



<Artículo metodológico>

Cálculo del intervalo de confianza para los coeficientes de correlación mediante sintaxis en SPSSJavier Santabárbara¹ 

Enviado: 08/03/2019. Aceptado: 23/05/2019. Publicado en prensa: 19/06/2019. Publicado: 01/07/2019

//Resumen

INTRODUCCIÓN: La asociación entre dos variables se expresa de forma analítica mediante el cálculo del estadístico llamado coeficiente de correlación. Una buena práctica de investigación requiere que, para tomar decisiones de investigación acertadas, deba reportarse no solo el valor p sino también el intervalo de confianza. Lamentablemente, el *software* estadístico más utilizado en el ámbito socioeducativo, SPSS, no proporciona en sus rutinas habituales el cálculo del intervalo de confianza para los coeficientes de correlación. Por tanto, los objetivos principales de este trabajo han sido calcular mediante sintaxis de SPSS el intervalo de confianza para los coeficientes de correlación a través de un caso práctico, destacando la importancia del reporte de un intervalo de confianza y proporcionando un modelo de redacción de los resultados de un estudio de correlación en un informe científico.

MÉTODO: Se proporcionan dos ficheros de sintaxis de SPSS (uno para muestras de pequeño tamaño y otro para muestras de gran tamaño) como herramienta útil para calcular el intervalo de confianza de un coeficiente de correlación.

RESULTADOS y DISCUSIÓN: El uso de los ficheros de sintaxis proporciona al investigador familiarizado con SPSS una forma rápida y sencilla de cálculo del intervalo de confianza para el coeficiente de correlación.

//Palabras clave

Correlación; Intervalo de confianza; Sintaxis SPSS.

//Datos del autor

¹ Departamento de Microbiología, Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad de Zaragoza, España. Autor para la correspondencia: jsantabarbara@unizar.es

//Referencia recomendada

Santabárbara, J. (2019). Cálculo del intervalo de confianza para los coeficientes de correlación mediante sintaxis en SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 12(2), 1–14. <http://doi.org/10.1344/reire2019.12.228245>

© 2019 Javier Santabárbara. Este artículo es de acceso abierto sujeto a la licencia Reconocimiento 4.0 Internacional de Creative Commons, la cual permite utilizar, distribuir y reproducir por cualquier medio sin restricciones siempre que se cite adecuadamente la obra original. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by/4.0/>





//Títol

Càlcul de l'interval de confiança per als coeficients de correlació mitjançant sintaxi en SPSS

//Resum

INTRODUCCIÓ: L'associació entre dues variables s'expressa de forma analítica mitjançant el càlcul de l'estadístic anomenat *coeficient de correlació*. Una bona pràctica d'investigació requereix que, per prendre decisions d'investigació encertades, es reporti no només el valor *p* sinó també l'interval de confiança. Lamentablement, el programari estadístic més utilitzat en l'àmbit socioeducatiu, SPSS, no proporciona en les seves rutines habituals el càlcul de l'interval de confiança per als coeficients de correlació. Els objectius principals d'aquest treball són calcular mitjançant sintaxi de SPSS l'interval de confiança per als coeficients de correlació a través d'un cas pràctic, destacar la importància de reportar un interval de confiança i proporcionar un model de redacció dels resultats d'un estudi de correlació en un informe científic.

MÈTODE: Es proporcionen dos fitxers de sintaxi de SPSS (un per a mostres de mida petita i un altre per a mostres de grans dimensions) com una eina útil per calcular l'interval de confiança d'un coeficient de correlació.

RESULTATS I DISCUSSIÓ: L'ús dels fitxers de sintaxi proporciona a l'investigador familiaritzat amb SPSS una forma ràpida i senzilla de calcular l'interval de confiança per al coeficient de correlació.

//Paraules clau

Correlació; Interval de confiança; Sintaxi SPSS.

//Title

Calculating the confidence interval of correlation coefficients using syntax in SPSS

//Abstract

INTRODUCTION: The association between two variables is expressed analytically by calculating the statistic called the correlation coefficient. In order to make sound research decisions, good research practice requires reporting not only the *p*-value but the confidence interval. While SPSS is the most commonly used statistical software in the socio-educational field, it does not provide the calculation of the confidence interval for the correlation coefficients in its usual routines. Therefore, the main objective of this study was to use a practical case to calculate by SPSS syntax the confidence interval for the correlation coefficients, highlighting the importance of reporting a confidence interval and providing a model for writing the results of a correlation study in a scientific report.

