Universidad Central de Venezuela
Facultad de Humanidades y Educación
Escuela de Educación
Cátedra: Métodos Cuantitativos
Asignatura: Estadística Aplicada a la Educación

Coeficientes Correlación: Phi, Contingencia, Biserial, Spearman

Prof: Johnnalid González G.

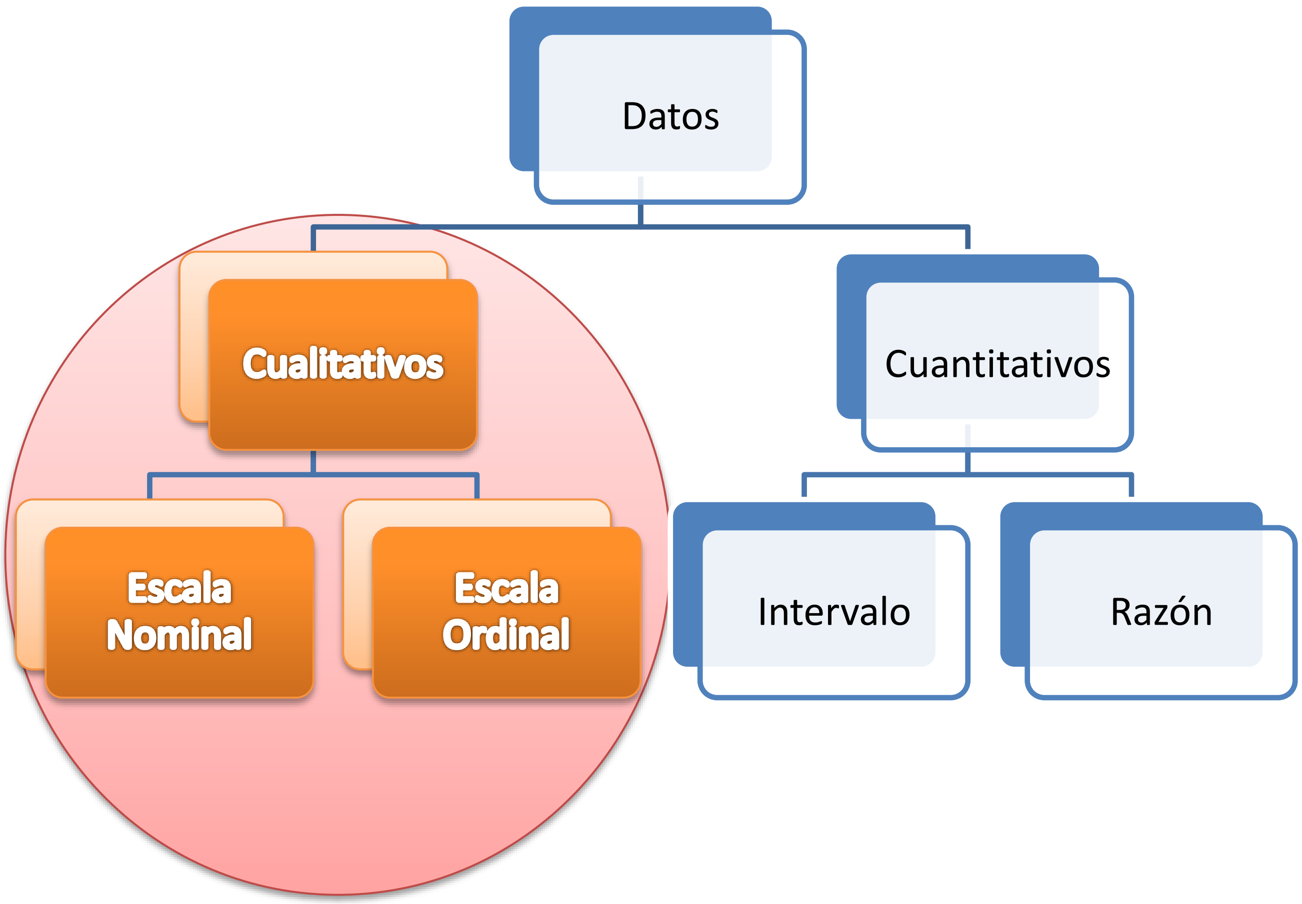
2018

Introducción

El presente tema pretende estudiar coeficientes de correlación, para determinar el grado de asociación de variables cualitativas, es decir, aquellas variables cuyas modalidades no se pueden cuantificar.

