UNIVERSIDAD NACIONAL EXPERIMENTAL DEL TÁCHIRA
VICE-RECTORADO ACADÉMICO
DECANATO DE DOCENCIA
DEPARTAMENTO DE MATEMÁTICA Y FÍSICA
UNIDAD CURRICULAR ESTADÍSTICA APLICADA A LA PSICOLOGÍA

TEMA 7. RELACIÓN DE VARIABLES CUALITATIVAS Y CUANTITATIVAS

En la investigación psicológica frecuentemente se miden numerosas variables en los individuos incluidos en el estudio. Muchas veces interesa determinar si existe relación entre algunas de esas variables (cualitativas o cuantitativas), o predecir el valor de una de ellas conociendo el valor de otras (regresión). En ocasiones interesa determinar si distintos instrumentos, métodos o personas obtienen valores similares cuando se mide una variable en las mismas unidades experimentales. Esos tres objetivos requieren métodos de análisis estadísticos distintos.

Dos variables se relacionan entre sí cuando el cambio de una categoría a otra (para variables cualitativas) o de un valor a otro (para variables cuantitativas), provoca una modificación en las categorías o valores de la otra variable; en caso contrario se dirán independientes. Para ello, nos apoyamos en un sistema de sistema de hipótesis, denominadas hipótesis nula (Ho) e hipótesis alternativa (Hi o Ha).

Así que a través del análisis estadístico se identifica una posible relación entre variables y según el tipo de variable puede llamarse asociación o correlación. Suele llamarse asociación cuando se busca la relación entre variables categóricas, de manera que sus categorías puedan colocarse en tablas de contingencia, y correlación cuando se relacionan variables en escala de intervalo o de razón o incluso, en escala ordinal y en vez de analizarla por tablas de contingencia, se les asignan rangos a las categorías convirtiéndose en numéricas.

Recordemos que, de acuerdo al tipo de variable, se desprende un determinado nivel de medición, a saber:

Ahora bien, estas relaciones entre las variables pueden ser:

1. Relaciones simples: entre dos o tres variables aleatorias en la cual se establece:

- Relación bivariada: relación entre dos variables (X, Y) donde X y Y son medidas en cualquier escala.
- Relación bivariada con influencia de una tercera variable que puede ser:
- Relación transicional: cuando la variable Y, dependiente en una relación bivariada (X, Y) es la independiente respecto a otra relación bivariada (Y, Z) [X→Y→Z].
- Interferencia de una o varias variables intervinientes: cuando una tercera variable (Z) provoca cambios importantes en la relación (X, Y).
- Correlación parcial: correlación entre dos variables eliminando los efectos de otra variable interviniente.

2. Relaciones múltiples: cuando se correlacionan múltiples variables aleatorias, generalmente existe una variable aleatoria dependiente y un vector aleatorio independiente, formado por las covariables e incluye variantes como:

- Correlación semiparcial: que permite conocer las contribuciones de las covariables sobre la variable dependiente eliminando el efecto de una o más covariables, según convenga.
- Correlaciones canónicas: cuando se correlacionan dos vectores aleatorios.

En una relación entre variables:

- Examinarse la intensidad por el valor del estadígrafo y el sentido por su signo.
- Conocer la significación de la relación, para ello siempre se deben realizar contrastes de hipótesis. En el caso de las variables cuantitativas se examina, además, la forma en que se relacionan, la cual puede observarse gráficamente en un plano cartesiano (diagrama de dispersión) o puede ser expresada matemáticamente. Esta expresión es importante para explicar e incluso, predecir lo que pasará en la variable dependiente en función de los cambios en la independiente.

Existen tablas o baremos que diversos autores han propuesto para la interpretación del coeficiente de asociación o correlación, tal como se menciona a continuación.

Baremo de interpretación del coeficiente de asociación o correlación

Valor	Significado
-1	Correlación negativa grande y perfecta
-0,9 a -0,99	Correlación negativa muy alta
-0,7 a -0,89	Correlación negativa alta
-0,4 a -0,69	Correlación negativa moderada
-0,2 a -0,39	Correlación negativa baja
-0,01 a -0,19	Correlación negativa muy baja
0	Correlación nula
0,01 a 0,19	Correlación positiva muy baja
0,2 a 0,39	Correlación positiva baja
0,4 a 0,69	Correlación positiva moderada
0,7 a 0,89	Correlación positiva alta

0,9 a 0,99	Correlación positiva muy alta
1	Correlación positiva grande y perfecta

Ahora bien, existe una gran variedad de coeficientes para corroborar la significación de una relación encontrada. A continuación, se describe la relación entre variables cualitativas, relación entre variables ordinales, relación entre variables cuantitativas y relación entre una variable continua y nominal.

7.1.1. Relación entre Variables Cualitativas

Para identificar la posible relación entre dos variables cualitativas hay que observar si la distribución de las categorías de una de las variables difiere en función de las de la otra, es decir, comparar las distribuciones condicionadas de una de las dos variables agrupadas en función de los valores de la otra. Si no hay relación entre las variables estas distribuciones deberían ser iguales.

Habitualmente se colocan los datos en una tabla de contingencia o de doble entrada, donde aparecen las frecuencias observadas (frecuencias absolutas conjuntas o número de casos que presentan simultáneamente las modalidades fila y columna) y se emplean métodos directos como el análisis de los residuos de la diferencia entre valores observados y esperados o la descomposición de la tabla en tablas de 2 × 2.

Tablas de Contingencia

Cuando se desea estudiar la relación o asociación entre dos variables cualitativas se utilizan las denominadas tablas de contingencia o distribución de frecuencias compuestas. Estas consisten en una tabla de doble de entrada, de forma que una variable ocupa las filas y la otra variable ocupa las columnas, y en cada casilla se incluye la frecuencia conjunta de ambas variables. Ejemplo:

		Variable 1		
		A	**B**	**Total**
Variable 2	**C**	n_{11}	n_{12}	$n_{11} + n_{12}$
	D	n_{21}	n_{22}	$N_{21} + n_{22}$
	Total	$n_{11} + n_{21}$	$n_{12} + n_{22}$	$n_{11} + n_{12} + n_{21} + n_{22}$

	Frecuencias conjuntas
	Frecuencias marginales

Así que, que las tablas de contingencia tienen el objetivo de representar en un resumen, la relación entre diferentes variables categóricas.

Estas tablas de contingencia al tener forma de matrices se describen según el orden 2 x 2, 2 x 3, 3 x 2, etc. El primer número nos indica la fila y el segundo número a la columna.

Inclinación profesional / Género	Lenguas Extranjeras	Licenciatura en Matemáticas	Enfermería	Total
Masculino (M)	14%	30%	10%	54%
Femenino (F)	16%	10%	20%	46%
Total	30%	40%	30%	100%

Tabla de contingencia 2 x 3

	Método comparativo		
Método candidato (Test)	*Positivo*	*Negativo*	*Total*
Positivo	a	b	a + b
Negativo	c	d	c + d
Total	a + c	b + d	n

Tabla de contingencia 2 x 2

Objetivos de una tabla de contingencia

1. Medir la interacción entre dos variables para conocer una serie de información de gran utilidad para comprender con mayor claridad los resultados de una investigación.
2. Mostrar cómo fue las respuestas de los encuestados que respondieron en ambas preguntas.
3. Ordenar la información recolectada para un estudio cuando los datos se encuentran divididos de forma bidimensional, esto significa a que se relaciona con dos factores cualitativos.
4. Analizar si hay una relación entre las variables cualitativas, ya sean dependientes o independientes.

Ventajas de realizar una tabla de contingencia

1. Facilita la lectura de los datos recolectados, ya que permite agruparlos cuando aún se encuentran sin procesar, lo que disminuye el margen de error al realizar un informe de investigación.
2. Es posible realizar gráficas que permitan visualizar la información fácilmente para su comprensión.
3. Permite ahorrar tiempo durante la correlación de variables.
4. Las tablas ofrecen resultados claros y precisos que permiten tomar mejores decisiones y crear estrategias basadas en datos.

Cuando usar una tabla de contingencia

La tabla de contingencia generalmente se realiza en datos categóricos, es decir que se pueden dividir en grupos mutuamente excluyentes (no pueden pertenecer a más de una categoría).

Un ejemplo de datos categóricos es el nivel de autoestima de una persona: alto, medio, bajo; los estilos de aprendizaje: auditivo, visual, kinestésico.

Uno de los principales usos de una tabla de contingencia es analizar la relación que existe entre los datos, las cuales no son fáciles de identificar. Es muy útil para estudiar las probabilidades condicionales entre dos eventos.

Importancia de hacer uso de una tabla de contingencia

La tabla ofrece un método simple de agrupar variables, que minimiza el potencial de confusión o error al proporcionar resultados claros. Además, una tabla puede ayudarnos a obtener grandes conocimientos de los datos sin procesar. Estas ideas no son fáciles de ver cuando los datos sin formato se organizan como una tabla.

Dado que la tabla de contingencia traza claramente las relaciones entre las preguntas categóricas, los investigadores pueden obtener información más profunda, que de otro modo se habría pasado por alto o habría tomado mucho tiempo descifrar de formas más complicadas de análisis estadístico.

La tabla de contingencia facilita la interpretación de los datos, lo cual es beneficioso para los investigadores que tienen un conocimiento limitado del análisis estadístico. Las personas no necesitan programación estadística para correlacionar variables categóricas.

La claridad que ofrece una tabla ayuda a los profesionales a evaluar su trabajo actual y trazar estrategias futuras.

Las tablas de contingencia permiten apreciar las relaciones que existe entre ambas variables, mediante la denominada coeficiente de contingencia (C). Se aplica cuando las tablas son diferentes de 2 x 2, para este tipo se recurre al coeficiente Phi (ϕ), Kappa o Cramer.

Coeficiente de Contingencia (C)

El coeficiente de contingencia C es una medida del grado de asociación o relación entre dos conjuntos de atributos. Es especialmente útil cuando hay una información clasificatoria (escala nominal) acerca de uno o ambos conjuntos de atributos.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \qquad 0 \le C \le 1 \quad \begin{cases} C = 0 \;\mapsto\; \text{Independencia} \\ C = 1 \;\mapsto\; \text{Asociación perfecta} \end{cases}$$

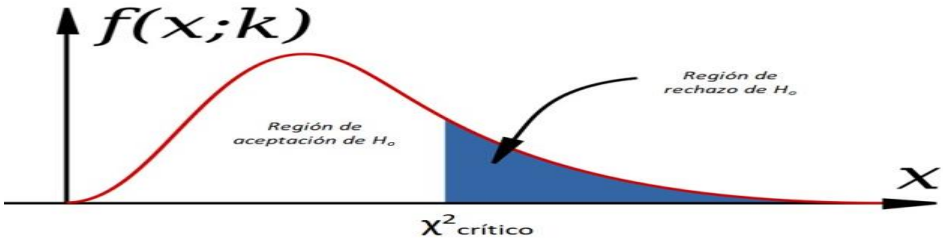
Cuanto mayor sea el valor del índice, mayor es el grado de asociación entre las variables. Su cálculo exige la obtención previa del índice conocido como ji-cuadrado (o chi-cuadrado) que se representa por X^2). El índice X^2 mide la discrepancia entre las frecuencias observadas (fo) y las frecuencias esperadas (fe) en el caso de que las variables fueran independientes. En este caso, se plantea un sistema de hipótesis:

Ho (Hipótesis nula): Las variables son independientes

Hi (Hipótesis alternativa): Existe una relación de dependencia

Para aceptar la Ho, el X^2 calculado debe ser menor al $X^2_{\alpha,gl}$, de no ser así, se rechaza la Ho.

En la siguiente grafica se aprecia la zona critica ($X^2_{\alpha,gl}$) en la cual se acepta o rechaza la Ho, todo dependerá donde caiga el valor de X^2 calculado. Por ejemplo, si el X^2 calculado cae en la zona sombreada se rechaza Ho, de lo contrario, se acepta Ho.



Para calcular el X^2 se utiliza la siguiente fórmula:

$$X^2 = \sum\sum \frac{(fo - fe)^2}{fe}$$

Siendo $\boldsymbol{fo}$ las frecuencias observadas y $\boldsymbol{fe}$ las frecuencias esperadas

Ejemplo. Supongamos que se quiere estudiar la comorbilidad, es decir, la tendencia de los trastornos a presentarse conjuntamente. Para ello se analizan 400 casos clínicos y se anotan, por un lado, cuál es el trastorno principal diagnosticado a cada uno y, por otro, si ese trastorno se presenta junto con un trastorno afectivo. La distribución de frecuencias conjuntas de las variables trastorno principal y presencia/ausencia de trastornos afectivos se muestra en la siguiente tabla.

Distribución de frecuencias conjuntas de las variables trastorno principal y presencia comórbida de trastorno afectivo

		Trastorno principal				
		Esquizofrenia	Alcoholismo	Trastorno de alimentación	Trastorno de personalidad	Trastorno obsesivo
Trastorno afectivo	Presencia	50	78	32	48	20
	Ausencia	50	42	48	12	20

Calcular:

a) El tipo de asociación entre las dos variables.

b) Demostrar si las variables son independientes.

