Punto cuatro

Una pareja de esposos salieron a jugar bolos y guardaron sus resultados para ver si existia una relación entre dichos resultados Use ρ de Pearson, el τ de kendall y el ρ de spearman para realizar una prueba de independencia entre los puntajes.

Esposo	147	158	131	142	183	151	196	129	155	158
Esposa	122	128	125	123	115	120	108	143	124	123

Coeficiente de correlación de Pearson (ρ de Pearson)

El coeficiente de correlación de Pearson mide la relación lineal entre dos variables cuantitativas. Evalúa tanto la dirección como la magnitud de esta relación, asignando un valor entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, donde las variables se mueven juntas en la misma dirección. Un valor de -1 muestra una correlación negativa perfecta, significando que las variables se mueven en direcciones opuestas. Un valor de 0 sugiere que no existe una relación lineal entre las variables.

Se puede calcular de la siguiente forma:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]^{1/2} \left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]^{1/2}}}$$

Donde:

- X: Puntaje del esposo.

- Y: Puntaje de la esposa.

- n: Número total de pares.

- $\bar{X}$: Promedio de los puntajes del esposo.

- $\bar{Y}$: Promedio de los puntajes de la esposa.

- $\bar{X} \times \bar{Y}$: El producto de los promedios de los puntajes del esposo y de la esposa.

Debido a la cantidad de datos, procederemos a calcular cada componente y desglosar la ecuación por partes para una mejor comprensión.

$$n = 10$$

$$\bar{X} = \frac{147+158+131+142+183+151+196+129+155+158}{10} = 155$$

$$\bar{Y} = \frac{122+128+125+123+115+120+108+143+124+123}{10} = 123.1$$

$$\bar{X} \times \bar{Y} = 155123.1 = 19080.5$$

$$\sum_{i=1}^{n} X_i Y_i - n\overline{XY} = [147 \cdot 122 - 10 \cdot 19080.5] + [158 \cdot 128 - 10 \cdot 19080.5] + [131 \cdot 125 - 10 \cdot 19080.5]$$

$$+ \ [142 \cdot 123 - 10 \cdot 19080.5] + [183 \cdot 115 - 10 \cdot 19080.5] + [151 \cdot 120 - 10 \cdot 19080.5]$$

$$+ \ [196 \cdot 108 - 10 \cdot 19080.5] + [129 \cdot 143 - 10 \cdot 19080.5] + [155 \cdot 124 - 10 \cdot 19080.5]$$

$$+ \ [158 \cdot 123 - 10 \cdot 19080.5] = -1372$$

$$\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = [147^2 - 10 \cdot 155^2] + [158^2 - 10 \cdot 155^2] + [131^2 - 10 \cdot 155^2] +$$

$$[142^2 - 10 \cdot 155^2] + [183^2 - 10 \cdot 155^2] + [151^2 - 10 \cdot 155^2] +$$

$$[196^2 - 10 \cdot 155^2] + [129^2 - 10 \cdot 155^2] + [155^2 - 10 \cdot 155^2] +$$

$$[158^2 - 10 \cdot 155^2] = 63.11894$$

$$\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 = \left(122^2 - 10 \cdot 123.1^2\right) + \left(128^2 - 10 \cdot 123.1^2\right) + \left(125^2 - 10 \cdot 123.1^2\right)$$

$$+ \ \left(123^2 - 10 \cdot 123.1^2\right) + \left(115^2 - 10 \cdot 123.1^2\right) + \left(120^2 - 10 \cdot 123.1^2\right)$$

$$+ \ \left(108^2 - 10 \cdot 155^2\right) + \left(143^2 - 10 \cdot 155^2\right) + \left(124^2 - 10 \cdot 155^2\right)$$

$$+ \ \left(123^2 - 10 \cdot 155^2\right) = 26.99815$$

Tenemos que

$$\frac{-1372}{[63.11894] \cdot [26.99815]} = -0.8051196$$

En R: -0.8051197

El coeficiente de Pearson de -0.805 indica una fuerte correlación negativa entre los puntajes de bolos del esposo y de la esposa. En términos prácticos, esto significa que, en general, cuando el puntaje del esposo aumenta, el puntaje de la esposa tiende a disminuir, y viceversa. Sin embargo, es importante recordar que Pearson mide relaciones lineales, por lo que este coeficiente está particularmente enfocado en cómo una variable aumenta o disminuye de manera lineal en relación con la otra.

