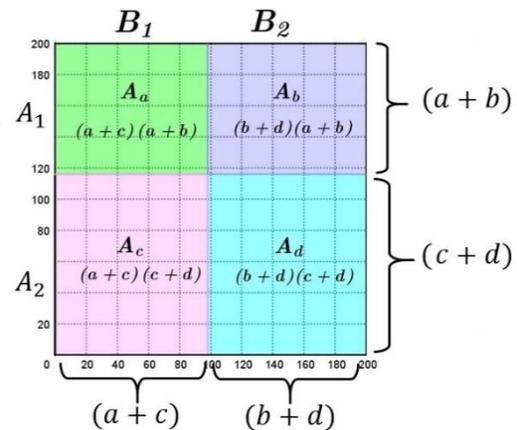


Figura 3: Modelo Geométrico de una tabla de contingencias 2X2

Si el modelo geométrico funciona, entonces:

Las probabilidades marginales por columna y fila pueden ser calculadas así:

$$P(A_1) = \frac{Aa+Ab}{T^2} \quad (44)$$

$$P(A_2) = \frac{Ac+Ad}{T^2} \quad (45)$$

$$P(B_1) = \frac{Aa+Ac}{T^2} \quad (46)$$

$$P(B_2) = \frac{Ab+Ad}{T^2} \quad (47)$$

Las probabilidades conjuntas pueden ser calculadas así:

$$P(A_1 \cap B_1) = Pa = \frac{Aa}{T^2} \quad (48)$$

$$P(A_1 \cap B_2) = Pb = \frac{Ab}{T^2} \quad (49)$$

$$P(A_2 \cap B_1) = Pc = \frac{Ac}{T^2} \quad (50)$$

$$P(A_2 \cap B_2) = Pd = \frac{Ad}{T^2} \quad (51)$$

Algunas de las probabilidades condicionales pueden ser calculadas así:

$$P(A_1|B_1) = \frac{Aa}{Aa+Ac} \quad (52)$$

$$P(A_2|B_1) = \frac{Ac}{Aa+Ac} \quad (53)$$

$$P(B_1|A_1) = \frac{Aa}{Aa+Ab} \quad (54)$$

$$P(B_2|A_1) = \frac{Ab}{Aa+Ab} \quad (55)$$

Ejemplo 01:

Utilizar el modelo geométrico propuesto para analizar la información de la siguiente tabla de contingencias y calcular las probabilidades que se solicitan a continuación:

Tabla 12: Tabla Ejemplo 01

TC	B_1	B_2	Totales
A_1	42	18	60
A_2	14	6	20
Totales	56	24	80

- 1) Las probabilidades marginales
- 2) Las probabilidades conjuntas
- 3) Las siguientes probabilidades condicionales

$$P(B_1|A_1); P(B_2|A_1); P(A_1|B_1) \text{ y } P(A_2|B_1)$$

- 4) Verificar las condiciones de independencia para los eventos A_1 y B_1
- 5) Repetir los cálculos tomando la información directamente de la tabla y verificar si difieren o no los resultados, si esto ocurre cuantificar el error de los resultados del modelo geométrico propuesto.

Desarrollo

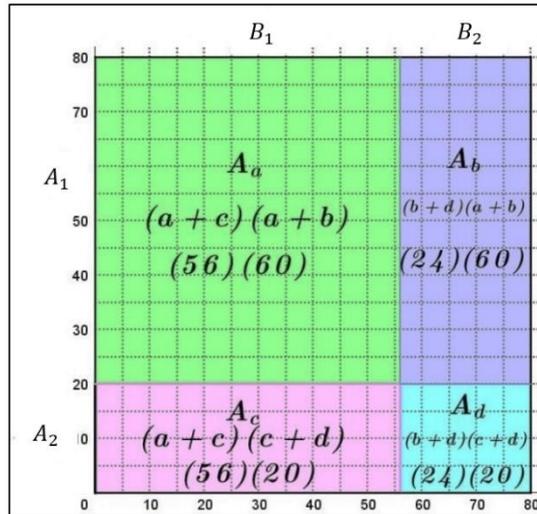


Figura 4: Modelo Geométrico aplicado al ejemplo 01

1) Probabilidades marginales

$$P(B_1) = \frac{Aa+Ac}{T^2} = \frac{(56)(60)+(56)(20)}{80^2} = \frac{(56)(60+20)}{(80)^2} = \frac{56(80)}{80^2} = \frac{56}{80} = 0,7$$

$$P(B_2) = \frac{Ab+Ad}{T^2} = \frac{(24)(60)+(24)(20)}{80^2} = \frac{(24)(60+20)}{(80)^2} = \frac{(24)(80)}{(80)^2} = \frac{24}{80} = 0,3$$

$$P(A_1) = \frac{Aa+Ab}{T^2} = \frac{(56)(60)+(24)(60)}{(80)^2} = \frac{(80)(60)}{(80)^2} = \frac{60}{80} = 0,75$$

$$P(A_2) = \frac{Ac+Ad}{T^2} = \frac{(56)(20)+(24)(20)}{(80)^2} = \frac{(80)(20)}{(80)^2} = \frac{20}{80} = 0,25$$

2) Probabilidades conjuntas utilizando la representación geométrica

$$P(A_1 \cap B_1) = Pa = \frac{Aa}{T^2} = \frac{(56)(60)}{(80)^2} = \frac{3360}{6400} = 0,525$$

$$P(A_1 \cap B_2) = Pb = \frac{Ab}{T^2} = \frac{(24)(60)}{80^2} = \frac{1440}{6400} = 0,225$$

$$P(A_2 \cap B_1) = Pc = \frac{Ac}{T^2} = \frac{(56)(20)}{(80)^2} = \frac{1120}{6400} = 0,175 \text{ y}$$

$$P(A_1 \cap B_2) = Pd = \frac{Ad}{T^2} = \frac{(24)(20)}{(80)^2} = \frac{480}{6400} = 0,075$$

La suma de las probabilidades conjuntas debe ser igual a 1

$$0,525 + 0,225 + 0,175 + 0,075 = 1.000$$

Antes de seguir adelante se debe mostrar que los cálculos de las probabilidades corresponden realmente a la información registrada en la tabla; para hacer esto se sugiere el siguiente razonamiento:

Si se multiplica el total de la tabla de contingencias por cada una de las probabilidades conjuntas, este producto debe coincidir con los valores observados.

$$T = 80$$

$$a = P(A_1 \cap B_1)(T) = P(A_1 \cap B_1)(80) = (0,525)(80) = 42$$

$$b = P(A_1 \cap B_2)(T) = P(A_1 \cap B_2)(80) = (0,225)(80) = 18$$

$$c = P(A_2 \cap B_1)(T) = P(A_2 \cap B_1)(80) = (0,175)(80) = 14$$

$$d = P(A_2 \cap B_2)(T) = P(A_2 \cap B_2)(80) = (0,075)(80) = 6$$

Los cálculos han pasado la prueba

3) Probabilidades Condicionales

$$P(B_1|A_1) = \frac{Aa}{Aa+Ab} = \frac{(56)(60)}{(56)(60)+(24)(60)} = \frac{(56)(60)}{(80)(60)} = \frac{56}{80} = 0,7$$

$$P(B_2|A_2) = \frac{Ab}{Aa+Ab} = \frac{(24)(60)}{(56)(60)+(24)(60)} = \frac{(24)(60)}{(80)(60)} = \frac{24}{80} = 0,3$$

$$P(A_1|B_2) = \frac{Aa}{Aa+Ac} = \frac{(56)(60)}{(56)(60)+(56)(20)} = \frac{(56)(60)}{(56)(80)} = \frac{60}{80} = 0,75$$

$$P(A_2|B_1) = \frac{Ac}{Aa+Ac} = \frac{(56)(20)}{(56)(60)+(56)(20)} = \frac{(56)(20)}{(56)(80)} = \frac{20}{80} = 0,25$$

