

SELECCIONES



De Pearson a Spearman

From Pearson to Spearman



Luis F Restrepo B¹*, Estad, Esp estad bioma; Julián González L², Esp Est, Esp Bioma, MS.

¹ Grupo Grica, Facultad de Ciencias Agrarias Universidad de Antioquia, Medellín, Colombia.

² Profesor Titular Universidad de Caldas, Facultad de Ciencias Exactas y Naturales, Manizales, Colombia.

luisitano@agronica.udea.edu.co

(Recibido: 5 octubre, 2005; aceptado: 26 abril, 2007).

Resumen

Este artículo trata acerca de los coeficientes de correlación de Pearson y Spearman los cuales son ampliamente utilizados en las ciencias agropecuarias con el fin de establecer relaciones entre variables generalmente de índole cuantitativo. Además contiene los supuestos fundamentales en los que se basa el método de Pearson: normalidad bivariada, linealidad en la interacción de las variables y la forma de programación en el paquete estadístico SAS; adicionalmente, la manera de interpretar las salidas derivadas del paquete estadístico las cuales oscilan entre $-1 \leq \rho \leq 1$.

Palabras clave: *coeficiente, correlación, supuestos.*

Summary

This article is about the Pearson and Spearman correlation coefficients which are widely used in agricultural sciences in order to establish the relationships between generally quantitative variables. It also contains the fundamental assumptions in which the Pearson method is based: bivariate normality, variable interaction linearity, as well as the programming Schedule in SAS software; in addition, the way in which SAS exits must be interpreted which oscillate between $-1 \leq \rho \leq 1$.

Key words: *assumptions, coefficient, correlation.*

Introducción

Los métodos de correlación de Pearson y Spearman son técnicas bivariadas que se emplean en el campo multivariado, en situaciones donde el

investigador animal quiere ver representaciones de la información derivadas de análisis matriciales con propiedades del álgebra lineal, que permiten establecer similitudes o disimilitudes entre las

* Autor para el envío de la correspondencia y la solicitud de separatas. Facultad de Ciencias Agrarias, Universidad de Antioquia, AA 1226, Medellín, Colombia. E-mail: luisitano@agronica.udea.edu.co

variables e individuos representados en dimensiones de menor valor, generalmente en planos o cubos (segunda y tercera dimensión) para esclarecer la variabilidad conjunta expresada en factores ortogonales que permiten tipificar lo que sucede con los datos. Un ejemplo particular es el referido al investigador en avicultura que podría mediante análisis de componentes principales por la técnica R proyectar todas las variables de índole cuantitativo relacionadas con el huevo, a fin de determinar la participación de cada variable dentro del factor abstracto que él debe discernir para poder comprender mejor lo que sucede con los aspectos morfométricos y de composición. También puede centrar su interés mediante análisis canónicos para evaluar las variables preponderantes en la morfometría del ave con relación a las variables tomadas en el huevo, y así conocer las incidencias particulares y colectivas de los factores sujetos en el estudio.

La técnica de correlación es fundamental en la explicación de muchos fenómenos relacionados con el campo animal y vegetal. A modo de ejemplo, se puede citar el interés que pueda tener el ganadero en relacionar la ganancia de peso con el consumo de alimento; la condición corporal con la incidencia de las variables climáticas; el intervalo entre partos con el número de partos, entre muchas otras variables.

En el campo de la ingeniería acuícola es de interés relacionar aspectos de la morfometría del pez asociada con las variables longitud estándar, ancho del cuerpo, longitud de la cabeza, altura de la cabeza, altura del opérculo, diámetro de los ojos, longitud cobertura ocular, apertura de la boca, grosor del labio, longitud base aleta dorsal, altura aleta dorsal, distancia aleta dorsal adiposa, longitud base aleta dorsal, longitud preanal, longitud aleta pélvica y la longitud aleta pectoral entre otras, con respecto a condiciones ambientales y de alimentación.

El experto en apicultura podría estar interesado en ver la asociación entre la producción de miel con respecto al tiempo que tarda la abeja en lograr extraer el polen de la flor, el tipo de flor y su edad vegetativa, las condiciones climáticas predominantes en el lugar de evaluación, la macro y micro fauna presentes.