Solución

Primer paso. Se calculan las frecuencias marginales. Para ello sumamos las frecuencias observadas horizontales y verticales, como se muestra en la tabla (en color rojo)

		Trastorno principal					
		Esquizofrenia	Alcoholismo	Trastorno de alimentación	Trastorno de personalidad	Trastorno obsesivo	**Total**
Trastorno afectivo	Presencia	50	78	32	48	20	**228**
	Ausencia	50	42	48	12	20	**172**
	Total	**100**	**120**	**80**	**60**	**40**	**400**

Segundo paso. Se calculan las frecuencias esperadas de cada celda (frecuencias conjuntas)

$$fe = \frac{f_{mc} * f_{mf}}{N}$$

Siendo f_{mc} = frecuencia marginal de la columna, f_{mf} = frecuencia marginal de la fila, N = total de datos.

Así, por ejemplo, se multiplica la frecuencia marginal de la columna total (esquizofrenia) por la frecuencia marginal de la fila total (presencia) y se divide por 400

$$fe = \frac{100 * 228}{400} = 57$$

Se multiplica la frecuencia marginal de la columna (alcoholismo) por la frecuencia marginal de la fila (presencia) y se divide por 400

$$fe = \frac{120 * 228}{400} = 78$$

Así sucesivamente, hasta completar el llenado de las frecuencias esperadas.

		Trastorno principal										
		Esquizofrenia		Alcoholismo		Trastorno de alimentación		Trastorno de personalidad		Trastorno obsesivo		Total
		fo	fe	fo	fe	fo	fe	Fo	fe	fo	fe	
Trastorno afectivo	Presencia	50	57	78	68,4	32	45,6	48	34,2	20	22,8	228
	Ausencia	50	43	42	51,6	48	34,4	12	25,8	20	17,2	172
	Total	100	100	120	120	80	80	60	60	40	40	400

Tercer paso. Se aplica la fórmula:

$$X^2 = \sum\sum \frac{(fo - fe)^2}{fe}$$

$$X^2 = \frac{(50-57)^2}{57} + \frac{(78-68{,}4)^2}{68{,}4} + \frac{(32-45{,}6)^2}{45{,}5} + \cdots + \frac{(20-17{,}2)^2}{17{,}2}$$

$$X^2 = 0{,}86 + 1{,}35 + 4{,}06 + 5{,}57 + 0{,}34 + 1{,}14 + 1{,}79 + 5{,}38 + 7{,}38 + 0{,}46 = 28{,}3$$

$$\boldsymbol{X^2 = 28,3}$$

Cuarto paso. Cálculo del coeficiente de contingencia

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

$$C = \sqrt{\frac{28{,}3}{28{,}3 + 400}} = 0{,}26$$

Quinto paso. Interpretamos el coeficiente de contingencia obtenido. Como el C = 0,26 revela que existe una asociación débil entre el trastorno afectivo y el trastorno principal entre los 400 pacientes atendidos.

Sexto paso. Verificamos si aceptamos o no la hipótesis nula (Ho) para un nivel de significancia dado. Para ello, se emplea la prueba no paramétrica X^2 para demostrar la significación de la asociación entre las variables estudiadas, recurriendo al siguiente sistema de hipótesis:

Ho: No existe asociación entre el trastorno afectivo y el trastorno principal

Hi: Existe asociación entre el trastorno afectivo y el trastorno principal

Alfa (α) = 0,05

Para ello, utilizamos la tabla de distribución de X^2.

Primero, calculamos los grados de libertad (ν)

$$gl = (F - 1)\, x\, (C - 1) = (2 - 1)\, x\, (5 - 1) = 1\, x\, 4 = 4$$

Segundo, en la tabla de distribución de X^2 buscamos el valor de probabilidad de X2 para un nivel de significancia de 0,05 y un grado de libertad de 4

	α									
ν	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515

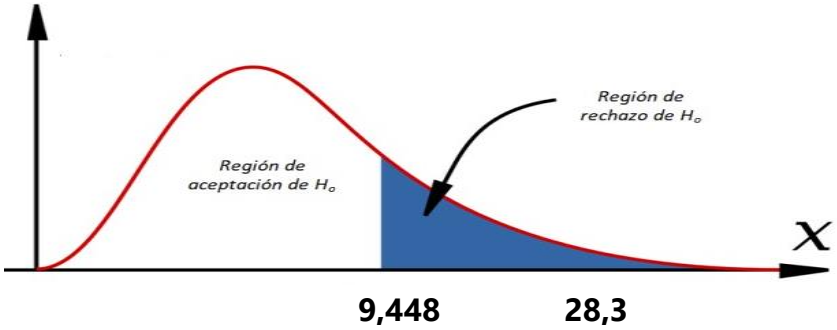
De acuerdo con la tabla se tiene un $X^2_{0,05,4} = 9,448$

Haciendo uso del Excel: **=INV.CHICUAD.CD(Probabilidad;Grados_de_libertad)**

=INV.CHICUAD.CD(0,05;4) = 9,448

Como el X^2 calculado (28,3) es mayor al X^2 critico (9,448) se rechaza Ho, es decir, no existe asociación entre el trastorno afectivo y el trastorno principal para un nivel de significancia del 5% o 0,05.

Veámoslo en una gráfica



Observe que el valor calculado de la X^2 cayó en la zona de región de rechazo de Ho.

Limitaciones de la prueba X^2

1. Para que sea válido se debe tener un tamaño muestral mayor de 40.

2. Si se tienen menos de 20 muestras el test no se puede aplicar.

3. Si no es posible obtener el número de muestras necesarias, y este número está entre 20 y 40, se tiene que cumplir la condición de que en todas las casillas los valores esperados deben ser superiores a 5.

Si algún valor esperado es demasiado pequeño, el sumando correspondiente a esa casilla, que es su contribución a la $X^2_{esperado}$, puede tener un valor desproporcionadamente grande; por ejemplo, supongamos que tenemos un valor observado de 110 y su valor esperado es de 100, su contribución a la $X^2_{esperado}$ será: $(110-100)^2/100=1$, mientras que en otro caso, el valor observado fuese de 11 y el valor esperado fuese de 1 su contribución a la $X^2_{esperado}$ sería $(11-1)^2/1=100$, a pesar de que en ambos casos la diferencia entre valores observados y esperados es la misma, la contribución a la X^2 es diez veces superior. Esta una de las razones para imponer el criterio restrictivo a valores bajos.

Actividad de autoevaluación

1. En un estudio se investigó la relación entre la ansiedad de ejecución (Baja, Media y Alta) y la realización correcta de una tarea (Sí, No) en 40 personas, obteniéndose los siguientes resultados

	Tarea realizada	
Ansiedad	Si	No
Baja	10	7
Media	15	18
Alta	12	8

(a) Determine el coeficiente de contingencia e interprete y (b) Existe asociación entre las variables involucradas en la investigación para un alfa del 10%.

2. En una investigación se estudió la aceptación o no del tratamiento psicológico por parte de pacientes que presentaban dos tipos de trastornos psicológicos. En la tabla se muestran los resultados del estudio:

Sujeto	**Aceptación del tratamiento**	**Trastorno psicológico**
1	Si	Depresión
2	Si	Depresión
3	No	Ansiedad
4	Si	Depresión
5	Si	Trastorno de personalidad
6	Si	Ansiedad
7	No	Trastorno de personalidad
8	No	Ansiedad
9	No	Depresión
10	Si	Trastorno de personalidad
11	Si	Depresión
12	Si	Trastorno de personalidad
13	No	Depresión
14	Si	Depresión
15	No	Trastorno de personalidad
16	Si	Trastorno de personalidad
17	No	Depresión
18	Si	Ansiedad
19	No	Depresión
20	Si	Trastorno de personalidad
21	Si	Depresión
22	Si	Depresión
23	No	Trastorno de personalidad
24	No	Depresión
25	Si	Trastorno de personalidad
26	Si	Depresión

27	No	Ansiedad
28	Si	Trastorno de personalidad

(a) Construya la tabla de contingencia; (b) determine el coeficiente de contingencia e interprete y (c) existe asociación entre las variables involucradas en la investigación.

3. Establezca un ejemplo cuya tabla de contingencia sea de la forma 3 x 2. Determine el coeficiente de contingencia e indique si existe o no asociación entre las variables analizadas.

Coeficiente de Correlación Phi (φ)

Se utiliza cuando las dos variables cualitativas nominales son dicotómicas o binarias: Verdadero o Falso, Si o No. Por tanto, se trata de cuadros de contingencia de 2 x 2, tal como se muestra a continuación.

		Y		
		$Y_1 = 1$	$Y_2 = 0$	**Total**
X	$X_1 = 1$	$n_{1,1}$	$n_{1,0}$	$\boldsymbol{n_{1,1} + n_{1,0}}$
	$X_2 = 0$	$n_{0,1}$	$n_{0,0}$	$\boldsymbol{n_{0,1} + n_{0,0}}$
Total		$\boldsymbol{n_{1,1} + n_{0,1}}$	$\boldsymbol{n_{1,0} + n_{0,0}}$	**N**

$n_{1,1}$ = verdaderos positivos $n_{1,0}$ = falsos positivos

$n_{0,1}$ = falsos negativos $n_{0,0}$ = verdaderos negativos.

1 = variable presente 0 = variable ausente

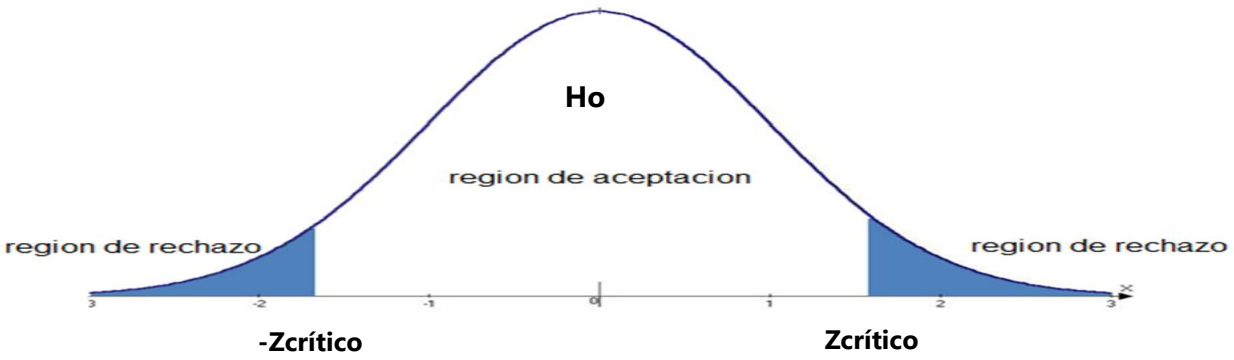
La fórmula del coeficiente Phi es:

$$\varphi = \frac{n_{1,1} * n_{0,0} * -n_{1,0} + n_{0,1}}{\sqrt{(n_{1,1} + n_{1,0}) * (n_{0,1} + n_{0,0}) * (n_{1,1} + n_{0,1}) * (n_{1,0} + n_{0,0})}} \quad -1 \leq \varphi \leq +1$$

Prueba de significación

Ho: No existe asociación entre las variables

Hi: Existe asociación entre las variables



Ejemplo. Supongamos que se realiza un estudio entre un grupo de hombres (1) y mujeres (0) para conocer cómo se relaciona con el fumar (1) o no fumar (0)

		Fumador	
		Fuma = 1	No Fuma = 0
Género	Hombres = 1	100	50
	Mujeres = 0	25	80

Determinar:

a) Grado de asociación entre las variables

b) Existe asociación entre las variables para un alfa de 0,05

Solución

a) Grado de asociación entre las variables

		Fumador		
		Fuma = 1	No Fuma = 0	
Género	Hombres = 1	100	50	150
	Mujeres = 0	25	80	105
		130	125	255

$$\varphi = \frac{100 * 80 - 50 * 25}{\sqrt{(150) * (105) * (130) * (125)}} = \frac{6750}{15998{,}05} = 0{,}422$$

Para la interpretación de coeficiente Phi se utiliza el siguiente baremo

El valor 0,422 indica que existe una asociación moderada entre las variables con sentido positivo. En el ejemplo, dentro de los hombres hay más fumadores que no fumadores, mientras que entre las mujeres hay más no fumadoras que fumadoras. Por tanto, concluimos que hay una tendencia moderada de asociación, que consiste en una mayor presencia del tabaquismo entre los hombres.

b) Existe asociación entre las variables para un alfa de 0,05

Ho: No existe asociación entre el género y el fumar

Hi: Existe asociación entre el género y el fumar

Alfa = 5% o 0,05

Para contrastar las hipótesis se recurre a la distribución normal:

Primero, calculamos el Zc

$$Zc = \sqrt{n}\varphi$$

$$Zc = \sqrt{255}\, x\, 0{,}422 = 6{,}74$$

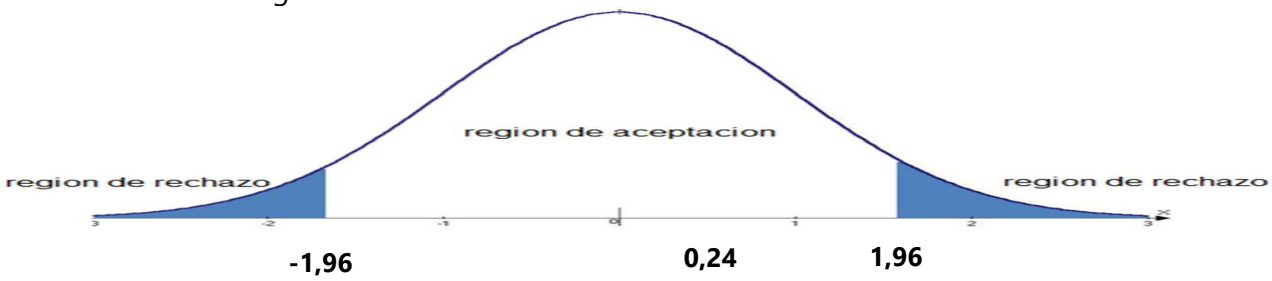
Segundo, para un α = 0,05 el $Z_{1-\alpha/2} = Z_{1-0{,}025} = Z_{0{,}975} = 1{,}96$

El valor de 1,96 se busca en la tabla de distribución normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Tercero, como el Zcalculado (0,422) es menor a 1,96 se acepta Ho. Es decir, no existe asociación entre el género y el fumar para un nivel de significación del 5%.

