**COEFICIENTE DE CORRELACIÓN TETRACÓRICO (rt)**

Báez, E. Barrios, G. Pernia, G, Salazar, D. Trejo, A.

Universidad Central De Venezuela

Junio, 2018

El coeficiente de correlación Tetracórico expresado como (rt) se utiliza cuando las dos variables X y Y son continuas dicotomizadas, de una manera más precisa, cuando las variables continuas han sido reducidas de manera artificial a dos categorías, así mismo, es apreciado como una aproximación a el coeficiente de correlación r de Pearson, debido a que si se presenta en las condiciones adecuadas da un coeficiente que es numéricamente semejante. Es importante señalar que el coeficiente Tetracórico (rt) requiere que las variables X y Y sean continuas, con una distribución normal y que estén relacionadas de forma lineal (Guilford, y Fruchter, 1984).

Con respecto al cálculo de la correlación tetracórica, en el caso de que en una de las casillas se presente una frecuencia que represente menos del 10% del total de las frecuencias, entonces se procede a calcular por medio de la fórmula del coeficiente Phi, aunque si se está realizando una investigación lo más indicado seria repetir la experiencia. De modo tal, como estadístico, el coeficiente Tetracórico (rt) posee menor confianza que el coeficiente r de Pearson, puesto que su empleo tiene poca justificación, ya que puede utilizarse en su lugar el coeficiente Phi (Chourio, J, 1987).

**Requisitos**

La aplicación de este coeficiente, según Hoffman, Stover, Iglesia y Fernández (2013) exigen tener en cuenta una serie de supuestos propios de variables paramétricas, como los siguientes:

* **Normalidad:** Esta correlación opera bajo el supuesto de normalidad multivariada de los ítems en análisis, la violación del mismo no trae consecuencias estadísticamente significativas sobre los resultados
* **Continuidad:** Se destaca que las variables observables categóricas (normal y nominal) por su propia naturaleza quedan excluidas del nivel de medición continuo (intervalo y razón) esto no supone un inconveniente para los análisis debido a que tetracórica es calculada a partir de variables latentes (las que si se suponen continuas), que subyacen a las observables.
* **Amplitud de la muestra:** Es importante el tamaño de la muestra, en donde el número de casos necesarios se encuentra estrechamente relacionado con la cantidad de reactivos incluidos en el análisis.
* **Número de reactivos:** Se recomienda en el caso de utilizar estadísticos no paramétricos, no superar los 30 debido a que los análisis que subyacen al procedimiento. Solo es posible exceder esa cantidad si se emplean métodos robustos.

También, según Hoffman y colaboradores (2013) y utilizando lo expuesto por Barajas y Pietro (2015) son necesarios otros requisitos tales como:

* Las variables tienen que ser de intervalo y razón para luego ser dicotomizadas
* Utilizar muestras grandes de 100 sujetos
* Ninguna de las celdas debe ser igual a 0 y tampoco un número menor al 10% del total de las frecuencias.
* Deben respetar el principio de linealidad
* Deben cumplir el principio de la distribución normal

**Paso para la aplicación de Tetracórica**

1. Se deben dicotomizar las dos variables a partir de su respectiva media o un punto de corte, de esta forma la distribución queda dividida en cuatro categorías o en una tabla de contingencia donde aquellos sujetos que tienen una puntuación por encima (1) y por debajo (0) de la media de cada una de las variables.

Según Abascal y Grande hay dos métodos para calcularlo:

* Método A: **rt= Cos ( 180 / 1+ √ad/cb) o rt = cos**

En donde "A" y “D” son valores de igual signo y “C” y “B” son de diferente signo.

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| W | (A) | (B) |
| Z | (C) | (D) |

**Ejemplo:**

Se realizó un estudio a 10 estudiantes que estudian medicina que cursan química, en el cual se les aplicó un examen para evaluar los conocimientos de la materia, se desea determinar si existe una relación entre las horas de estudio y las notas adquiridas.

|  |  |  |
| --- | --- | --- |
| **Sujetos** | **Horas de estudio (X)** | **Notas (Y)** |
| 1 | 2 | 5 |
| 2 | 3 | 5 |
| 3 | 6 | 18 |
| 4 | 6 | 9 |
| 5 | 7 | 8 |
| 6 | 7 | 14 |
| 7 | 7 | 14 |
| 8 | 10 | 15 |
| 9 | 10 | 16 |
| 10 | 10 | 17 |

**Paso #1**

Sacar la media para X: ; 68/10= 6.8

Sacar la media para Y: En este caso seria 10 por ser la nota mínima aprobatoria.