METHOD: Two SPSS syntax files are provided, one for small samples and one for large samples, as a useful tool to calculate the confidence interval of a correlation coefficient.

RESULTS and DISCUSSION: The use of syntax files provides the researcher familiar with SPSS with a quick and easy way to calculate the confidence interval for a correlation coefficient.

//Keywords

Correlation; Confidence interval; SPSS syntax.

1. Introducción: correlación

El concepto de relación o correlación se refiere al grado de dependencia existente entre dos o más variables. Este artículo se limita únicamente a considerar el caso de dos variables aleatorias (correlación *simple*).

Una relación lineal *positiva* entre X e Y significa que los valores de las dos variables varían de forma parecida: los individuos con valores grandes en X tienden a puntuar alto en Y , y los individuos con valores pequeños en X tienden a puntuar bajo en Y . Una relación lineal *negativa* significa que los valores de ambas variables varían justamente al revés: los individuos con valores pequeños en X tienden a puntuar alto en Y , y los individuos con valores grandes en X tienden a puntuar bajo en Y .

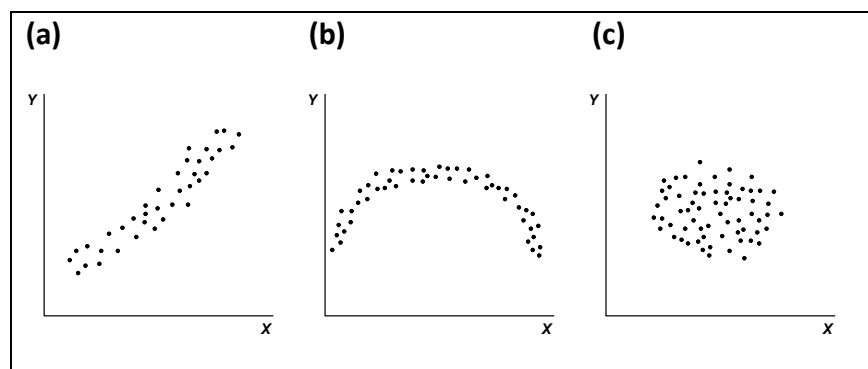
¿Cómo puede identificarse la relación funcional entre dos variables?

Diagrama de dispersión

El diagrama de dispersión proporciona, quizá, el medio más simple y útil para estudiar la relación entre dos variables X e Y . Se forma marcando un punto en un sistema de ejes cartesianos para cada par de observaciones (x_i, y_i) con $i = 1, 2, \dots, n$. Los valores de X se indican sobre el eje horizontal y los de Y sobre el vertical. Al conjunto de puntos representados en el diagrama de dispersión se le suele llamar *nube de puntos*. Al observarla, se aprecia el patrón indicativo de la relación funcional entre las dos variables.

En la Figura 1 se han representado algunos diagramas de dispersión, observándose en (a) una relación *lineal* o rectilínea entre las variables X e Y , que además, es positiva.

Figura 1. Distintos diagramas de dispersión



En cuanto a (b), la relación entre X e Y no es lineal, sino *curvilínea*, mientras que en (c) *no hay relación* entre ambas variables.

Así, cuando se construya un diagrama de dispersión se inspecciona su (Baldi y Moore, 2013; Gerstman, 2015):

- forma (p. e., línea recta = lineal, curvilínea, aleatoria...),



- dirección (tendencia creciente = asociación positiva, tendencia decreciente = asociación negativa, sin tendencia = no asociación), y
- fuerza/intensidad (cómo de próximos se encuentran los puntos adheridos a una línea recta imaginaria).

Hasta aquí se ha tratado la relación entre dos variables desde el punto de vista gráfico, pero ¿cómo podría expresarse estadísticamente dicha relación?

2. Justificación y objetivos del presente trabajo

La asociación entre dos variables cuantitativas se expresa de forma analítica mediante el cálculo del estadístico llamado coeficiente de correlación, como se verá en los apartados 3 y 4.

En la investigación educativa, la significación estadística y el intervalo de confianza son dos caras de una misma moneda; se complementan, pero no se sustituyen. Una buena práctica de investigación requiere que, para tomar decisiones de investigación acertadas, ambas partes deban ser consideradas. Sin embargo, el *software* estadístico más utilizado en el ámbito socioeducativo, SPSS, no proporciona en sus rutinas habituales el cálculo del intervalo de confianza para los coeficientes de correlación.