La técnica de la correlación ordinal puede aplicarse también a mediciones, si a estos valores se les asigna un número de acuerdo a su cuantía y característica. Ej: los resultados de una prueba (estadística, inglés u otra asignatura) pueden ser representados por medio de números ordinales que indican las posiciones o lugares ocupado por quienes la presentaron.

Estructura de las Escalas de Medición



Correlación Ordinal

Cuando los valores de dos variables están representados por números ordinales, para calcular la magnitud de la posible relación entre ellas, se utilizan los llamados coeficientes de correlación ordinal o de rango, siendo los de mayor uso, el de Spearman y el de Kendall.

Fuente: Chourio, Hugo.2011:175.

Correlación Ordinal

Cuando los valores de dos variables están representados por números ordinales, para calcular la magnitud de la posible relación entre ellas, se utilizan los llamados coeficientes de correlación ordinal o de rango, siendo los de mayor uso, el de Spearman y el de Kendall.

Fuente: Chourio, Hugo.2011:175.

Coeficiente de Spearman

Aspectos a considerar:

✓Es el coeficiente más empleado en los métodos de correlación por rangos
✓Se recomienda usar este método, con datos entre 25 o 30 o menos
✓Las variables son medidos en escalas ordinales
✓Es más fácil y rápido de calcular que el Coeficiente de Correlación de Pearson.

Propiedades:

✓Toma valores entre $-1 < r_s <+1$
✓El coeficiente rs es un caso particular de ρ_{xy}
✓Si calculamos el coeficiente de correlación de Pearson entre dos variables X e Y, y el coeficiente de correlación de Spearman para las mismas puntuaciones pero transformadas en rangos, ambos coeficientes se aproximan en valor según aumenta el número de sujetos n.

Coeficiente de Spearman

APLICACIÓN DEL COEFICIENTE DE SPEARMAN:

Un hecho común en los estudios de investigación es que las respuestas a las preguntas en que los investigadores están más interesados sólo se pueden medir con escalas ordinales o incluso nominales.

Por ejemplo:

Si estuvieran interesados en estudiar el consumo de café entre hombres y mujeres. En este caso, aplicar el coeficiente de correlación de Pearson a los datos podría suponer que estas medidas de sexo tienen propiedades de escala de intervalo o de razón y posiblemente producirá resultados engañosos o exagerados.

Coeficiente de Spearman

Definición:

Es un caso especial del Coeficiente de Correlación de Pearson, cuando los valores se presentan como los primeros números consecutivos.

Si los rangos se tratan como puntajes u otros valores, y no hay rangos empates, entonces es cierto que: $\rho = r_s$, entonces la fórmula a emplear sería:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n+1)(n-1)}, -1 \leq r_s \leq +1$$

Coeficiente de Spearman

Pasos para el cálculo del Coeficiente de Spearman:

1) Verificar que los datos estén dados en rangos o posiciones

2) Establecer la diferencia entre los rangos ocupados por cada variable de estudio.

3) Elevar al cuadrado cada una de estas diferencias

4) Aplicar la fórmula

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n+1)(n-1)}, -1 \leq r_s \leq +1$$

5) Interpretar el resultado

Coeficiente de Spearman

Ejemplo 1:

Se desea conocer el grado de relación entre las posiciones que ocuparon 10 atletas que tomaron parte en dos pruebas de 100 (X_i) y 200 (Y_i) mts planos. Los resultados se muestran a continuación:

Atleta	A	B	C	D	E	F	G	H	I	J
100	1	2	4	3	5	6	7	8	10	9
200	2	1	3	4	6	5	7	8	9	10

Coeficiente de Spearman

Procedimiento:

1) Los valores están dados directamente en rangos

2) Establecer la diferencia entre los lugares ocupados por cada atleta:
   $d_i = X_i - Y_i$

Atleta	A	B	C	D	E	F	G	H	I	J
100 (X_i)	1	2	4	3	5	6	7	8	10	9
200 (Y_i)	2	1	3	4	6	5	7	8	9	10
di	-1	1	1	-1	-1	1	0	0	1	-1

Coeficiente de Spearman

Procedimiento:

3) Elevar al cuadrado cada diferencia y finalmente sumar el cuadrado de tales diferencias. $\sum d_i^2$.