**Paso #2**

Se van a ubicar a los sujetos por sus puntajes, es decir, si el puntaje del sujeto para la variable X es inferior a su punto de corte o a la media recibe un signo negativo (-) o si el puntaje del sujeto es mayor al punto de corte recibirá un signo positivo (+); esto también pasará con la variable Y, luego de que estén todos los puntajes con los signos se llevara a cabo ubicarlos por cuadrantes según el signo de ambas variables.

|  |  |  |
| --- | --- | --- |
| HORAS DE ESTUDIO | NOTAS | |
| **0** | **1** |
| **0** | (A) | (B) |
| **1** | (C) | (D) |

|  |  |  |
| --- | --- | --- |
| **Sujetos** | **Horas de estudio (X)** | **Notas (Y)** |
| 1 | - | - |
| 2 | - | - |
| 3 | - | + |
| 4 | - | - |
| 5 | + | - |
| 6 | + | + |
| 7 | + | + |
| 8 | + | + |
| 9 | + | + |
| 10 | + | + |

**Paso #3**

Luego se distribuyen en los cuatro cuadrantes los puntajes de los sujetos y se toma de cada uno de los cuadrantes la frecuencia de sujetos que se encuentran en cada uno, para luego aplicar la fórmula de **rt= Cos ( 180 / 1+ √ad/cb) ó rt =cos**

|  |  |  |  |
| --- | --- | --- | --- |
| NOTAS  HORAS DE ESTUDIO | 0 | 1 | TOTAL |
| 0 | 3(A) | 1(B) | **A+B=4** |
| 1 | 1(C) | 5(D) | **C+D=6** |
| TOTAL | **A+C=4** | **B+D=6** | **N=10** |

Cuadrante

* A: 3; B:1; C:1; D:5

Se aplica la fórmula:

rt= cos

rt= cos 36,94

rt= 0.80

Este método no requiere el uso de tablas de correlación para obtener el valor **rt**, sin embargo es necesario el uso de una calculadora científica ya que implica el cálculo de un coseno y una raíz cuadrada.

* Método B diagonal de Davidoff y Cohen:

Cuando **ad>bc**

rt**= ad/bc**

* Se busca el valor rt que corresponde en una tabla de correlación Tetracórica

En caso de que **bc>ad** se calcula:

**rt=bc/ad**

* Se busca el valor rt que corresponde en una tabla de correlación Tetracórica y se le coloca signo negativo al rt correspondiente.

**Ejemplo:**

Se administró a una muestra de 500 personas, una escala de actitud hacia la legalización de las drogas. Los participantes debían indicar si estaban de acuerdo (1) o en desacuerdo (0) con los ítems. Dos de los presentados fueron los siguientes:

A) Pienso que las drogas deberían ser legales

B) La legalización de las drogas provocaría una situación de descontrol en la sociedad

Los investigadores suponen que una distribución normal bivariante subyace a las puntuaciones de ambos ítems. Los datos se muestran a continuación:

|  |  |  |  |
| --- | --- | --- | --- |
| Drogas legales  Descontrol social | 0 | 1 | total |
| 0 | 125(A) | 125(B) | A+B=250 |
| 1 | 200(C) | 50(D) | C+D=250 |
| total | A+C=325 | B+D=175 | N=500 |

**Paso #1** Multiplicar

**A x D**= 125 x 50= 6250

**B x C**= 125 x 200= 25000

Entonces se busca el valor rt que corresponde en una tabla de correlación Tetracórica, en el caso de que BC sea mayor a AD

**b x c ≥ a x d** por lo tanto el resultado de rt es menor a 0 o negativo.

**Paso #2** Calcular C y R o el cociente

C== = 4; Este resultado se buscará una aproximación del 4 (en este caso) en la tabla de correlación Tetracórica. Luego de haber encontrado los limites inferior y superior, se calculará nuestro siguiente cociente denominado R.

R=

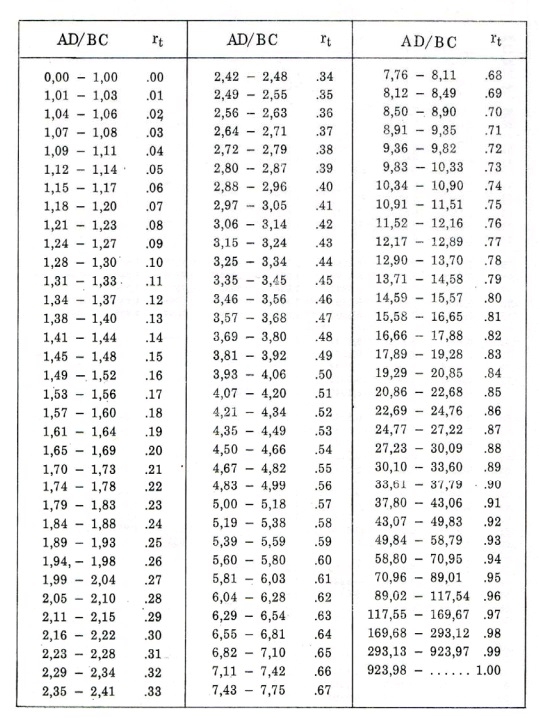
Valores hallados en la tabla:

Linf: 3,93 R= = 0,005

Lsup: 4,06

**Paso #3** En la tabla en el lado derecho de los límites esta la columna rt que en este caso es 0,5; lo sumaremos con nuestro resultado en R que fue 0,005 quedando:

rt= 0,5+0,005= 0,50 ; En este método recordemos que se le coloca negativo el signo, por lo tanto quedaría -0,505.



