

# Cálculo del tamaño de muestra necesario para estimar el coeficiente de correlación de Pearson mediante sintaxis en SPSS

Javier Santabárbara<sup>1</sup> 

Artículo metodológico. Recibido: 11/10/2020. Aceptado: 09/12/2020. Publicación avanzada: 18/12/2020. Publicado: 04/01/2021.

## Resumen

**INTRODUCCIÓN.** El estudio de la asociación entre dos variables requiere del cálculo del estadístico llamado coeficiente de correlación. En la fase de diseño de este tipo de estudios, debe realizarse la estimación del número mínimo de participantes a seleccionar para asegurar que podremos realizar inferencias estadísticas con una potencia adecuada. Desafortunadamente, SPSS no ofrece el cálculo del tamaño muestral para los estudios de asociación de variables cuantitativas. Por tanto, el objetivo principal de este trabajo ha sido calcular mediante sintaxis de SPSS el tamaño de la muestra necesario para estimar el coeficiente de correlación a través de un caso práctico.

**MÉTODO.** Se proporciona un fichero de sintaxis de SPSS para el cálculo del tamaño muestral en estudios de asociación.

**RESULTADOS y DISCUSIÓN.** El uso del fichero de sintaxis servirá de apoyo metodológico para todo investigador usuario de SPSS interesado en el cálculo del tamaño muestral para la estimación del coeficiente de correlación.

## Palabras clave

Correlación; Tamaño muestral; Sintaxis SPSS.

## Datos del autor

<sup>1</sup> Universidad de Zaragoza, España. Departamento de Microbiología, Pediatría, Radiología y Salud Pública. Facultad de Medicina. Contacto para la correspondencia: [jsantabarbara@unizar.es](mailto:jsantabarbara@unizar.es)

## Referencia recomendada

Santabárbara, J. (2021). Cálculo del tamaño de muestra necesario para estimar el coeficiente de correlación de Pearson mediante sintaxis en SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 14(1), 1–7. <https://doi.org/10.1344/reire2021.14.132565>

© 2021 El autor. 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 <https://creativecommons.org/licenses/by/4.0/>



#### Títol (català)

Càlcul de la mida de mostra necessària per estimar el coeficient de correlació de Pearson mitjançant sintaxi en SPSS

#### Resum

**INTRODUCCIÓ.** L'estudi de l'associació entre dues variables requereix el càlcul de l'estadístic anomenat «coeficient de correlació». En la fase de disseny d'aquest tipus d'estudis s'ha de fer l'estimació del nombre mínim de participants que cal seleccionar, per assegurar que podem fer inferències estadístiques amb una potència adequada. Desafortunadament, SPSS no ofereix el càlcul de la grandària mostral per als estudis d'associació de variables quantitatives. Per tant, l'objectiu principal d'aquest treball ha estat calcular, mitjançant sintaxi de SPSS, la mida de la mostra necessària per estimar el coeficient de correlació a través d'un cas pràctic.

**MÈTODE.** Es proporciona un fitxer de sintaxi de SPSS per al càlcul de la grandària mostral en estudis d'associació.

**RESULTATS i DISCUSSIÓ.** L'ús del fitxer de sintaxi servirà de suport metodològic per a qualsevol investigador usuari d'SPSS interessat en el càlcul de la grandària mostral per a l'estimació del coeficient de correlació.

#### Paraules clau

Correlació; Mida de la mostra; Sintaxi SPSS.

#### Title (English)

Calculation of the sample size required to estimate the Pearson correlation coefficient using SPSS syntax

#### Abstract

**INTRODUCTION.** The study of the association between two variables requires the calculation of the statistic called the correlation coefficient. In the design phase of this type of study, the minimum number of participants to be selected must be estimated to ensure that the statistical inferences made have adequate power. Unfortunately, SPSS does not offer the calculation of the sample size for the association studies of quantitative variables. Therefore, the main aim of this study, using SPSS syntax, is to calculate, the sample size necessary to estimate the correlation coefficient through a practical case.

**METHOD.** An SPSS syntax file is provided for the calculation of the sample size in association studies.

**RESULTS and DISCUSSION.** The use of the syntax file will serve as methodological support for all SPSS users and researchers interested in calculating the sample size in order to estimate the correlation coefficient.

#### Keywords

Correlation; Sample size; SPSS syntax.