Coeficiente ρ de Spearman

La correlación de Spearman evalúa cómo se relacionan dos variables basándose en el orden de sus valores, no en su magnitud real. Asigna rangos y utiliza estas posiciones para calcular la asociación, identificando tanto relaciones lineales como no lineales. La muestra de tamaño n, cada observación se representa como un par (X_i, Y_i). Se asigna un rango $R(X_i)$ a cada X_i, y $R(Y_i)$ a cada Y_i, en función de su posición relativa dentro de los conjuntos de valores de X y Y respectivamente. Para datos no numéricos, los rangos se otorgan basados en categorías de calidad. En caso de empates, se asigna el promedio de rangos correspondiente, un procedimiento alineado con el test de Mann-Whitney.

$$\frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{n} R^2(x_i) - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^{n} R^2(y_i) - n\left(\frac{n+1}{2}\right)^2\right)}}$$

Como se puede ver en la tabla de datos, los datos del esposo muestran resultados iguales, debido a que el número 158 aparece en dos ocasiones. Se creó una tabla para ordenar los rangos y manejar la duplicidad del número 158 en las posiciones 7 y 8.

x	y	R(x)	R(y)	R(x)R(y)
147	122	4.0	4.0	16.00
158	128	7.5	9.0	67.50
131	125	3.0	8.0	24.00
142	123	3.0	5.5	16.50
183	115	9.0	2.0	18.00
151	120	5.0	3.0	15.00
196	108	10.0	1.0	10.00
129	143	1.0	10.0	10.00
155	124	6.0	7.0	42.00
158	123	7.5	5.5	41.25

En situaciones donde el valor 158 se presenta múltiples veces, procedemos a reemplazarlo por el promedio de los rangos que debería ocupar, resultando en 7.5 a partir del cálculo $\frac{7+8}{2}$. Así, sustituimos 158 por 7.5 en los lugares que corresponden.

$$\sum_{i=0}^{n} R^2(X_i) = 4.0^2{+}7.5^2{+}2.0^2{+}3.0^2{+}9.0^2{+}5.0^2{+}10.0^2{+}1.0^2{+}6.0^2{+}7.5^2 = 384.5$$

$$\sum_{i=0}^{n} R^2(y_i) = 4.0^2{+}9.0^2{+}8.0^2{+}5.5^2{+}2.0^2{+}3.0^2{+}1.0^2{+}10.0^2{+}7.0^2{+}5.5^2 = 384.5$$

$$\sum_{i=0}^{n} R(x_i)R(y_i) = 16.00{+}67.50{+}16.00{+}16.50{+}18.00{+}15.00{+}10.00{+}10.00{+}42.00{+}41.25 = 252.25$$

$$\frac{252.25 - 10\left(\frac{10+1}{2}\right)^2}{\sqrt{(384.5 - 10\left(\frac{10+1}{2}\right)^2) \cdot (384.5 - 10\left(\frac{10+1}{2}\right)^2)}} = 0.6128049$$

En R: -0.6128049

El coeficiente ρ de Spearman de -0.613 indica una correlación negativa moderada a fuerte, similar a la de Pearson pero basada en rangos en lugar de valores exactos. Al igual que Kendall, Spearman es más robusto ante datos no normales o la presencia de valores atípicos. Este coeficiente sugiere que, en términos de rangos, existe una tendencia negativa moderada a fuerte entre los puntajes de bolos de la pareja.

Coeficiente τ de Kendall.

El coeficiente τ de Kendall cuantifica la correlación entre dos series de datos examinando la concordancia y la discordancia en el ordenamiento de sus valores. Se obtiene mediante la diferencia entre los números de pares concordantes y discordantes, lo que refleja tanto la intensidad como la dirección de la relación entre las variables.

Cuando se presentan empates, el criterio de comparación es el siguiente:

si $X_i = X_j$, la comparación entre estos pares se omite. En cambio, si $Y_i = Y_j$ pero al mismo tiempo $X_i \neq X_j$, entonces dicho par se considera como parcialmente concordante y parcia

$$\tau = \frac{N_c - N_d}{N_c + N_d'}$$

Donde N_c representa el número de pares concordantes y N_d el número de pares discordantes.

