LECTURA 09: INTRODUCCIÓN A LA ESTADÍSTICA NO PARAMÉTRICA (PARTE II). PRUEBA DE CORRELACIÓN DE SPEARMAN

TEMA 20: PRUEBA DE CORRELACIÓN DE SPEARMAN

1. INTRODUCCIÒN:

El coefciente de correlación de Spearman es una prueba no paramétrica cuando se desea quiere medir la relación entre dos variables y no se cumple el supuesto de normalidad en la distribución de tales valores.

El coeficiente de correlación de Spearman de designa por r_s .

2. PASOS PARA LLEVAR ACABO LA PRUEBA DEL COEFICIENTE DE CORRELACIÓN DE SPEARMAN:

1. Formulación de hipótesis:

a) Prueba unilateral derecha

H_o: X e Y son mutuamente independientes.

H_1: Existe una tendencia a formar parejas entre los valores grandes de X e Y.

b) Prueba unilateral izquierda

H_o: X e Y son mutuamente independientes.

H_1: Existe una tendencia de los valores grandes de X a formar parejas con los valores pequeños de Y.

c) Prueba bilateral

H_o: X e Y son mutuamente independientes.

H_1: X e Y no son mutuamente independientes.

Las hipótesis unilaterales indicadas en los incisos b) y c) se utilizan, respectivamente, cuando el investigador desea saber si es posible concluir que las variables están directa o inversamente relacionadas. Las hipótesis especificadas en el inciso a) conducen a una prueba bilateral, y se utilizan cuando se desea descubrir cualquier desviación de la independencia.

Elaborado por : Mg. Carmen Barreto R.
Fecha : Marzo 2011
Versión : 2

2. Nivel de significancia: α

3. Estadística de prueba:

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

4. Establecimiento de los criterios de decisión:

Si n está entre 4 y 30, se compara el valor calculado de r_s, con los valores críticos, r_s^* de la tabla P. Para la prueba bilateral, se rechaza Ho en el nivel de significación α si r_s, es mayor que r_s^* o menor que $-r_s^*$ está en la intersección de la columna encabezada por α/2 y el renglón que corresponde a n. Para la prueba unilateral derecha que especifica una correlación directa, se rechaza Ho en el nivel de signficaciòn α si r_s es mayor que r_s^* para α y n. La hipótesis nula Ho, se rechaza en la prueba unilateral izquierda en el nivel de signficaciòn α si r_s es menor que $-r_s^*$ para α y n.

Opcionalmente también puede trabajar con la tabla Q para una prueba bilateral.

Si n > 30, se puede calcular:

$$z = r_s\sqrt{n-1}$$

y utilizar la tabla I y II para obtener los valores críticos.

5. Cálculos:

Se calcula el valor de:

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

o:

$$z = r_s\sqrt{n-1}$$

Pasos para hallar r_s:

a) Clasificar por jerarquìa los valores de X desde 1 hasta n (el número de parejas de valores de X e Y en la muestra). Clasificar por jerarquía los valores de Y desde 1 hasta n.

b) Calcular d_i, para cada pareja de observaciones, restando la jerarquía de Y_i de la jerarquía de X_i.

c) Elevar al cuadrado cada d_i y calcular $\sum d_i^2$, la suma de los valores elevados al cuadrado.

Finalmente calcular :

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

6. Decisión:

Se acepta o rechaza H_o.

OBSERVACIONES:

Las observaciones de igual valor numèrico plantean un problema: el uso de la Tabla P es estrictamente válido solo cuando hay dos valores iguales (a menos que se emplee algún procedimiento aleatorio para cambiar los que sean iguales. Sin embargo, en la práctica, con frecuencia se utiliza la tabla después de que se ha utilizado algún otro método para manejar los valores numèricamente iguales. Si el número de valores iguales es grande, puede utilizarse la siguiente corrección por valores iguales:

$$T = \frac{t^3 - t}{12}$$

donde t es el número de observaciones de igual valor numérico para alguna jerarquia particular. Cuando se utiliza este factor de corrección, r_s, se calcula a partir de:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d_i^2}{2\sqrt{\sum x^2 \sum y^2}}$$

Elaborado por : Mg. Carmen Barreto R.
Fecha : Marzo 2011
Versión : 2

Donde:

$$\sum x^2 = \frac{n^3 - n}{12} - \sum T_x$$

$$\sum y^2 = \frac{n^3 - n}{12} - \sum T_y$$

Además:

T_x= es la suma de los valores de T para diversas jerarquías del valor númerico igual en X.