Por tanto, los objetivos principales de este trabajo fueron (a) calcular mediante sintaxis de SPSS el intervalo de confianza para el coeficiente de correlación (para muestras pequeñas o grandes) a través de un caso práctico que complementa el trabajo de Reguant-Álvarez, Vilà-Baños y Torrado-Fonseca (2018) sobre la asociación de dos variables, (b) exponer con claridad la importancia del reporte de un intervalo de confianza y (c) presentar un modelo de redacción de los resultados de un estudio de correlación en un informe científico.

Cabe destacar que en el presente artículo se muestran dos tipos de cálculo del intervalo de confianza:

- Método “exacto” para utilizar con muestras pequeñas (menos de 30 observaciones, $n < 30$). Con fines meramente pedagógicos se mostrará su uso para el coeficiente de correlación de Pearson (apartado 3).
- Método “asintótico” para utilizar con muestras grandes (igual o más de 30 observaciones, $n \geq 30$). Con fines meramente pedagógicos se mostrará su uso para el coeficiente de correlación de Spearman (apartado 4).

3. Coeficiente de correlación lineal de Pearson

Definición y expresión matemática

Para una interpretación adecuada de la relación entre dos variables, lo ideal sería disponer de una medida de la relación entre ellas que no esté influenciada por las unidades en que vienen medidas. Una forma de conseguir este objetivo es dividir la covarianza (medida de la variabilidad conjunta) por el producto de las desviaciones típicas de las variables, ya que así se obtiene un coeficiente

adimensional denominado coeficiente de correlación lineal de Pearson que se simboliza con r y cuya expresión matemática es:

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

donde s_{xy} es la covarianza entre X e Y , s_x es la desviación típica de X , y s_y la de Y .

Nótese que el coeficiente de correlación lineal de Pearson en la población será siempre desconocido y se estimará mediante el obtenido en la muestra: r .

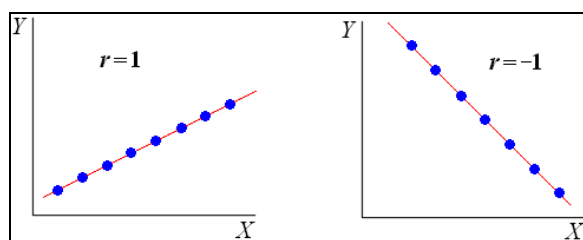
Una expresión equivalente (Martínez-González, Sánchez-Villegas, Toledo Atucha y Faulin Fajardo, 2014) es:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Los valores de r varían en el intervalo $[-1, 1]$ y su interpretación es la siguiente:

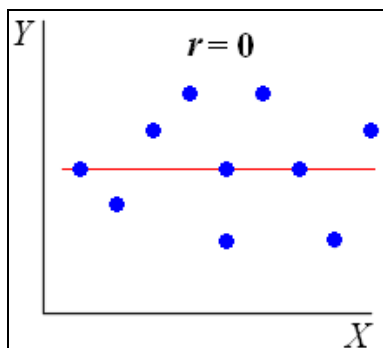
1. Si $|r| = 1$ entonces hay una relación lineal perfecta entre X e Y .

Figura 2. Relación lineal perfecta entre X e Y



2. Si $r = 0$ entonces no hay relación lineal entre ambas variables, diciéndose que son incorreladas.

Figura 3. Ausencia de relación lineal entre X e Y

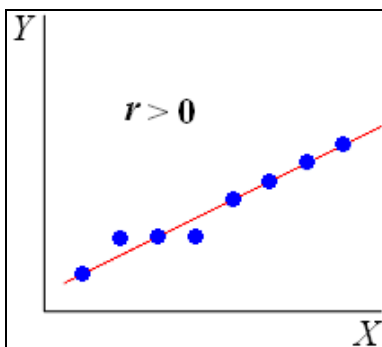


Esto no quiere decir que X e Y no estén asociadas, sino simplemente que, de existir asociación entre ambas, no será lineal, pudiendo ser parabólica, exponencial, logarítmica, etc.



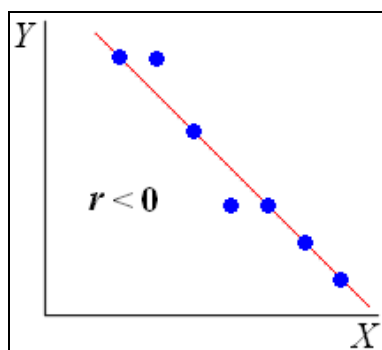
- Si $r > 0$ entonces hay una relación lineal positiva entre X e Y , de modo que cuando X crece (o decrece) entonces Y crece (o decrece).

