

Universidad de Chile
Facultad de Ciencias Sociales (FACSO)
Departamento de Sociología



UNIVERSIDAD DE CHILE

Guía de Asociación entre variables (Pearson y Spearman en SPSS)

Ayudantía Estadística I 2014

Equipo de ayudantes:

Ignacio Díaz

Carolina García

Magdalena León

Felipe Ruiz

Francisca Torres

Docentes: Paulina Lizama, Giorgio Boccardo,

Noviembre de 2014

Supuestos de la Estadística inferencial

Muestreos probabilísticos

Uno de los supuestos básicos de la estadística paramétrica, es decir, que permite contrastar hipótesis referidas a algún parámetro (la noción de parámetro siempre remite al valor del estadístico en la población); es el cumplimiento de normalidad y el uso de muestreo aleatorios.

Existen distintos tipos de muestreos probabilísticos (en este contexto, sinónimo de aleatorio), que serán estudiados con mayor profundidad en Estadística III. Por ahora, la idea es que manejen los conceptos básicos asociados al Muestreo Aleatorio Simple (MAS).

Pongamos un ejemplo. Si quisiéramos calcular cuantas horas semanales dedican en promedio los estudiantes de Pregrado de la Universidad de Chile a actividades académicas, sería muy engorroso encuestar a cada uno de los estudiantes, considerando en tamaño de la población. Para ello podríamos tomar una muestra de manera aleatoria a partir de un listado de los estudiantes (MAS), es decir, seleccionar estudiantes al azar (como en un tómbola), asegurando que cada uno de los elementos tiene la misma probabilidad de ser elegido (equiprobable), y que esta probabilidad es conocida y distinta de cero.

Las combinaciones de elementos posibles para conformar la muestra (supongamos que seleccionaremos una muestra de 500 casos (n) y que el tamaño de la población (N) es de 30.000 alumnos) son infinitas. Dado que estamos calculando un parámetro en la población con sólo una parcialidad del total de estudiantes y que todas las muestras arrojan un valor distinto, es decir, los resultados son o pueden ser distintos de una muestra a otra para la misma población, siempre existe un error asociado que denominamos error de muestreo y que representa un error desconocido por el investigador.

Sin embargo, el error de muestreo es sistemático y predecible, y mediante el cálculo del error típico podemos acercarnos a conocer la precisión que posee nuestro estimador (valor de media y/o otro estadístico calculado a partir de los datos de la muestra) para dar cuenta del parámetro. A dicho error, inherente a todo diseño, debemos agregar el error no muestral, que tiene que ver, por ejemplo, con problema de aplicación del instrumento, de codificación y de digitación de los datos. Estos se dan muy fuertemente cuando realizamos censos a la población y tenemos una enorme número de encuestados.

Ambos errores suponen una dificultad para conocer el valor exacto del parámetro en la población. La experiencia enseña que la probabilidad de errores no muestrales en el proceso es mayor que el error de muestreo (Vivanco, 2005). En este sentido, el ocupar muestras bien diseñadas cuando tenemos poblaciones grandes presenta una ventaja esencial que guarda relación con que trabajamos con un error asociado menor que el producido por un censo.

Normalidad

Para poder entender como funciona la estadística paramétrica, es fundamental tener claro el concepto de curva normal (ver Guía 7 de ayudantía) y la necesidad de que este se cumpla para poder realizar estimaciones a la población. La manera de comprobar normalidad univariante será desarrollada en otro apartado de la misma guía.

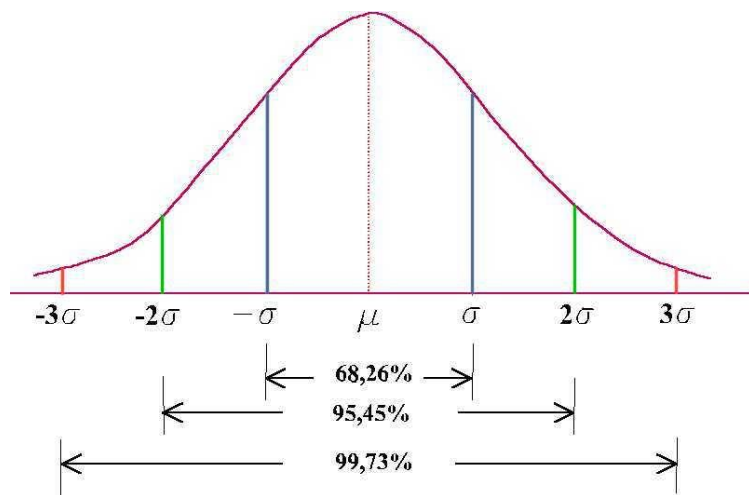
Nivel de confianza

Para entender mejor el concepto de error y nivel de confianza, es necesario entender el concepto de distribución muestral.