Atleta	A	B	C	D	E	F	G	H	I	J
100 (X_i)	1	2	4	3	5	6	7	8	10	9
200 (Y_i)	2	1	3	4	6	5	7	8	9	10
di	-1	1	1	-1	-1	1	0	0	1	-1
di^2	1	1	1	1	1	1	0	0	1	1

$= \sum d_i^2 = 8$

Coeficiente de Spearman

4) Aplicar la fórmula

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n+1)(n-1)} = 1 - \frac{6*8}{10*11*9} = 0,95$$

5) **Interpretación**: Los atletas que lograron las mejores posiciones en la prueba de 100 metros planos, tienden en forma muy alta a obtener las mejores posiciones en la prueba de 200 metros planos.

Coeficiente de Spearman

Ejemplo 2:

Calcular e interpretar el Coeficiente de Correlación de Spearman, entre los resultados obtenidos por un grupo de estudiantes de tercer año en dos pruebas objetivas finales de lapso, Castellano (X_i) y Cs. Biológicas (Y_i).

Estudiantes	A	B	C	D	E	F	G
Castellano	48	47	46	46	45	43	43
Cs Biológicas	25	25	19	12	12	12	11

Coeficiente de Spearman

Procedimiento:

1) Se convierte los puntajes en posiciones, de la siguiente manera: en la Variable X_i (Castellano), el estudiante A obtuvo la mayor puntuación, entonces se le asigna la posición 1 y asi sucesivamente, en el caso de los estudiantes C y D, tienen las mismas puntuaciones, es decir se tiene un "empates", **se deben sumar los lugares que les tocarían si no estuviesen empates y se divide entre el número de valores iguales,** de la misma forma se le aplicaría a los estudiantes F y G.

Estudiantes	A	B	C	D	E	F	G
Castellano (X_i)	48	47	46	46	45	43	43
Cs Biológicas (Y_i)	25	25	19	12	12	12	11
($X_i´$)	1	2	3,5	3,5	5	6,5	6,5
($Y_i´$)	1,5	1,5	3	5	5	5	7

Coeficiente de Spearman

Procedimiento:

2) Establecer la diferencia entre los lugares ocupados por cada atleta: $d_i = X_i´ - Y_i´$.

3) Elevar al cuadrado cada diferencia y finalmente sumar el cuadrado de tales diferencias. $\sum d_i^2$.

Estudiantes	A	B	C	D	E	F	G
($X_i´$)	1	2	3,5	3,5	5	6,5	6,5
($Y_i´$)	1,5	1,5	3	5	5	5	7
di	-0,5	0,5	0,5	-1,5	0	1,5	-0,5
di^2	0,25	0,25	0,25	2,25	0	2,25	0,25

$= \sum d_i^2 = 5,5$

Coeficiente de Spearman

4) Aplicar la fórmula

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n+1)(n-1)} = 1 - \frac{6 * 5,5}{7 * 8 * 6} = 0,90$$

5) **Interpretación**: Los estudiantes que lograron las mejores notas en la prueba de Castellano, tienden muy altamente a obtener las mejores calificaciones en la prueba de Ciencias Biológicas.

Coeficiente de Correlación Biserial

Definición:

El coeficiente de correlación biserial, se utiliza cuando queremos conocer la correlación existente entre dos variables, de las cuales una ha sido considerada como escala de intervalos y la otra resulta ser una variable dicotómica (significa que toma dos modalidades. Ej: sexo, si o no, etc).

Observación: Cuando la variable continua dicotomizada, o ambas variables, se desvían demasiado de la distribución normal, el valor calculado del coeficiente de correlación r_b , es mayor que la unidad.