Veámoslo en la gráfica:



Prueba diagnóstica. A través de la tabla 2 x 2 se pueden calcular los atributos de la prueba diagnóstica, como sigue:

1. Sensibilidad: Probabilidad de que un enfermo sea identificado correctamente por la prueba, es decir, que tenga una prueba positiva. Son los enfermos con prueba positiva de entre todos los enfermos. Se obtiene como sigue:

$$\text{Sensibilidad} = \frac{n_{1,1}}{n_{1,1} + n_{0,1}}$$

2. Especificidad: Probabilidad de que un individuo sin la enfermedad sea identificado correctamente por la prueba, es decir, que tenga una prueba negativa. Son los sanos con prueba negativa de entre todos los sanos. Se obtiene así

$$\text{Especificidad} = \frac{n_{0,0}}{n_{1,0} + n_{0,0}}.$$

3. Valor predictivo positivo: Se enuncia como la capacidad que tiene una prueba, cuando es positiva, de predecir que el paciente tiene la enfermedad y se puede estimar dividiendo a los verdaderos positivos entre los verdaderos y falsos positivos:

$$\text{VPP} = \frac{n_{1,1}}{n_{1,1} + n_{1,0}}$$

4. Valor predictivo negativo: Es la capacidad de una prueba diagnóstica, cuando es negativa, de predecir que el paciente no tiene la enfermedad y se estima dividiendo a los verdaderos negativos entre los falsos y verdaderos negativos:

$$\text{VPN} = \frac{n_{0,0}}{n_{0,1} + n_{0,0}}$$

5. Prevalencia: Es el índice de individuos que padecen una cierta enfermedad dentro del total de un grupo de personas en estudio:

$$\text{Prevalencia} = \frac{n_{1,1} + n_{0,1}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}}$$

6. Exactitud o eficiencia de una prueba diagnóstica utiliza todos los valores de la tabla 2 X 2 y se obtiene dividiendo la suma de los verdaderos positivos con los verdaderos negativos entre la suma de todos los valores, de la siguiente manera:

$$\text{Exactitud} = \frac{n_{1,1} + n_{0,0}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}}$$

		Enfermedad			
		Presente	**Ausente**		
Prueba diagnóstica	+	$n_{1,1}$	$n_{1,0}$	$\boldsymbol{n_{1,1} + n_{1,0}}$	**VPP** $\frac{n_{1,1}}{n_{1,1} + n_{1,0}}$
	-	$n_{0,1}$	$n_{0,0}$	$\boldsymbol{n_{0,1} + n_{0,0}}$	**VPN** $\frac{n_{0,0}}{n_{0,1} + n_{0,0}}$
Total		$\boldsymbol{n_{1,1} + n_{0,1}}$	$\boldsymbol{n_{1,0} + n_{0,0}}$	**N**	**Prevalencia** $\frac{n_{1,1} + n_{0,0}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}}$
		Sensibilidad $\frac{n_{1,1}}{n_{1,1} + n_{0,1}}$	**Especificidad** $\frac{n_{0,0}}{n_{1,0} + n_{0,0}}$		**Exactitud** $\frac{n_{1,1} + n_{0,0}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}}$

Nota. Tanto la sensibilidad como la especificidad proporcionan información acerca de la probabilidad de obtener un resultado concreto (positivo o negativo) en función de la verdadera condición del enfermo con respecto a la enfermedad.

- Verdadero positivo: el paciente tiene la enfermedad y el test es positivo
- Falso positivo: el paciente no tiene la enfermedad, pero el resultado del test es positivo
- Verdadero negativo: el paciente no tiene la enfermedad y el test es negativo
- Falso negativo: el paciente tiene la enfermedad, pero el resultado del test es negativo

Parámetro	Fórmula	Definición
Sensibilidad	a/(a+c)	Proporción de pacientes con la enfermedad que tendrán test positivo
Especificidad	d/(b+d)	Proporción de pacientes sin la enfermedad que tendrán test negativo
Valor predictivo positivo	a/(a+b)	Probabilidad de que el paciente tenga la enfermedad dado que el test es positivo
Valor predictivo negativo	d/(c+d)	Probabilidad de que el paciente no tenga la enfermedad dado que el test es negativo
Likelihood ratio (+)	sensibilidad/(1-especificidad)	Describe cuántas veces es más probable que reciba un resultado determinado una persona con la enfermedad que una persona sin la enfermedad
Likelihood ratio (-)	(1-sensibilidad)/especificidad	
Exactitud	(a+d) / (a+b+c+d)	La probabilidad de que el resultado del test prediga correctamente la presencia o ausencia de la enfermedad
Odds ratio diagnóstico	(a/c) / (b/d)	Razón entre la odds de estar enfermo si la prueba da positivo y la odds de no estar enfermo si la prueba da negativo

Ejemplo. El Instituto de Investigaciones Psicológica pretende validar una nueva prueba, más simple que las tradicionales, para el diagnóstico de trastorno de depresión de los pacientes atendidos. En una muestra de 500 pacientes atendidos por el Instituto se administran dos pruebas (tradicional y la nueva) para el diagnóstico de depresión. Los resultados se muestran en la siguiente tabla:

		Diagnóstico tradicional de depresión		
		Positivo	**Negativo**	**Total**
Versión nueva de la escala de depresión	**Positivo**	125	50	175
	Negativo	25	300	325
	Total	150	350	500

$$\text{Sensibilidad} = \frac{n_{1,1}}{n_{1,1} + n_{0,1}} = \frac{125}{150} = 0{,}833 = 83{,}3\%$$

Existe un 83,3% de probabilidad de que una persona con depresión sea identificada correctamente por la nueva versión de la escala de depresión.

$$\text{Especificidad} = \frac{n_{0,0}}{n_{1,0} + n_{0,0}} = \frac{300}{350} = 0{,}857 = 85{,}7\%.$$

Existe un 85,7% de probabilidad de que una persona sin depresión sea identificada correctamente por la nueva versión de la escala de depresión.

$$\text{VPP} = \frac{n_{1,1}}{n_{1,1} + n_{1,0}} = \frac{125}{175} = 0{,}714 = 71{,}4\%$$

Existe un 71,4% de probabilidad de que la nueva versión de la escala de depresión logre predecir que una persona tiene depresión.

$$\text{VPN} = \frac{n_{0,0}}{n_{0,1} + n_{0,0}} = \frac{300}{325} = 0{,}923 = 92{,}3\%$$

Existe un 93,3% de probabilidad de que la nueva versión de la escala de depresión logre predecir que una persona tiene depresión.

$$\text{Prevalencia} = \frac{n_{1,1} + n_{0,1}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}} = \frac{150}{500} = 0{,}30 = 30\%$$

Existe un 30% de probabilidad de que la persona sea depresiva.

$$\text{Exactitud} = \frac{n_{1,1} + n_{0,0}}{n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}} = \frac{425}{500} = 0{,}85 = 85\%$$

Existe un 85% de probabilidad de que la nueva versión de la escala de depresión sea efectiva para predecir correctamente la presencia o ausencia de personas depresivas.

Uso de la tabla 2x2 en epidemiologia

En epidemiología, los estudios de cohortes basados en una tabla de 2x2 son los más sencillos para utilizar el contraste de hipótesis Chi-cuadrado. En estos estudios partimos de dos grupos de población, el primer grupo formado por los individuos expuestos a un factor, y el segundo grupo formado por individuos no expuestos a ese mismo factor; el objetivo del estudio es conocer si esta exposición favorece o perjudica la aparición de un evento

En los estudios de cohortes la medida de asociación más utilizada es la conocida como **riesgo relativo** (RR). El RR es el cociente entre la incidencia de los individuos expuestos y la incidencia de los no expuestos. Puede tomar valores desde cercanos a 0 hasta infinito.

- Un valor de 1 indica que no hay asociación entre el efecto y el factor de riesgo.

- Un valor superior a 1 indica que la presencia del factor de riesgo se asocia con el evento.

- Un valor inferior a 1 indica una asociación inversa entre el evento y el factor de riesgo, es decir que este último actuará como factor de protección frente a la aparición del evento.

Ejemplo.

Supongamos que tenemos una población dividida en dos grupos homogéneos. El primer grupo está formado por individuos fumadores (expuestos), mientras que el segundo grupo está formado por individuos no fumadores (no expuestos) y queremos valorar si en el grupo de fumadores aumenta la incidencia de cáncer de pulmón o no (recordad que la incidencia son los casos nuevos que aparecen en un periodo de tiempo). Se relacionaría de la siguiente manera:

	Fumadores	No fumadores	Total
Cáncer	44	26	70
No cáncer	15	68	83
Total	59	94	153

La incidencia (I) de los expuestos sería $I_e = {}^{44}/_{59} = 0{,}75$, mientras que la incidencia de los no expuestos seria $I_{ne} = {}^{26}/_{94} = 0{,}58$. El riesgo relativo (RR) no es más que una medida de asociación entre la presencia de un factor y la aparición o no de un evento, y matemáticamente se relaciona como el cociente entre las dos incidencias de los grupos

$$RR = \frac{le}{lne} = \frac{0{,}78}{0{,}28} = 2{,}68$$

Un RR de 2,68 quiere decir que el grupo de fumadores tiene 2,68 más posibilidades de padecer cáncer de pulmón que el grupo de no fumadores, luego podríamos concluir que el fumar es un factor de riesgo para la aparición de cáncer de pulmón.

Como cualquier parámetro estimado, el RR debe informarse siempre acompañado de su intervalo de confianza (IC). Para determinar el IC previamente se calcula el error estándar (ee) y el RR y se propone un valor de probabilidad (usualmente 95%).

Como se trata de una variable binominal, el error estándar de una proporción es el siguiente:

$$ee = \sqrt{\frac{1}{a} - \frac{1}{(a+b)} + \frac{1}{c} - \frac{1}{(c+d)}}$$

El IC del RR sería:

$$IC95\% = RR \ \pm 1{,}96 \ x \ ee$$

En caso de utilizarse una probabilidad diferente de 95% se busca en la tabla de distribución normal el valor de z correspondiente a dicha probabilidad.

Las letras a, b c y d se corresponden a las casillas de la tabla de 2 x 2

a	b
c	d

Siguiendo con el ejemplo anterior, determinar el error estándar y el intervalo de confianza a un nivel de confianza del 95%

$$ee = \sqrt{\frac{1}{a} - \frac{1}{(a+b)} + \frac{1}{c} - \frac{1}{(c+d)}}$$

$$ee = \sqrt{\frac{1}{44} - \frac{1}{(70)} + \frac{1}{15} - \frac{1}{(83)}} = 0{,}25$$

$$IC95\% = RR \ \pm 1{,}96 \ x \ ee$$

$$IC95\% = 2{,}68 \ \pm 1{,}96 \ x \ 0{,}25 =$$

$$IC95\% = (2{,}19 - 3{,}17)$$

Para un nivel de confianza del 95% se concluye que el grupo de fumadores tiene entre 2,19 y 3,17 más posibilidades de padecer cáncer de pulmón que el grupo de no fumadores.

Otro elemento a considerar en los estudios epidemiológicos es el **odds ratio** (OR), el cual permite determinar la existencia de una asociación entre una variable respuesta y un factor de exposición, y la cuantificación del efecto de la exposición sobre la respuesta.

El odds ratio se define como los odds de los enfermos entre los expuestos dividido entre los odds de los enfermos entre los que tienen la exposición ausente, es decir:

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

La OR se interpreta así:

OR <1 indica una asociación protectora, lo que significa que es poco probable que ocurra el evento.

OR =1 indica que no hay asociación entre ambas variables.

OR > 1 indica que hay una asociación, siendo más fuerte como mayor sea el número.