El ingeniero agrónomo puede mediante las técnicas de correlación establecer patrones de asociación entre las variables del suelo representadas en elementos mayores o menores, con respecto al desarrollo de un cultivo. En el caso particular de pastos y forrajes, se podría evaluar la asociación entre las condiciones físico químicas del pasto y la producción de carne o leche de un hato en particular.

El experto en clínica veterinaria puede en un momento dado relacionar la cantidad suministrada de un fármaco para anestesiarse a un animal y su correlación con el pulso cardíaco, la edad, la condición corporal y el peso del canino entre otras variables de interés.

En nutrición animal se puede asociar la cantidad de alimento suministrada cuando se elabora un bloque nutricional como suplemento con respecto a la conversión y ganancia de peso en las distintas etapas del proceso de desarrollo del animal.

Cuando se trabaja con fauna silvestre se mide la presencia o ausencia de determinada especie en relación a las variables ambientales presentes y dominantes en un ecosistema, además se evalúan las variables morfométricas y ambientales para establecer patrones que permitan evaluar la factibilidad de la especie para subsistir en el medio.

En avicultura se puede asociar el peso del huevo en relación a la cantidad de alimento suministrada, la edad del ave, la longitud del oviducto, la cantidad de agua tomada por las aves, densidad del galpón, humedad relativa, temperaturas máxima y mínima.

En el cultivo del banano mediante las técnicas de correlación se puede medir la intensidad de la relación de las variables precipitación, temperatura, velocidad del viento, humedad relativa, con la altura de la planta, grosor del fruto, longitud del fruto, porcentaje de pérdida, calidad del fruto; además, se pueden asociar elementos del suelo como nitrógeno, fósforo, potasio y algunos elementos menores.

En general son múltiples los casos particulares donde las técnicas de correlación son ampliamente utilizadas, no sólo para evaluar el grado de

dependencia entre ellas, sino también para ver la bondad de ajuste de los modelos de serie de tiempo en la evaluación de las auto correlaciones y en la convalidación de algunos supuestos.

Coefficiente de correlación

Un coeficiente de correlación, mide el grado de relación o asociación existente generalmente entre dos variables aleatorias. No es conveniente identificar correlación con dependencia causal, ya que, si hay una semejanza formal entre ambos conceptos, no puede deducirse de esto que sean análogos (9, 15); en efecto es posible que haya una alta correlación entre dos acontecimientos y que sin embargo, no exista entre ellos relación de causa o efecto; por ejemplo cuando dos acontecimientos tienen alguna causa común, pueden resultar altamente asociados y no son el uno causa del otro. Cabe recordar que el coeficiente fluctúa entre $-1 \leq \rho \leq 1$.

En un estudio médico se estableció que al comparar el consumo de carbón con la mortalidad, en periodos análogos, se encontró alta correlación. Esto no quiere decir que el consumo de carbón sea la causa de las muertes, sino que tanto el aumento de consumo de carbón, como el aumento de la mortalidad se producen en las épocas de frío más intenso. En conclusión, un coeficiente de correlación por sí mismo no puede probar ni desmentir una relación causal entre variables. La relación causa – efecto es posible definirla sólo a través de la comprensión de la relación natural que exista entre las variable y esto no debe manifestarse sólo por la existencia de una fuerte asociación. El análisis de correlación es técnicamente neutral (7, 12, 13).

Coefficiente de correlación de Pearson

Tiene como objetivo medir la fuerza o grado de asociación entre dos variables aleatorias cuantitativas que poseen una distribución normal bivariada conjunta. El coeficiente se define por la siguiente fórmula (14):

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad -1 \leq \rho \leq 1$$

Cuando $\rho=+$ la relación es directa entre las variables. Si $\rho=-$ la relación es inversa y si $\rho=0$ son independientes. Dicho coeficiente se puede expresar en términos de su estadístico como (8):

$$\gamma = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right]}}$$

$$-1 \leq \gamma \leq 1$$

El coeficiente de correlación de Pearson es la media geométrica entre las pendientes de los modelos de regresión lineal simple Y/X, X/Y así:

$$y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Donde:

β_0 = intercepto del modelo.