4) Condiciones para verificar independencia

1) $P(A) = P(A|B)$

$P(A) = 0,75$ y $P(A|B) = 0,75$

2) $P(B|A) = P(B)$

$P(B) = 0,7$ y $P(B|A) = 0,7$

3) $P(A \cap B) = P(A)P(B)$

$P(A \cap B) = 0,525$ y $P(A)P(B) = (0,75)(0,70) = 0,525$

Se cumplen las condiciones de independencia

5) Cálculos de probabilidad tomado la información directamente de la tabla

Probabilidades conjuntas

$$P(A_1 \cap B_1) = Pa = \frac{a}{T} = \frac{a}{a+b+c+d} = \frac{42}{80} = 0,525$$

$$P(A_1 \cap B_2) = Pb = \frac{b}{T} = \frac{b}{a+b+c+d} = \frac{18}{80} = 0,225$$

$$P(A_2 \cap B_1) = Pc = \frac{c}{T} = \frac{c}{a+b+c+d} = \frac{14}{80} = 0,175$$

$$P(A_2 \cap B_2) = Pd = \frac{d}{T} = \frac{d}{a+b+c+d} = \frac{6}{80} = 0,075$$

Probabilidades Marginales

$$P(B) = \frac{a+c}{a+b+c+d} = \frac{a+c}{T} = \frac{56}{80} = 0,70$$

$$P(\sim B) = \frac{b+d}{a+b+c+d} = \frac{b+d}{T} = \frac{24}{80} = 0,30$$

$$P(A) = \frac{a+b}{a+b+c+d} = \frac{a+b}{T} = \frac{60}{80} = 0,75$$

$$P(\sim A) = \frac{c+d}{a+b+c+d} = \frac{c+d}{T} = \frac{20}{80} = 0,25$$

Probabilidades condicionales

$$P(B|A) = \frac{a}{a+b} = \frac{42}{60} = 0,70 \quad P(\sim B|A) = \frac{b}{a+b} = \frac{18}{60} = 0,30$$

$$P(A|B) = \frac{a}{a+c} = \frac{42}{56} = 0,75 \quad P(\sim A|B) = \frac{c}{a+c} = \frac{14}{56} = 0,25$$

Condiciones para verificar independencia

$$P(A) = P(A|B)$$

$$P(A) = 0,75 \text{ y } P(A|B) = 0,75$$

$$P(B|A) = P(B)$$

$$P(B) = 0,7 \text{ y } P(B|A) = 0,7$$

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap B) = 0,525 \text{ y } P(A)P(B) = (0,75)(0,70) = 0,525$$

Se cumplen las condiciones de independencia

No se presentan diferencias entre los resultados del modelo propuesto y los resultados obtenidos tomando la información directamente de la tabla.

Ejercicio:

Utilizar el modelo geométrico propuesto para analizar la información de las siguientes tablas de contingencias y calcular las probabilidades que se solicitan a continuación:

Tabla 13: Ejercicio propuesto 01

TC	B	~B	Totales
A	20	12	32
~A	30	18	48
Totales	50	30	80

Tabla 14: Ejercicio propuesto 02

TC	B	~B	Totales
A	14	16	30
~A	21	24	45
Totales	35	40	75

- 1) Las probabilidades marginales
- 2) Las probabilidades conjuntas
- 3) Las siguientes probabilidades condicionales

$$P(B|A); P(\sim B|A); P(A|B) \text{ y } P(\sim A|B)$$

- 4) Verificar las condiciones de independencia para los eventos A y B
- 5) Repetir los cálculos utilizando directamente la información de la tabla y verificar si difieren o no con los resultados obtenidos con el modelo geométrico propuesto, si esto ocurre, se debe cuantificar el error o diferencia.

Ejemplo 02:

Utilizar el modelo geométrico propuesto para representar la siguiente tabla de contingencias y además calcular las probabilidades solicitadas utilizando los dos métodos en paralelo:

Tabla 15: Información ejemplo 02

TC	B_1	B_2	Totales
A_1	36	12	48
A_2	8	24	32
Totales	44	36	80

- 1) Las probabilidades marginales
- 2) Las probabilidades conjuntas
- 3) Las siguientes probabilidades condicionales
 $P(B_1|A_1)$; $P(B_2|A_1)$; $P(A_1|B_1)$ y $P(A_2|B_1)$
- 4) Verificar las condiciones de independencia para los eventos A_1 y B_1
- 5) Repetir los cálculos tomando directamente la información de la tabla y verificar si difieren o no con los resultados obtenidos utilizando el modelo geométrico, si esto ocurre, cuantificar el error o diferencia.

Desarrollo

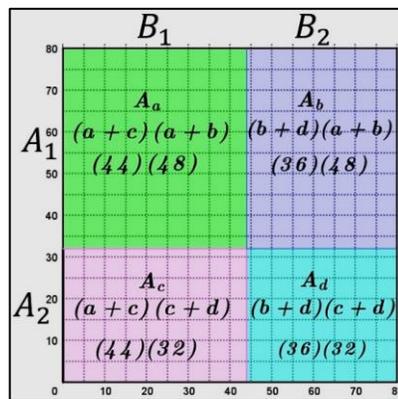


Figura 5: Modelo Geométrico aplicado al ejemplo 02

1) Cálculos con el modelo geométrico

Probabilidades marginales

$$P(B_1) = \frac{Aa+Ac}{T^2} = \frac{(44)(48)+(44)(32)}{80^2} = \frac{(44)(48+32)}{80^2} = \frac{(44)(80)}{80^2} = \frac{44}{80} = \frac{11}{20} = 0,55$$

$$P(B_2) = \frac{Ab+Ad}{T^2} = \frac{(36)(48)+(36)(32)}{80^2} = \frac{(36)(80)}{80^2} = \frac{36}{80} = \frac{9}{20} = 0,45$$

$$P(A_1) = \frac{Aa+Ab}{T^2} = \frac{(44)(48)+(36)(48)}{80^2} = \frac{(80)(48)}{80^2} = \frac{48}{80} = \frac{3}{5} = 0,60$$

$$P(A_2) = \frac{Ac+Ad}{T^2} = \frac{(44)(32)+(36)(32)}{80^2} = \frac{(80)(32)}{(80)^2} = \frac{32}{80} = 0,40$$

Probabilidades conjuntas

$$P(A_1 \cap B_1) = Pa = \frac{Aa}{T^2} = \frac{(44)(48)}{(80)^2} = \frac{2112}{6400} = 0,330$$

$$P(A_1 \cap B_2) = Pb = \frac{Ab}{T^2} = \frac{(36)(48)}{80^2} = \frac{1728}{6400} = 0,270$$

$$P(A_2 \cap B_1) = Pc = \frac{Ac}{T^2} = \frac{(44)(32)}{(80)^2} = \frac{1408}{6400} = 0,220$$

$$P(A_2 \cap B_2) = Pd = \frac{Ad}{T^2} = \frac{(36)(36)}{(80)^2} = \frac{1152}{6400} = 0,180$$

La suma de todas las probabilidades conjuntas debe ser igual a 1

$$0,330 + 0,270 + 0,220 + 0,180 = 1.000$$

Probabilidades Condicionales

$$P(B_1|A_1) = \frac{Aa}{Aa+Ab} = \frac{(44)(48)}{(44)(48)+(36)(48)} = \frac{(44)(48)}{(80)(48)} = \frac{44}{80} = 0,55$$

$$P(B_2|A_1) = \frac{Ab}{Aa+Ab} = \frac{(36)(48)}{(44)(48)+(36)(48)} = \frac{(36)(48)}{(80)(48)} = \frac{36}{80} = 0,45$$

$$P(A_1|B_1) = \frac{Aa}{Aa+Ac} = \frac{(44)(48)}{(44)(48)+(44)(32)} = \frac{(44)(48)}{(44)(80)} = \frac{48}{80} = \frac{3}{5} = 0,60$$

$$P(A_2|B_1) = \frac{Ac}{Aa+Ac} = \frac{(44)(32)}{(44)(48)+(44)(32)} = \frac{(44)(32)}{(44)(80)} = \frac{32}{80} = \frac{2}{5} = 0,40$$

Condiciones para verificar independencia de los eventos A_1 y B_1

$$\checkmark P(A_1) = P(A_1|B_1)$$

$$P(A_1) = 0,60 \quad \text{y} \quad P(A_1|B_1) = 0,60$$

$$\checkmark P(B_1) = P(B_1|A_1)$$

$$P(B_1) = 0,55 \quad \text{y} \quad P(B_1|A_1) = 0,55$$

$$\checkmark P(A_1 \cap B_1) = P(A_1)P(B_1)$$

$$P(A_1 \cap B_1) = 0,330 \quad \text{y} \quad P(A_1)P(B_1) = (0,60)(0,55) = 0,330$$

Los eventos A_1 y B_1 son independientes de acuerdo a las probabilidades calculadas con el modelo geométrico

- 2) Cálculos de probabilidad utilizando el método tradicional, tomado directamente los valores registrados en cada una de las celdas de la tabla de contingencias.