Coeficiente de Correlación Biserial

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

$$r_b = \frac{\bar{x}_p - \bar{x}_q}{st} * \frac{p * q}{y} \quad \text{ó} \quad r_b = \frac{\bar{x}_p - \bar{x}_t}{St} * \frac{p}{y}$$

Donde:

$\bar{x}_p$ Media aritmética de la categoría “p”

st Desviación típica total

p es la proporción de observaciones de unas de dos modalidades

y Altura de la ordenada que separa en la curva normal a las proporciones “p” y “q”

Coeficiente de Correlación Biserial

Ejemplo 1:

Un grupo de bachilleres aspirantes a FACES de la UCV, fue sometido a una prueba objetiva de conocimientos generales **(PCG: X_i)** y a un test de aptitud **(Y_i)** hacia las carreras vigentes en esa Facultad. Se desea saber qué relación existe entre los puntajes obtenidos en la prueba y los resultados del test. El test ha sido dicotomizado así: Con Vocación (CV) y Sin Vocación (SV). Los resultados fueron los siguientes:

ASP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
PCG	54	53	60	44	39	34	36	43	49	62	66	46	44	49	56	48
TEST	28	24	30	17	20	16	18	22	22	28	30	22	17	18	15	24
Asp	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
PCG	45	35	65	54	48	56	49	57	54	40	49	50	43	68	44	68
TEST	20	16	25	27	26	29	22	26	22	14	18	23	22	29	16	28

Coeficiente de Correlación Biserial

Pasos para el cálculo del Coeficiente de Correlación Biserial:

1) Determinar las características de las variables
2) Se distribuye la variable continua no dicotomizada en intervalo de clase
3) Se calcula el rango y el número de clases
4) Construir la tabla de distribución de la variable Y_i
5) Se calcula el factor dicotomizador de la variable Y_i , para obtener las dos categorías, a la cual, llamaremos "p" y "q".
6) Se construye la tabla de Dicotomización de la variable Y_i y operaciones

Dicotomización de la variable Y_i y operaciones

Coeficiente de Correlación Biserial

Cálculos:

Paso 5:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{714}{32} = 22{,}31 ptos.$$

Se calcula el Coeficiente de Correlación Biserial

$$r_b = \frac{\bar{x}_p - \bar{x}_t}{St} * \frac{p}{y} = \frac{58{,}28 - 50{,}75}{9{,}06} * \left(\frac{0{,}44}{0{,}3944}\right) = 0{,}93$$

Interpretación: Los bachilleres que lograron las mejores puntuaciones en la prueba de conocimientos generales, tienden muy altamente a obtener las mejores calificaciones en el test y viceversa.

Coeficiente de Correlación Phi (ϕ)

Aspectos a considerar:

✓Se utiliza cuando las dos variables son dicotómicas: Verdadero o Falso, Si o No.

✓Este coeficiente al igual que el Coeficiente de Correlación Biserial Puntual, son coeficientes producto-momento, del tipo de Pearson, aun cuando no posean la misma precisión de este último.

✓Es el más utilizado en análisis de ítems.

✓Cuando las dos variables se reparten por igual, los límites máximos del coeficiente de correlación Phi se encuentra entre -1 y +1.

Coeficiente de Correlación Phi (ϕ)

La fórmula a emplear sería:

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(C+D)(B+D)(A+C)}}, -1 \leq \phi \leq +1$$

Donde las letras A,B, C y D, representan las frecuencias de la siguiente tabla de doble entrada:

(+ -) B	(+ +) A	(A + B) p	Las proporciones son:
(- -) D	(- +) C	(C+ D) q	p=(A+B)/n q=(C+D)/n
(B + D) q´	(A + C) p´	n	p´=(A+C)/n q´=(B+D)/n

Coeficiente de Correlación Phi (ϕ)

Ejemplo:

Diez estudiantes presentaron un examen de Sociología y se desea conocer la relación que existe entre las repuestas a los ítems 7 y 20. los resultados observados fueron los siguientes: “0 = no respondió” y “1 = respondió mal”

Alumno	A	B	C	D	E	F	G	H	I	J
Ítem 7 (X_i)	0	1	1	0	0	1	1	1	0	1
Ítem 20 (Y_i)	0	0	1	0	1	1	1	1	0	1

Coeficiente de Correlación Phi (ϕ)

Procedimiento:

1) Se considera la tabla de los signos con sus respectivas letras.
2) Se colocan signos: positivo a una categoría y negativo a la otra. En este ejercicio se le colocó el signo positivo (+) al "1" y negativo (-) al "0".