β_1 = pendiente del modelo, cambio esperado en y por unidad de cambio en x.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Por el método de los mínimos cuadrados ordinarios

$$\text{Ahora } X_i = \beta_0' + \beta_1' Y_i + \epsilon_i'$$

$$\hat{\beta}_1' = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

$$\gamma = \sqrt{\hat{\beta}_1 \times \hat{\beta}_1'}$$

$$\gamma = \sqrt{\left[\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \left[\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right]}$$

$$\gamma = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) \left(\sum_{i=1}^n (Y_i - \bar{Y})^2 \right)}}$$

Supuestos que fundamentan al coeficiente de correlación:

1. La distribución conjunta de las variables (X, Y) debe ser normal bivariada (8), y corresponde a:

$$f = (X, Y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{0.5}} e^{-\frac{1}{2(1-\rho^2)}A}$$

$$A = \left[\left(\frac{X-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{X-\mu_x}{\sigma_x} \right) \left(\frac{Y-\mu_y}{\sigma_y} \right) + \left(\frac{Y-\mu_y}{\sigma_y} \right)^2 \right]$$

$-\infty < X, Y < \infty$

2. En términos prácticos para validar dicho supuesto se debe observar que cada variable se distribuya en forma normal (11, 15), si una sola de las variables se desvía de la normalidad, tampoco es normal la distribución conjunta.
3. Debe existir una relación de tipo lineal entre las variables (X, Y).
4. Para cada valor de X, hay una subpoblación de valores de Y normalmente distribuidas.
5. Las subpoblaciones de valores Y tienen varianza constante.
6. Los promedios de las subpoblaciones de Y tienen ubicación en la misma línea recta.
7. Las subpoblaciones de X tienen varianza constante.
8. Las medias de las subpoblaciones de X se encuentran en la misma línea recta.
9. Para cada valor de Y hay una subpoblación de valores X que están normalmente distribuidas (18).

Prueba de hipótesis

Bajo la hipótesis Ho: $\rho = 0$

Se tiene $t = \frac{\gamma}{\sqrt{\frac{1-\gamma^2}{n-2}}} \sim t_{\alpha/2}(n-2)$

Si se desea contrastar

Ho: $\rho = \rho_0$ hipótesis nula,

H₁: $\rho \neq \rho_0$ hipótesis alternativa

Se debe aplicar la estadística:

$$Z = \frac{1}{2}LN \frac{1+\gamma}{1-\gamma} - \sim N \left(\frac{1}{2}LN \frac{1+\rho}{1-\rho}; \sqrt{\frac{1}{\eta-3}} \right)$$

Ahora: $\frac{1}{2}LN \frac{1+\gamma}{1-\gamma} - \frac{1}{2}LN \frac{1+\rho_0}{1-\rho_0} \sim N \left(0, \sqrt{\frac{1}{\eta-3}} \right)$

$$\frac{1}{2}LN \frac{1+\gamma}{1-\gamma} - \frac{1}{2}LN \frac{1+\rho_0}{1-\rho_0}$$

Luego $Z' = \frac{\text{[Ecuación anterior]}}{\sqrt{\frac{1}{\eta-3}}} \sim N(0,1)$

Si $|Z'| \geq Z_{1-\frac{\alpha}{2}}$ se rechaza Ho, se acepta H₁,

hay relación estadística entre las variable. Si se acepta Ho las variables son independientes.

Cabe anotar que el coeficiente de correlación de pearson puede ser empleado utilizando un factor de ponderación Wi, el cual efectúa un ajuste a la media aritmética y por ende al coeficiente de asociación. Esta ponderación se debe aplicar, cuando el investigador quiera darle un peso específico a cada individuo que interviene en el estudio.

Cuando una variable es dicotómica y la otra continua, el coeficiente de correlación de pearson es igual al biseral puntual, si ambas variables son dicotómicas, Pearson es igual al coeficiente phi (ϕ).

El coeficiente de correlación de pearson no se debe extrapolar más allá del rango de los valores observados cuando se efectúa inferencia. Para construir el intervalo de confianza asociado con Pearson, ambas variables se deben distribuir en forma normal y para predecir se utiliza la técnica propuesta por Rubens (16).

Aplicación

Se evaluaron 44 ejemplares de trucha en un estudio de índole acuícola, con el fin de relacionar la longitud del cuerpo y su ancho. En la salida de la

figura 1, se muestra el coeficiente de correlación de Pearson y los supuestos asociados a la técnica, donde los datos fueron suministrados por un docente del área.