Probabilidades marginales observadas

$$P(B_1) = \frac{a+c}{a+b+c+d} = \frac{44}{80} = 0,55 \quad P(B_2) = \frac{b+d}{a+b+c+d} = \frac{36}{80} = 0,45$$

$$P(A_1) = \frac{a+b}{a+b+c+d} = \frac{48}{80} = 0,60 \quad P(A_2) = \frac{c+d}{a+b+c+d} = \frac{32}{80} = 0,40$$

Probabilidades conjuntas observadas

$$P(A_1 \cap B_1) = Pa = \frac{a}{T} = \frac{a}{a+b+c+d} = \frac{36}{80} = 0,45$$

$$P(A_1 \cap B_2) = Pb = \frac{b}{T} = \frac{b}{a+b+c+d} = \frac{12}{80} = 0,15$$

$$P(A_2 \cap B_1) = Pc = \frac{c}{T} = \frac{c}{a+b+c+d} = \frac{8}{80} = 0,10$$

$$P(A_2 \cap B_2) = Pd = \frac{d}{T} = \frac{d}{a+b+c+d} = \frac{24}{80} = 0,300$$

Probabilidades condicionales observadas

$$P(B_1|A_1) = \frac{a}{a+b} = \frac{42}{60} = 0,70 \quad P(B_2|A_1) = \frac{b}{a+b} = \frac{18}{60} = 0,30$$

$$P(A_1|B_1) = \frac{a}{a+c} = \frac{42}{56} = 0,75 \quad P(A_2|B_1) = \frac{c}{a+c} = \frac{14}{56} = 0,25$$

Condiciones para verificar independencia de los eventos A_1 y B_1 de acuerdo a la información de los datos observados.

$$\checkmark P(A_1) = P(A_1|B_1)$$

$$P(A_1) = 0,60 \quad \text{y} \quad P(A_1|B_1) = 0,75$$

No se cumple

$$\checkmark P(B_1) = P(B_1|A_1)$$

$$P(B_1) = 0,55 \quad \text{y} \quad P(B_1|A_1) = 0,70$$

No se cumple

$$\checkmark P(A_1 \cap B_1) = P(A_1)P(B_1)$$

$$P(A_1 \cap B_1) = 0,45 \quad \text{y} \quad P(A_1)P(B_1) = (0,60)(0,55) = 0,33$$

No se cumple

No se cumple ninguna condición de independencia.

Se presentan diferencias entre los resultados obtenidos con el método tradicional, y los calculados a partir del modelo geométrico. Para establecer que tanto se diferencian estos resultados se deben calcular los valores esperados utilizando las probabilidades dadas por el modelo geométrico.

Valores esperados de acuerdo al modelo geométrico

$$a_e = P(A_1 \cap B_1)(T) = P(A_1 \cap B_1)(80) = (0,330)(80) = 26,400$$

$$b_e = P(A_1 \cap B_2)(T) = P(A_1 \cap B_2)(80) = (0,270)(80) = 21,600$$

$$c_e = P(A_2 \cap B_1)(T) = P(A_2 \cap B_1)(80) = (0,220)(80) = 17,600$$

$$d_e = P(A_2 \cap B_2)(T) = P(A_2 \cap B_2)(80) = (0,180)(80) = 14,400$$

Para cuantificar el error se deben determinar las diferencias entre los valores observados en la tabla de contingencias original, y los valores esperados dados por el modelo geométrico.

$$\Delta a = (a - a_e) = 36 - 26,400 = 9,600 \quad \Delta b = (b - b_e) = 12 - 21,600 = -9,600$$

$$\Delta c = (c - c_e) = 8 - 17,600 = -9,600 \quad \Delta d = (d - d_e) = 24 - 14,400 = 9,600$$

Resumir los errores de los valores esperados en relación a los observados en una tabla

Tabla 16: Tabla resumen de los errores entre los valores esperados y observados

TC	B_1	B_2	Totales
A_1	Δa_e	Δb_e	$\Delta a_e + \Delta b_e$
A_2	Δc_e	Δd_e	$\Delta c_e + \Delta d_e$
Totales	$\Delta a_e + \Delta c_e$	$\Delta b_e + \Delta d_e$	$\Delta a_e + \Delta b_e + \Delta c_e + \Delta d_e$

Tabla 17: Resumen de los errores entre los valores esperados y observados
ejemplo 2

TC	B_1	B_2	Totales
A_1	9,600	-9,600	0
A_2	-9,600	9,600	0
Totales	0	0	0

Si se tiene en cuenta el error sin importar el signo, el error total en relación a los valores esperados de acuerdo al modelo geométrico es de: $(4)(9,600) = 38,4$

$$EVMG = 38,4$$

Ahora bien, se ha calculado el error entre los datos observados y los esperados, pero, ¿cuál es el error entre las probabilidades observadas y las probabilidades esperadas?

$$\Delta P(a) = (P(a) - P(a_e)) = (0,450 - 0,330) = 0.12$$

$$\Delta P(b) = (P(b) - P(b_e)) = (0,150 - 0,270) = -0.12$$

$$\Delta P(c) = (P(c) - P(c_e)) = (0,100 - 0,220) = -0.12$$

$$\Delta P(d) = (P(d) - P(d_e)) = (0,300 - 0,180) = 0.12$$

En una tabla se resumen los errores de las probabilidades esperadas en relación a los valores de las probabilidades observadas

Tabla 18: Resumen de los errores entre las probabilidades esperadas y las observadas

TC	B_1	B_2	Totales
A_1	$\Delta P(a)$	$\Delta P(b)$	$\Delta P(a) + \Delta P(b)$
A_2	$\Delta P(c)$	$\Delta P(d)$	$\Delta P(c) + \Delta P(d)$
Totales	$\Delta P(a) + \Delta P(c)$	$\Delta P(b) + \Delta P(d)$	$\Delta P(a) + \Delta P(b) + \Delta P(c) + \Delta P(d)$

Tabla 19: Resumen de los errores entre las probabilidades esperadas y las observadas del ejemplo 2

TC	B_1	B_2	Totales
A_1	0.120	-0.120	0
A_2	-0.120	0.120	0
Totales	0	0	0

Si sólo se tiene en cuenta el error sin importar el signo, el error total es:

$$(4)(0.120) = 0.48$$

A esto lo llamaremos error de las probabilidades calculadas con el modelo geométrico en relación con las probabilidades observadas directamente en la tabla de contingencias.

$$EPMG = 0.48$$

Los errores totales que se han obtenido son una primera aproximación a medidas o coeficientes que permiten establecer el grado de relación o asociación, entre las variables de una tabla de contingencias 2X2; cuando se ha establecido que no se cumplen las condiciones de independencia, y además se observa que existen diferencias entre los valores calculados con las probabilidades del modelo geométrico (valores esperados) y los valores observados.

3.4 Applets de GeoGebra primera fase

En la página oficial de GeoGebra se diseña un libro interactivo con los siguientes capítulos:

1) Modelo Geométrico

Muestra cómo se puede representar cualquier tabla de contingencias 2X2 en un cuadrado dividido en cuatro regiones.

2) Applet para evaluar probabilidades conjuntas y marginales

Permite a los estudiantes adquirir habilidad para calcular las probabilidades conjuntas y marginales en una tabla de contingencias.

3) Applet para calcular probabilidades condicionales

Permite a los estudiantes adquirir habilidad para calcular probabilidades condicionales.

3.5 Formalización de la Asociación en Tablas de Contingencias 2x2

PROBLEMA PROPUESTO

En la facultad de educación de una prestigiosa universidad se propone que el estado debe financiar programas de aprendizaje de una segunda lengua en todos los colegios, sobre todo en los grados conformados por estudiantes entre los 7 y 12 años de edad. Este programa puede contribuir en beneficiar a los estudiantes en aspectos como los siguientes:

- 1) Mejorar procesos asociados a la memoria
- 2) Facilidad para interactuar en diferentes contextos (académicos, culturales, etc.)
- 3) Mayores oportunidades a nivel educativo y laboral
- 4) Mayor desarrollo de habilidades comunicativas

Para sustentar estadísticamente su propuesta la universidad realizó una investigación que involucraba diferentes pruebas y entrevistas, con el fin de evaluar el grado de desarrollo de las habilidades y beneficios ya mencionados. Los resultados se resumen en las siguientes tablas de contingencias:

Tabla 20: Tablas de Contingencias Problema Propuesto

Grupo 01: Bachilleres graduados entre 1998 y 2003				Grupo 02: Bachilleres graduados entre 2004 y 2010			
TC	B_1	B_2	Totales	TC	B_1	B_2	Totales
A_1	189	126	315	A_1	360	10	370
A_2	111	74	185	A_2	8	122	130
Totales	300	200	500	Totales	368	132	500

A_1 : Bachiller de colegio bilingüe con énfasis en el aprendizaje de una segunda lengua desde primaria.