Figura 4. Relación lineal positiva entre X e Y



- Si $r < 0$ entonces la relación entre X e Y es negativa, esto es, cuando X crece (o decrece) entonces Y decrece (o crece).

Figura 5. Relación lineal negativa entre X e Y

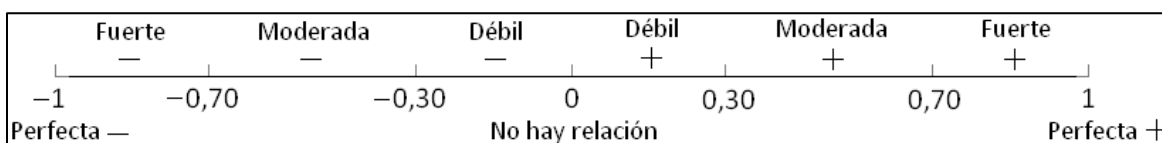


Ahora bien, ¿cómo se interpreta el valor del coeficiente de correlación de Pearson?

Interpretación

En la práctica, se encuentran coeficientes de correlación de Pearson que no están próximos a 0, 1 o -1, por lo que se necesitará algún criterio que aclare la intensidad de la relación entre las variables según el valor de su coeficiente de correlación. Se propone, por ejemplo, la escala mostrada en la figura siguiente con el fin de interpretar r (Martínez González *et al.*, 2014):

Figura 6. Interpretación del coeficiente de correlación lineal de Pearson



Según esta escala, un coeficiente de correlación de, por ejemplo, 0,36 puede interpretarse como una correlación positiva «moderada».

Se ha calculado una estimación puntual del verdadero coeficiente de correlación de Pearson en la población (ρ), pero la Inferencia Estadística permite una estimación de ρ mucho más informativa: por intervalo.

Intervalo de confianza para el coeficiente de correlación lineal de Pearson

En la investigación educativa, la significación estadística y el intervalo de confianza son dos caras de una moneda; se complementan, pero no se sustituyen. Una buena práctica de investigación requiere que, para tomar decisiones de investigación acertadas, ambas partes deben ser consideradas (Fan, 2001).

Así, puede calcularse un intervalo de confianza (exacto) para el coeficiente de correlación de Pearson. Lamentablemente, *SPSS* no ofrece el cálculo del intervalo de confianza para el coeficiente de correlación de Pearson, luego se proporciona el archivo de sintaxis *intervconf_exacto.sps* para tal objetivo (basado en la expresión matemática original de Jeyaratman (1992) y recogida en Gerstman, 2015), cuyo contenido es el que se muestra en la Figura 7.

Figura 7. Archivo de sintaxis *intervconf_exacto.sps*

```
* Introducimos el coeficiente de correlación (r), el tamaño de la muestra (n), y el nivel de confianza
del intervalo (conf).

data list free / r n conf.
begin data.
-0,635 37 0,95
end data.

*Calculamos t.

compute t = idf.t(1-((1-conf)/2),n-2).

* Calculamos w.

compute w = sqrt((t**2)/((t**2)+n-2)).

* Calculamos los límites inferior y superior del intervalo.

compute inf_r = (r-w)/(1-r*w).
compute sup_r = (r+w)/(1+r*w).

formats r conf to sup_r (f10.3) / n (f8).
list r n conf inf_r sup_r.
```

Caso práctico

Para presentar el cálculo del intervalo de confianza para el coeficiente de correlación de Pearson, se utilizará el ejemplo de Reguant-Álvarez *et al.* (2018) en el que estudiaron si existe una relación entre las calificaciones finales obtenidas por cada estudiante y el número de inasistencias que acumula

durante el semestre en una muestra de 37 estudiantes, ambas variables escalares. El resultado se muestra en la Tabla 1.

Tabla 1. Coeficiente de correlación de Pearson

| | | Cantidad de inasistencia durante el semestre | Calificaciones finales |
|--|------------------------|--|------------------------|
| Cantidad de inasistencia durante el semestre | Correlación de Pearson | 1 | -,635** |
| | Sig. (bilateral) | | ,000 |
| | N | 37 | 37 |
| Calificaciones finales | Correlación de Pearson | -,635** | 1 |
| | Sig. (bilateral) | ,000 | |
| | N | 37 | 37 |