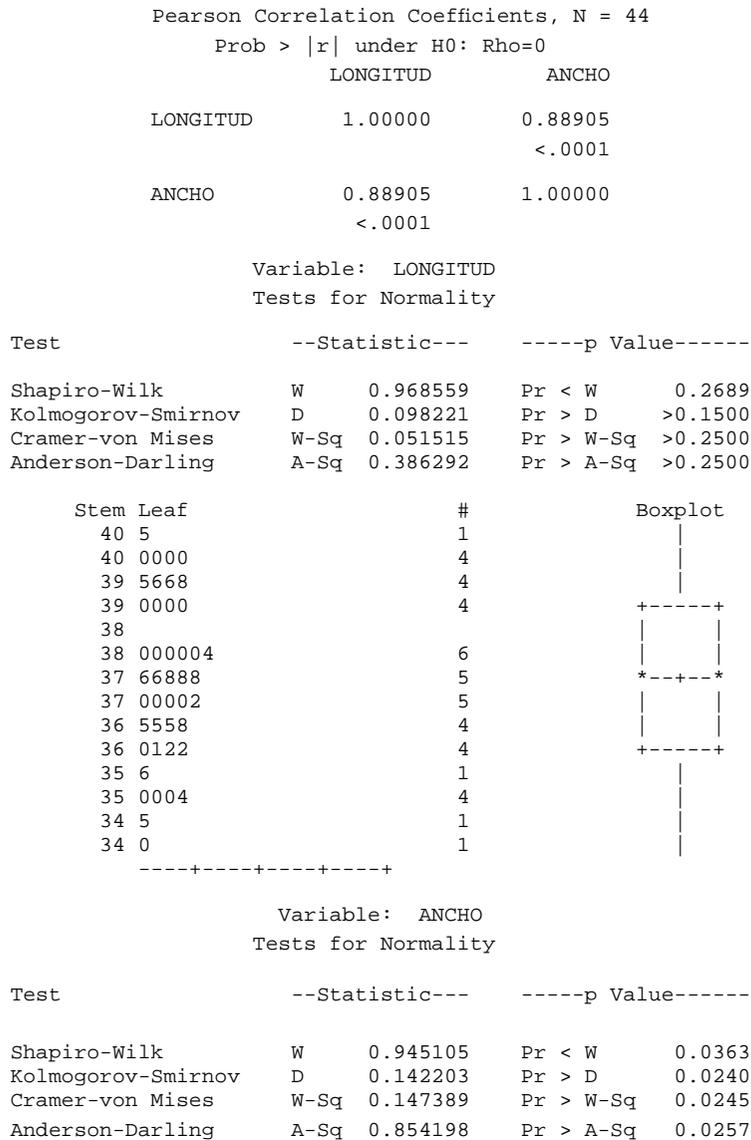


Figura 1. Salidas estadísticas del coeficiente de correlación de Pearson.

Interpretación

Al observar el valor p de prueba asociado con el estadístico de Pearson se aprecia relación entre la longitud y ancho dado que $p < 0.05$; cabe anotar que el supuesto de normalidad bivariada no se cumplió

debido a que la variable ancho no se ajusta a la distribución gaussiana o normal como se aprecia en las pruebas de normalidad de Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling donde el $p < 0.05$. La figura 2 muestra una tendencia lineal.