A_2 : Bachiller de colegio no bilingüe sin énfasis en el aprendizaje de una segunda lengua.

B_1 : Bachiller que en las pruebas y entrevistas muestra gran desarrollo de las habilidades objeto de estudio y la posibilidad de mejores oportunidades a nivel académico o laboral.

B_2 : Bachiller que en las pruebas y entrevistas no muestra gran desarrollo de las habilidades objeto de estudio, y además reconoce que tiene pocas oportunidades a nivel académico o laboral.

Con el modelo geométrico propuesto representar las tablas de contingencias de los dos grupos, y además calcular las probabilidades solicitadas a continuación:

- 1) Las probabilidades conjuntas
- 2) Las probabilidades marginales
- 3) Las siguientes probabilidades condicionales

$$P(A_1|B_1); P(A_2|B_2); P(B_1|A_1) \text{ y } P(B_2|A_2);$$

- 4) Verificar las condiciones de independencia para los eventos A_1 y B_1
- 5) Verificar las condiciones de independencia para los eventos A_2 y B_2
- 6) Repetir los cálculos tomando directamente los datos registrados en la tabla de contingencias y verificar si difieren o no los resultados, si esto ocurre cuantificar el error entre los resultados obtenidos con el modelo geométrico y los calculados con el método tradicional.

Contestar:

- 1) ¿La universidad puede sustentar su propuesta con los resultados obtenidos por su investigación?
- 2) ¿Existe alguna relación entre ser bachiller de colegio bilingüe con énfasis en el aprendizaje de una segunda lengua desde primaria, y tener la posibilidad de mejores oportunidades a nivel académico o laboral, junto con los demás beneficios objeto de estudio de la investigación de la universidad?
- 3) ¿Es posible cuantificar la relación ya mencionada, en un rango entre 0 y 1, en donde 0 indica que no existe ningún tipo de relación y 1 que la reacción existe y es muy fuerte?

Después de que los estudiantes han calculado las probabilidades solicitadas, y han propuesto respuestas a las preguntas planteadas, se presentan en clase algunas medidas de asociación que pueden ser utilizadas para analizar si las variables involucradas en la tabla de contingencias tienen o no tienen alguna relación de dependencia.

Para estimar la asociación se utilizan las siguientes medidas en el desarrollo de ejercicios en clase y con los algunos Applets de GeoGebra

1) Riesgo relativo por columnas

$$RR_C = \frac{P(A_1/B_1)}{P(A_1/B_2)} = \frac{a/(a+c)}{b/(b+d)} = \frac{a(b+d)}{b(a+c)}$$

2) Riesgos relativos por filas

$$RR_F = \frac{P(B_1/A_1)}{P(B_1/A_2)} = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)}$$

3) Riesgos relativos cruzados

$$RC = \frac{ad}{bc}$$

4) Cálculo Chi-Cuadrado para evaluar las distancias entre los valores observados y esperados

$$\chi^2 = \sum \frac{(e-o)^2}{e}$$

5) Phi de Pearson

Al formalizar en la clase la teoría sobre asociación en tablas de contingencias, y las formas de evaluar tanto la dependencia de las variables como el grado de asociación se utiliza el coeficiente de Phi de Pearson.

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad \text{o} \quad \Phi = \frac{ad-bc}{\sqrt{(a+c)(a+b)(b+d)(c+d)}}$$

Puesto que, de acuerdo al modelo geométrico, $(a+c)(a+b) = A_a$ y $(b+d)(c+d) = A_d$, es posible escribir este coeficiente de una forma más simple

$$\Phi = \frac{ad-bc}{\sqrt{(A_a)(A_d)}} \quad (56)$$

Para calificar el grado de asociación entre las variables de la tabla de contingencias se propone a los estudiantes la siguiente escala o tabla de valores.

Tabla 21: Valores para la Interpretación del Phi de Pearson

Coefficiente de Asociación	Interpretación de la asociación entre las variables cualitativas de la tabla de contingencia
-1.000	Indirecta y perfecta
-1 a -0.500	Indirecta y alta
-0.499 a 0	Indirecta y moderada
0	Ninguna. Total independencia
0.00 a 0.499	Directa y moderada
0.500 a 1.00	Directa y alta
1.000	Directa y perfecta

3.6 Applets de GeoGebra segunda fase

Una vez los estudiantes estén familiarizados con los cálculos de los riesgos relativos, el cálculo de Chi-Cuadrado y de Phi de Pearson, tendrán acceso a los siguientes Applets:

- 1) Applet para evaluar riesgos relativos y Phi de Pearson

Permite a los estudiantes adquirir habilidad para calcular riesgos relativos y Phi de Pearson en una tabla de contingencias.

- 2) Applet para calcular riesgos relativos y analizar la asociación

Permite a los estudiantes adquirir habilidad en el cálculo de riesgos relativos, coeficiente Phi de Pearson, y evaluar si los datos de una tabla de contingencia presentan o no asociación y si es así, determinar si la asociación es directa, inversa, alta o moderada.

3) Applet para calcular Chi-Cuadrado y evaluar independencia

Permite a los estudiantes adquirir habilidad para calcular Chi-Cuadrado y evaluar la independencia en una tabla de contingencias.

3.7 Tareas enviadas al Moodle del Colegio

Los estudiantes deberán enviar evidencias, una captura de pantalla, del desarrollo de las diferentes actividades desarrolladas en GeoGebra al Moodle del colegio

3.8 Solución de un caso hipotético

Después de formalizar las medidas de estimación de la asociación, se entregará a los estudiantes un problema que debe ser analizado con lo aprendido y en el cual se pide determinar si existe o no relación entre las variables de la tabla de contingencias, además se solicita cuantificar el grado de asociación utilizando el Phi de Pearson.

EVALUACIÓN SOBRE TABLAS DE CONTINGENCIA

De acuerdo a la siguiente información conteste las preguntas 1 a 3

Se ha realizado un estudio estadístico a quinientas personas, mayores de 25 años, sobre el cuidado de su condición física y su vida social, en relación a episodios de alto nivel de estrés o ansiedad. Las dos preguntas de la encuesta fueron las siguientes:

- 1) ¿Asiste al gimnasio con regularidad (3 a 4 veces por semana), se reúne frecuentemente con sus amigos o conocidos en reuniones sociales, viaja fuera de la ciudad, por lo menos una vez al mes, va al cine?

Si ____ No____

- 2) ¿En los últimos 3 años ha sentido altos niveles de estrés o ansiedad en el trabajo o en casa, en tres o más ocasiones por año? Si ____ No ____

Tabla 22: Problema Evaluación Tablas de Contingencias 2X2

Tabla de Contingencias	B_1	B_2	Totales
A_1	255	78	
A_2	108	59	
Totales			

A_1 = Asiste al gimnasio con regularidad (3 a 4 veces por semana), se reúne con sus amigos en reuniones sociales, viaja fuera de la ciudad los fines de semana o va al cine con frecuencia

A_2 = No asiste al gimnasio con regularidad, menos de dos veces a la semana, se reúne muy ocasionalmente con sus amigos en reuniones sociales, viaja con poca frecuencia fuera de la ciudad los fines de semana, o va al cine en muy pocas ocasiones.

B_1 = En los últimos 3 años no ha sufrido o sentido episodios que le indique que sufre de altos niveles de estrés, o ansiedad en el trabajo o en casa, en tres ocasiones o más por año.

B_2 = En los últimos 3 años ha sufrido o sentido episodios que le indique que sufre de altos niveles de estrés o ansiedad en el trabajo o en casa, en tres ocasiones o más por año.

Contestar:

- 1) ¿Se puede afirmar que sufrir de estrés o ansiedad se relaciona muy estrechamente con el cuidado de la condición física y la vida social?
- 2) ¿Qué tanto más probable es que sufra de estrés la persona que descuida su condición física y su vida social, comparada con una persona preocupada por su condición física y con una vida social más activa?
- 3) ¿Es válido concluir que existe una muy alta independencia entre sufrir episodios de ansiedad y no mantener una vida social activa o ejercitarse con frecuencia?