** . La correlación es significativa en el nivel 0,01 (2 colas)

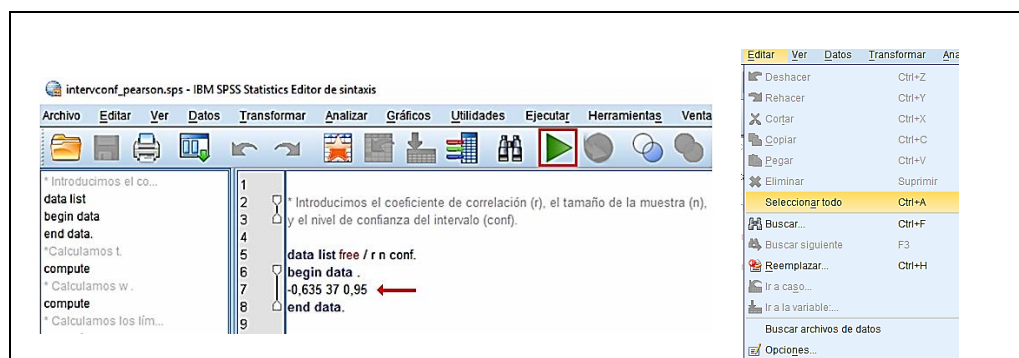
Fuente: Reguant-Álvarez *et al.* (2018).

Así, en la muestra de 37 estudiantes, el valor del coeficiente de correlación de Pearson (r) es $-0,635$ ($p < 0,001$), esto es, existe una relación negativa, significativa y de intensidad negativa, concluyendo que en la medida que la inasistencia aumenta, las calificaciones disminuyen *significativamente*.

Ahora bien, sería interesante no solo reportar una estimación puntual del grado de asociación sino también por intervalo, esto es: entre que valores esperaríamos que variara el verdadero coeficiente de correlación de Pearson con una confianza del 95% (o cualquier otra).

En el fichero de sintaxis se introducirá el valor del *coeficiente de correlación de Pearson* ($-0,635$), el tamaño muestral (37) y el nivel de confianza (en tanto por uno: 0,95) según se disponen en la Figura 8 (ver flecha). A continuación, para seleccionar el texto del archivo de sintaxis, elegiremos el procedimiento *Editar* y, dentro de él, *Seleccionar todo* (Figura 8, dcha.):

Figura 8. Archivo de sintaxis *intervconf_exacto.sps*



Una vez seleccionada la totalidad del texto, se pulsa el botón [Play] (recuadro rojo de la Figura 8, izda.) para ejecutar el programa e inmediatamente aparecerá en el *Visor de resultados* el intervalo de confianza buscado (Tabla 2):



Tabla 2. Intervalo de confianza del 95% para el coeficiente de correlación lineal de Pearson

| <i>r</i> | <i>n</i> | <i>conf</i> | <i>inf_r</i> | <i>sup_r</i> |
|----------|----------|-------------|--------------|--------------|
| -,635 | 37 | ,950 | -,796 | -,391 |

Con fines meramente pedagógicos se ha utilizado el método “exacto” en el cálculo del intervalo de confianza a pesar de tener más de 30 observaciones muestrales. Opcionalmente, se podría aplicar el intervalo de confianza “asintótico” (ver apartado epígrafe 4.3), obteniendo un resultado muy similar (-0,795 a -0,392), ya que ambos métodos convergen para tamaños de muestra grandes ($n > 30$).

¿Siempre se puede calcular el intervalo de confianza para coeficiente de correlación lineal de Pearson?

Supuestos de aplicación del coeficiente de correlación lineal de Pearson

Para poder aplicar el coeficiente de correlación lineal de Pearson (y realizar una estimación por intervalo y un contraste de hipótesis sobre su valor) se han de cumplir los siguientes supuestos:

1. Ambas variables en estudio han de ser cuantitativas (continuas).
2. Relación lineal entre ambas variables.
3. Normalidad de ambas variables en estudio.
4. Las observaciones han de ser independientes, es decir, solo hay una observación de cada variable para cada individuo.

¿Qué ocurre cuando no se cumplen las condiciones anteriores?

Se recurre al coeficiente de correlación no paramétrico de Spearman (apartado 4).

Presentación de los resultados

El reporte de la correlación lineal debería incluir los siguientes elementos (Cumming y Calin-Jageman, 2017):

- Estadísticos descriptivos de las variables X e Y .
- Diagrama de dispersión.
- El tamaño de la muestra para el cálculo del coeficiente de correlación lineal (debido a valores perdidos puede no coincidir con el tamaño de la muestra en estudio). Si se violara el supuesto de normalidad para alguna de las variables, no reportar el intervalo de confianza.
- Coeficiente de correlación lineal junto con el intervalo de confianza para el coeficiente poblacional.