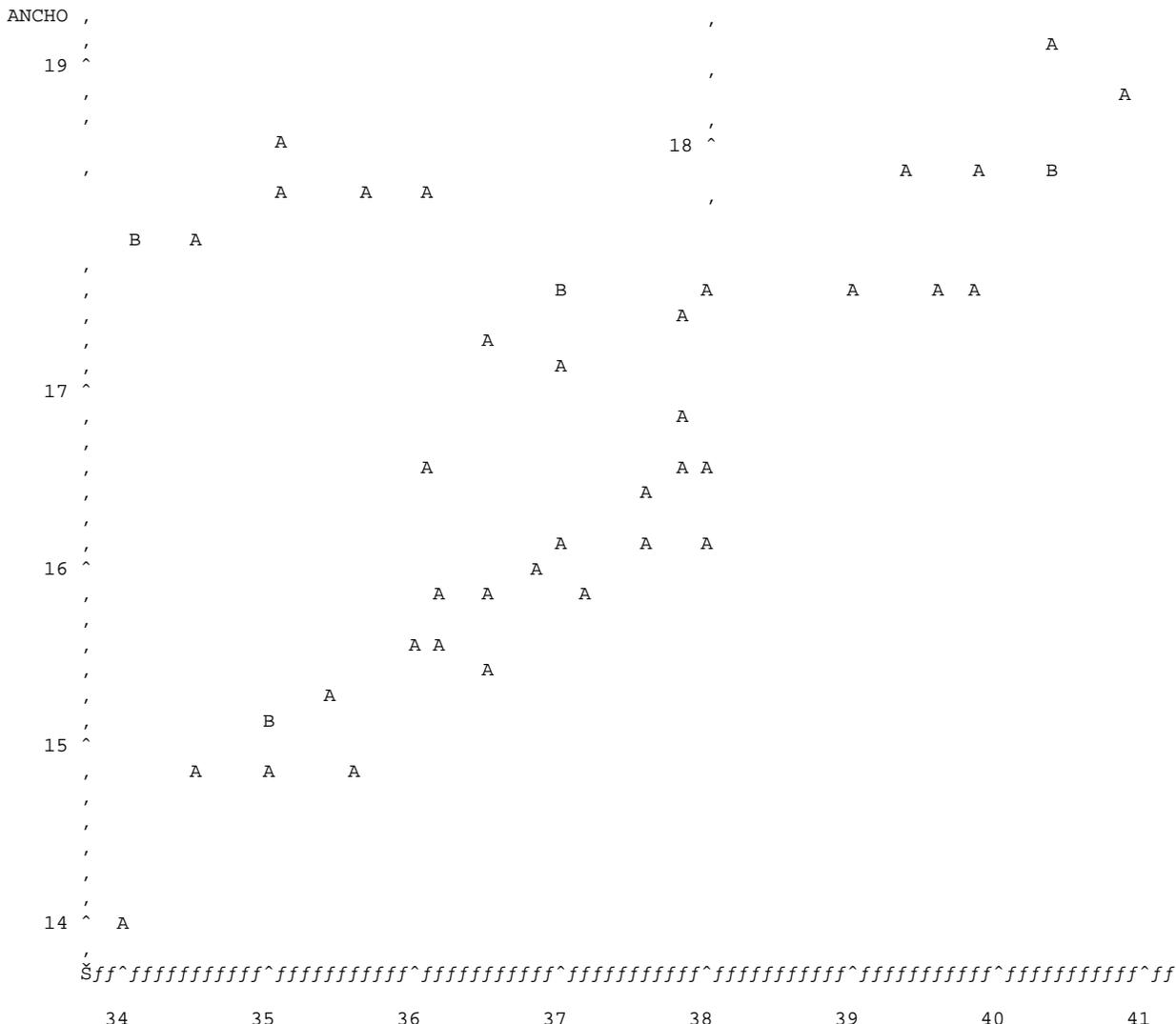


Figura 2. Diagrama de dispersión.

Obs	N	R	CONF	LINF	LSUP	METODO
1	44	0.88905	0.95	0.80030	0.93711	PEARSON
2	44	0.88905	0.99	0.76077	0.94839	PEARSON

Figura 3. Intervalos de confianza de Rubens.

Lo anterior indica que con un 95% de confiabilidad el coeficiente de correlación fluctúa entre 0.80030 y 0.93711

Coefficiente de correlación de Spearman

El coeficiente de correlación de Spearman es un coeficiente no paramétrico alternativo al coeficiente de correlación de Pearson cuando este no cumple los supuestos (3, 4, 19). Charles Spearman contribuyó

al análisis del factor, a la teoría de la inteligencia, elaboró una prueba de la teoría mental (17).

Se define el coeficiente de correlación de rangos de Spearman como el coeficiente de correlación lineal

entre los rangos $R_i(x)$ y $R_i(y)$, en la fórmula de Pearson se reemplaza X_i por $R_i(x)$ y Y_i por $R_i(y)$ quedando:

$$\gamma_s = \frac{\sum_{i=1}^n [R_i(x) - \bar{R}(x)](R_i(y) - \bar{R}(y))}{\sqrt{\left[\sum_{i=1}^n (R_i(x) - \bar{R}(x))^2 \right] \left[\sum_{i=1}^n (R_i(y) - \bar{R}(y))^2 \right]}}$$

Se tiene $R(x) = R(y) =$

$$\begin{aligned} \frac{\sum_{i=1}^n \frac{R_i(x)}{\eta}}{\eta} &= \frac{\sum_{i=1}^n \frac{R_i(y)}{\eta}}{\eta} \\ &= \frac{1+2+3+4+\dots+n}{n} = \frac{n+1}{2} \end{aligned}$$