Alumno	Ítem 7	Ítem 20				
A	-0	-0		(+ -) B*	(+ +) A*****	(A + B) p=6
B	+1	-0		(- -) D***	(- +) C*	(C+ D) q=4
C	+1	+1		(B + D) q´=4	(A + C) p´=6	n=10
D	-0	-0				
E	-0	+1				
F	+1	+1				
G	+1	+1				
H	+1	+1				
I	-0	-0				
J	+1	+1				

Coeficiente de Correlación Phi (ϕ)

La fórmula a emplear sería:

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(C+D)(B+D)(A+C)}} = \frac{5*3-1}{\sqrt{6*4*6*4}} = 0{,}58$$

Interpretación:

Los estudiantes que respondieron correctamente al ítem 7 tienden moderadamente a acertar el ítem 20 o viceversa.

Coeficiente de Contingencia

Definición:

Este coeficiente se aplica para variables nominales, se presenta en forma de tabla de doble entrada con variables que expresan “atributos”, donde la í-esima fila y j-esima columna, denominada O_{ij}, describen cada una de las frecuencias observadas asociadas a los atributos.

Características:

- ✓No existe relación entre las variables, por tanto diremos que estas tendrán una proporción similar.
- ✓Se utiliza para evitar el efecto del tamaño de la muestra.
- ✓En una tabla de dos filas por columna es recomendable realizar la corrección de Yates.

Propiedades:

- ✓Su valor se encuentra entre -1 y +1
- ✓Mide la intensidad de la relación
- ✓El valor de C depende del número de filas y columnas de la tabla de contingencia construido para su calculo.

Coeficiente de Contingencia

Su fórmula es:

$$\chi^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(e_{ij} - o_{ij})^2}{e_{ij}}$$

Donde:
χ^2 = Chi – Cuadrado
e_{ij} = frecuencia esperada por fila y columna
O_{ij} = frecuencia observada por fila y columna

$$e_{ij} = \frac{f_f * f_c}{n}$$

Donde:
f_f = Frecuencias marginal por fila
f_c = Frecuencias marginal por columna
n = Número total de observaciones

Interpretación:

Si $\chi^2 = 0$, entonces, hay independencia entre las variables
Si $\chi^2 > 0$, entonces, hay mayor grado de asociación entre variables
Si $\chi^2 < 0$, entonces, hay menor grado de asociación entre variables

Observación: El mayor inconveniente que tiene este coeficiente es que es proporcional al número de observaciones, y por tanto no tiene una cota, por lo que no es muy adecuado su uso.

Coeficiente de Contingencia

Ejemplo:

Se desea determinar si existe relación entre el sexo y la especialidad cursada para los alumnos que estudian en el Magisterio.

Sexo	Especialidad Académica				Total
	Ciencias	Humanas	Lengua	Preescolar	
Hombre	70	60	36	12	178
Mujer	40	54	39	38	171
Total	110	114	75	50	349

Coeficiente de Contingencia

Paso 1:

Calcular las frecuencias esperadas para la i-ésima fila y la j-ésima columna simultáneamente, mediante la siguiente fórmula:

$$e_{ij} = \frac{f_f * f_c}{n}$$

Paso 2:

Calcular el valor del chi cuadrado, mediante la siguiente fórmula:

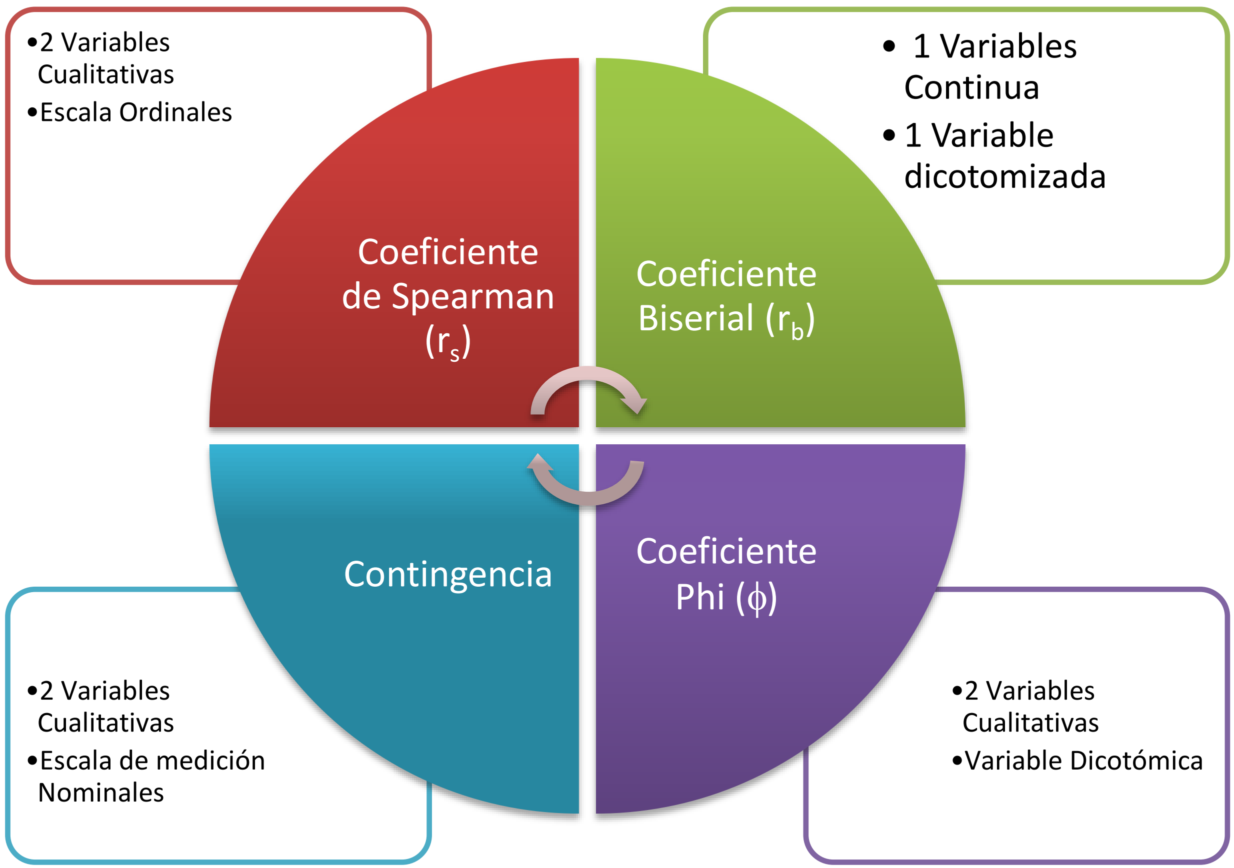
$$\chi^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(e_{ij} - o_{ij})^2}{e_{ij}}$$

Paso 3:

Calcular el valor de "C":

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

Resumen de Coeficientes de Correlación Ordinal



Reglas prácticas acerca de la fuerza de los Coeficientes de Correlación

Rango de Coeficiente	Descripción de la Fuerza
± 0.81 a ± 1.0	Muy fuerte
± 0.61 a ± 0.8	Fuerte
± 0.41 a ± 0.6	Moderada
± 0.21 a ± 0.4	Débil
± 0.00 a ± 0.2	Ninguna

Lista de Coeficientes según el tipo de variable

Variables o Escalas	Coeficientes
Dos cuantitativas (intervalo o de razón)	Coeficientes de Correlación de Pearson
cuantitativas y ordinal o dos ordinales	Coeficientes de Correlación de Spearman
dos nominales o nominal y cuantitativa	Coeficiente de Contingencia
cuantitativas y cuantitativa con dicotomia artificial	Coeficientes de Correlación Biseral
cuantitativas y cuantitativa con dicotomia auténtica	Coeficientes de Correlación Punto Biseral
dos nominales de dicotomia auténtica	Coeficiente Phi(ϕ)
dos de dicotomia artificial (con "n" mayor de 100)	Correlacion Tetracónica