- Si se desea, el *valor p* del contraste de hipótesis para el coeficiente de correlación poblacional.
- Una interpretación del coeficiente de correlación lineal y su intervalo de confianza, pero evitando en todo momento hablar de causalidad.
- Descripción de los métodos utilizados para comprobar los supuestos de aplicación del coeficiente de correlación (en el apartado de análisis estadístico).

Los resultados realizados en este capítulo se podrían presentar de la siguiente manera (Field, 2018):

Tabla 3. Coeficiente de correlación de Pearson

| | Calificaciones finales | | | |
|--|------------------------|----------|-------------------|----------|
| | <i>n</i> | <i>r</i> | IC 95 % | <i>p</i> |
| Cantidad de inasistencia durante el semestre | 37 | -0,635 | (-0,796 a -0,391) | < 0,001 |

Y se escribiría en el apartado de resultados:

«En nuestra muestra de 37 estudiantes, el valor del coeficiente de correlación de Pearson (r) entre la cantidad de inasistencia en el semestre y la calificación final es -0,635. El signo negativo del coeficiente indica que, a medida que aumenta la inasistencia, disminuye la calificación final. La magnitud absoluta del coeficiente es 0,635 (siendo el mínimo posible 0 y el máximo 1), revelando que la intensidad de la asociación entre ambas variables es *moderada*, ya que es el 63,5% de la máxima posible. Por tanto, existe una moderada relación lineal negativa entre ambas variables. Tenemos una confianza del 95% en que el *verdadero* coeficiente de correlación lineal de Pearson entre la cantidad de inasistencia y la calificación final, en la población de estudiantes, varía entre -0,796 y -0,391. Finalmente, como este intervalo de confianza no contiene el valor cero, concluimos que la asociación encontrada es estadísticamente significativa (con un nivel de significación de 0,05). El coeficiente de correlación es estadísticamente significativo a un nivel del 1 por mil ($p < 0,001$). En consecuencia, consideraremos que existe una probabilidad muy pequeña de que la asociación encontrada entre la cantidad de inasistencia y la calificación final se deba únicamente al azar».

4. Coeficiente de correlación de Spearman

Definición

El coeficiente de correlación de Spearman (r_s o *rho*) es un estimador *no paramétrico* que se utiliza en aquellos casos en que las variables examinadas son ordinales, no cumplen necesariamente criterios de normalidad o el diagrama de dispersión no muestra una relación funcional lineal entre las variables. Esto es, cuando no es aplicable el coeficiente de correlación lineal de Pearson (r).

A diferencia de este, el *coeficiente de correlación de Spearman* estima la intensidad de la relación entre dos variables no necesariamente lineal (con la excepción de una función curvilínea, en la que ambos coeficientes valdrán cero).

Notemos que el verdadero coeficiente de correlación de Spearman en la población será siempre desconocido y se estimará mediante el obtenido en la muestra: *rho*.

Expresión matemática

La forma más sencilla de calcular el coeficiente de correlación de Spearman (*rho*) es transformar los valores originales de las variables en sus rangos y, con ellos, calcular el coeficiente de correlación de Pearson (Triola, 2013), en el supuesto de que existan empates. Si no existen empates entre los rangos para ninguna de las dos variables, la correlación se puede obtener calculando la diferencia entre los rangos de *X* e *Y* (*d*), y aplicando la siguiente expresión (Triola, 2013):

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

El coeficiente de correlación de Pearson (*r*) cuantifica la relación *lineal* entre dos variables cuantitativas *X* e *Y*, y de alguna forma responde a la pregunta: cuándo *X* aumenta (disminuye), ¿lo hace también *Y* en una cantidad *constante*? Por el contrario, dado que el coeficiente de correlación de Spearman (*rho*) únicamente se basa en los rangos de los valores, luego, de alguna forma, únicamente responde a la pregunta: cuando *X* aumenta (disminuye), ¿lo hace también *Y* (en cualquier cantidad)?

Intervalo de confianza para el coeficiente de correlación de Spearman

No se ahondará en el procedimiento matemático para el cálculo del intervalo de confianza (asintótico) para el coeficiente de correlación de Spearman; aquellos interesados pueden consultar Martínez-González *et al.* (2014).

Para poder calcular el intervalo de confianza del 95% para el coeficiente de correlación de Spearman poblacional con *SPSS*, se proporciona el archivo de sintaxis *intervconf_asintotico.sps*, cuyo contenido es el que aparece en la Figura 9.