Cuando hablamos de distribución muestral, nos referimos a una descripción matemática de todos los resultados posibles del muestreo y la probabilidad de cada uno, a partir de un muestreo repetido (Ritchey, 2008). Se obtienen (teóricamente) las medias (u otro estadístico) de todas las muestras posibles de tamaño n extraídas de una población de tamaño N .

El Teorema del límite central y la Ley de los grandes números establecen que la distribución de medias muestrales extraídas de forma aleatoria, independientemente de la distribución de la variable en la población, se aproxima a la distribución normal a medida que aumenta el tamaño de la muestra aumenta. (Vivanco, 2005)

En otras palabras, esto quiere decir que si calculo la media de todas las muestras posibles de ser extraídas de una población, esas medias van a tender a distribuirse de manera normal; es decir, la mayoría de las medias se concentrarán en torno al valor de la media de las medias, y la probabilidad de obtener valores lejanos de la media de las medias irá descendiendo a medida que nos vamos alejando del centro de la curva (media, mediana, moda)



La curva normal que se presenta en la imagen es una distribución muestral estandarizada, donde μ representa a la media poblacional y en donde las unidades propias de la variable

(notas, peso, edad, etc.) se transforman en un puntaje Z (unidades de desviación estándar o varoles sigma σ). Esto permite comparar variables con distintas unidades de medida, o con la misma unidad de medida, pero con distintas medias y desviaciones estándar.

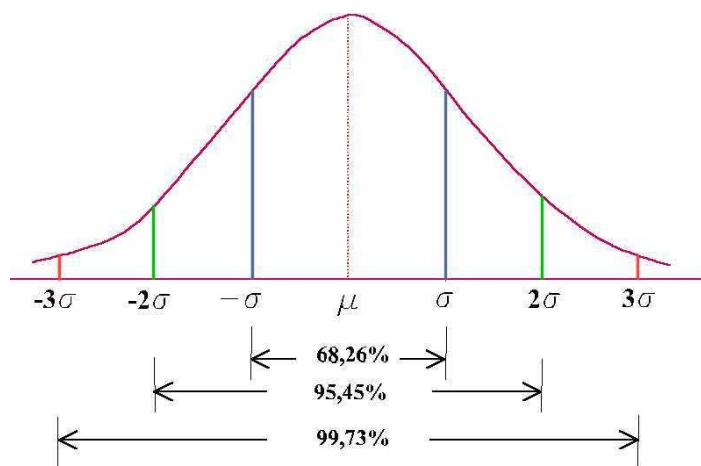
Como ya vimos, la desviación estándar es una medida de dispersión que describe como las puntuaciones de una variable se dispersan a lo largo de su distribución. En rigor, corresponde a un promedio del cuadrado de las puntuaciones de desviación de los individuos en la muestra. Las puntuaciones estandarizadas corresponden a la distancia estandarizada de un valor X hacia la media. La puntuación Z indica la dirección de la puntuación (- ó +) y la distancia hacia la media.

Cuando queremos inferir a la población nunca logramos tener la certeza absoluta de que estamos estimando con perfecta precisión el parámetro. Para conocer la precisión de nuestra estimación, el investigador debe elegir un nivel de confianza en función de la exactitud que desea obtener. A mayor nivel de confianza, mayor el tamaño de la muestra (lo que implica una serie de problemas por los costos que supone la realización de muestreos de tamaño muy elevado). Por convención, en ciencias sociales se utiliza un nivel de confianza de 95%.

Nivel de confianza p (%)	Factor de cobertura k_p
68,27	1
90	1,645
95	1,960
95,45	2
99	2,576
99,73	3

El nivel de confianza se sustenta en una distribución normal y tiene asociado un coeficiente de confianza o puntaje crítico (en la tabla, el factor de cobertura) que equivale al puntaje Z que demarca el límite del área señalada por el nivel de confianza.

Si volvemos a mirar la distribución normal antes señalada, las flechas que se encuentran debajo del gráfico y que señalan %, corresponde al área que encuentra comprendida entre distintas unidades de desviación estándar o valores sigma. Como estamos hablando de una curva estandarizada, ello quiere decir que esos porcentajes son válidos para toda distribución de variables normales.