Se considera que hay concordancia si $\frac{Y_j - Y_i}{X_j - X_i} > 0$, y discordancia si $\frac{Y_j - Y_i}{X_j - X_i} < 0$. En el caso de que $\frac{Y_j - Y_i}{X_j - X_i} = 0$, el par se clasifica como mitad concordante y mitad discordante. Si $X_i = X_j$, entonces no se realiza ninguna comparación entre los pares. Finalmente, para facilitar el cálculo de N_c (número de coincidencias) y N_d (número de discrepancias), es beneficioso ordenar primero las observaciones (X_i, Y_i) por los valores ascendentes de X y, después, hacer lo mismo con los valores ascendentes de Y. Este procedimiento simplifica la comparación descendente de cada valor de Y. De acuerdo con esta metodología, definimos las reglas siguientes: se asigna un signo + indicando concordancia si $\frac{Y_j - Y_i}{X_j - X_i} > 0$. Si $\frac{Y_j - Y_i}{X_j - X_i} < 0$, entonces se asigna un signo $-$ para señalar discordancia. Cuando $\frac{Y_j - Y_i}{X_j - X_i} = 0$, lo cual indica igualdad perfecta, se asigna un signo 0 que refleja una concordancia y discordancia a partes iguales, añadiendo 0.5 tanto a la cuenta de concordancia como a la de discordancia.

Par (129,143)

$$\frac{129 - 131}{143 - 125} = \frac{-}{+} = -$$

$$\frac{129 - 142}{143 - 123} = \frac{-}{+} = -$$

$$\frac{129-147}{143-122}=\frac{-}{+}=-$$
$$\frac{129-151}{143-120}=\frac{-}{+}=-$$
$$\frac{129-155}{143-124}=\frac{-}{+}=-$$
$$\frac{129-158}{143-123}=\frac{-}{+}=-$$
$$\frac{129-158}{143-128}=\frac{-}{+}=-$$
$$\frac{129-183}{143-115}=\frac{-}{+}=-$$
$$\frac{129-196}{143-108}=\frac{-}{+}=-$$

Concordancias=0 Discordancias=9

Par (131,125)

$$\frac{131-142}{125-125}=\frac{-}{+}=-$$
$$\frac{131-1123}{125-123}=\frac{-}{+}=-$$
$$\frac{131-147}{125-122}=\frac{-}{+}=-$$
$$\frac{131-151}{125-120}=\frac{-}{+}=-$$
$$\frac{131-155}{125-124}=\frac{-}{+}=-$$
$$\frac{131-158}{125-123}=\frac{-}{+}=-$$
$$\frac{131-158}{125-128}=\frac{-}{-}=+$$
$$\frac{131-183}{125-115}=\frac{-}{+}=-$$
$$\frac{131-196}{125-108}=\frac{-}{+}=-$$

Concordancias=1 Discordancias=7

Par (142,123)

$$\frac{142-147}{123-122}=\frac{-}{+}=-$$

$$\frac{142-151}{123-120}=\frac{-}{+}=-$$
$$\frac{142-155}{123-124}=\frac{-}{-}=+$$
$$\frac{142-158}{123-123}=\frac{-}{0}=0$$
$$\frac{142-158}{123-128}=\frac{-}{-}=+$$
$$\frac{142-183}{123-115}=\frac{-}{+}=-$$
$$\frac{142-196}{123-108}=\frac{-}{+}=-$$

Concordancias=2.5 Discordancias=4.5

Par(147,122)

$$\frac{147-151}{122-120}=\frac{-}{+}=-$$
$$\frac{147-155}{122-124}=\frac{-}{-}=+$$
$$\frac{147-158}{122-123}=\frac{-}{-}=+$$
$$\frac{147-158}{122-128}=\frac{-}{-}=+$$
$$\frac{147-183}{122-115}=\frac{-}{+}=-$$
$$\frac{147-196}{122-108}=\frac{-}{+}=-$$

Concordancias=3 Discordancias=3

Par(151,120)

$$\frac{151-155}{120-124}=\frac{-}{-}=+$$
$$\frac{151-158}{120-123}=\frac{-}{-}=+$$
$$\frac{151-158}{120-128}=\frac{-}{-}=+$$
$$\frac{151-183}{120-115}=\frac{-}{+}=-$$
$$\frac{151-196}{120-108}=\frac{-}{+}=-$$

Concordancias=3 Discordancias=2

Par(155,124)

$$\frac{155-158}{124-123}=\frac{-}{+}=-$$
$$\frac{155-158}{124-128}=\frac{-}{-}=+$$
$$\frac{155-183}{124-115}=\frac{-}{+}=-$$
$$\frac{155-196}{124-108}=\frac{-}{+}=-$$

Concordancias=1 Discordancias=3

Par(158,123)

$$\frac{158-158}{123-128}=\frac{0}{-}=0$$
$$\frac{158-183}{123-115}=\frac{-}{+}=-$$
$$\frac{158-196}{123-108}=\frac{-}{+}=-$$

Concordancias=0.5 Discordancias=2.5

Par(158,128)

$$\frac{158-183}{128-115}=\frac{-}{+}=-$$
$$\frac{158-196}{128-108}=\frac{-}{+}=-$$