Por la propiedad telescópica

Ahora:

$$\sum R_i^2(x) = \sum R_i^2(y) = \frac{\eta(\eta+1)(2\eta+1)}{6}$$

por la misma propiedad citada anteriormente.

$$\begin{aligned} \sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 &= \sum_{i=1}^n R_i^2(x) - \eta \bar{R}^2(x) \\ &= \frac{\eta(\eta+1)(2\eta+1)}{6} - \eta \left(\frac{\eta+1}{2} \right)^2 = \frac{\eta^3 - \eta}{12} \end{aligned}$$

De igual forma,

$$\sum [R_i(y) - \bar{R}(y)]^2 = \frac{\eta^3 - \eta}{12}$$

Definamos $d_i = R_i(x) - R_i(y)$ esto es la diferencia de rangos entre las variables.

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \sum_{i=1}^n [R_i(x) - R_i(y)]^2 = \sum_{i=1}^n ([R_i(x) - \bar{R}(x)] - [R_i(y) - \bar{R}(y)])^2 \\ &= \sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2 - 2 \sum_{i=1}^n [R_i(x) - \bar{R}(x)][R_i(y) - \bar{R}(y)] \\ &= \frac{\sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2 - \sum_{i=1}^n d_i^2}{2} = \frac{\eta^3 - \eta}{12} - \frac{\sum_{i=1}^n d_i^2}{2} \end{aligned}$$

Por lo tanto el coeficiente de correlación de Spearman resulta como:

$$\begin{aligned} \gamma_s &= \frac{\frac{\eta^3 - \eta}{12} - \frac{\sum_{i=1}^n d_i^2}{2}}{\sqrt{\left(\frac{\eta^3 - \eta}{12} \right) \left(\frac{\eta^3 - \eta}{12} \right)}} = 1 - 6 \frac{\sum_{i=1}^n d_i^2}{\eta^3 - \eta} \\ -1 &\leq \gamma_s \leq 1 \end{aligned}$$

La asociación crítica del contraste la formarán valores de γ_s próximos a 1 o a -1 determinándose a partir de la relación:

$$p([\gamma_s] \geq K / H_0) = \alpha$$

Cuando $n \geq 10$ se debe emplear la distribución asintótica propuesta por Kendall dada a través de la variable aleatoria

$$t = \frac{\gamma_s}{\sqrt{\frac{1 - \gamma_s^2}{n - 2}}} \sim t_{\alpha/2}(\eta - 2)$$

Cuando aparecen datos repetidos en la variable aleatoria X o Y se recomienda aplicar la fórmula:

$$\gamma_s = \frac{\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2 - \sum_{i=1}^n d_i^2}{2 \sqrt{\left(\sum_{i=1}^n X_i^2 \right) \left(\sum_{i=1}^n Y_i^2 \right)}}$$

Teniendo que $\sum X_i^2 = \frac{\eta^3 - \eta}{12} = T_x'$

$$T_x = \frac{\sum T_x^3 - T_x'}{12}$$

$$d_i = X_i' - Y_i'$$

$$\sum Y_i^2 = \frac{\eta^3 - \eta}{12} = T_y$$

$$T_y = \frac{\sum T_y^3 - T_y'}{12}$$

T_x' se define como el número de veces que se repite un valor en la variable X, igual definición para Y

γ_s mide la tendencia de X, Y a relacionarse en forma monótona creciente o decreciente. Al medir el grado de asociación de forma monótona entre las variables X, Y, γ_s no se encuentra restringido a descubrir sólo una asociación lineal entre las variables.

Ventajas del coeficiente de Spearman

1. Al ser Spearman una técnica no paramétrica es libre de distribución probabilística (2, 5, 9).
2. Los supuestos son menos estrictos. Es robusto a la presencia de outliers (es decir permite ciertos desvíos del patrón normal). La manifestación de una relación causa-efecto es posible sólo a través de la comprensión de la relación natural que existe entre las variable y no debe manifestarse sólo por la existencia de una fuerte correlación (1, 5).

Ejemplo

Con base en la información de las 44 truchas y ante el no cumplimiento de los supuestos del coeficiente de correlación de Pearson se aplicó la técnica no paramétrica de Spearman dando como resultado la siguiente salida (véase Figura 4).

```
Spearman Correlation Coefficients, N = 44
Prob > |r| under H0: Rho=0

                LONGITUD      ANCHO
LONGITUD      1.00000      0.89692
                <.0001
ANCHO          0.89692      1.00000
                <.0001
```

Figura 4. Salida estadística del programa SAS del coeficiente de Spearman.