Es decir, para toda variable con distribución normal, entre -1σ y $+1 \sigma$ (unidades de desviación estándar), se encuentra el 68,26% de los valores de los distintos casos. Para toda variable con distribución normal, entre -2σ y $+2 \sigma$ (unidades de desviación estándar),

se encuentra el 95,45% de los valores de los distintos casos y así sucesivamente.

Cuando inferimos a la población existen dos alternativas.

1. Podemos utilizar la estimación puntual, que define un estadístico como la estimación del parámetro poblacional. Es decir, siguiendo el ejemplo anterior, se señala que “Los estudiantes de Pregrado de la Universidad de Chile, dedican en promedio 60 horas a la semana a realizar actividades académicas, con un 95% de confianza.”
2. Establecer en torno a un estadístico un **intervalo de confianza** para estimar en términos probabilísticos el parámetro. Para el mismo ejemplo, señalaríamos que “Los estudiantes de Pregrado de la Universidad de Chile, dedican en promedio entre 58 y 62 horas a la semana a realizar actividades académicas, con un 95% de confianza”

La primera alternativa señalada prácticamente no se ocupa porque desconoce el error asociado a la estimación del parámetro. Entonces, en general, a partir de un nivel de confianza dado, se construirán intervalos de confianza que señalan una cota superior e inferior que permite inferir el parámetro poblacional. Un nivel de confianza de 95% asignado a un intervalo de confianza indica que tenemos un 95% de probabilidad de que la muestra extraída haya generado un intervalo que contiene al parámetro poblacional.

Por otro lado, el nivel de confianza en la estimación presenta como complementario la probabilidad de error en la estimación, conocido como error alfa (α), que representa la probabilidad de equivocarse en la estimación que estamos haciendo. Es decir, que la muestra que seleccioné de origen a un intervalo de confianza que no contiene al parámetro poblacional.

Nivel de confianza+nivel de error=100%

Para un 95% de confianza, tenemos un 5% de probabilidad de error; para un 99%, tenemos un 1% probabilidad de error y así para los distintos niveles de confianza utilizados.

Correlación lineal simple – coeficientes de asociación.

Los coeficientes de asociación son valores numéricos que permiten cuantificar el grado de ajuste y de relación lineal entre dos variables.

Nos centraremos en dos de ellos: el **coeficiente de correlación de Pearson**, y el **coeficiente de correlación de Spearman**. El primero es un coeficiente paramétrico, es decir, infiere sus resultados a la población real, lo que hace necesario que la distribución de nuestra muestra se asemeje a la distribución real, es decir, que haya normalidad. Esta es la mayor diferencia entre ambos coeficientes, que el de Pearson es paramétrico, y requiere que se cumpla el supuesto de normalidad en las variables, mientras que el de

Spearman es no paramétrico, pues la distribución muestral no se ajusta a una distribución conocida, por lo que los estimadores muestrales no son representativos de los parámetros poblacionales.

A raíz de esta primera diferencia, se van limitando aún más los usos que se le pueden dar a cada coeficiente. Como el de Pearson es paramétrico y requiere de normalidad univariante, sólo podrá calcularse en variables cuantitativas, con niveles de medición intervalar o de razón. Por otro lado, si nuestras variables cuantitativas no cumplen con el supuesto de normalidad (no se distribuyen de acuerdo a la curva normal), o son variables de tipo cualitativo (ordinal), sólo queda usar el coeficiente de correlación de Spearman.

Más adelante se verá cuán similares, en términos de interpretación, son uno y otro, sólo que es primordial tener cuidado en el tipo de variables que se busca correlacionar, y si cumplen o no el supuesto de normalidad univariante.

Como ya vimos, para comprobar si existe o no normalidad univariante existen distintas formas: revisar el gráfico de histograma, la significación estadística de la prueba K-S y el cálculo del Z de simetría y Z de curtosis.

Para llevar a cabo el posterior análisis de correlaciones con coeficiente de asociación, tomaremos dos variables de la encuesta CEP de noviembre-diciembre del 2012. La primera variable es la MBP04 “¿Cómo calificaría Ud. Su actual situación económica?”, cuyas categorías de respuesta son:

- 1=Muy mala
- 2=Mala
- 3=Ni buena ni mala
- 4=Buena
- 5=Muy buena
- 8=No sabe
- 9=No responde