La OR es el cociente entre una probabilidad y su complementaria. Es la razón de la probabilidad de que ocurra algo frente a la probabilidad de que no ocurra.

$$OR = \frac{P}{1-P}$$

Para el ejemplo que se viene realizando, el OR sería:

	Fumadores	No fumadores	Total
Cáncer	44	26	70
No cáncer	15	68	83
Total	59	94	153

La OR de padecer cáncer si es fumador

$$Odds = \frac{a}{b} = \frac{44}{26} = 1{,}69$$

La OR de no padecer cáncer si es fumador

$$Odds = \frac{c}{d} = \frac{15}{68} = 0{,}22$$

$$OR = \frac{1{,}69}{0{,}22} = 7{,}68$$

El valor 7,68 significa que los fumadores tienen aproximadamente 8 veces más probabilidad de padecer de cáncer que los no fumadores.

Al igual que en el caso del RR, al ser un parámetro de estimación la OR debe expresarse con su intervalo de confianza. Su fórmula es la siguiente:

$$ee = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$IC95\% = lnOR \pm t_{a/2}\, x\, ee$$

Actividad de autoevaluación

1. La tabla muestra el número de atletas que hacen estiramientos antes del ejercicio y cuántos tuvieron lesiones durante el año pasado.

	Lesión durante el año pasado	Ninguna lesión durante el año pasado	Total
Hace estiramientos	55	295	350
No hace estiramientos	231	219	450

Total	286	514	800

Determine el coeficiente Phi e interprete.

2. Se quiere estudiar el efecto de dos fármacos en el tratamiento de una enfermedad infecciona. Para ello, disponemos de un grupo de pacientes infectados, distribuyéndolos al azar dos grupos de tratamiento. Los datos del estudio se resumen en la siguiente tabla:

	Tratamiento A	Tratamiento B	Total
Si mejora	23	33	56
No mejora	12	18	30
Total	35	51	86

Determine el coeficiente Phi e interprete.

3. . Se selecciono al azar 40 personas que asistieron a consulta psicológica con el objetivo de evaluar un test de ansiedad, obteniendo los siguientes datos:

	Con ansiedad	Sin ansiedad	Total
Positivo	17	14	31
Negativo	3	6	9
Total	20	20	40

Calcule: Sensibilidad, especificidad, VPP, VPN, prevalencia y exactitud.

Determine el riesgo relativo de padecer ansiedad y el intervalo de confianza al 95%.

Determine el odds ratio

4. Se ha observado que los estudiantes que inician sus estudios de maestría presentan dificultad en el primer semestre por lo que algunos de ellos deciden retirarse voluntariamente. A continuación, se presentan los resultados realizados a 20 estudiantes de la maestría en Psicología que se matricularon en el semestre 2020-I y que abandonaron el curso de Herramientas de Análisis Cuantitativo. Determine

a) Coeficiente de correlación Phi

b) Para un alfa del 10% realice la respectiva prueba de contraste

Estudiante	Estado civil	Permanencia
1	0	0
2	1	1
3	0	1
4	0	0
5	1	1
6	1	0
7	0	0
8	1	1
9	0	0
10	0	1
11	0	0

12	1	1
13	0	0
14	0	0
15	0	0
16	1	1
17	1	1
18	0	1
19	1	0
20	0	0
21	0	0
22	1	0
23	1	1

Estado civil: 1 = no casado 0 = casado

Permanencia: 1 = permanece en el curso hasta el final 0 = abandona el curso

Representaciones gráficas

Con respecto a las representaciones gráficas, se utilizan dos tipos: diagramas de barras conjuntos y rectángulos partidos. Veamos estos dos tipos de gráficos en relación con el ejemplo de los trastornos afectivos y trastornos principales.

1. Un *diagrama de barras conjunto* se confecciona en esencia como el diagrama de barras sencillo, pero disponiendo barras de diferentes colores (o tonos de gris) para cada categoría de la segunda variable.

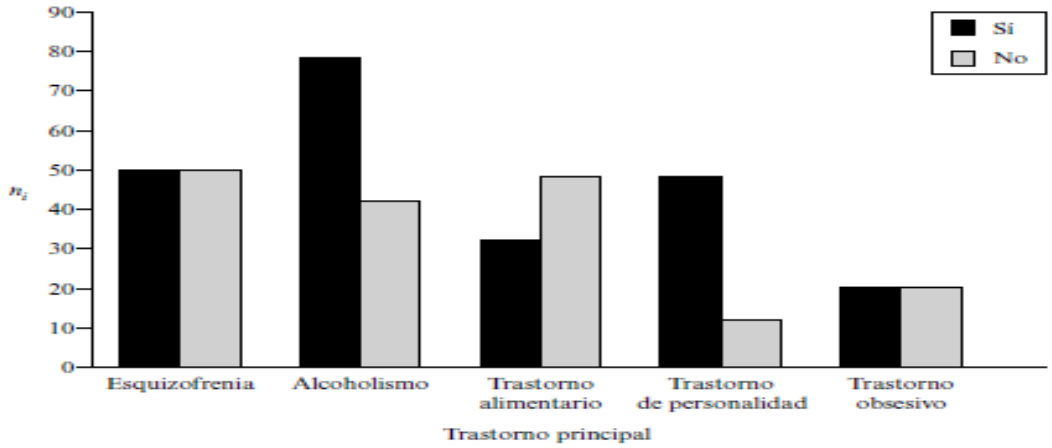


Gráfico 1. Diagrama de barras conjunto de las variables trastorno principal y presencia de trastorno afectivo (SÍ/NO)

2. Un *diagrama de rectángulos partidos* se construyen barras, que aquí llamaremos rectángulos, para cada categoría de una de las variables. Estos rectángulos son todos de las mismas dimensiones, pero cada uno se subdivide en parcelas en función de las frecuencias relativas en la otra variable.

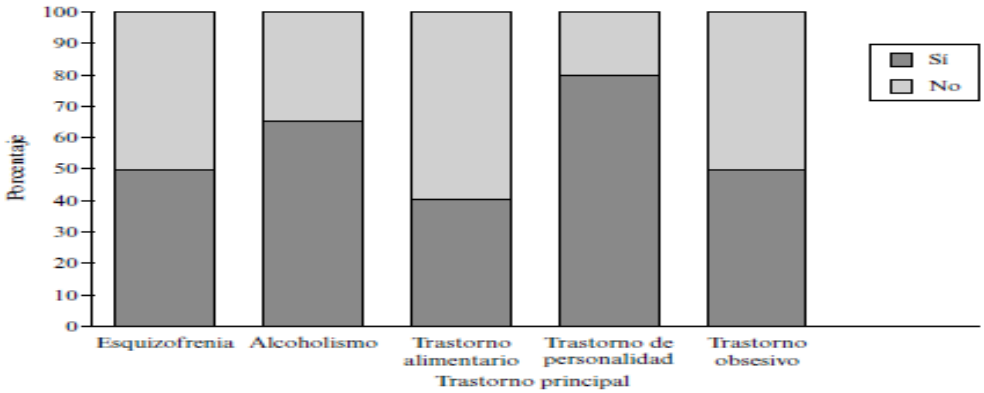


Gráfico 2. Diagrama de rectángulos partidos de las variables trastorno principal (abscisas) y presencia de trastorno afectivo (ordenadas)

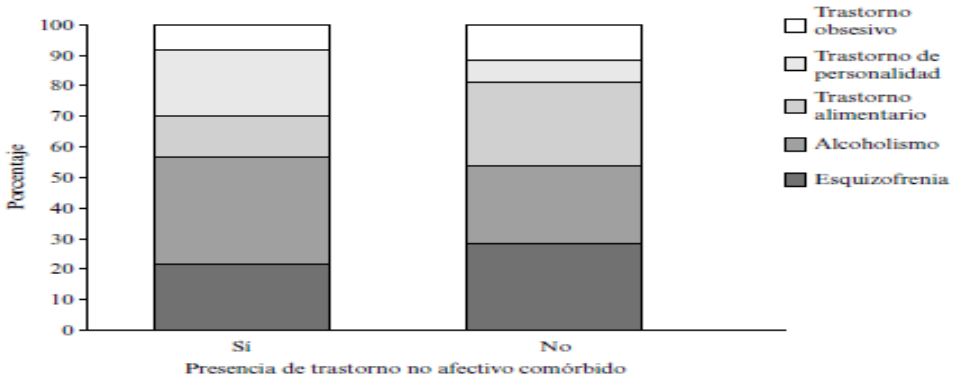


Gráfico 3. Diagrama de rectángulos partidos de las variables trastorno principal (ordenadas) y presencia de trastorno afectivo (abscisas)

7.1.2. Relación entre Variables Ordinales

Una relación entre dos variables ordinales se da cuando los cambios en el orden de las categorías de una variable influyen en el orden de las categorías de la otra variable. En este caso se emplean coeficientes no paramétricos que trabajan con rangos y se basan en el concepto de inversión y no-inversión.

Una relación positiva (directa) o de "no inversión" habla de un predominio de asociación entre los rangos altos de una variable con los rangos altos de la otra al igual que entre los rangos bajos de ambas variables. En cambio, una relación negativa (inversa) o de "inversión" habla de un predominio de asociación entre los rangos altos de una variable con los bajos de la otra. Si las variables son independientes habrá tantas inversiones como no inversiones y el coeficiente valdrá cero; sin embargo, un valor de cero no implica necesariamente que sean independientes.

Entre estos coeficientes están:

1. Rho de Spearman empleado para comparar dos conjuntos de rangos ordenados en una muestra o dos grupos con los rangos ordenados de varias unidades de análisis.

2. Tau b de Kendall empleado para comparar dos rangos cuando se tiene un par de rangos por cada unidad de observación. Es solo aplicable en tablas de contingencia cuadradas (igual

número de filas que de columnas) y si ninguna frecuencia marginal tiene valor cero en sus casillas; no llega a valer 1 si la tabla no es cuadrada.

3. Tau c de Kendall que supera las dificultades del tau b.

4. Gamma de Goodman y Kruskal.

5. d de Somers. Este tiene una versión simétrica que coincide con el Tau-b de Kendall y una asimétrica, modificación del Gamma, que considera a las variables como dependiente e independiente.

El Gamma se emplea para identificar la relación entre dos variables cuando al menos una variable es ordinal. Spearman y Kendall pueden emplearse cuando ambas son ordinales o en escala intervalar.

Todos estos coeficientes arrojan resultados entre -1 y 1, excepto la versión asimétrica del d de Sommer. Todos se basan en los conceptos de inversión y no inversión, la diferencia entre ellos está en el tratamiento que dan a los empates (cuando los pares no son de tipo inversión ni no inversión ya que los rangos en ambas variables coinciden).

Para conocer la concordancia entre dos o más observadores es posible emplear una versión del coeficiente de Kendall y el W de concordancia de Kendall.

Coeficiente de Correlación de Spearman o de Rangos (r_s)

El coeficiente de correlación de Spearman (r_s) es una medida no paramétrica de asociación lineal que utiliza los rangos, números de orden, de cada grupo de sujetos y compara dichos rangos. Es decir, se utiliza para conocer el grado y el sentido de la relación que existe entre dos variables medidas en un nivel ordinal.

Si la medición de ambas variables se encuentra en un nivel de intervalos, pero no se cumple con las condiciones que requiere la aplicación de la prueba de correlación de Pearson, la prueba de Spearman en una buena alternativa pues tiene casi la misma potencia. Por tanto, no se requiere una distribución normal de los datos para la correlación de Spearman.

La fórmula es la siguiente:

$$\boldsymbol{r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}}$$

Siendo **di** la diferencia entre el valor ordinal de X y el valor ordinal de Y; **n** es el número de observaciones registradas.

El coeficiente toma valores entre -1 y +1. Un valor cercano a 0 indica que las variables apenas están relacionadas.

Ventajas del coeficiente de correlación de Spearman

Hay algunas ventajas al usar $\boldsymbol{r_s}$ en vez de $\boldsymbol{r}$.

1. No suponemos que la relación fundamental entre *X* y Y es lineal, por lo tanto, cuando los datos poseen una relación curvilínea distinta, el coeficiente de correlación de Spearman o de Rangos (r_s) probablemente será más confiable que la medida convencional.

2. El uso del coeficiente de correlación de rangos es el hecho de que no se hacen suposiciones de normalidad respecto a las distribuciones de *X* y *Y*.

3. Cuando no somos capaces de hacer mediciones numéricas significativas; sin embargo, se pueden establecer rangos. Tal es el caso, por ejemplo, cuando diferentes jueces clasifican a un grupo de individuos de acuerdo con algún atributo. El coeficiente de correlación de rangos se puede utilizar en esta situación como una medida de la consistencia de los dos jueces.

Prueba de significación

Para probar la hipotesis nula de que $\rho = 0$ utilizando el coeficiente de correlación de Spearman se necesita considerar la distribución muestral de los valores r_s, con base en la suposición de que no hay correlación Vs. hipotesis alternativa de que $\rho > 0$ de que si hay correlación entre las variables involucradas.

En la tabla de correlación de Spearman aparecen valores críticos. Esta tabla es similar a la tabla de valores críticos para la distribución *t*, excepto por la columna izquierda, que ahora proporciona el número de pares de observaciones en vez de los grados de libertad.