Figura 9. Archivo de sintaxis *intervconf_asintotico.sps*

* Introducimos el coeficiente de correlación de Spearman (*rho*), el tamaño de la muestra (*n*), y el nivel de confianza del intervalo (*conf*).

```
data list free / rho n conf.
begin data.
0,505 37 0,95
end data.
```

* Transformamos *r* en el valor *z*: *fz*.
compute *fz* = 0.5*ln((1+rho)/(1-rho)).

(Continúa en la siguiente página)

```

* Calculamos el error de estimación: sez.
compute sez = 1/sqrt(n-3).

* Calculamos el valor crítico de la curva normal: critz.
compute critz = abs(idf.normal((1 - conf)/2,0,1)).

* Calculamos los límites inferior y superiores para fz.
compute lo_fz = fz - critz*sez.
compute hi_fz = fz + critz*sez.

* Deshacemos la transformación fz en r.
compute inf_rho = (exp(2*lo_fz) - 1)/(exp(2*lo_fz) + 1).
compute sup_rho = (exp(2*hi_fz) - 1)/(exp(2*hi_fz) + 1).

formats rho conf to sup_rho (f10.2) / n (f8).

list rho n conf inf_rho sup_rho.

```

Caso práctico

Continuando con el ejemplo de Reguant-Álvarez *et al.* (2018), se desea estudiar cómo se relacionan dos variables de escala ordinal: la asistencia a clase, cuyas categorías son: muy irregular (1), irregular (2) y regular (3); y la nota obtenida, cuyas categorías son: suspenso (1), aprobado (2), notable (3) y excelente (4).

Estas autoras obtuvieron el siguiente resultado (Tabla 5):

Tabla 5. Coeficiente de correlación de Spearman

| | | Asistencia ordinal | Nota examen ordinal |
|---------------------|----------------------------|--------------------|---------------------|
| Asistencia ordinal | Correlación de correlación | 1 | ,505** |
| | Sig. (bilateral) | | ,001 |
| | N | 37 | 37 |
| Nota examen ordinal | Correlación de correlación | ,505** | 1 |
| | Sig. (bilateral) | ,001 | |
| | N | 37 | 37 |

** . La correlación es significativa en el nivel 0,01 (2 colas)

Fuente: Reguant-Álvarez *et al.* (2018).

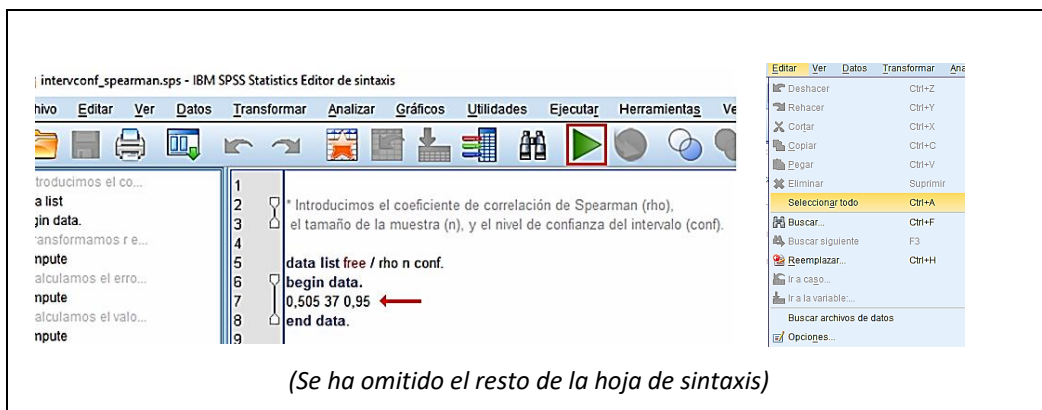
Esto es, en la muestra de 37 estudiantes, el valor del coeficiente de correlación de Spearman (*rho*), que es 0,505 ($p = 0,001$), concluyendo que en la medida que la asistencia a clase es regular, los resultados de aprendizaje mejoran *significativamente*.

J. Santabàrbara. Cálculo del intervalo de confianza para los coeficientes de correlación mediante sintaxis en SPSS

Ahora bien, sería interesante no solo reportar una estimación puntual del grado de asociación sino también su estimación por intervalo: entre que valores se esperaría que variara el verdadero coeficiente de correlación de Spearman con una confianza del 95% (o cualquier otra).

En el fichero de sintaxis se introducirá el valor del coeficiente de correlación (-0,505), el tamaño de la muestra (37), y el nivel de confianza del intervalo (0,95). Se elegirá después el procedimiento *Edición* y, dentro de él, *Seleccionar todo*.