Daremos por perdidos los valores 8 y 9. Como es una variable ordinal de al menos 5 categorías de respuesta, la tomaremos como una variable cuantitativa ordinal, para así poder calcular normalidad y utilizarla para el coeficiente de asociación de Pearson. Hay que tener muy en cuenta que el uso de esta variable es meramente para fines de ejemplificar la comprobación de supuestos. Si bien hemos establecido que con un mínimo de 5 categorías de respuesta una variable ordinal (nunca nominal) puede ser tomada como intervalar, ésta sigue siendo mayoritariamente cualitativa, por lo que hay un gran sesgo a la hora de calcular normalidad con estas variables (recordar que **sólo las variables cuantitativas** pueden tener una distribución normal). Las variables más idóneas, entonces, para llevar a cabo una correlación con un coeficiente de Pearson, serán siempre las

realmente intervalares o de razón. Si se toman ordinales como intervalares, hay que asumir el riesgo de que, aun cuando las pruebas arrojen normalidad, habrá un sesgo por el hecho de que la variable es cualitativa.

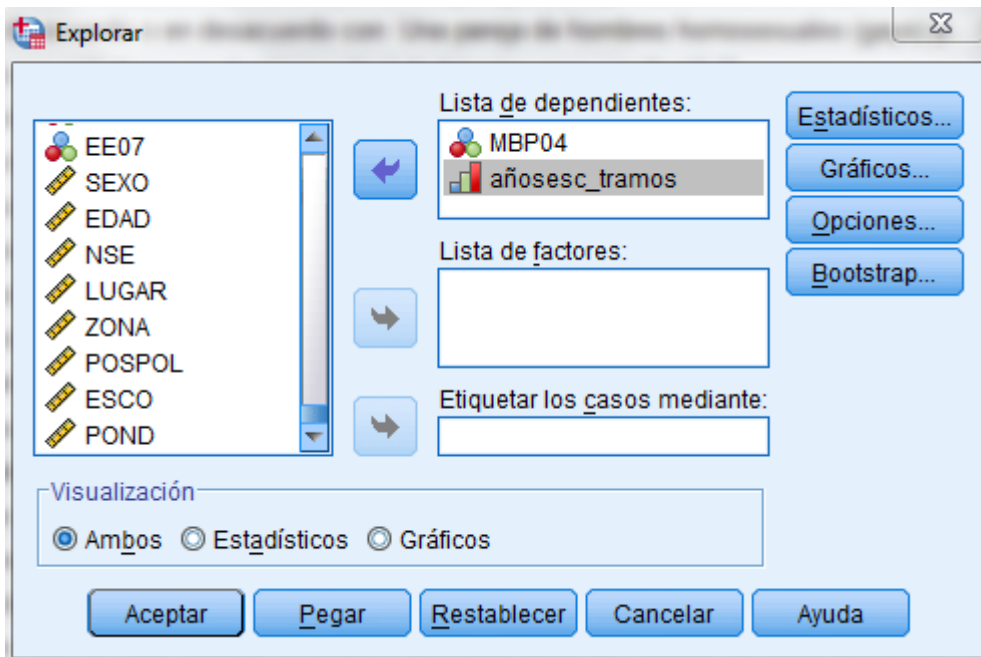
La siguiente variable que usaremos es una recodificación de la variable DDP03 “¿Cuántos años de estudios aprobados tiene Ud.?”. Esta variable escalar sólo incluye los valores 0=Nunca estudió y 99=No responde. Al ver las frecuencias de las respuestas, vemos que, dando por perdido el 99, el rango es de 0 a 80, siendo éste un valor muy extremo (y casi imposible en la realidad), que dificultaría mucho el cumplimiento de la normalidad.

Es por esto que se recodificó en 5 categorías de respuesta:

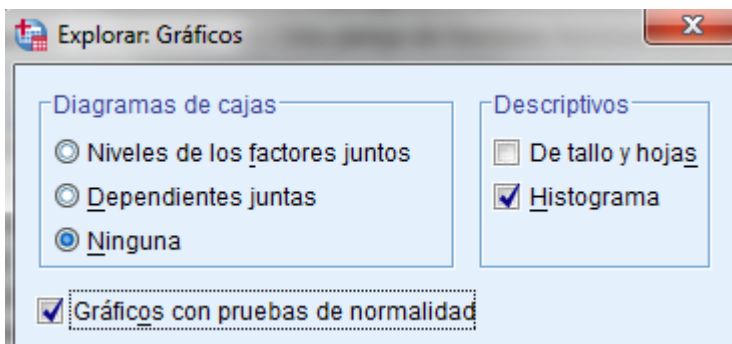
- 1= Nunca estudió
- 2=1-8 años de estudios (Enseñanza Básica)
- 3=9-12 años de estudios (Enseñanza Media)
- 4=13-18 años de estudios (Educación Superior)
- 5=19 o más años de estudios (Postgrado)

De esta forma, los tramos tienen una justificación en términos de ciclos de estudio, y, como es ordinal y tiene 5 categorías de respuesta, también la usaremos como cuantitativa (asumiendo todo lo explicitado anteriormente).