4. Conclusiones y recomendaciones

4.1 Conclusiones

Las actividades propuestas permitieron a los estudiantes de grado once, del Colegio Saludcoop Sur I.E.D. jornada tarde, familiarizarse con los conceptos de tablas de contingencias, asociación entre variables cuantitativas, probabilidad conjunta, marginal, condicional, e independencia estadística entre dos eventos.

Además, en el desarrollo de la propuesta se diseñaron objetos virtuales de aprendizaje, en GeoGebra, que facilitaron a los estudiantes el análisis de la información registrada en diferentes tablas de contingencia y la elaboración de diferentes representaciones gráficas como son diagramas apilados, adosados y el modelo geométrico, para calcular las probabilidades esperadas de eventos estadísticamente independientes en tablas de contingencia 2×2 .

El modelo geométrico ofrece una primera aproximación a una medida de distanciamiento de los valores teóricos ideales esperados para eventos estadísticamente independientes en relación a valores observados en una tabla de contingencias 2×2 .

En el proceso de análisis de la información tomada para construir las tablas de contingencias, se destacó la importancia que tienen las tablas de contingencias para el estudio estadístico de variables cualitativas.

4.2 Recomendaciones

Para futuras aplicaciones se recomienda:

- Diseñar previamente problemas o situaciones muy simples que puedan ser representadas con facilidad en los cuadernos o en hojas milimetradas pequeñas.
- Tener en cuenta que no todos los estudiantes tiene acceso a un computador todo el tiempo y se deben buscar otras alternativas.
- En lo posible el diseño de los Applets debe considerar pantallas de computador pequeñas, e incluso que funcionen en celulares.
- Tener presente que las primeras situaciones o problemas deben diseñarse de tal forma que las predicciones del Modelo Geométrico coincidan exactamente con los cálculos que se pueden realizar con el método tradicional; posteriormente se presenta a los estudiantes un ejercicio diferente que servirá de pretexto para iniciar conjeturas sobre la relación entre la independencia y el grado de asociación entre las variables categóricas objeto de estudio.

A. Anexo: Cuestionario Evaluación Diagnóstica

A. Anexo: Cuestionario Evaluación Diagnóstica

Se realiza una encuesta a 260 solteros, 400 casados y 340 viudos o divorciados sobre los aspectos que más contribuyen a su felicidad. Los resultados se muestran en la siguiente tabla. Ejercicio adaptado del libro de Estadística Elemental de Freund John E y Simon Gary A., página 361(1992)

Tabla 23: Tabla de Contingencias Evaluación Diagnóstica

	Solteros	Casados	Divorciados o viudos
Amigos y vida social	140	200	160
Trabajo/Estudio	40	120	120
Salud y condición física	80	80	60

Teniendo en cuenta la información de la tabla de contingencias, contestar:

- 1) Si se desea cuantificar la posibilidad de seleccionar una persona que asegure que se siente más feliz cuando está con sus amigos departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?
- 2) Si se desea cuantificar la posibilidad de seleccionar una persona soltera que asegure que se siente más feliz cuando está con sus amigos departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?
- 3) Si se desea cuantificar la posibilidad de seleccionar una persona divorciada o viuda que asegure que se siente más feliz cuando está con sus amigos

departiendo en reuniones sociales, en el estadio o en paseos. ¿El procedimiento sugerido es?

- 4) Si se desea cuantificar la posibilidad o calcular la probabilidad de seleccionar una persona divorciada o viuda que asegure que se siente más feliz cuando está estudiando o está trabajando ¿El procedimiento sugerido es?
- 5) Si se sabe que la persona seleccionada está casada ¿cuál es la probabilidad que se sienta más feliz cuando asiste al gimnasio o practica algún deporte?
- 6) Si se ha escogido una persona que se siente un poco más feliz en su trabajo o estudiando ¿cuál es la probabilidad de que sea soltera?
- 7) Si se ha escogido una persona que se siente más feliz cuando está departiendo con sus amigos en reuniones sociales ¿cuál es la probabilidad de que sea casada?
- 8) Si se sabe que la persona seleccionada piensa que el trabajo o el estudio es lo más importante para lograr su felicidad ¿cuál es la probabilidad de que no sea casada?
- 9) ¿La probabilidad de seleccionar una persona que no prefiera estar con sus amigos en reuniones sociales o en paseos para sentirse feliz, es?
- 10) ¿Cuál es la probabilidad de seleccionar una persona que no prefiera realizar actividad deportiva alguna para sentirse feliz?

B. Anexo: Libro de GeoGebra

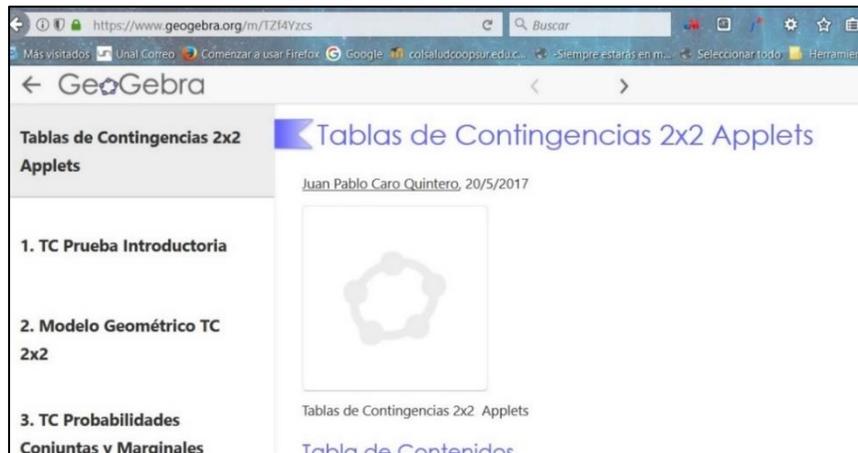


Figura 6: Libro de GeoGebra

Dirección electrónica <https://ggbm.at/TZf4Yzcs>

Prueba Introdutoria TC 2017

Inicio

	Muj	Hom
Mat	1	1
LC	1	1
Dan	1	1
Soc	1	1

Calificación

0.0651

Perfiles por Materia

Perfiles por Sexo

1103 Juan

	Frec. Obs Mat	Frec. Obs Sexo	Frec. Obs General
Mat	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
LC	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
Dan	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
Soc	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$

Completar la tabla de contingencias si:

- 1) El 25% de los estudiantes prefieren la clase de sociales
- 2) Tres de cada cinco estudiantes son mujeres. 3 de cada 8 hombres prefieren la clase de Danzas
- 3) De los estudiantes que prefieren las matemáticas 120 son mujeres
- 4) Un tercio de las mujeres prefiere las clases de Lengua Castellana, y otro tercio prefiere la clase de Sociales
- 5) La cuarta parte de los hombres prefieren la clase de Lengua Castellana