## 1. Introducción: importancia del cálculo del tamaño muestral

En ocasiones, en las investigaciones que estudian a los seres humanos, no es posible seleccionar todos los elementos que forman parte de una población. En estos casos, en la fase de planificación de la investigación debe determinarse el tamaño muestral necesario para su ejecución. No realizar este proceso puede conducirnos a dos situaciones diferentes: la primera consecuencia podría ser, que realizásemos el estudio sin el número adecuado de individuos, con lo cual no podríamos estimar los parámetros de forma precisa y además no encontraríamos diferencias significativas cuando en la realidad sí existen –escenario de escasa potencia estadística–; y la segunda, que podríamos estudiar un número innecesario de individuos, lo cual lleva implícito no solo la pérdida de tiempo e incremento de recursos innecesarios, sino que, además, la calidad del estudio, dado dicho incremento, puede verse afectada en sentido negativo (Santabárbara et al., 2015).

En la fase del cálculo de este tamaño muestral, es fundamental saber qué tipo de estadístico utilizaremos para llevar a cabo nuestro estudio (diferencia de medias, diferencia de proporciones, coeficiente de correlación, etc.). Es importante destacar que el cálculo de tamaño muestral será distinto para cada tipo de escenario (Santabárbara et al., 2015).

## 2. Justificación y objetivos

El presente trabajo se focaliza en los estudios de asociación entre variables cuantitativas. Para analizar la asociación entre dos variables de naturaleza cuantitativa, deberemos calcular el coeficiente de correlación, siendo el más extendido el de Pearson –el lector interesado en la aplicación de esta técnica estadística puede consultar Reguant-Álvarez et al. (2018), así como Santabárbara (2019). A modo de recordatorio, en los estudios de asociación, se pone a prueba la hipótesis nula ( $H_0$ ) de no relación entre las variables (coeficiente de correlación poblacional igual a cero:  $\rho = 0$ ) frente a la alternativa de existencia de asociación ( $H_1: \rho \neq 0$ ). Así, en el análisis del contraste de hipótesis, la atención principal está puesta en  $\alpha$ , la probabilidad de cometer un error de tipo I (rechazar la hipótesis nula cuando es verdadera), junto con  $\beta$ , la probabilidad de cometer un error de tipo II (no rechazar la hipótesis nula cuando es falsa). Hay una razón que justifica esta diferencia. Para un contraste dado,  $\alpha$  es un solo número asignado por el investigador antes de realizar el contraste. Es una medida del riesgo aceptable de rechazar una hipótesis nula verdadera. Por otra parte,  $\beta$  puede adoptar uno de muchos valores. Supóngase que se pretende probar la hipótesis nula de que un parámetro de alguna población (p.e.,  $\rho$ ) es igual a algún valor específico (p.e., 0). Si  $H_0$  ( $\rho = 0$ ) es falsa y no se rechaza, se comete el error de tipo II. Si el valor supuesto del parámetro no es el valor correcto, el valor de  $\beta$  (la probabilidad de cometer un error de tipo II) depende de los siguientes factores: 1) valor correcto del parámetro de interés, 2) valor supuesto del parámetro, 3) valor de  $\alpha$ , y 4) tamaño de la muestra ( $n$ ) (Pardo et al., 2009).

En caso de que no se realice una adecuada estimación del tamaño muestral necesario, el investigador podría llevarse la desagradable sorpresa de no encontrar una asociación estadísticamente significativa cuando analice sus datos, dado que quizá no se disponga de suficiente potencia estadística para ello.

En el escenario de la investigación educativa, si disponemos del *software* estadístico SPSS, nos encontraremos ante la imposibilidad de realizar la estimación del tamaño de la muestra para el coeficiente de correlación de Pearson. Por lo tanto, el objetivo principal del presente trabajo es calcular mediante sintaxis de SPSS la estimación del tamaño de muestra necesario para llevar a cabo un estudio de asociación

de variables, en el que se plantea la utilización del coeficiente de correlación de Pearson, utilizando los resultados del trabajo previo de Reguant-Álvarez et al. (2018).