Concordancias=0 Discordancias=2

Par(183,115)

$$\frac{183-196}{115-108}=\frac{-}{+}=-$$

Concordancias=0 Discordancias=1

Par(196,108)

Como es el último por ende ambas son cero.
Concordancias=0 Discordancias=0

Tenemos que:

X_i, Y_i	Pares Concordantes	Pares Discordantes
(129, 143)	0	9
(131, 125)	1	7
(142, 123)	2.5	4.5
(147, 122)	3	3
(151, 120)	3	2
(155, 124)	1	3
(158, 123)	0.5	2.5
(158, 128)	0	2
(183, 115)	0	1
(196, 108)	0	0

$$\tau = \frac{N_c - N_d}{N_c + N_d} = \frac{11 - 34}{11 + 34} = -0,5111111$$

En R: -0.5227273

El cálculo manual del coeficiente de Tau de Kendall difiere ligeramente del obtenido con R, sin embargo, ambos indican una correlación negativa de intensidad moderada de -0.523 también sugiere una correlación negativa, aunque no tan fuerte como la indicada por Pearson. Kendall mide la concordancia en los ordenamientos de los datos entre dos variables, lo que significa que puntajes más altos de un miembro de la pareja tienden a asociarse con puntajes más bajos del otro miembro, pero con menos énfasis en la linealidad de esa relación. Este coeficiente es útil para entender las tendencias generales en los datos que pueden no ser estrictamente lineales.

Conclusión

Los tres coeficientes muestran una tendencia negativa en la relación entre los puntajes de bolos de la pareja, sugiriendo que no hay una dependencia directa en el sentido de que puntajes altos de uno impliquen puntajes altos del otro; de hecho, la tendencia es lo contrario. Esto podría interpretarse como que cuando uno de los dos tiene un buen día en el juego, el otro tiende a no tenerlo, aunque la interpretación exacta podría variar dependiendo de otros factores no considerados en este análisis.

Prueba de Independencia

Se llevaran a cabo pruebas de hipótesis utilizando las correlaciones de Pearson, Spearman y Kendall para determinar si existe independencia entre los puntajes de bolos de una pareja de esposos.

Prueba de Independencia de Pearson

```
        Pearson's product-moment correlation

data:  esposo and esposa
t = -3.8394, df = 8, p-value = 0.004951
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9521021 -0.3559160
sample estimates:
       cor
-0.8051197
```

$$H_0 : \rho = 0$$

Los puntajes son independientes

$$H_1 : \rho \neq 0$$

Los puntajes no son independientes

$$p{-}valor = 0.004951 < 0.05$$

Rechazamos H_0

Prueba de Independencia de Spearman

```
        Spearman's rank correlation rho

data:  esposo and esposa
S = 266.11, p-value = 0.05961
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6128049
```

$$H_0 : \rho_s = 0$$

Los puntajes son independientes

$$H_1 : \rho_s \neq 0$$

Los puntajes no son independientes

$$p{-}valor = 0.05961 > 0.05$$

No rechazamos H_0

Prueba de Independencia de Kendall

```
        Kendall's rank correlation tau

data:  esposo and esposa
z = -2.0737, p-value = 0.03811
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau
-0.5227273
```

$$H_0 : \tau = 0$$

Los puntajes son independientes,

$$H_1 : \tau \neq 0$$

Los puntajes no son independientes,

$$p{-}valor = 0.03811 < 0.05$$

Rechazamos H_0,

Al llevar a cabo los análisis de correlación mediante las metodologías de Pearson, Spearman y Kendall, se han obtenido resultados que nos permiten establecer distintas inferencias sobre la relación entre las puntuaciones de los individuos en estudio. El test de Pearson resultó en un p-valor de 0.004951, indicando con claridad la existencia de una correlación lineal negativa significativa y justificando el rechazo de la hipótesis nula de independencia. Por el contrario, el método de Spearman generó un p-valor de 0.05961, que supera el nivel de significación establecido de 0.05, lo cual significa que no hay evidencia suficiente para descartar la hipótesis nula y, por ende, no se puede afirmar la presencia de una relación monótona significativa. En el caso del test de Kendall, el p-valor obtenido fue de 0.03811, que está por debajo del umbral de significancia y sugiere una correlación negativa significativa en términos de rango. Estos hallazgos indican que, según los tests de Pearson y Kendall, las puntuaciones no son independientes y comparten una relación negativa. No obstante, la ausencia de evidencia significativa en la prueba de Spearman para un nivel de confianza del 95la naturaleza de los datos o la sensibilidad inherente a cada test puedan influir en la interpretación de la relación entre las variables.