Como la distribución de los valores r_s es simétrica alrededor de cero cuando $\rho = 0$, el valor r_s que deja un área de α a la izquierda es igual al negativo del valor r_s que deja un área de α a la derecha. Para una hipotesis alternativa bilateral la región critica de tamaño α cae igualmente en las dos colas de la distribución. Para una prueba en la que la hipotesis alternativa es negativa, la región critica está completamente en la cola izquierda de la distribución y, cuando la hipotesis alternativa es positiva, la región critica se coloca por completo en la cola derecha.

Ejemplo. A continuación, se muestra las calificaciones de 15 estudiantes obtenidas de dos asignaturas: matemática (X_1) y física (X_2):

X_1	X_2
14	14
17	10
8	8
17	15
10	6
10	4
12	10
10	8
20	10
9	7
10	9
17	4
13	14
10	12
17	5

Solución

Procedimiento manual para el cálculo del coeficiente de correlación de Spearman

Paso 1. Se ordenan de menor a mayor los datos de cada variable

X_1	X_2
8	4
9	4
10	5
10	6
10	7
10	8
10	8
12	9
13	10
14	10
17	10
17	12
17	14
17	14
20	15

Paso 2. Se enumera en forma decreciente los datos de cada variable

X_1	X_2
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15

Paso 3. Se promedian los rangos empatados según los datos originales. Por ejemplo, en la variable X_1 se observa que aparece 5 veces el valor 10, los cuales ocupan, una vez ordenados los datos, las

posiciones 3, 4, 5, 6, 7. Se suman y se promedian, 25/5 = 5. Este valor se coloca en las respectivas casillas que ocupa la calificación 10. Se realiza el mismo procedimiento para aquellos rangos empatados encada una de las variables.

X_1	X_2
1	1,5
2	1,5
5	3
5	4
5	5
5	6,5
5	6,5
8	8
9	10
10	10
12,5	10
12,5	12
12,5	13,5
12,5	13,5
15	15

Paso 4. Los valores originales de cada variable se sustituyen por los rangos correspondientes. Para ello, volvemos a los datos sin ordenar y se coloca a cada dato su correspondiente rango de acuerdo a lo realizado en el paso anterior.

X_1	Rango X_1	X_2	Rango X_2
14	10	14	13,5
17	12,5	10	10
8	1	8	6,5
17	12,5	15	15
10	5	6	4
10	5	4	1,5
12	8	10	10
10	5	8	6,5
20	15	10	10
9	2	7	5
10	5	9	8
17	12,5	4	1,5
13	9	14	13,5
10	5	12	12
17	12,5	5	3

Paso 5. Se elabora la siguiente tabla y se calcula el desvío al cuadrado

Rango X_1	Rango X_2	d	d^2
10	13,5	-3,5	12,25
12,5	10	2,5	6,25
1	6,5	-5,5	30,25
12,5	15	-2,5	6,25
5	4	1	1
5	1,5	3,5	12,25
8	10	-2	4
5	6,5	-1,5	2,25
15	10	5	25
2	5	-3	9
5	8	-3	9
12,5	1,5	11	121
9	13,5	-4,5	20,25
5	12	-7	49
12,5	3	9,5	90,25
			Σ = 398

Paso 6. Calculamos el coeficiente de correlación de Spearman

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$r_s = 1 - \frac{6 * 398}{15 * (15^2-1)} = 0{,}289$$

Paso 7. Interpretación del coeficiente de correlación obtenido.

El valor 0,289 obtenido evidencia un grado de asociación positiva baja entre las asignaturas matemática (X_1) y física (X_2). El hecho que el r_s dio positivo indica que a puntajes altos en X_1 se asocia a puntajes altos en X_2.

Para comprobar si existe una correlación o no entre los puntajes de matemática y los puntajes de física se establece el siguiente sistema de hipótesis:

Ho: $\rho = 0$

Hi: $\rho > 0$.

Asumimos un valor de significación, por ejemplo 0,05.

Buscamos en la tabla de correlación de Spearman el valor crítico para n = 15 y un α = 0,05 lo cual arroja un valor de 0,441.

Interpretación. Dado que el valor critico (0,441) resultó ser mayor al valor calculado (0,289) se concluye que no hay una correlación significativa entre los puntajes de ambas asignaturas para un nivel de significación del 5%.

Nota. Cuando n excede a los valores dados en la tabla de correlación de Spearman se puede probar si existe una correlación significativa calculando $Z = r_s\sqrt{n-1}$ y comparando con los

valores críticos de la distribución normal estándar que se presentan en dicha tabla.

Actividad de autoevaluación

En la siguiente tabla se muestran los datos de 10 mujeres. Indagar si existe asociación estadísticamente significativa entre el coeficiente intelectual y el tamaño del cerebro para un alfa del 1%.

Sujeto	Coeficiente intelectual	Tamaño del cerebro (miles de mega pixeles en imagen de escáner)
1	133	817
2	137	952
3	99	929
4	138	991
5	92	854
6	132	834
7	132	865
8	98	879
9	92	834
10	135	791

2. En la siguiente tabla se presentan las calificaciones registradas de 10 estudiantes en un examen de medio curso y las del examen final en un curso de cálculo:

Sujeto	Examen de medio curso	Examen final
1	84	73
2	98	63
3	91	87
4	72	66
5	86	78
6	93	78
7	80	91
8	0	0
9	92	88
10	87	77

a) Calcule el coeficiente de correlación de Spearman.

b) Pruebe la hipotesis nula de que $\rho = 0$ en comparación con la hipotesis alternativa de que $\rho > 0$. Utilice $\alpha = 0.025$.

7.1.3. Relación entre Variables Cuantitativas

La relación entre dos variables cuantitativas puede medirse por la covariación, pero esta depende de las unidades de medida de las variables y no está acotada, por lo que se prefiere el uso de los coeficientes de correlación que permiten medir la fuerza y la dirección de la asociación entre ambas variables.

Uno de los más utilizados es el coeficiente de correlación producto-momento de *Pearson (r),* coeficiente paramétrico que solo puede calcularse para variables con niveles de medición intervalar o de razón.

Los valores de este coeficiente oscilan de -1 a 1, siendo los valores extremos los que indican la mayor correlación y el 0 la ausencia de correlación. El signo del coeficiente indica el sentido de la relación. Ante un signo positivo se dirá que la relación es directa (las variables cambian en el mismo sentido) y ante uno negativo se dirá que es inversa (a medida que aumenta una disminuye la otra).

El gráfico más adecuado para apreciar la relación entre dos variables numéricas es el diagrama de dispersión.

En el caso de darse una correlación lineal con dependencia entre las variables es posible estimar la recta que mejor describe esta relación entre ambas variables mediante la regresión lineal simple y mediante el coeficiente de determinación que evaluará el ajuste de los datos a la recta obtenida.

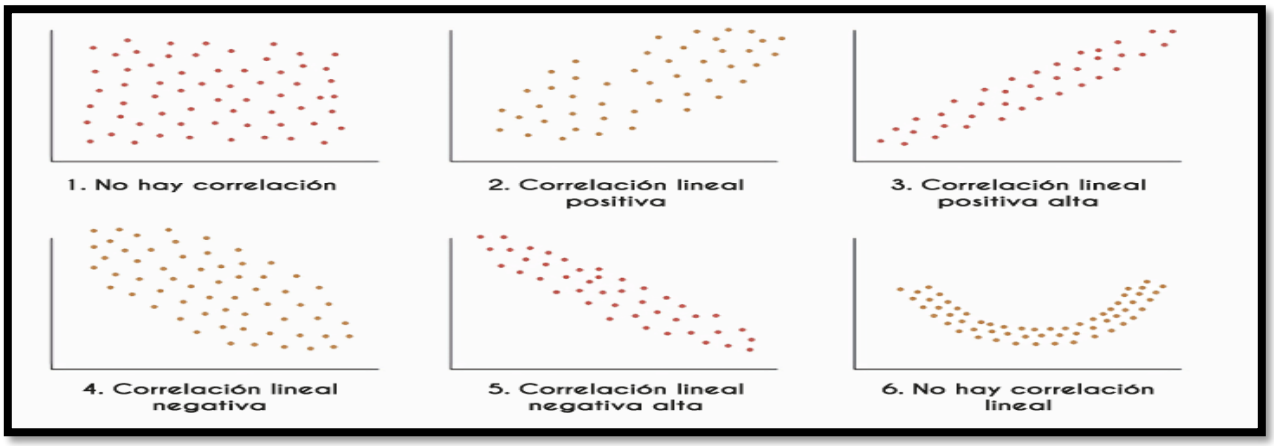
Diagrama de dispersión

Cuando sobre cada individuo de una población se observan simultáneamente dos características cuantitativas X e Y , se dice que se está observando una variable estadística bidimensional, que se representa por (X, Y). La representación gráfica más usual es el diagrama de dispersión o nube de puntos, que consiste en situar en un sistema de ejes coordenados los puntos que resultan de tomar en el eje horizontal los valores de una de las variables y en el eje vertical los valores de la otra.

El diagrama de dispersión consiste en plotear los pares de valores de las variables de cada unidad de observación en el plano cartesiano, formando una nube de puntos que puede adoptar diferentes formas las cuales ofrecen idea del tipo de relación (asemeja una recta cuando existe correlación lineal o una curva para una relación curvilínea) o sin forma específica que indica la independencia entre las variables.

¿**Cuál es el objetivo del diagrama de dispersión?** Es conocer la correlación (o causalidad) entre variables. Esto se consigue trazando una línea de mejor ajuste, a la que se conoce como línea de tendencia o regresión. Esta línea representa la solución matemática de la relación entre variables. La línea de regresión puede revelar tres tipos de relaciones. Si los valores de Y aumentan en una función de X, hay una correlación positiva (aumenta). Si los valores de Y se reducen en una función de X, hay una correlación negativa (disminuye). Si los puntos de datos son aleatorios, no hay correlación entre variables. Las correlaciones también pueden expresarse como curvas. Estas líneas con tendencia de curva suelen ser líneas de segundo (cúbico) o tercer orden (cuadrático).

En la Figura se observa diferentes tipos de diagrama de dispersión, indicando el tipo de correlación.



Diagramas de dispersión y tipos de correlación

Por ejemplo, un investigador realizó un estudio para determinar la relación entre el número promedio de cigarros consumidos en cientos por persona (X), y la tasa de mortalidad por cáncer de pulmón en 15 localidades, en muertes por cien mil habitantes (Y), como se muestra en la siguiente tabla

Cigarrillos	Mortalidad
X	Y
18,20	17,05
25,82	19,80
18,24	15,98
28,60	22,07
31,10	22,83
33,60	24,55
40,46	27,27
28,27	21,10
20,10	16,80
27,91	22,80
26,18	20,30
22,12	18,00
21,84	16,84
23,40	18,70
21,58	24,45

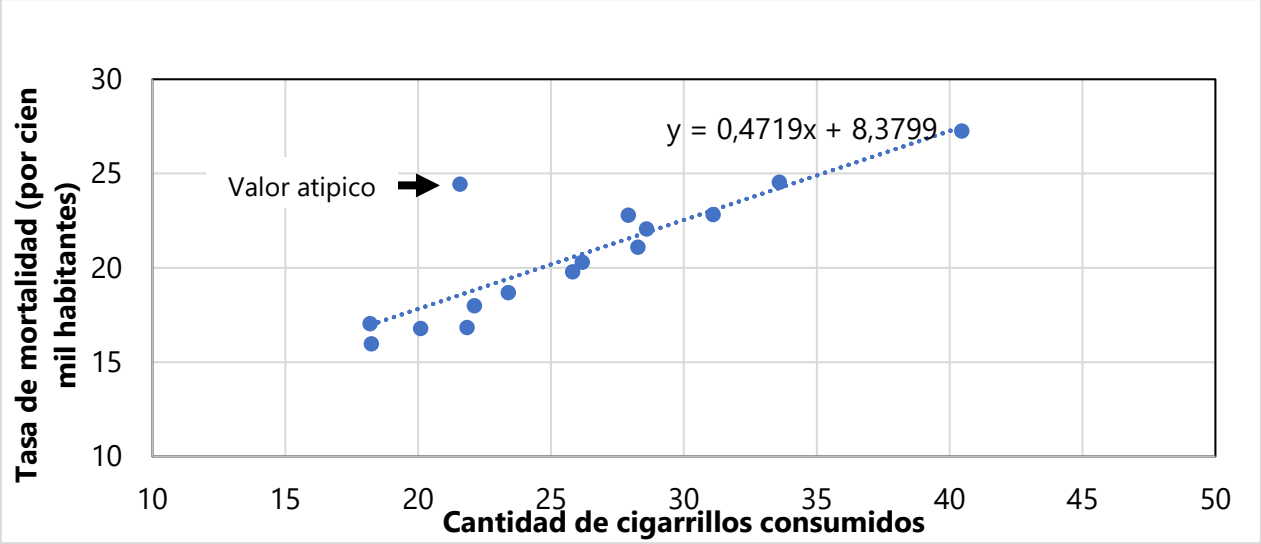


Gráfico 1. Relación entre la cantidad de cigarrillos consumidos en cientos por persona y la tasa de mortalidad por cáncer de pulmón en 15 localidades, en muertes por cien mil habitantes

En el gráfico de dispersión se puede observar que conforme se incrementa la cantidad de cigarrillos consumidos, se incrementa de forma lineal positiva, el índice de mortalidad por cada 100000 habitantes. Es decir, se puede representar la asociación entre esas variables, con una línea recta. Sin embargo, se observa un valor atípico. También en la gráfica se incluye la ecuación de regresión lineal: **Y = 8,38 + 0,47X**

Coeficiente de correlación lineal de Pearson

El método más común de determinar si existe asociación **lineal** entre dos variables cuantitativas continuas es el Análisis de Correlación de Pearson. Con este método se obtiene el Coeficiente de Correlación de Pearson, usualmente representado por la letra **R**. Como suele utilizarse una muestra, lo que se obtiene en realidad es un estimado del coeficiente de correlación poblacional, **r**.