Figura 7: Applet Introdutorio Información de Tablas de Contingencia

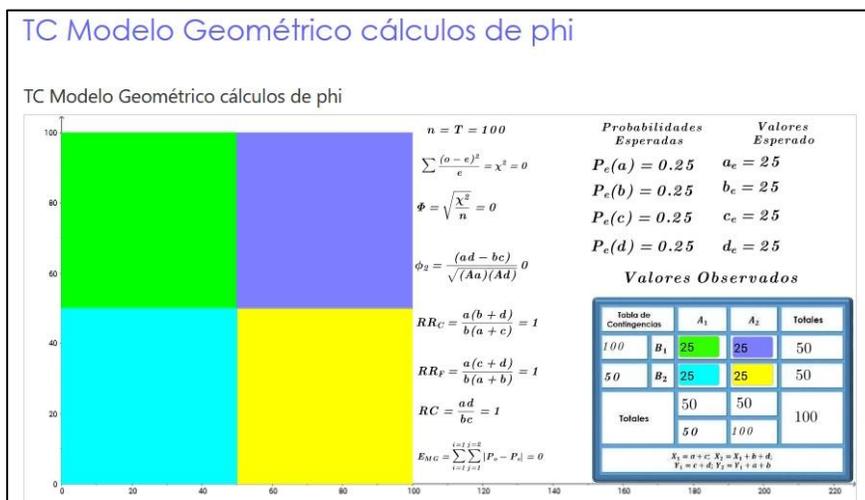


Figura 8: Modelo Geométrico

Capítulos del libro GeoGebra

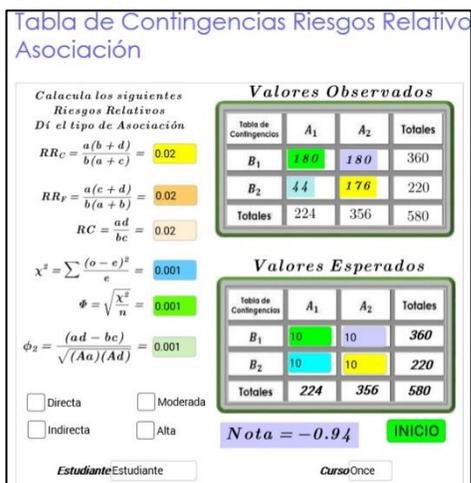


Figura 10: Applet para el cálculo de Riesgos Relativos

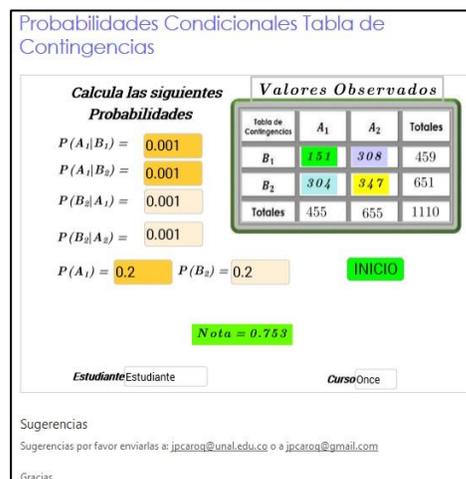


Figura 9: Applet para el cálculo de probabilidad condiciones

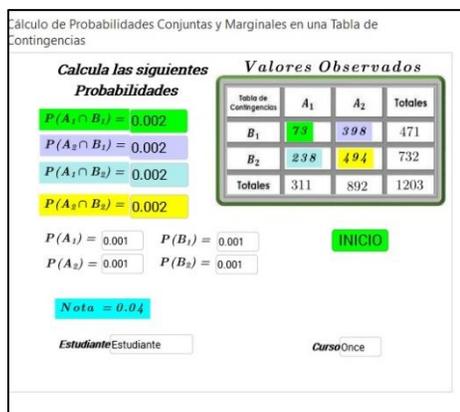


Figura 11: Applet para calcular probabilidades conjuntas y marginales

C. Anexo: Análisis Algebraico del Modelo Geométrico

A continuación se desarrolla un análisis algebraico para determinar la equivalencia entre el modelo geométrico propuesto, y los cálculos de probabilidades tomado directamente la información de la tabla de contingencias, modelo tradicional, con variables categóricas estadísticamente independientes.

Dada la siguiente tabla de contingencias, determinar si es posible establecer alguna relación entre el modelo propuesto y el método tradicional al momento de calcular las probabilidades conjuntas.

Tabla 24: Para el análisis algebraico del modelo geométrico

TC	B	~B	Totales
A	a	b	$a + b$
~A	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d = T$

C1. Primer Procedimiento Propuesto

Tabla 25: Comparativo cálculo de probabilidades procedimiento 1

Cálculo de la Probabilidad Conjunta Método tradicional	Cálculo de la Probabilidad Conjunta Modelo geométrico propuesto
$P(A \cap B) = Pa = \frac{a}{T} = \frac{a}{(a+b+c+d)}$	$P(A \cap B) = Pa = \frac{Aa}{T^2} = \frac{(a+c)(a+b)}{(a+b+c+d)^2}$

Puesto que con los dos modelos, el tradicional y el geométrico, se obtienen los mismos resultados cuando los valores de la tabla de contingencia son independientes, debe ser posible equipararlos.

$$\frac{a}{(a+b+c+d)} = \frac{(a+c)(a+b)}{(a+b+c+d)^2}$$

$$\frac{a(a+b+c+d)^2}{(a+b+c+d)} = (a+c)(a+b)$$

$$a(a+b+c+d) = (a+c)(a+b)$$

$$a^2 + ab + ac + ad = a^2 + ab + ac + bc$$

$$a^2 - a^2 + ab - ab + ac - ac + ad = bc$$

$$ad = bc \quad (57)$$

$$\frac{a}{c} = \frac{b}{d}$$

$$a:c :: b:d$$

C2. Segundo Procedimiento Propuesto

En este procedimiento alternativo se considera que lo que se registra en la tabla de contingencias no son los valores observados, sino sus probabilidades de ocurrencia, de tal modo que al sumarlas se tiene como resultado 1, dicho de otro modo, esta nueva tabla muestra todas las probabilidades conjuntas y marginales de todos los eventos consignados en la tabla original.

Tabla 26: Tabla de contingencias construida a partir de probabilidades

TC	B	~B	Totales
A	a	b	$a + b$
~A	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d = T = 1$

Tabla 27: Comparativo cálculo de probabilidades procedimiento 2

Cálculo de la Probabilidad Conjunta Método tradicional	Cálculo de la Probabilidad Conjunta Modelo geométrico propuesto
$P(A \cap B) = Pa = \frac{a}{T} = \frac{a}{(a+b+c+d)}$	$P(A \cap B) = Pa = \frac{Aa}{T^2} = \frac{(a+c)(a+b)}{(a+b+c+d)^2}$
Se considera que	Se considera que
$T = a + b + c + d = 1$	$T = a + b + c + d = 1$
$P(A \cap B) = a$	$P(A \cap B) = (a + c)(a + b)$

Puesto que con los dos modelos, el tradicional y el geométrico, se obtienen los mismos resultados cuando los valores de la tabla de contingencia son independientes, debe ser posible equipararlos

$$a = (a + c)(a + b)$$

$$a = a^2 + ab + ac + bc$$

$$a = a(a + b + c) + bc$$

Al sustituir $(a + b + c)$ por $(1 - d)$ tenemos:

$$a = a(1 - d) + bc$$

$$a = a - ad + bc$$

$$a - a + ad = bc$$

$$ad = bc$$

$$\frac{a}{c} = \frac{b}{d}$$

$$a:c :: b:d$$

Como con los dos procedimientos hemos logrado los mismos resultados, esto nos sugiere que:

Si se establece que los datos observados presentan la proporción geométrica $a:c :: b:d$, posiblemente estamos frente a un caso en donde la tabla de contingencias muestra datos estadísticamente independientes.

C3. Cálculo de Probabilidades con el Modelo Geométrico y su relación con los Cálculos del Método Tradicional

Para todos los cálculos se debe tener en cuenta que $ad = bc$, puesto que el modelo geométrico funciona con tablas de contingencia cuando sus eventos son estadísticamente independientes.

Probabilidades Marginales

$$P(A) = \frac{Aa+Ab}{T^2} = \frac{(a+b)(a+b)+(b+d)(a+b)}{(a+b+c+d)^2} = \frac{(a+b+c+d)(a+b)}{(a+b+c+d)^2} = \frac{(a+b)}{(a+b+c+d)} = \frac{(a+b)}{T}$$

$$P(\sim A) = \frac{Ac+Ad}{T^2} = \frac{(a+c)(c+d)+(b+d)(c+d)}{(a+b+c+d)^2} = \frac{(a+b+c+d)(c+d)}{(a+b+c+d)^2} = \frac{(c+d)}{(a+b+c+d)} = \frac{(c+d)}{T}$$

$$P(B) = \frac{Aa+Ac}{T^2} = \frac{(a+c)(a+b)+(a+c)(c+d)}{(a+b+c+d)^2} = \frac{(a+c)(a+b+c+d)}{(a+b+c+d)^2} = \frac{(a+c)}{(a+b+c+d)} = \frac{(a+c)}{T};$$

$$P(\sim B) = \frac{Ab+Ad}{T^2} = \frac{(b+d)(a+b)+(b+d)(c+d)}{(a+b+c+d)^2} = \frac{(b+d)(a+b+c+d)}{(a+b+c+d)^2} = \frac{(b+d)}{(a+b+c+d)} = \frac{(b+d)}{T};$$