Se encontró relación directa entre la longitud y el ancho del pez ya que el $p < 0.05$

```
intervalo de Rubens
3 44 0.89692 0.95 0.81378 0.94167 SPEARMAN
4 44 0.89692 0.95 0.81378 0.94167 SPEARMAN
```

Figura 5. Intervalo de confianza de Rubens para el coeficiente de Spearman.

Con un 95% se estima que el coeficiente oscila entre 0.81378 y 0.94167

Procedimiento SAS para establecer los coeficientes de correlación de pearson validando supuestos:

```
DATA BASE;
INPUT X Y;
CARDS;
DATOS
DATOS
.
.
.DATOS
;
PROC CORR PEARSON ;
VAR X Y;
TITTLE 'VALIDACION DE SUPUESTOS';
PROC PLOT;
PLOT Y*X;
PROC UNIVARIATE NORMAL PLOT;
VAR X Y;
RUN;
```

Procedimiento SAS coeficiente de correlación de spearman:

```
DATA BASE;
INPUT X Y;
CARDS;
DATOS
DATOS
.
.
.DATOS
;
PROC CORR SPEARMAN;
VAR X Y;
RUN;
```

Intervalo de confianza por la aproximación de Rubén (6)

```
OPTIONS LINESIZE=75 PAGESIZE=54 NODATE;
DATA BASE;
/* N=MUESTRA, R=CORRELACION, CONF=NIVEL*/
INPUT N R CONF;
U=-PROBIT((1-CONF)/2);
RTILDE=R/SQRT(1-R*R);
A=2*N-3-U*U;
B= RTILDE*SQRT((2*N-3)*(2*N-5));
C=(2*N-5-U*U)*RTILDE*RTILDE-2*U*U;
Y1=(2*B-SQRT(4*B*B-A*A*C))/(2*A);
Y2=(2*B+SQRT(4*B*B-4*A*C))/(2*A);
/* LL=LIMITE INFERIOR, UL=LIMITE SUPERIOR */
LL=Y1/SQRT(1+Y1*Y1);
UL=Y2/SQRT(1+Y2*Y2);
CARDS;
DATOS
;
PROC PRINT;
VAR N R CONF LL UL;
RUN;
```

La correlación fue utilizada por primera vez por Francis Galton, aunque su discípulo Karl Pearson fue quien estudió en profundidad sus propiedades. La correlación de rangos fue introducida por el psicólogo Charles Edward Spearman al intentar construir una teoría de la inteligencia (10, 14, 17).

Los coeficientes de correlación de Pearson y Spearman

Estos coeficientes también son apropiados para evaluar la relación entre variables ordinales representadas en tablas de contingencia.

El coeficiente de correlación de Pearson se define:

$$\gamma = \frac{SS_{rc}}{\sqrt{SS_r SS_c}} = \frac{V}{W}$$

$SS_{rc} = V = \sum \sum \eta_{ij} (R_i - \bar{R})(C_j - \bar{C})$ covarianza entre variables.

$$SS_r = \sum \sum \eta_{ij} (R_i - \bar{R})^2 \text{ Desviación típica de r.}$$

$$SS_c = \sum \sum \eta_{ij} (C_j - \bar{C})^2 \text{ Desviación típica de c.}$$

R_i, C_j representan los valores referidos a las filas y a la columna

Para probar el coeficiente de correlación de Pearson asintóticamente, se elabora una prueba estadística estandarizada γ^* bajo la hipótesis nula de que el coeficiente de correlación es igual a cero.

$$\gamma^* = \frac{\gamma}{\sqrt{VAR(\gamma)}}$$

$$VAR(\gamma) = \frac{\sum \sum \eta_{ij} (\bar{R}_i - R)^2 (C_j - \bar{C})^2 - SS_{rc}^2 / r}{SS_r SS_c}$$

La varianza asintótica es derivada de una muestra con distribución multinomial representada en una tabla de contingencia.