Dos aspectos importantes del coeficiente de correlación son su magnitud y su signo. La magnitud refleja la intensidad de la asociación entre las dos variables; el valor absoluto de la magnitud puede variar entre cero y uno. Valores cercanos a cero indican que las variables no están asociadas, es decir, que el valor de una variable es independiente del valor de la otra.

El signo, por su parte, refleja cómo están asociados los valores de ambas variables. Si el signo es positivo indica que a valores altos de una variable corresponden valores altos de la otra, o a valores bajos de una variable corresponden valores bajos de la otra. Si el signo es negativo, indica que a valores altos de una variable corresponden valores bajos de la otra. Es decir, el sigo positivo indica que los valores de ambas variables cambian en el mismo sentido, mientras que el signo negativo indica que cambian en sentido contrario.

Por ejemplo, puede evaluarse la relación entre las horas de estudio empleado por el estudiante y la calificación obtenida en una prueba. Si el valor calculado, referido a un coeficiente de correlación es positivo se dice que existe una relación positiva o directa entre ambas variables, es decir, cuanto mayor sea las horas de estudio mayor es la calificación; pero, si es negativo, la relación es negativa o inversa, y si es cero no existe una relación.

Entre las características del coeficiente de correlación de Pearson tenemos

1. $-1 \leq \rho \leq 1$

2. Los valores de ρ cercanos a 0 indican una correlación lineal débil entre X e Y.

3. Los valores de ρ cercanos a +1indican una fuerte correlación lineal positiva entre X e Y.

4. Los valores de ρ cercanos a -1indican una fuerte correlación lineal negativa entre X e Y.

¿Cómo se calcula el coeficiente de correlación lineal?

Para el caso del coeficiente de correlación poblacional o coeficiente de correlación de momento-producto o correlación de Pearson se denota con la letra $\boldsymbol{\rho}$ (rho), y en el caso de la muestra con la letra ***r***. Este tipo de coeficiente se utiliza para medir el grado de asociación lineal entre las variables X e Y. Se expresa de la siguiente manera

$$\mathbf{r} = \frac{\boldsymbol{Cov(X,Y)}}{\boldsymbol{Sx\,.\,Sy}} = \frac{\mathbf{n}\sum \mathbf{XY} - \sum \mathbf{X}\sum \mathbf{Y}}{\sqrt{\mathbf{n}\sum \mathbf{X^2} - (\sum \mathbf{X})^2}\sqrt{\mathbf{n}\sum \mathbf{Y^2} - (\sum \mathbf{Y})^2}}$$

Cov (X,Y) es la covarianza de ambas variables X, Y

Sx y Sy son las desviaciones estándares de las variables X y Y respectivamente.

En el caso de ejemplo anterior, el coeficiente de correlación se calcula de acuerdo con los siguientes pasos:

Primero elaboramos la siguiente tabla:

Cigarrillos	Mortalidad			
X	Y	X^2	Y^2	XY
18,20	17,05	331,24	290,70	310,31
25,82	19,80	666,67	392,04	511,24
18,24	15,98	332,70	255,36	291,48
28,60	22,07	817,96	487,08	631,20
31,10	22,83	967,21	521,21	710,013
33,60	24,55	1128,96	602,70	824,88
40,46	27,27	1637,01	743,65	1103,34
28,27	21,10	799,19	445,21	596,497
20,10	16,80	404,01	282,24	337,68
27,91	22,80	778,97	519,84	636,348
26,18	20,30	685,39	412,09	531,454
22,12	18,00	489,29	324,00	398,16
21,84	16,84	476,99	283,59	367,7856
23,40	18,70	547,56	349,69	437,58
21,58	24,45	465,70	597,80	527,631
Σ = 387,42	**Σ = 308,54**	**Σ =10528,85**	**Σ = 6507,21**	**Σ = 8215,60**

Segundo aplicamos la fórmula

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{15 \text{ x } 8215{,}60 - 387{,}42 \text{ x } 308{,}54}{\sqrt{15 \text{ x } 10528{,}85 - (387{,}42)^2}\sqrt{15 \text{ x } 6507{,}21 - (308{,}54)^2}}$$

$$r = \frac{123233{,}94 - 119534{,}57}{\sqrt{7838{,}51}\ x\ \sqrt{2411{,}22}} = \frac{3999{,}37}{4347{,}46} = 0{,}85$$

$$r = 0{,}85$$

Tercero interpretamos el resultado de r. Teniendo en cuenta la tabla de baremo señalada anteriormente, el valor de r = 0,85 cae en el rango de 0,70 a 0,89, por lo tanto, la correlación entre la cantidad de cigarrillos consumidos en cientos por persona y la tasa de mortalidad por cáncer de pulmón en 15 localidades, en muertes por cien mil habitantes presenta una relación positiva alta.

Para tener en cuenta: Covarianza es el valor a través del cual se refleja en qué cuantía don variables cualesquiera varían de forma conjunta respecto de sus medias aritméticas. Así, esta medida nos permite conocer cómo se comportan las variables en cuestión respecto de otras variables.

$$\sigma_{XY} = \frac{\sum(Xi - \bar{X})(Yi - \bar{Y})}{N}$$

$$S_{XY} = \frac{\sum(Xi - \bar{X})(Yi - \bar{Y})}{n - 1}$$

Actividad de autoevaluación

1. Elabore el diagrama de dispersión y calcule el coeficiente de correlación de Pearson para las variables X e Y en la siguiente muestra de 10 sujetos

Sujeto	1	2	3	4	5	6	7	8	9	10
X	2	5	7	3	4	5	5	6	1	4
Y	5	1	6	6	5	4	3	2	1	5

2. Llevado a cabo un estudio sobre la relación entre la motivación de logro con diferentes facetas de la satisfacción laboral en una muestra de 50 trabajadores, se obtuvo la siguiente matriz de correlaciones:

	SS	*SH*	*SR*	*MC*	*OP*
SS		0,82	0,61	0,42	−0,49
SH			0,35	0,15	−0,15
SR				0,75	0,31
MC					0,45
OP					

Donde: SS: Satisfacción con el sueldo obtenido.

SH: Satisfacción con el horario realizado.
SR: Satisfacción con el reconocimiento obtenido.
MC: Motivación por la consecución de objetivos.
OP: Oportunidades de promoción.

Responda las siguientes cuestiones:

a) ¿Qué variable correlaciona más con MC?

b) ¿Qué variable correlaciona menos con SH?

c) ¿Cuál es la mayor correlación lineal encontrada?

d) ¿SR se relaciona más con SS o con MC?

e) ¿Cómo se interpreta la correlación lineal negativa entre OP y SS?

3. Medidas las variables X: Rendimiento académico e Y: Tiempo dedicado al ocio (horas/semana), se obtuvieron los siguientes resultados en dos secciones de Psicología:

	Sección 1					**Sección 3**				
Sujeto	1	2	3	4	5	1	2	3	4	5
X	4	7	2	8	9	1	4	3	5	5
Y	30	22	27	20	17	35	31	29	28	23

A partir de estos datos, conteste a las siguientes cuestiones:

a) Indique cuál de las dos muestras dedica un mayor número de horas semanales a su tiempo de ocio.

b) Calcule el coeficiente de correlación de Pearson en cada muestra y diga en cuál de las dos muestras existe mayor correlación lineal entre el rendimiento académico y el tiempo dedicado al ocio.

c) Elabore una representación gráfica conjunta del diagrama de dispersión de ambas muestras.

7.1.4. Relación entre una variable continua y una nominal

Las técnicas empleadas para estudiar la relación entre una variable continua y una nominal varían en dependencia de la variable cualitativa nominal y las características de las muestras en estudio.

Si la variable cualitativa es politómica se pueden emplear el coeficiente de correlación, que no supone linealidad y cuyo cuadrado puede interpretarse, si el diseño lo permite, como la proporción de varianza de la variable cuantitativa que es explicada por la variable categórica o el índice F de Cohen cuyo valor mide la intensidad de la asociación.

Si la variable es dicotómica se podrán emplear el coeficiente biserial puntual, los índices Delta de Glass, g de Hedges, d de Cohen o el índice d, modificación del índice d de Cohen para el caso de una medición repetida en dos momentos para un mismo grupo.

Estos índices son muy empleados para determinar el llamado tamaño del efecto que se produce en grupos bajo diferentes tratamientos en estudios experimentales o en el metaanálisis de investigaciones cuyo efecto cuantitativo se obtuvo por la media.

A veces las categorías de una variable cualitativa se forman a partir del empleo en la investigación de más de una muestra, en este caso lo que se pretende es buscar diferencias entre ellas. Si las diferentes categorías de la variable politómica representan los niveles donde se mide la variable cuantitativa se debe emplear el análisis de varianza paramétrico o el no paramétrico de Kruskal Wallis. Dada una dicotomía condicionada por dos muestras independientes se pueden emplear las pruebas de comparación de media en muestras independientes t de student sin o con la aproximación de Welch, en dependencia de si existe o no homocedasticidad de varianzas, respectivamente; o sus alternativas no paramétricas, la U de Mann y Withney, las rachas de Wald Wolfowitz, entre otros. Si la dicotomía es condicionada por muestras pareadas se debe emplear la t de student para muestras pareadas o sus alternativas no paramétricas el test de los signos, el de rangos con signos de Wilcoxon, entre otros.

El empleo de una prueba paramétrica siempre dependerá del cumplimiento de supuestos como la independencia, normalidad y homocedasticidad de las varianzas.[14] El supuesto de independencia puede corroborarse mediante el test de Durbin Watson, el de normalidad a través de técnicas gráficas como el histograma, los gráficos de probabilidad normal P-P y de cuantiles normales Q-Q o con técnicas estadísticas como las pruebas de bondad del ajuste de Kolmogorov-Smirnov, Shapiro-Wilk, D'Agostino, Lilliefors, entre otros o el cálculo de los coeficientes de simetría y de curtosis, mientras que la igualdad de varianzas se verifica con los test de Levene, Barttle o C de Cochran.

Los gráficos empleados para representar esta información son los polígonos de frecuencia, donde se podrá comparar el grado de solapamiento de las distribuciones de la variable cuantitativa condicionadas por las categorías de la cualitativa y en el análisis exploratorio de los datos se emplean las gráficas de cajas y bigotes o las de barras de error múltiples con la distribución de la variable cuantitativa condicionada a la variable cualitativa (una caja o barra por cada categoría de la variable nominal).

Correlación Biserial - Puntual

La correlación biserial puntual (r_{bp}) se utiliza cuando se tiene una variable auténticamente dicotómica y una variable (X) cuantitativa continua, no se distribuyen normalmente, por lo que no es necesario conocer el valor de y.

Las fórmulas son las siguientes (cualquiera se puede utilizar):

$$r_{bp} = \frac{\bar{x}_p - \bar{x}_q}{\sigma_X} * \sqrt{p * q} \qquad -1 < r_{bp} < 1$$

$$r_{bp} = \frac{\bar{x}_p - \bar{x}}{\sigma_X} * \sqrt{\frac{p}{q}}$$

$\bar{x}_p$ = media de la muestra que acertaron el ítem (X = 1) – variable dicotómica

$\bar{x}_q$ = media de la muestra que no acertaron el ítem (X = 0) – variable dicotómica

$\bar{x}$ = media de todos los casos en la variable X

p = proporción de personas que acertaron en el ítem – variable dicotómica

q = proporción de personas que no acertaron en el ítem – variable dicotómica

σ_X = desviación estándar de los puntajes totales pertenecientes a la muestra con los valores de la variable continua (X).

Si $\bar{x}_p > \bar{x}_q$ y la diferencia $\bar{x}_p - \bar{x}_q$ es positiva. Esto quiere decir que los sujetos que puntúen alto en Y (variable cuantitativa) tenderán a pertenecer a la modalidad p. Los que puntúen bajo en Y tenderán a pertenecer a la modalidad q.

Si $\bar{x}_p < \bar{x}_q$ y la diferencia $\bar{x}_p - \bar{x}_q$ es negativo la interpretación se realiza a la inversa: a puntuación alta en Y le corresponderá la modalidad q y, a puntuación baja en X, la modalidad p.