Probabilidades Condicionales

$$P(A|B) = \frac{Aa}{Aa+Ac} = \frac{(a+c)(a+b)}{(a+c)(a+b)+(a+c)(c+d)} = \frac{(a+c)(a+b)}{(a+b+c+d)(a+c)} = \frac{a^2+ab+ac+bc}{(a+b+c+d)(a+c)}$$

$$P(A|B) = \frac{a^2 + ab + ac + ad}{(a + b + c + d)(a + c)} = \frac{a(a + b + c + d)}{(a + b + c + d)(a + c)} = \frac{a}{(a + c)}$$

$$P(\sim A|B) = \frac{Ac}{Aa+Ac} = \frac{(a+c)(c+d)}{(a+c)(a+b)+(a+c)(c+d)} = \frac{ac+ad+c^2+cd}{(a+c)(a+b+c+d)}$$

$$P(\sim A|B) = \frac{ac+bc+c^2+cd}{(a+c)(a+b+c+d)} = \frac{c(a+b+c+d)}{(a+c)(a+b+c+d)} = \frac{c}{(a+c)}$$

$$P(B|A) = \frac{Aa}{Aa+Ab} = \frac{(a+c)(a+b)}{(a+c)(a+b)+(b+d)(a+b)} = \frac{a^2+ab+ac+bc}{(a+b)(a+b+c+d)}$$

$$P(B|A) = \frac{a^2+ab+ac+ad}{(a+b)(a+b+c+d)} = \frac{a(a+b+c+d)}{(a+b)(a+b+c+d)} = \frac{a}{(a+b)}$$

$$P(\sim B|A) = \frac{Ab}{Aa+Ab} = \frac{(b+d)(a+b)}{(a+c)(a+b)+(b+d)(a+b)} = \frac{ab+bd+ad+bd}{(a+b+c+d)(a+b)}$$

$$P(\sim B|A) = \frac{ab+bd+bc+bd}{(a+b+c+d)(a+b)} = \frac{b(a+b+c+d)}{(a+b+c+d)(a+b)} = \frac{b}{(a+b)}$$

Probabilidades Conjuntas

$$P(A \cap B) = Pa = \frac{Aa}{T^2} = \frac{(a+c)(a+b)}{(a+b+c+d)^2} = \frac{a^2+ab+ac+bc}{(a+b+c+d)^2} = \frac{a^2+ab+ac+ad}{(a+b+c+d)^2}$$

$$P(A \cap B) = \frac{a(a+b+c+d)}{(a+b+c+d)^2} = \frac{a}{(a+b+c+d)} = \frac{a}{T}$$

$$P(A \cap \sim B) = Pb = \frac{Ab}{T^2} = \frac{(b+d)(a+b)}{(a+b+c+d)^2} = \frac{ab+b^2+ad+bd}{(a+b+c+d)^2} = \frac{ab+b^2+bc+bd}{(a+b+c+d)^2}$$

$$P(A \cap \sim B) = \frac{b(a+b+c+d)}{(a+b+c+d)^2} = \frac{b}{(a+b+c+d)} = \frac{b}{T}$$

$$P(\sim A \cap B) = Pc = \frac{Ac}{T^2} = \frac{(a+c)(c+d)}{(a+b+c+d)^2} = \frac{ac+ad+c^2+cd}{(a+b+c+d)^2} = \frac{ac+bc+c^2+cd}{(a+b+c+d)^2}$$

$$P(\sim A \cap B) = \frac{c(a+b+c+d)}{(a+b+c+d)^2} = \frac{c}{(a+b+c+d)} = \frac{c}{T}$$

$$P(\sim A \cap \sim B) = Pd = \frac{Ad}{T^2} = \frac{(b+d)(c+d)}{(a+b+c+d)^2} = \frac{bc+bd+cd+d^2}{(a+b+c+d)^2} = \frac{ad+bd+cd+d^2}{(a+b+c+d)^2}$$

$$P(\sim A \cap \sim B) = \frac{d(a+b+c+d)}{(a+b+c+d)^2} = \frac{d}{(a+b+c+d)} = \frac{d}{T}$$

D. Fórmula simplificada para calcular el coeficiente de Phi de Pearson

Comprobación de la equivalencia entre las siguientes expresiones para calcular el coeficiente Phi de Pearson

$$\Phi = \sqrt{\frac{\chi^2}{n}} \quad \text{y} \quad \Phi = \frac{ad-bc}{\sqrt{(a+c)(a+b)(b+d)(c+d)}} \quad \text{son equivalentes}$$

Para el desarrollo de este procedimiento se sustituye n por T como la suma de las frecuencias absolutas de la tabla de contingencias 2X2

$$n = T = a + b + c + d$$

Tabla de Contingencias, datos observados

	B_1	B_2	Total
A_1	a	b	$a + b$
A_2	c	d	$c + d$
Total	$a + c$	$b + d$	$T = a + b + c + d$

Los valores esperados de la tabla de contingencias se calculan utilizando las siguientes fórmulas:

$$a_e = \frac{(a+c)(a+b)}{T}$$

$$b_e = \frac{(b+d)(a+b)}{T}$$

$$c_e = \frac{(a+c)(c+d)}{T}$$

$$d_e = \frac{(b+d)(c+d)}{T}$$

Cálculo de χ^2

$$\chi^2 = \sum \frac{(o-e)^2}{e} = \frac{(a-a_e)^2}{a_e} + \frac{(b-b_e)^2}{b_e} + \frac{(c-c_e)^2}{c_e} + \frac{(d-d_e)^2}{d_e}$$

$$\frac{(a - a_e)^2}{a_e}$$

$$\frac{(b - b_e)^2}{b_e}$$

$$a_e = \frac{(a + c)(a + b)}{T}$$

$$b_e = \frac{(b + d)(a + b)}{T}$$

$$\frac{\left(a - \frac{(a + c)(a + b)}{T}\right)^2}{\frac{(a + c)(a + b)}{T}}$$

$$\frac{\left(b - \frac{(b + d)(a + b)}{T}\right)^2}{\frac{(b + d)(a + b)}{T}}$$

$$\frac{\left(\frac{aT - (a + c)(a + b)}{T}\right)^2}{\frac{(a + c)(a + b)}{T}}$$

$$\frac{\left(\frac{bT - (b + d)(a + b)}{T}\right)^2}{\frac{(b + d)(a + b)}{T}}$$

$$\frac{\frac{(aT - (a + c)(a + b))^2}{(T)^2}}{\frac{(a + c)(a + b)}{T}}$$

$$\frac{\frac{(bT - (b + d)(a + b))^2}{(T)^2}}{\frac{(b + d)(a + b)}{T}}$$

$$\frac{T(aT - a^2 - ab - ac - bc)^2}{(a + c)(a + b)(T)^2}$$

$$\frac{T(bT - ab - b^2 - ad - bd)^2}{(b + d)(a + b)(T)^2}$$

$$\frac{(a(T - a - b - c) - bc)^2}{(a + c)(a + b)T}$$

$$\frac{(b(T - a - b - d) - ad)^2}{(b + d)(a + b)T}$$

Sustitución

$$(T - a - b - c) = d$$

$$\frac{(ad - bc)^2}{(a + c)(a + b)T}$$

$$\frac{(c - c_e)^2}{c_e}$$

$$c_e = \frac{(a + c)(c + d)}{T}$$

$$\frac{\left(c - \frac{(a + c)(c + d)}{T}\right)^2}{\frac{(a + c)(c + d)}{T}}$$

$$\frac{\left(\frac{cT - (a + c)(c + d)}{T}\right)^2}{\frac{(a + c)(c + d)}{T}}$$

$$\frac{\frac{(cT - (a + c)(c + d))^2}{(T)^2}}{\frac{(a + c)(c + d)}{T}}$$

$$\frac{T(cT - ac - ad - c^2 - cd)^2}{(a + c)(c + d)(T)^2}$$

Sustitución

$$(T - a - b - d) = c$$

$$\frac{(bc - ad)^2}{(b + d)(a + b)T}$$

$$\frac{(d - d_e)^2}{d_e}$$

$$d_e = \frac{(b + d)(c + d)}{T}$$

$$\frac{\left(d - \frac{(b + d)(c + d)}{T}\right)^2}{\frac{(b + d)(c + d)}{T}}$$

$$\frac{\left(\frac{dT - (b + d)(c + d)}{T}\right)^2}{\frac{(b + d)(c + d)}{T}}$$

$$\frac{\frac{(dT - (b + d)(c + d))^2}{(T)^2}}{\frac{(b + d)(c + d)}{T}}$$

$$\frac{T(dT - bc - bd - cd - d^2)^2}{(b + d)(c + d)(T)^2}$$

$$\frac{(c(T - a - c - d) - ad)^2}{(a + c)(c + d)T}$$

Sustitución

$$(T - a - c - d) = b$$

$$\frac{(cb - ad)^2}{(a + c)(c + d)T}$$

$$\frac{(d(T - b - c - d) - bc)^2}{(b + d)(c + d)T}$$

Sustitución

$$(T - b - c - d) = a$$

$$\frac{(ad - bc)^2}{(b + d)(c + d)T}$$

$$\chi^2 = \frac{(ad - bc)^2}{(a + c)(a + b)T} + \frac{(bc - ad)^2}{(b + d)(a + b)T} + \frac{(cb - ad)^2}{(a + c)(a + b)T} + \frac{(ad - bc)^2}{(b + d)(c + d)T}$$