Figura 10. Archivo de sintaxis *intervconf_asintotico.sps*



Una vez seleccionado todo el texto, se pulsa el botón [Play] (ver recuadro rojo de la Figura 10, izda.) para ejecutar el programa, e inmediatamente aparecerá en el *Visor* el intervalo de confianza buscado (Tabla 6):

Tabla 6. Intervalo de confianza del 95% para el coeficiente de correlación lineal de Spearman

| <i>rho</i> | <i>n</i> | <i>conf</i> | <i>inf_rho</i> | <i>sup_rho</i> |
|------------|----------|-------------|----------------|----------------|
| ,505 | 37 | ,950 | ,216 | ,712 |

Presentación de los resultados

Los resultados realizados en este estudio se podrían presentar de la siguiente manera (Field, 2018):

Tabla 7. Presentación del coeficiente de correlación de Spearman

| | Nota examen (ordinal) | | | |
|----------------------|-----------------------|------------|-----------------|----------|
| | <i>n</i> | <i>rho</i> | IC 95 % | <i>p</i> |
| Asistencia (ordinal) | 37 | 0,505 | (0,216 a 0,712) | 0,001 |

Y se escribiría en el apartado de resultados:

«En nuestra muestra de 37 estudiantes, el valor del coeficiente de correlación de Spearman (*rho*) entre la asistencia y la nota en el examen es 0,505; indicando que la intensidad de la asociación entre ambas variables es *moderada*. Tenemos una confianza del 95% en que el *verdadero* coeficiente de correlación lineal de Spearman entre la asistencia y la nota en el examen varía entre 0,216 y 0,712 en la población de estudiantes. Finalmente, como este intervalo de confianza no contiene el valor

cero, concluimos que la asociación encontrada es estadísticamente significativa a un nivel del 1 por mil ($p = 0,001$). Por tanto, consideraremos que existe una probabilidad muy pequeña ($p = 0,001$) de que la asociación encontrada entre estas dos variables se deba únicamente al azar».

5. Conclusión

A pesar de que SPSS solo arroja el resultado del coeficiente de correlación y el *valor p* del contraste de significación, el cálculo del intervalo de confianza con los ficheros proporcionados en este trabajo (disponibles en material suplementario) es un proceso relativamente sencillo y rápido. Así, el presente trabajo supone una ayuda metodológica a los investigadores del campo de las ciencias de la educación que deseen reportar no solo la intensidad de la asociación entre dos variables (coeficiente de correlación) sino también entre qué valores varía dicho coeficiente con una confianza determinada (intervalo de confianza).

<Referencias bibliográficas>

- Baldi, B., y Moore, D. S. (2013). *The practice of statistics in the life sciences* (3ª ed.). Nueva York: McMillan Learning.
- Cumming, G., y Calin-Jageman, R. (2017). *Introduction to new statistics. Estimation, open science, & beyond*. Nueva York: Routledge.
- Fan, X. (2001). Statistical significance and effect size in education research: two sides of a coin. *The Journal of Educational Research*, 94(5), 275–282. <http://dx.doi.org/10.1080/00220670109598763>
- Gerstman, B. (2015). *Basic Biostatistics: Statistics for Public Health Practice* (2ª ed.). Burlington (MA): Jones and Bartlett.
- Field, A. (2018). *Discovering statistics using SPSS* (5ª ed.). Londres: SAGE Pub.
- Jeyaratman, S. (1992). Confidence intervals for the correlation coefficient. *Statistics and Probability Letters*, 15(5), 389–393. [https://dx.doi.org/10.1016/0167-7152\(92\)90172-2](https://dx.doi.org/10.1016/0167-7152(92)90172-2)
- Martínez-González, M. A., Sánchez-Villegas, A., Toledo Atucha, E. A., y Faulin Fajardo, J. (2014). *Bioestadística amigable* (3ª ed.). Barcelona: Elsevier.
- Reguant-Álvarez, M., Vilà-Baños, R., y Torrado-Fonseca, M. (2018). La relación entre dos variables según la escala de medición con SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 11(2), 45–60. <http://dx.doi.org/10.1344/reire2018.11.221733>
- Triola, M. F. (2013). *Estadística* (11ª ed.). México: Pearson Educación.

<Archivos complementarios>

Sintaxis SPSS para el cálculo del intervalo de confianza exacto de una correlación (muestras pequeñas): <http://revistes.ub.edu/index.php/REIRE/rt/suppFiles/28245/0>

Sintaxis SPSS para el cálculo del intervalo de confianza asintótico de una correlación (muestras grandes): <http://revistes.ub.edu/index.php/REIRE/rt/suppFiles/28245/1>