Ejemplo. Supongamos que se aplicó una encuesta a un grupo de 10 empleados de la empresa DATA con la finalidad de conocer si existe correlación entre la satisfacción personal y el rendimiento laboral (si / no). Se asume que ambas variables no siguen una distribución normal. Los datos se muestran a continuación:

Rendimiento laboral	**Satisfacción personal (X)**
0	55
0	57
1	60
1	62
0	57
0	59
1	59
1	60
0	56
1	58

Si = 1 No = 0

Paso 1. Calculamos la media por grupo (p → 1 q → 0)

$$\bar{x}_p = \frac{60 + 62 + 59 + 60 + 58}{5} = \frac{299}{5} = 59{,}8 \qquad \bar{x}_q = \frac{55 + 57 + 57 + 59 + 56}{5} = \frac{284}{5} = 56{,}8$$

Paso 2. Calculamos los valores de p y q

$$p = \frac{n_p}{N} = \frac{5}{10} = 0{,}5 \qquad q = \frac{n_q}{N} = \frac{5}{10} = 0{,}5$$

Paso 3. Calculamos la desviación estándar de la variable cuantitativa

X	$\boldsymbol{X^2}$
55	3025
57	3249
60	3600

62	3844
57	3249
59	3481
59	3481
60	3600
56	3136
58	3364
Σ = 583	**Σ = 34029**

$$\bar{X} = \frac{\sum X}{N} = \frac{583}{10} = 58{,}3$$

$$\sigma_X = \sqrt{\frac{\sum X^2}{N} - \bar{X}^2} = \sqrt{\frac{34029}{10} - (58{,}3)^2} = 2{,}00$$

Paso 4. Calculamos el coeficiente de correlación biserial – puntual

$$r_{bp} = \frac{\bar{x}_p - \bar{x}_q}{\sigma_X} * \sqrt{p * q}$$

$$r_{bp} = \frac{59{,}8 - 56{,}8}{2{,}00} * \sqrt{0{,}5 * 0{,}5}$$

$$r_{bp} = 0{,}775$$

Existe una correlación positiva alta entre la satisfacción personal y el rendimiento laboral en los empleados de la empresa DATA; es decir, en la medida que aumenta el rendimiento laboral aumenta la satisfacción laboral del empleado.

Correlación biserial

El coeficiente de correlación biserial (r_b), se utiliza cuando se desea conocer la correlación existente entre dos variables, de las cuales, una ha sido considerada como escala de intervalos o de razón y la otra resulta ser una variable dicotómica o dicotomizada. Ambas variables siguen una distribución normal.

Las fórmulas empleadas para calcular el coeficiente de correlación biserial son las siguientes:

$$\boldsymbol{r_b = \frac{\bar{x}_p - \bar{X}}{\sigma_x} * \frac{p * q}{y}}$$

$\bar{x}_p$ = puntuación media para pares de datos de la categoría p

$\bar{x}_q$ = puntuación media para pares de datos de la categoría q

p = proporción de pares de datos para la categoría p

q = proporción de pares de datos para la categoría q

σ_x = desviación estándar de la variable cuantitativa

y = altura de la ordenada que separa en la curva normal a las proporciones p y q

Ejemplo. Se ha medido la estatura (X) y el peso (Y) a un grupo de 10 individuos dividiéndolos según el peso en obesos (peso superior a la mediana) y delgados (peso inferior a la mediana). Se desea conocer si existe alguna relación entre peso y altura. Los resultados de la estatura se muestran a continuación

Solución. Como una de las variables cuantitativas ha sido dicotomizada (como es el caso del peso) se aplica el coeficiente de correlación biserial.

Paso 1. Se ordenan los datos de menor a mayor y se determina la mediana.

155 – 160 – 164 – 166 – 168 – 169 – 170 – 172 – 174 – 176

Se encuentra entre los datos X5 y X6. Por tanto, las primeras cinco estaturas el peso de los individuos se consideran delgados (p) y las otras cinco estaturas el peso de los individuos se consideran obesos (q).

X	155	160	164	166	168	169	170	172	174	176
Y	0	0	0	0	0	1	1	1	1	1

Paso 2. Calculamos las medias de p y q

$$\bar{x}_p = \frac{169 + 170 + 172 + 174 + 176}{5} = \frac{861}{5} = 172{,}2 \quad \bar{x}_q = \frac{155 + 160 + 164 + 166 + 168}{5} = \frac{813}{5} = 162{,}6$$

Paso 3. Calculamos p y q

$$p = \frac{n_p}{N} = \frac{5}{10} = 0{,}5 \qquad q = \frac{n_q}{N} = \frac{5}{10} = 0{,}5$$

Paso 4. Calculamos la desviación estándar de la variable cuantitativa

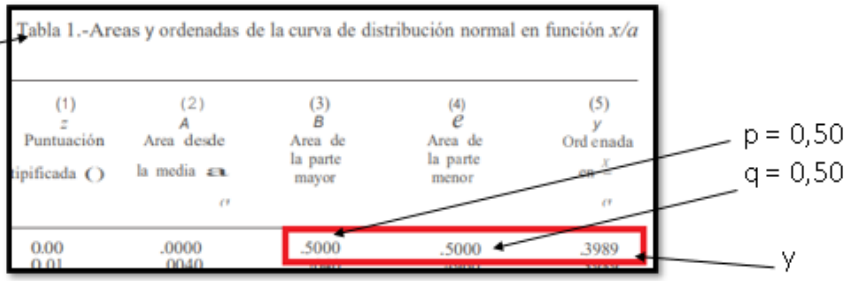
X	**Y**	$\boldsymbol{X^2}$
155	0	24025
160	0	25600
164	0	26896
166	0	27556
168	0	28224
169	1	28561
170	1	28900
172	1	29584
174	1	30276
176	1	30976
Σ = 1674		**Σ = 280598**

$$\bar{X} = \frac{\sum X}{N} = \frac{1674}{10} = 167{,}4$$

$$\sigma_X = \sqrt{\frac{\sum X^2}{N} - \bar{X}^2} = \sqrt{\frac{280598}{10} - (167{,}4)^2} = 6{,}09$$

Paso 5. Buscamos **y** que es la altura de la ordenada que separa en la curva normal a las proporciones p y q

Ver Libro de Ramón Pérez Juste. Estadística aplicada a las ciencias sociales



Paso 6. Calculamos el coeficiente de correlación biserial

$$r_b = \frac{\bar{x}_p - \bar{x}_q}{\sigma_x} * \frac{p * q}{y}$$

$$r_b = \frac{172{,}2 - 162{,}6}{6{,}09} * \frac{0{,}5 * 0{,}5}{0{,}3989} = 0{,}988$$

Existe una relación positiva muy alta entre la estatura y el peso, es decir, en la medida que aumenta la estatura aumenta el peso de la persona.

Actividad de autoevaluación

1. Tenemos las puntuaciones totales de 10 sujetos que han contestado a una prueba objetiva (X) cuya calificación oscila entre 1 y 10; además sabemos si han acertado (1) o fallado (0) un ítem Y. Suponga que las variables siguen una distribución normal. Qué resultados esperaría sino siguieran una distribución normal.

X	Y
1	0
1	1
2	0
2	0
3	1
4	1
6	1
6	1
8	1
10	1

RESUMEN DE TIPOS DE COEFICIENTES DE CORRELACIÓN

Variable 1	Variable 2	Coeficiente de correlación	Fórmula
Nominal dicotómica	Nominal dicotómica	Phi (Ø)	$\varphi = \frac{n_{1,1} * n_{0,0} * -n_{1,0} + n_{0,1}}{\sqrt{(n_{1,1} + n_{0,1}) * (n_{1,0} + n_{0,0}) * (n_{1,1} + n_{1,0}) * (n_{0,1} + n_{0,0})}}$
Nominal (2 o más)	Nominal (2 o más)	V de Cramer (V)	$V = \sqrt{\frac{X^2}{n(min[c,r]-1)}}$ N: Número total de frecuencia. Min: Menor número de categorías entre filas (r) y columnas (c), menos 1.
Nominal	Nominal	Lambda (λ) o coeficiente de predictibilidad de Guttman	
Nominal categórica	Nominal categórica	Kappa (K)	$\mathrm{k} = \frac{(\mathrm{p_o} - \mathrm{p_e})}{(1 - \mathrm{p_e})}$ p_o: acuerdo relativo observado entre los evaluadores p_e: probabilidad hipotética de acuerdo al azar
Nominal dicotómica	Nominal dicotómica	Q de Yule (Q)	$Q(x,y) = \frac{ad - bc}{ad + bc}$
Nominal dicotómica	Nominal dicotómica	Riesgo relativo (RR) y Odd Ratio (OR)	$RR = \frac{[a/(a+b]}{[c/(c+d)]}$ $OR = \frac{\left[a/(a+b)\right]\left[b/(a+b)\right]}{\left[c/(c+d)\right]\left[d/(c+d)\right]} = \frac{a/b}{c/d}$
Categórica (3 o más)	Categórica (3 o más)	Coeficiente de contingencia (C)	$C = \sqrt{\frac{X^2}{X^2 + n}}$

Ordinal	Ordinal	Spearman (r_s)	$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$
Ordinal	Ordinal	Gamma o Gamma de Goodman y Kruskal (γ)	$\gamma = \frac{fc - fi}{fc + fi}$ fc = número de concordancias fi = número de discordancias
Ordinal	Ordinal	Tau-b- de Kendall y Tau-c- de Kendall (τ)	$\tau = {C - D}/{C + D}$ C es el número de pares concordantes D es el número de pares discordantes $\tau = \frac{c - d}{\frac{n(n-1)}{2}}$
Ordinal	Ordinal	W de concordancia de Kendall	$W = \frac{n\sum_{j=1}^{n}(\sum R_j)^2 - \sum_{j=1}^{n}(\sum R_j)}{\frac{m^2 n(n^2-1)}{12}}$ W = coeficiente de concordancia W de Kendall n = número de objetos (o individuos) a quienes se están asignando rangos m = número de jueces Rj = Suma de los rangos asignados al j-esimo
Cuantitativa (no normal)	Dicotómica o dicotomizada	Biserial – puntual (r_{bp})	$r_{bp} = \frac{\bar{x}_p - \bar{x}_q}{\sigma_X} * \sqrt{p * q}$ $\quad$ $r_{bp} = \frac{\bar{x}_p - \bar{x}}{\sigma_X} * \sqrt{\frac{p}{q}}$
Cuantitativa (normal)	Dicotómica o dicotomizada	Biserial (r_b)	$r_b = \frac{\bar{x}_p - \bar{x}_q}{\sigma_x} * \frac{p * q}{y}$ $\quad$ $r_b = \frac{\bar{x}_p - \bar{X}}{\sigma_x} * \frac{p * q}{y}$
Cuantitativa (normal)	Cuantitativa (normal)	Correlación de Pearson	$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$

Valores críticos del coeficiente de correlación de rangos de Spearman

n	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900			
6	0.829	0.886	0.943	
7	0.714	0.786	0.893	
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

Areas y ordenadas de la curva de distribución normal en función x/a

(1) z Puntuación tipificada	(2) A Area desde la media $\frac{a}{a}$	(3) B Area de la parte mayor	(4) e Area de la parte menor	(5) y Ordenada en $\frac{x}{a}$
0.00	.0000	.5000	.5000	.3989
0.01	.0040	.5040	.4960	.3989
0.02	.0080	.5080	.4920	.3989
0.03	.0120	.5120	.4880	.3988
0.04	.0160	.5160	.4840	.3986
0.05	.0199	.5199	.4801	.3984
0.06	.0239	.5239	.4761	.3982
0.07	.0279	.5279	.4721	.3980
0.08	.0319	.5319	.4681	.3977
0.09	.0359	.5359	.4641	.3973
0.10	.0398	.5398	.4602	.3970
0.11	.043M	.5438	.4562	.3965
0.12	.0478	.5478	.4522	.3961
0.13	.0517	.5517	.4483	.3956
0.14	*.0551*	.5557	.4443	.3951
0.15	.0596	*.5596*	.4404	.3945
0.16	.0636	.5636	.4364	.3939
0.17	.0675	*.5615*	.4325	.3932
0.18	.0714	.5714	.4286	.3925
0.19	.0753	.5753	.4247	.3918
0.20	.0793	.5793	.4207	.3910
0.21	.0832	.5832	.4168	.3902
0.22	.0871	.5871	.4129	.3894
0.23	.0910	.5910	.4090	.3885
0.24	.0948	.5948	.4052	.3876
0.25	.0987	*.5981*	.4013	.3867
0.26	.1026	.6026	.3974	.3857
0.27	.1064	.6064	.3936	.3847
0.28	.1103	.6103	.3897	.3836
0.29	.1141	.6141	.3859	.3825
0.30	.1179	.6179	.3821	.3814
0.31	.1217	.6217	.3783	.3802
0.32	.1255	.6255	.3745	.3790
0.33	.1293	.6293	.3707	.3778
0.34	.1331	.6331	.3669	.3765
0.35	.1368	.6368	.3632	.3752
0.36	.1406	.6406	.3594	.3739
0.37	.1443	.6443	.3557	.3725
0.38	.1480	.6480	.3520	.3712
0.39	.1517	.6517	.3483	.3697
0.40	.1554	.6554	.3446	.3683
0.41	.1591	.6591	.3409	.3668
0.42	.1628	.6628	.3372	.3653
0.43	.1664	.6664	.3336	.3637
0.44	.1700	.6700	.3300	.3621