**Ejemplo**

Se desea conocer la relación entre el nivel de estrés y si fuman o no a grupo de 120 sujetos, para ello se aplica un instrumento que mide la tendencia al estrés físico (que reporta valores por encima de lo estándar)

Los resultados se agrupan en cuadrantes representativos para un plano cartesiano según Guilford, quedando como: A (++), B (-+), C (+-) y D (--)

En los resultados se asigna un valor binario arbitrario respecto a la dicotomizacion de las variables. Para este ejercicio se tomaron en cuenta las respuestas que coincidieron en presencia de estrés y la cantidad de cigarrillos y las que no, por lo que se toma las opciones si y no.

Tabla para el cuadrante Tetracórico

|  |  |  |  |
| --- | --- | --- | --- |
| **Fuman** | **No** | **Si** | **Total** |
| **Estrés** |
| **Si** | 15(B) | 60(A) | 75 |
| **No** | 30(D) | 15(C) | 45 |
| **Total** | 45 | 75 | 120 |

A= 60, B= 15, C= 15 y D=30

Se aplica la fórmula:

**Interpretación**

La interpretación del coeficiente tetracórico es similar a la de Pearson, es decir, se toma en cuenta si hay relación entre las variables, el modulo y el signo. Por lo que se debe expresar si el valor es distinto a 0, que indica que existe asociación entre las variables.

El modulo del coeficiente que sugiere el nivel o fuerza con la que se relacionan esas variables, que circulan entre -1 y 1, tomando como referencia los siguientes intervalos: si esta entre 0,1 y 0,3; se puede indicar que la asociación es baja. Si el valor esta entre 0,4 y 0,6 sugiere que la asociación es moderada. Si el valor esta entre 0,7 y 0,9, se puede indicar que la asociación es fuerte. Teóricamente si el valor es 1 o -1 se considera que la relación es perfecta.

Por último se debe interpretar el signo del coeficiente, si el valor es negativo, indica que la relación entre las variables es inversamente proporcional, y si es positivo, se puede indicar que la relación es directamente proporcional.

En el ejemplo anterior se puede interpretar el resultado de la siguiente forma:

El coeficiente Tetracórico muestra un valor de 0,68 que sugiere una relación moderada y directamente proporcional. Es decir, a medida que aumenta la presencia de sujetos con estrés también hay tendencia a encontrar sujetos que fumen.

**Limitaciones:**

* Una limitación que hay que tomar en cuenta es que debido a la dicotomización artificial que debemos realizar eligiendo un punto de corte, se perderá una cantidad de información en gran o menor medida lo que puede llegar a afectar la correlación de las variables (Guilford, y Fruchter, 1984)
* No se puede utilizar para estudiar variables dicotómicas (Guilford, y Fruchter, 1984). Esto debido a que se necesitan los valores crudos para aplicar las formulas correspondientes.
* Se debe evitar la situación de que hayan más cantidad de puntajes en X que en Y o viceversa, esto debido a que al tener más cantidad de puntajes en una de las dos variables el error de estimación es mucho más alto (Guilford, y Fruchter, 1984).

**Referencias Bibliográficas**

Abascal, E y Grande, I. (2014). *Fundamentos y Técnicas de investigación comercial*. (12ª Ed) Madrid: Edit. Esic.

Bajaras, I; Prieto, M. (2015). *Análisis factorial con variables categóricas.* Tesis de grado. Universidad de Valladolid, España

Chourio, J. (1987). *Etadistica II*. (1era Ed) Caracas: Edit Biosfera.

Guilford , J., y Fruchter, B. (1984). *Estadística aplicada a la Psicología y a la Educación*. Mc Graw Hill.

Hoffman, A; Stover, J; Iglesia, G; Fernández, M. (2013). Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios. *Ciencias Psicológicas, 7, (2).*

S.A. (2014). *Estadística descriptiva II*. Recuperado: junio del 2018, de Proclapar Sitio web:<http://www.proclapar.com/2014/04/04/estadistica-descriptiva-ii-ejercicio-24-coeficiente-de-correlacion-tetracorica-3/#disqus_thread>