### 3. Tamaño de la muestra para el coeficiente de correlación lineal de Pearson

#### 3.1. Definición y expresión matemática

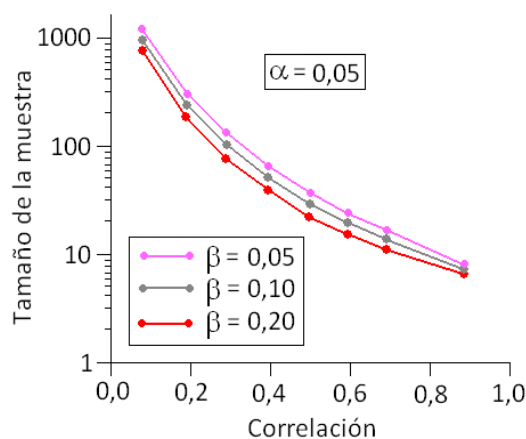
Para obtener el número *mínimo* ( $n$ ) de sujetos necesarios en un estudio de correlación debe fijarse el valor del coeficiente de correlación que se considera de relevancia para el investigador ( $\rho$ ), así como la probabilidad de cometer un error de tipo I ( $\alpha$ ) y la probabilidad de cometer un error de tipo II ( $\beta$ ), siendo el contraste bilateral (habitual), y aplicar la fórmula (Hulley et al., 2013):

$$n = \left( \frac{z_{1-\alpha/2} + z_{1-\beta}}{\frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right)} \right)^2 + 3$$

Así, en la Figura 1 podremos observar cómo varía el tamaño de la muestra necesario en un estudio de correlación para un nivel de significación –probabilidad de error de tipo I– fijado en el 5 % (habitual) según los valores del coeficiente de correlación de significancia para el investigador y la probabilidad de error de tipo II en 5 %, 10 % y 20 % –habitual– (Norman y Streiner, 2014).

**Figura 1**

*Tamaño de la muestra necesario fijado el nivel de significación*



*Nota.* Norman y Streiner (2014) proponen esta fórmula *aproximada* para los valores habituales ( $\alpha = 0,05$ ,  $1 - \beta = 0,8$  y contraste bilateral):

$$n = \frac{8}{r^2} + 2$$

#### 3.2. Fichero de sintaxis de SPSS

Desafortunadamente, SPSS no dispone de la rutina de cálculo del tamaño de muestra necesario para el coeficiente de correlación de Pearson, luego se proporciona el archivo de sintaxis nominado

*tamañomuestra\_correlacion.sps* para tal objetivo –basado en la expresión matemática original de Hulley et al. (2013)–, cuyo contenido se muestra en la Figura 2.

### Figura 2

*Archivo de sintaxis tamañomuestra\_correlacion.sps*

---

```
* Introducimos el coeficiente de correlación (r),  
el nivel de significación (sig) y la potencia del contraste  
(potencia).  
  
data list free / r sig potencia.  
begin data.  
-0,6 0,05 0,8  
end data.  
  
* Transformamos r en el valor z: fz.  
compute fz = 0.5*ln((1+r)/(1-r)).  
  
* Calculamos el valor crítico de la curva normal para el nivel de  
significación: critz_sig.  
compute critz_sig = abs(idf.normal(sig/2,0,1)).  
  
* Calculamos el valor crítico de la curva normal para la potencia:  
critz_potencia.  
compute critz_potencia = abs(idf.normal(1-potencia,0,1)).  
  
* Calculamos el tamaño de la muestra (n).  
compute n = rnd((((critz_sig + critz_potencia)/fz)**2)+3,1,0).  
  
formats r (f10.2) sig potencia (f10.2) / n (f8).  
  
* Mostramos el tamaño de muestra necesario.  
list r sig potencia n.
```