Para comprobar el cumplimiento del supuesto de normalidad univariante, llevaremos a cabo las tres pruebas de normalidad ya vistas en otras ayudantías, y ya mencionadas más arriba (tanto las gráficas como las estadísticas). Para esto iremos a la opción Analizar→Estadísticos descriptivos→Explorar. Buscamos y seleccionamos las dos variables que vamos a usar:



En la opción 'Gráficos' seleccionaremos la opción Ninguna en la sección de Diagrama de Cajas; en la sección Descriptivos desmarcaremos la opción De tallo y hojas y marcaremos Histograma (para así poder realizar la primera prueba de normalidad univariada); por último, marcamos la opción Gráficos con pruebas de normalidad, que nos entregarán los valores de simetría y curtosis que nos permitirán calcular las Z respectivas, y también la prueba de Kolmogorov-Smirnov.



Le damos pegar y ejecutamos (ctrl+R) la siguiente sintaxis:

```
EXAMINE VARIABLES=MBP04 añosec_tramos
/PLOT HISTOGRAM NPLOT
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```


Resumen de procesamiento de casos

	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
¿Cómo calificaría Ud. su actual situación económica?	1557	99,6%	7	0,4%	1564	100,0%
¿Cuántos años de estudios aprobados tiene Ud.? (TRAMOS)	1557	99,6%	7	0,4%	1564	100,0%

Esta tabla nos muestra el resumen de los casos. Vemos que en las dos variables hay sólo 7 casos perdidos, por lo que se trabajará con un total de 1557 casos válidos.

Descriptivos

		Estadístico	Error estándar
¿Cómo calificaría Ud. su actual situación económica?	Media	2,97	,019
	95% de intervalo de confianza para la media	Límite inferior Límite superior	2,93 3,00
	Media recortada al 5%	2,98	
	Mediana	3,00	
	Varianza	,586	
	Desviación estándar	,765	
	Mínimo	1	
	Máximo	5	
	Rango	4	
	Rango intercuartil	0	
	Asimetría	-,150	,062
	Curtosis	,068	,124
¿Cuántos años de estudios aprobados tiene Ud.? (TRAMOS)	Media	3,0157	,02159
	95% de intervalo de confianza para la media	Límite inferior Límite superior	2,9733 3,0580

Media recortada al 5%	3,0035	
Mediana	3,0000	
Varianza	,726	
Desviación estándar	,85205	
Mínimo	1,00	
Máximo	5,00	
Rango	4,00	
Rango intercuartil	2,00	
Asimetría	,092	,062
Curtosis	-,619	,124

La tabla de estadísticos descriptivos nos muestra varios valores que permiten tener un primer acercamiento a la variable (la estamos *explorando*), como la media, la mediana, la desviación estándar y el rango. Además entrega los valores de la asimetría y la curtosis, con los que podemos calcular el Z de simetría y el Z de curtosis, respectivamente:

$$z_{simetría} = \frac{simetría}{\sqrt{6/N}}$$

$$z_{curtosis} = \frac{curtosis}{\sqrt{24/N}}$$

Como ya vimos, si el valor Z da entre -1,96 y 1,96 ($\pm 1,96$ como el valor crítico más común, aunque puede ampliarse a $\pm 2,58$ si se aumenta el nivel de confianza), entonces se considera que la variable es normal. Veamos cómo se comportan nuestras dos variables.

La variable “¿Cómo calificaría Ud. Su actual situación económica?” tiene un valor de asimetría de -0,150 y un valor de curtosis de 0,068.

$$z_{s1} = \frac{simetría}{\sqrt{6/N}} = \frac{-0,150}{\sqrt{6/1557}} = -2,416$$

$$z_{c1} = \frac{curtosis}{\sqrt{24/N}} = \frac{0,068}{\sqrt{24/1557}} = 0,547$$