Sustitución

$$(ad - bc)^2 = (bc - ad)^2$$

$$\chi^2 = \frac{(ad - bc)^2}{(a + c)(a + b)T} + \frac{(ad - bc)^2}{(b + d)(a + b)T} + \frac{(ad - bc)^2}{(a + c)(a + b)T} + \frac{(ad - bc)^2}{(b + d)(c + d)T}$$

Factorizando $(ad - bc)^2$ y T , tenemos:

$$\chi^2 = \left(\frac{(ad - bc)^2}{T} \right) \left(\frac{1}{(a + c)(a + b)} + \frac{1}{(b + d)(a + b)} + \frac{1}{(a + c)(c + d)} + \frac{1}{(b + d)(c + d)} \right)$$

Sumando las fracciones utilizando la propiedad asociativa, tenemos:

$$\chi^2 = \left(\frac{(ad - bc)^2}{T} \right) \left(\frac{(b + d) + (a + c)}{(a + c)(a + b)(b + d)} + \frac{(b + d) + (a + c)}{(a + c)(c + d)(b + d)} \right)$$

Sustitución

$$T = a + b + c + d$$

$$\chi^2 = \left(\frac{(ad - bc)^2}{T} \right) \left(\frac{T}{(a+c)(a+b)(b+d)} + \frac{T}{(a+c)(c+d)(b+d)} \right)$$

Se factoriza nuevamente T y simplificando

$$\chi^2 = \left(\frac{T(ad - bc)^2}{T} \right) \left(\frac{1}{(a+c)(a+b)(b+d)} + \frac{1}{(a+c)(c+d)(b+d)} \right)$$

$$\chi^2 = \left(\frac{(ad - bc)^2}{1} \right) \left(\frac{1}{(a+c)(a+b)(b+d)} + \frac{1}{(a+c)(c+d)(b+d)} \right)$$

$$\chi^2 = \left(\frac{(ad - bc)^2}{1} \right) \left(\frac{(c+d) + (a+b)}{(a+c)(a+b)(b+d)(c+d)} \right)$$

Sustitución

$$T = a + b + c + d$$

$$\chi^2 = \left(\frac{(ad - bc)^2}{1} \right) \left(\frac{T}{(a+c)(a+b)(b+d)(c+d)} \right)$$

$$\chi^2 = \left(\frac{T(ad - bc)^2}{(a+c)(a+b)(b+d)(c+d)} \right)$$

$$\Phi = \sqrt{\frac{\chi^2}{T}} = \sqrt{\frac{\frac{T(ad - bc)^2}{(a+c)(a+b)(b+d)(c+d)}}{\frac{T}{1}}} = \sqrt{\frac{T(ad - bc)^2}{T(a+c)(a+b)(b+d)(c+d)}}$$

$$\Phi = \sqrt{\frac{(ad - bc)^2}{(a + c)(a + b)(b + d)(c + d)}} = \frac{ad - bc}{\sqrt{(a + c)(a + b)(b + d)(c + d)}}$$

Utilizando las definiciones de las áreas del modelo geométrico se llega a la siguiente expresión simplificada

$$\Phi = \frac{ad - bc}{\sqrt{(A_a)(A_d)}}$$

E. Actividades registradas en el Moodle del colegio

⟨Evaluación Probabilidades en Tablas de Contingencia **Asociación de variables cualitativas en tablas de contingencias**

Actividades para analizar la asociación entre variables cualitativas en una tabla de contingencias 2X2

-  Asociación de variables cualitativas en tablas de contingencias
-  PROBABILIDADES CONDICIONALES EN UNA TABLA DE CONTINGENCIAS 2X2, E INDEPENDENCIA ENTRE SUS EVENTOS 503.6KB documento PDF
-  Modelo Geométrico Tablas de Contingencia
-  Probabilidades Conjuntas y Marginale en Tablas de Contingencia 2X2
-  Calculo de Probabilidades Condicionales en Tablas de Contingencia
-  Practica cálculo de χ^2 en Tablas de Contingencia 2X2
-  TC Riesgos Relativos, χ^2 y Φ de Pearson
-  Tablas de Contingencia 2X2: Cálculo de los Riesgos Relativos, χ^2 , Φ de Pearson y análisis del tipo de asociación

Figura 12: Tareas en Moodle sobre tablas de contingencia

Bibliografía

- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2008). *Estadística para Administración y Economía*. México: Cengage Learning Editores, S.A.
- Anderson, D. R., Sweeney, D. J., Williams, T., Camm, J. D., & Cochran, J. J. (2015). *Estadística para Negocio y Economía*. México: Cengage Learning.
- Arteaga, P., Batanero, C., Contreras, J. M., & Cañadas, G. (marzo de 2011). Las Tablas y Gráficos Estadísticos como Objetos Culturales. *Numero Revista de Didáctica de las Matemáticas*, 76, 55-67.
- Batanero, C., Contreras, J., Díaz, C., & Roa, R. (2013). El sentido estadístico y su desarrollo. *Números Revista de Didáctica de las Matemática*, 7-18.
- Batanero, C. (2002). *Los Retos de la Cultura Estadística*. Buenos Aires, 2002.
- Camarero Rioja, L. A. (2014). *Acerca de las medidas de asociación en investigación social: un viejo problema que conviene no olvidar*. (U. N. Distancia, Ed.) Obtenido de http://www2.uned.es/dpto-sociologia-l/departamento_sociologia/luis_camarero/asociacion.pdf
- Canavos, G. C. (1988). *Probabilidades y Estadística Aplicaciones y métodos*. Naucalpan de Juárez, México, México: McGraw-Hill Interamericana de México, S.A.
- Cañadas, G. (2010). *Las tablas de contingencia en la formación de profesionales de Psicología*. Universidad de Granada.

- Cañadas, G. (2013). *Comprensión intuitiva y aprendizaje formal de las tablas de contingencia en alumnos de Psicología*. Granada, España: Universidad de Granada. Obtenido de <http://hdl.handle.net/10481/26383>
- Contreras, J. M., Cañadas, G., Gea, M. M., & Arteaga, P. (2012). Comprensión de Frecuencias y Representaciones Asociadas a las Tablas de Contingencia. En F. España Pérez, & M. B. Sépulveda Lucena (Ed.), *XIV Congreso de enseñanza y aprendizaje de las matemáticas "Diversidad y Matemáticas"* (págs. 129-138). Málaga: Sociedad Andaluza de Educación Matemática THALES. Obtenido de <http://thales.cica.es/xivceam/actas/pdf/actas.pdf>
- Contreras, J., Cañadas, G., & Artea, P. (Junio de 2013). Problemática y recursos en la interpretación de las tablas. *Unión Revista Iberoamericana de Educación Matemática*, 86 - 95.
- Díaz, C., & De la Fuente, E. (Diciembre de 2005). Conflictos semióticos en el cálculo de. *Biaix* 24, 85-91.
- Duval, R. (1999). *Semiosis y Pensamiento Humano*. Cali: Universidad de Valle Instituto de Educación y Pedagogía Grupo de Educación Matemática.
- Freund, J. E., & Simon, G. A. (1992). *Estadística Elemental*. Naucalpan de Juárez, México, México: Pearson Prentice Hall.
- Godino, J. D. (2003). *Teoría de las Funciones Semióticas*. Granada: Universidad de Granada.
- Goldino, G., & Janvier, C. (1998). Representacion and the psychology of. *Journal of Mathematics Behaviour*, 1-4.
- Otero, J. V., & Medina Mora, E. (2005). Análisis de datos cualitativos. Obtenido de https://www.uam.es/personal_pdi/economicas/eva/pdf/tab_conting.pdf
- Sanchez Rivero, M. (1998). *Modelización Estadística de Tablas de Contingencia: Aplicación Al Análisis de la Demanda Turística Española*. BADAJOZ,, España: Universidad de Extremadura.

Spiegel, M. R., & Stephens, L. J. (2009). *Estadística*. México: McGraw-Hill Companies, Inc.

Wild, C., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical*, 67(3), 221-248. Obtenido de <http://www.jstor.org/stable/1403699> .