(1) z Puntuación tipificada	(2) A Area desde la media $\frac{a}{a}$	(3) B Area de la parte mayor	(4) e Area de la parte menor	(5) y Ordenada en $\frac{x}{a}$
0.45	.1736	.6736	.3264	.360.5
0.46	.1772	.6772	*.3228*	.3.589
0.47	.1808	.6808	.3192	.3572
0.48	.1844	.6844	.3156	.3.5.5.5
0.49	.1879	.6879	.3121	.3538
0..50	.1915	.6915	.3085	.3521
0.51	.1950	*.6950*	.3050	.3503
0.52	.1985	.6985	.3015	.3485
0..53	.2019	.7019	.2981	.3467
0.54	.2054	.7054	.2946	.3448
0.55	.2088	.7088	.2912	.3429
0.56	.2123	.7123	.2877	.3410
0..57	.21.57	.7157	.2843	.3391
0..58	.2190	.7190	.2810	.33'12
0..59	.2224	.7224	.2776	.3352
0.60	.2257	.72.57	.2743	.3332
0.61	.2291	.7291	.2709	.3312
0.62	.2324	.7324	.2676	.329
0.63	.2357	.7357	.2643	.3271
0.64	.2389	.7389	.2611	.3251
0.65	.2422	.7422	.2578	.323C
0.66	.2454	.7454	.2546	.3209
0.67	.2486	.7486	.2514	.3181
0.68	.2517	.7517	.2483	.316é
0.69	.2549	.7.549	.2451	.3144
0.70	.2580	.7.580	.2420	.312
0.71	.2611	.7611	.2389	.3101
0.72	.2642	.7642	.2358	.3079
0.73	*.2613*	.7673	.2327	.3056
0.74	.2704	.7704	.2296	.3034
0.7.5	.2734	.7734	.2266	.3011
0.76	.2764	.7764	.2236	.2989
0.77	.2794	.7794	.2206	.2966
0.78	.2823	.7823	.2117	.2943
0.79	.2852	.7852	.2148	.2920
0.80	.2881	.7881	.2119	.2897
0.81	.2910	.7910	.2090	.2874
0.82	.2939	.7939	.2061	.28.50
0.83	.2967	.7967	.2033	.2827
0.84	.2995	.7995	.2005	.2803
0.85	.3023	.8023	.1977	.2780
0.86	.3051	.80.51	.1949	.27.56
0.87	.3078	.8078	.1922	.2732
0.88	.3106	.8106	.1894	.2709
0.89	.3133	.8133	.1867	.2685

(1) Puntuación tipificada	(2) Area desde la media a	(3) Area de la parte mayor	(4) Area de la parte menor	(5) Ordenada en
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0193	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
0.98	.3365	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444
1.00	.3413	.8413	.1587	.2420
1.01	.3438	.8438	.1562	.2396
1.02	.3461	.8461	.1539	.2371
1.03	.3485	.8485	.1515	.2347
1.04	.3508	.8508	.1492	.2323
1.05	.3531	.8531	.1469	.2299
1.06	.3554	.8554	.1446	.2275
1.07	.3577	.8577	.1423	.2251
1.08	.3599	.8599	.1401	.2227
1.09	.3621	.8621	.1379	.2203
1.10	.3643	.8643	.1357	.2179
1.11	.3665	.8665	.1335	.2155
1.12	.3686	.8686	.1314	.2131
1.13	.3708	.8708	.1292	.2107
1.14	.3729	.8729	.1271	.2083
1.15	.3749	.8749	.1251	.2059
1.16	.3770	.8770	.1230	.2036
1.17	.3790	.8790	.1210	.2012
1.18	.3810	.8810	.1190	.1989
1.19	.3830	.8830	.1170	.1965
1.20	.3849	.8849	.1151	.1942
1.21	.3869	.8869	.1131	.1919
1.22	.3888	.8888	.1112	.1895
1.23	.3907	.8907	.1093	.1872
1.24	.3925	.8925	.1075	.1849
1.25	.3944	.8944	.1056	.1826
1.26	.3962	.8962	.1038	.1804
1.27	.3980	.8980	.1020	.1781
1.28	.3997	.8997	.1003	.1758
1.29	.4015	.9015	.0985	.1736
1.30	.4032	.9032	.0968	.1714
1.31	.4049	.9049	.0951	.1691
1.32	.4066	.9066	.0934	.1669
1.33	.4082	.9082	.0918	.1647
1.34	.4099	.9099	.0901	.1626

(1) Puntuación tipificada	(2) Arca desde la media	(3) Arca de la parte mayor	(4) Arca de la parte menor	(5) Ordenada en
1.35	.4115	.911 S	.0885	.1604
1.36	.4131	.9131	.0869	.1.582
1.37	.4147	.9147	.0853	.1.561
1.38	.4162	.9162	.0838	.1.539
1.39	.4177	.9177	.0823	.1.518
1.40	.4192	.9192	.0808	.1497
1.41	.4207	.9207	.0793	.1476
1.42	.4222	.9222	.0778	.14.56
1.43	.4236	.9236	.0764	.1435
1.44	.42.51	.92.51	.0749	.141.5
1.45	.4265	.9265	.0735	.1394
1.46	.4279	.9279	.0721	.1374
1.47	.4292	.9292	.0708	.1354
1.48	.4306	.9306	.0694	.1334
1.49	.4319	.9319	.0681	.1315
1.50	.4332	.9332	.0668	.129.5
1.S1	.4345	.9345	.06.5.5	.1276
1.52	.43.57	.9357	.0643	.12.57
1.53	.4370	9370	.0630	.1238
1.54	.4382	.9382	.0618	.12!9
1.55	.4394	.9394	.0606	.1200
1.56	.4406	.9406	.0.594	.1182
1.57	.4418	.9418	.0582	.1163
1.58	.4429	.9429	.0571	.1145
1.59	.4441	.9441	.0559	.1127
1.60	.4452	.9452	.0548	.1109
1.61	.4463	.9463	.0.537	.1092
1.62	.4474	.9474	.0526	.1074
1.63	.4484	.9484	.0516	.1051
1.64	.4495	.9495	.0505	.1040
1.65	.4505	.950.5	.0495	.1023
1.66	.4515	.9515	.0485	.1006
1.67	.4525	.9525	.0415	.0989
1.68	.4535	.9535	.046.5	.0973
1.69	.4545	.9545	.0455	.09.57
1.70	.4554	.9.554	.0446	.0940
1.71	.4564	.9.564	.0436	.0925
1.72	.4573	.9573	.0427	.0909
1.73	.4582	.9582	.0418	.0893
1.74	.4591	.9591	.0409	.0878
1.15	.4599	.9599	.0401	.0863
1.76	.4608	.9608	.0392	.0848
1.77	.4616	.9616	.0384	.0833
1.78	.4625	.9625	.0375	.0818
1.79	.4633	.9633	.0367	.0804

(1) Puntuación tipificada	(2) Area desde la media a	(3) Area de la parte mayor	(4) Area de la parte menor	(5) Ordenada en
1.80	.4641	.9641	.0359	.0790
1.81	.4649	.9649	.0351	.0775
1.82	.4656	.9656	.0344	.0761
1.83	.4664	.9664	.0336	.0748
1.84	.4671	.9671	.0329	.07 4
1.85	.4648	.9678	.0322	.0721
1.86	.4686	.9686	.0314	.0707
1.87	.4693	.9693	.0307	.0694
1.88	.4699	.9699	.0301	.0681
1.89	.4706	.9706	.0294	.0669
1.90	.4713	.9713	.0287	.0656
1.91	.4719	.9719	.0281	.0644
1.92	.4726	.9726	.0274	.0632
1.93	.4732	.9732	.0268	.0620
1.94	.4738	.9738	.0262	.0608
1.95	.4744	.9744	.0256	.0596
1.96	.4750	.9750	.0250	.0584
1.97	.4756	.9756	.0244	.0573
1.98	.4761	.9761	.0239	.0562
1.99	.4767	.9767	.0233	.0551
2.00	.4772	.9772	.0228	.0540
2.01	.4778	.9778	.0222	.0529
2.02	.4783	.9783	.0217	.0519
2.03	.4788	.9788	.0212	.0508
2.04	.4793	.9793	.0207	.0498
2.05	.4798	.9798	.0202	.0488
2.06	.4803	.9803	.0197	.0478
2.07	.4808	.9808	.0192	.0468
2.08	.4812	.9812	.0188	.0459
2.09	.4817	.9817	.0183	.0449
2.10	.4821	.9821	.0179	.0440
2.11	.4826	.9826	.0174	.0431
2.12	.4830	.9830	.0170	.0422
2.13	.4834	.9834	.0166	.0413
2.14	.4838	.9838	.0162	.0404
2.15	.4842	.9842	.0158	.0396
2.16	.4846	.9846	.0154	.0387
2.17	.4850	.9850	.0150	.0379
2.18	.4854	.9854	.0146	.0371
2.19	.4857	.9857	.0143	.0363
2.20	.4861	.9861	.0139	.0355
2.21	.4864	.9864	.0136	.0347
2.22	.4868	.9868	.0132	.0339
2.23	.4871	.9871	.0129	.0332
2.24	.4875	.9875	.0125	.0325

(1) Puntuación tipificada	(2) Area desde la media a	(3) Area de la parte mayor	(4) Area de la parte menor	(5) Ordenada en
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
1.:n	A%9%	.9 9%	.0\01	.0110
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.OOJ5	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
2.57	.4949	.9949	.0051	.0147
2.58	.4951	.9951	.0049	.0143
2.59	.4952	.9952	.0048	.0139
2.60	.4953	.9953	.0047	.0136
2.61	.4955	.9955	.0045	.0132
2.62	.4956	.9956	.0044	.0129
2.63	.4957	.9957	.0043	.0126
2.64	.4959	.9959	.0041	.0122
2.65	.4960	.9960	.0040	.0119
2.66	.4961	.9961	.0039	.0116
2.67	.4962	.9962	.0038	.0113
2.68	.4963	.9963	.0037	.0110
1.69	.4964	.9964	.0036	.0107

(1) z Puntuación tipificada ( )	(2) A Area desde la media a_{G}	(3) B Area de la parte mayor	(4) e Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
2.70	.4965	.9965	.0035	.0104
2.71	.4966	.9966	.0034	.0101
2.72	.4967	.9967	.0033	.0099
2.73	.4968	.9968	.0032	.0096
2.74	.4969	.9969	.0031	.0093
2.75	.4970	.9970	.0030	.0091
2.76	.4971	.9971	.0029	.0088
2.77	.4972	.9972	.0028	.0086
2.78	.4973	.9973	.0027	.0084
2.79	.4974	.9974	.0026	.0081
2.80	.4974	.9974	.0026	.0079
2.81	.4975	.9975	.0025	.0077
2.82	.4976	.9976	.0024	.0075
2.83	.4977	.9977	.0023	.0073
2.84	.4977	.9977	.0023	.0071
2.85	.4978	.9978	.0022	.0069
2.86	.4979	.9979	.0021	.0067
2.87	.4979	.9979	.0021	.0065
2.88	.4980	.9980	.0020	.0063
2.89	.4981	.9981	.0019	.0061
2.90	.4981	.9981	.0019	.0060
2.91	.4982	.9982	.0018	.0058
2.92	.4982	.9982	.0018	.0056
2.93	.4983	.9983	.0017	.0055
2.94	.4984	.9984	.0016	.0053
2.95	.4984	.9984	.0016	.00.SJ
2.96	.4985	.9985	.0015	.0050
2.97	.4985	.9985	.0015	.0048
2.98	.4986	.9986	.0014	.0047
2.99	.4986	.9986	.0014	.0046
3.00	.4987	.9987	.0013	.0044
3.01	.4987	.9987	.0013	.0043
3.02	.4987	.9987	.0013	.0042
3.03	.4988	.9988	.0012	.0040
3.04	.4988	.9988	.0012	.0039
3.05	.4989	.9989	.0011	.0038
3.06	.4989	.9989	.0011	.0037
3.07	.4989	.9989	.0011	.0036
3.08	.4990	.9990	.0010	.0035
3.09	.4990	.9990	.0010	.0034
3.10	.4990	.9990	.0010	.0033
3.11	.4991	.9991	.0009	.0032
3.12	.4991	.9991	.0009	.0031
3.13	.4991	.9991	.0009	.0030
3.14	.4992	.9992	.0008	.0029

(1) - Puntuación tipificada ( )	(2) A Area desde la media $a_{(1}$	(3) B Area de la parte mayor	(4) e Area de la parte menor	15) y Ordénada en $\frac{x}{-}$
3.15	.4992	.9992	.0008	.0028
3.16	.4992	.9992	.0008	.0027
3.17	.4992	.9992	.0008	.0026
3.18	.4993	.9993	.0007	.0025
3.19	.4993	.9993	.0007	.0025
3.20	.4993	.9993	.0007	.0024
3.21	.4993	.9993	.0007	.0023
3.22	.4994	.9994	.0006	.0022
3.23	.4994	.9994	.0006	.0022
3.24	.4994	.9994	.0006	.0021
3.30	.4995	.9995	.0005	.0017
3.40	.4997	.9997	.0003	.0012
3.50	.4998	.9998	.0002	.0009
3.60	.4998	.9998	.0002	.0006
3.70	.4999	.9999	.0001	.0004