---

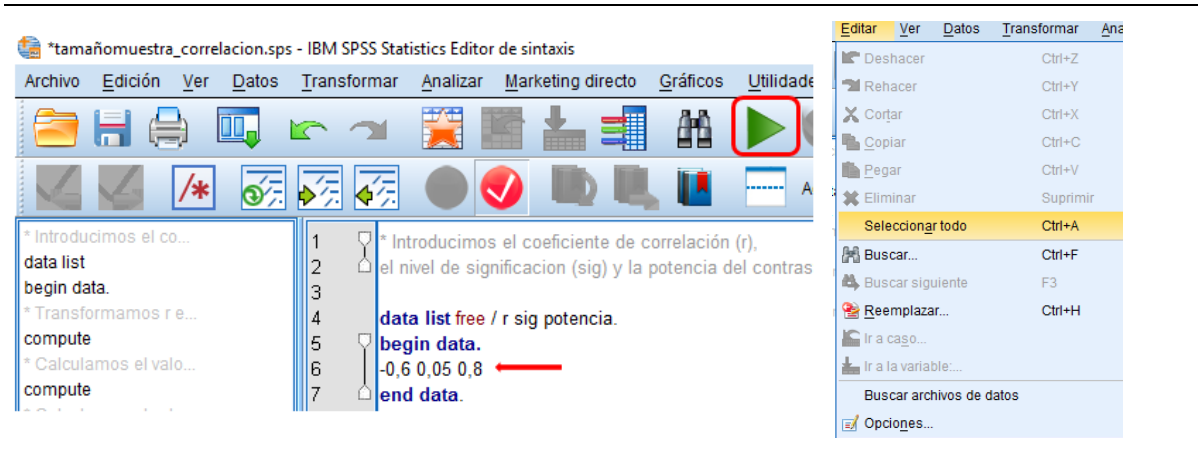
### 3.3. Caso práctico

Para ejemplificar el procedimiento de cálculo mediante sintaxis de SPSS nos basamos en el estudio de Reguant-Álvarez et al. (2018) el cual trataba la asociación entre las calificaciones finales obtenidas por cada estudiante y el número de inasistencias que acumula durante el semestre. A continuación se mostrará cómo estimar el tamaño de muestra necesario para llevar a cabo un estudio de asociación. Supongamos que, en la fase de diseño del estudio, estos investigadores postularon que podrían concluir que existía una asociación entre ambas variables si el coeficiente de correlación de Pearson hallado fuese, como mínimo, de -0,6. Además, estaban dispuestos a asumir una probabilidad de cometer un error de tipo I ( $\alpha$ ) de 0,05 y la de tipo II ( $\beta$ ) de 0,2 (potencia:  $1-\beta = 0,8$ ).

En el fichero de sintaxis se introducirá el valor del *coeficiente de correlación de Pearson (-0,6)*, el nivel de significación (en tanto por uno: 0,05) y la potencia del contraste (en tanto por uno: 0,8) según se disponen en la Figura 3 (izda., ver flecha). A continuación, para seleccionar el texto del archivo de sintaxis, elegiremos el procedimiento *Editar* y, dentro de él, *Seleccionar todo* (Figura 3, dcha.):

**Figura 3**

*Archivo de sintaxis tamaño muestra\_correlacion.sps.*



Una vez seleccionada la totalidad del texto, se pulsa el botón [Play] (recuadro rojo de la Figura 3, izda.) para ejecutar el programa e inmediatamente aparecerá en el *Visor de resultados* el intervalo de confianza buscado (Tabla 1):

**Tabla 1**

*Tamaño de muestra necesario para el coeficiente de correlación lineal de Pearson*

<i>r</i>	<i>sig</i>	<i>potencia</i>	<i>n</i>
-,60	,05	,80	19

Luego, nuestras investigadoras podrían concluir que, con una muestra mínima de 19 estudiantes podrían llevar a término su estudio de asociación, una vez tenidas en cuenta las condiciones de aplicación de dicho parámetro (Norman y Streiner, 2014).

## 4. Conclusión

La estimación del tamaño muestral con el fichero proporcionado en este trabajo (disponible en material suplementario) servirá de ayuda metodológica a los investigadores del campo de las ciencias de la educación durante la fase de diseño de sus trabajos para evitar que el análisis de los datos concluya que el coeficiente de correlación calculado no alcanza la significación estadística quizá por el escaso tamaño muestral utilizado.

## Archivo complementario

Sintaxis SPSS para el cálculo del tamaño de la muestra necesario para el coeficiente de correlación de Pearson: <https://revistes.ub.edu/index.php/REIRE/rt/suppFiles/32565/0>

## Referencias

- Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., y Newman, T. B. (2013). *Designing clinical research: an epidemiologic approach* (4.ª ed.). Lippincott Williams & Wilkins.
- Norman, G. R., y Streiner, D. L. (2014). *Biostatistics: The bare essentials* (4.ª ed.). People's Medical Publishing House.
- Pardo, A., Ruiz, M., y San-Martin, R. (2009). *Análisis de datos en ciencias sociales y de la salud I*. Síntesis.
- 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. <https://doi.org/10.1344/reire2018.11.221733>
- Santabàrbara, 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. <https://doi.org/10.1344/reire2019.12.228245>
- Santabàrbara, J., López-Antón, R., Rubio-Aranda, E., Lobo-Escolar, E., y Marcos-Aragüés, G. (2015). *Cálculo del tamaño de la muestra en estudios biomédicos*. Prensas de la Universidad de Zaragoza.