Vemos que el valor del Z de simetría se escapa de los rangos de -1,96 y +1,96, por lo que la distribución de esta variable no sería normal (no se asemeja a la distribución real de la población), sin embargo, el valor no es tan lejano (unos 0,5 bajo el rango), y aún está dentro del criterio mayor de $\pm 2,58$, por lo que sí sería normal si se estuviese trabajando

con un mayor nivel de confianza (lo que implicaría reducir el error alfa y los niveles de significación, así como aumentar el tamaño de la muestra).

En cuanto a la variable sobre los años de escolaridad, se ve una asimetría de 0,092 y una curtosis de -0,619.

$$z_{s2} = \frac{\text{simetría}}{\sqrt{6/N}} = \frac{0,092}{\sqrt{6/1557}} = 1,482$$

$$z_{c1} = \frac{\text{curtosis}}{\sqrt{24/N}} = \frac{-0,619}{\sqrt{24/1557}} = -4,986$$

El valor del Z de simetría de esta variable es de 1,48, que cae dentro del rango del valor crítico ($\pm 1,96$), lo que quiere decir que esta variable sí tiene un comportamiento y una distribución normal.

La siguiente tabla que arroja SPSS es la de los Tests de normalidad, en específico la de Kolmogorov-Smirnov. Como ya hemos visto, para que esta prueba arroje existencia de normalidad en la variable, el estadístico tiene que tener una significación mayor a 0,05.

Pruebas de normalidad

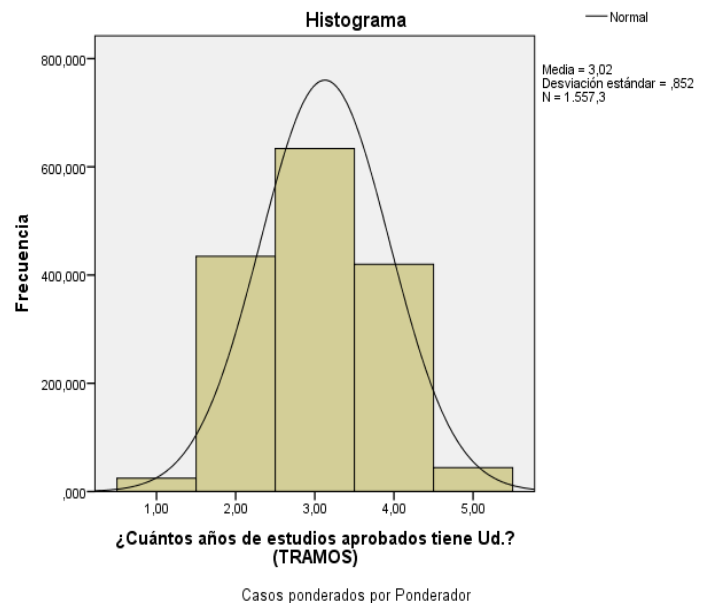
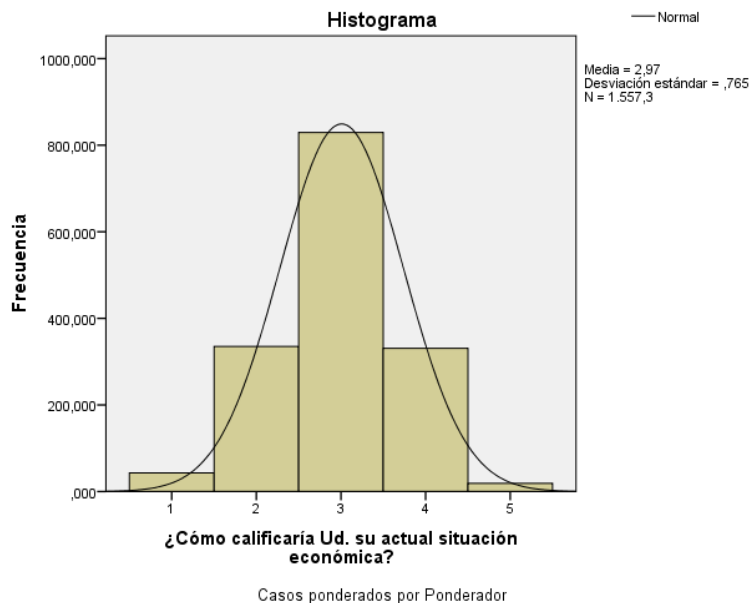
	Kolmogorov-Smirnov ^a		
	Estadístico	gl	Sig.
¿Cómo calificaría Ud. su actual situación económica?	,275	1557	,000
¿Cuántos años de estudios aprobados tiene Ud.? (TRAMOS)	,209	1557	,000

a. Corrección de significación de Lilliefors

Como podemos observar, ninguna de las dos variables tiene un valor p superior a 0,05, por lo que ninguna sería normal según esta prueba. Sin embargo, hay que recordar que el test K-S sólo es válido para muestras de entre 50 y 1000 casos, y la nuestra tiene 1557, por lo que podría ocurrir que, dado el alto tamaño muestral, la prueba arroje un valor que indica que no hay normalidad cuando sí la hay, así como también puede pasar que el estadístico esté en lo correcto.

Lo más seguro sería sólo ocupar este test en muestras del tamaño correcto, y, para muestras más grandes, valerse de las otras dos pruebas.

Por último están los histogramas de ambas variables. Se ve que en los dos hay una



distribución bastante cercana a la normal, con un valor central que acumula a la mayor cantidad de respuestas (moda), y valores extremos con muy pocos casos. Es decir, el análisis gráfico nos diría que las variables sí son normales. De nuevo, hay que tener en cuenta que se está trabajando con una variable de sólo 5 categorías de respuesta, por lo que, si bien los casos se acumulan en el centro, la distribución tiene un sesgo.

De este modo, podemos concluir que las dos variables sí tienen normalidad univariante y que sí sirven para medir asociación entre ellas con el coeficiente de correlación de Pearson. Es cierto que no se pudo comprobar normalidad con la prueba K-S, debido al tamaño muestral, pero la apariencia normal de los histogramas de ambas variables indican que éstas sí tendrían una distribución bastante normal. Además, con los Z de simetría se apoya esta normalidad, a pesar de que el valor de la variable sobre la situación económica se escape del intervalo $\pm 1,96$, éste podría agrandarse hasta $\pm 2,58$ si se aumenta el nivel de confianza.