El coeficiente de correlación de Spearman se define:

$$\gamma_s = \frac{V}{W}$$

$$V = \sum_i \sum_j \eta_{ij} R_{(i)} C_{(j)}$$

$$W = \frac{1}{12} \sqrt{FG}$$

$$F = \eta^3 - \sum_{i=1} \eta_i^3$$

$$G = \eta^3 - \sum_j \eta_j^3$$

$$R_{(i)} = R_i - \frac{\eta}{2}$$

$$C_{(i)} = C_i - \frac{\eta}{2}$$

$$\gamma_s \text{ tiene } \frac{1}{\eta^2 W^4} \sum \sum \eta_{ij} (Z_{ij} - \bar{Z})^2$$

$$\bar{Z} = \frac{WV_{ij} - VW_{ij}}{W}$$

$$V_{ij} = \eta \left(R_{ij} C_{(j)} + \frac{1}{2} \sum \eta_{il} C_{(L)} + \frac{1}{2} \sum \eta_{kj} R_{(k)} + \sum \sum \eta_{kl} R_{(k)} + \sum \sum n_{kl} R_{(K)} \right)$$

$$W_{ij} = \frac{-\eta}{96W} (F\eta_j^2 + G\eta_i^2)$$

Para efectuar la prueba asintótica para el coeficiente de correlación de Spearman se usa γ_s^* el cual se define:

$$\gamma_s^* = \frac{\gamma_s}{\sqrt{VARQ(\gamma_s)}}$$

$$VARQ(\gamma_s) = \frac{1}{\eta^2 W^2} \sum_i \sum_j \eta_{ij} (V_{ij} - \bar{V})^2$$

$$\bar{V} = \sum_i \sum_j \frac{\eta_{ij} V_{ij}}{\eta}$$

La varianza es derivada de una distribución multinomial referida a una tabla de contingencia.

El procedimiento SAS para encontrar el coeficiente de correlación de Pearson y Spearman referido a tablas de contingencia es:

DATA BASE;
INPUT X Y;
CARDS;
DATOS
;
PROC FREQ;
TABLES X*Y/MEASURES CHISQ;
RUN;

Referencias

- Altman DG. Practical statistics for medical research, C and Hall London, 1991. 611p.
- Bartholomew DJ. Spearman and the origin and development of test theory. *British J Mat Stat Psychol* 1995; 48:211-220.
- Burt C. Experimental test of general Inteligence. *British J Psychol* 1909; 3:94-177.
- Cattell RB, Spearman CE. *International encyclopedia of the social sciences*, D.E Sills ed New York, 1968; 15:108-111.
- Conover WJ, *Practical nonparametric statistical* 3ra ed New York, John Wiley, 1998. 578p.
- Dallas E. *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores. Mexico, D.F. 2000. 566p.
- Dawin-Saunders B, Trapp RG. *Bioestadística Medica*. 2da ed México, el Manual Moderno, 1996. 86p.
- Fanavos C. *Probabilidad y Estadística Aplicaciones y Métodos*. Editorial MCgraw-Hill. Mexico, D.F. 1988. 651p.
- Joe H. *Multivariate models and dependence conceptsts*. Chapman and Hall/CRC, Boca Ralton. New York, 1997. 395p.
- Levy P, Spearman's C. Contribution to Test Theory. *J Roy Statl Soc* 1995; 48:221-235.
- Looney SW, TR Jr. Use of the correlation coefficient with normal probability plots. *J Roy Stat Soc* 1985; 39:75-79.
- Martín AA, Luna JD. *Bioestadística para las ciencias de la salud*, 4ta ed Madrid, Norma, 1993. 114p.
- Milton JS, Toscos JO. *Estadística para biología y ciencias de la salud*, interamericana Mcgraw- Hill Madrid, 2001. 186p.
- Ostle B. *Estadística aplicada*, editorial científico técnica. Mexico D.F, 1980. 629p.
- Pita FS. *Correlación frente a la Causalidad*. JANO 1996; 4:59-260.
- Ruben H. Some new results on the distribution of the sample correlation coefficients. *J Roy Stat Soc* 1966.
- Spearman CE. General inteligence objectively determined and measured. *Am J Psicol* 1909; 15:201-293.
- Wayne WD, *Estadística con aplicaciones a las ciencias sociales y a la educación*, Mcgraw-Hill. Mexico, D.F, 1990. 504p.
- Zimymerman DW, Williams RH. Properties of the Spearman correction for attnuation for normal and realistic non-normal distributions. *Appl Psychol Measurm* 1997; 21:253-270.