T_y= es la suma de los valores de T para diversas jerarquías del valor númerico igual enY.

Muchos investigadores señalan que a menos que sea excesivo el número de cantidades iguales, la corrección produce una diferencia muy pequeña en el valor de r_s. Cuando el número de valores iguales es pequeño, puede seguirse el procedimiento habitual de asignar a las observaciones de igual valor númerico la media de las jerarquías que intervienen y proceder con los pasos anteriores.

Ejemplo 1:

En un estudio de la relación entre la edad y los resultados del electroencefalograma /EEG), se recopilaron datos en 20 personas con edades entre 20 y 60 años. La Tabla Nº 3 muestra las edades y un valor de rendimiento del EEG particular para cada una de esas 20 personas. Los investigadores pretenden saber si es posible concluir que este rendimiento del EEG particular tiene relación inversa con la edad a un nivel de signficancia α=0.05.

Tabla Nº 3

Número de individuo	Edad (X)	Valor resultante del EEG(Y)
1	20	98
2	21	75
3	22	95
4	24	100
5	27	99
6	30	65
7	31	64
8	33	70
9	35	85
10	38	74
11	40	68
12	42	66
13	44	71
14	46	62
15	48	69
16	51	54
17	53	63
18	55	52
19	58	67
20	60	55

Solución:

1. **Formulación de hipótesis:**

   Prueba unilateral izquierda

   H_o: El rendimiento del EEG y la edad son mutuamente independientes.

   H_1: Existe una tendencia del rendimiento del EEG a disminuir con la edad.

2. **Nivel de significancia: α=0.05**

3. **Estadística de prueba:**

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

4. Establecimiento de los criterios de decisión.

La hipótesis nula H_o se rechaza en la prueba unilateral izquierda en el nivel de signficación α=0.05 si r_s es menor que $-r_s^*$ para α=0.05 y n=20.

En la tabla P observamos que para α=0.05 y n=20, $-r_s^* = -0.3789$.

5. Cálculos:

Pasos para calcular el valor de r_s:

a) Clasificar por jerarquìa los valores de X desde 1 hasta 20. Clasificar por jerarquía los valores de Y desde 1 hasta 20

b) Calcular d_i, para cada pareja de observaciones, restando la jerarquía de Y_i de la jerarquía de X_i. Ver Tabla Nº 4.

c) Elevar al cuadrado cada d_i y calcular $\sum d_i^2$, la suma de los valores elevados al cuadrado.

Tabla Nº 4.- Jerarquia para los datos de la Tabla Nº 3.

Número de individuo	Edad (X)	Valor resultante del EEG(Y)	d_i	d_i^2
1	1	18	-17	289
2	2	15	-13	169
3	3	17	-14	196
4	4	20	-16	256
5	5	19	-14	196
6	6	7	-1	1
7	7	6	1	1
8	8	12	-4	16
9	9	16	-7	49
10	10	14	-4	16
11	11	10	1	1
12	12	8	4	16
13	13	13	0	0
14	14	4	10	100
15	15	11	4	16
16	16	2	14	196

17	17	5	12	144
18	18	1	17	289
19	19	9	10	100
20	20	3	17	289
				$\sum d_i^2 = 2340$

Se calcula el valor de:

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

$$r_s = 1 - \frac{6\sum_{i=1}^{20} d_i^2}{20(20^2 - 1)}$$

$$r_s = -0.76$$

6. Decisión:

Como $r_s = -0.76 < -0.3789$, se rechaza H_o. Lo que quiere decir que las variables se encuentran inversamente relacionadas.

Elaborado por : Mg. Carmen Barreto R.
Fecha : Marzo 2011
Versión : 2