Con la normalidad univariante mayormente comprobada en las dos variables, pasaremos ahora a calcular la correlación entre ambas con el coeficiente de asociación de Pearson.

Coeficiente de correlación de Pearson en SPSS

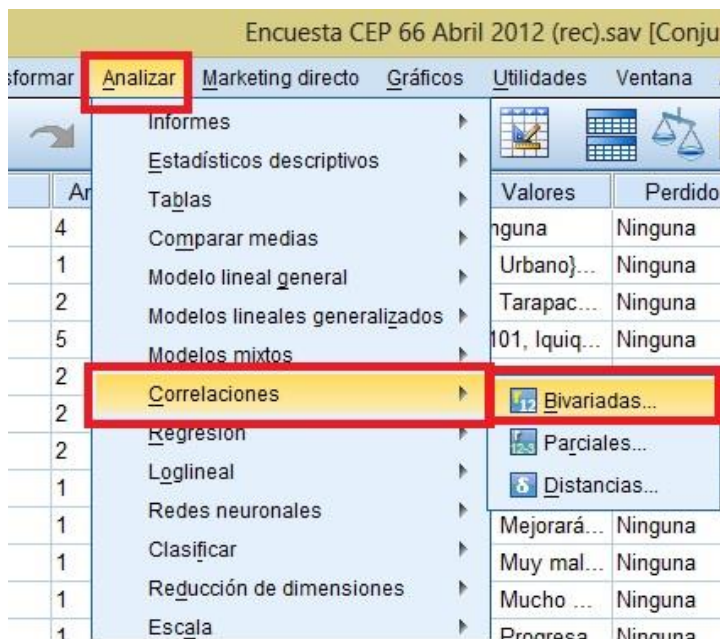
El coeficiente de correlación de Pearson permite medir la fuerza y la dirección de la asociación de dos variables cuantitativas aleatorias con una distribución bivariada conjunta. En este caso sólo buscaremos comprobar que cada una por sí solas presenten una distribución normal univariada. Los valores de la correlación de Pearson van desde -1

hasta 1, siendo los valores extremos los que indican mayor correlación entre variables, y siendo el 0 el punto que indica la no existencia de correlación. El signo positivo o negativo del coeficiente indica si la relación es directa (positivo) o inversa (negativo). La correlación no implica causalidad o dependencia.

Para la interpretación de los resultados, tanto para Pearson y Spearman hay que considerar lo siguiente: Si el coeficiente de correlación arrojado va entre 0 y 0,2, entonces la correlación es mínima; si va entre 0,2 y 0,4, es una correlación baja; si va entre 0,4 y 0,6, entonces es una correlación moderada, ya entre 0,6 y 0,8 es una correlación buena; finalmente, entre 0,8 y 1, es una correlación muy buena. Esto mismo aplica en negativo.

Como ya se ha comprobado que las variables a utilizar presentan normalidad univariada, procederemos a mostrar cómo obtener el coeficiente de correlación de Pearson en SPSS.

Primero, vamos a 'analizar' > 'correlaciones' > 'bivariadas'

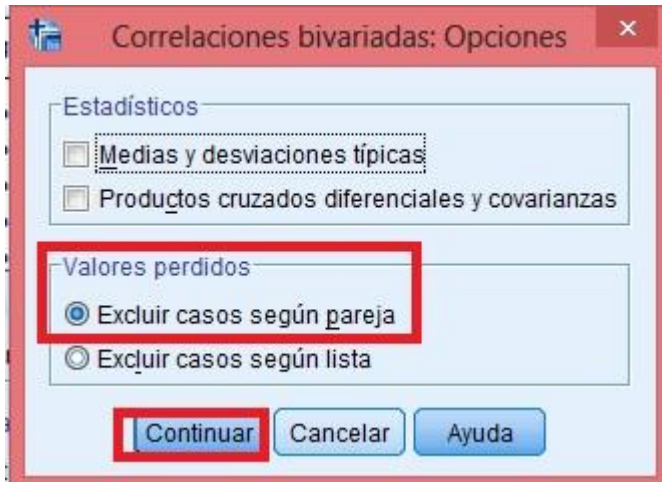


Luego nos aparece un cuadro de diálogo. Aquí ingresamos las variables que vamos a utilizar como es de costumbre, seleccionándolas de la lista de variables del costado izquierdo, y apretando el botón que contiene una flecha para ingresarlas al análisis. En este caso, como hemos visto, trabajaremos con las variables 'MBP04' y la reconfigurada 'añoscsc_tramos'. Ahora, dentro de las opciones posibles, marcamos dentro de los coeficientes de correlación la casilla que dice 'Pearson'. En prueba de significación marcamos 'bilateral', lo que quiere decir que nuestra prueba de significación sobre nuestro coeficiente de correlación será de dos colas, sin especificar la dirección del efecto de la correlación que vamos a encontrar, pueden encontrarse valores extremos hacia

ambas direcciones, al contrario de la opción 'unilateral'. Además, marcamos la casilla de 'marcar las correlaciones significativas', que añade un asterisco en la tabla de frecuencias de los resultados cuando la significación de la correlación sea bajo 0,05, y dos asteriscos cuando sea 0,01, lo que permite saber más rápidamente si es que existe correlación.



Tras esto, seleccionamos el botón 'Opciones', ubicado a la derecha del cuadro. Allí verificamos que esté sólo seleccionada la opción 'excluir casos según pareja', en la sección de valores perdidos, lo que hará que se excluyan de cada análisis los casos que tengan valores perdidos en cualquiera de las dos variables o en ambas variables utilizadas en el cálculo del estadístico.



Luego damos clic en 'continuar', y luego en 'pegar', con lo que se nos debería abrir la ventana de sintaxis, presentándose la siguiente sintaxis para el análisis:

```

CORRELATIONS

/VARIABLES=MBP04 añosesc_tramos

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE.

```

Luego seleccionamos la sintaxis y damos clic en el botón 'play'. A continuación se presentan los resultados:

Correlaciones

	¿Cómo calificaría Ud. su actual situación económica?	¿Cuántos años de estudios aprobados tiene Ud.? (TRAMOS)
¿Cómo calificaría Ud. su actual situación	Correlación de Pearson	1
	Sig. (bilateral)	,172** ,000

económica?	N	1564	1563
¿Cuántos años de estudios aprobados tiene Ud.? (TRAMOS)	Correlación de Pearson	,172**	1
	Sig. (bilateral)	,000	
	N	1563	1563

** . La correlación es significativa al nivel 0,01 (bilateral).

Para interpretar estos datos entonces, vemos que el coeficiente de correlación de Pearson entre estas dos variables es de 0,172. Como los valores de Pearson van de -1 a 1, siendo el 0 el indicador de que no existe correlación, vemos que entre estas dos variables existe una correlación mínima, ya que no es 0, pero es muy cercano. La dirección de la correlación es positiva, es decir, es directa, por lo tanto al aumentar los años de escolaridad, mejoraría la autopercepción de la situación económica, y viceversa.

Otro aspecto que debemos considerar para la interpretación es la significación, que en su interpretación está estrechamente vinculado al nivel de confianza y al error alfa (α). Si bien por ahora no ahondaremos en pruebas de hipótesis, es importante que tengan clara la mecánica para discriminar entre asociación y no asociación entre las variables.

El error alfa es equivalente al nivel de significación. Un nivel de significación del 5% (significación=0,05) significa que, al señalar que existe asociación entre las variables (o rechazar que no existe asociación), tenemos un 5% de probabilidad de equivocarnos. Como estamos trabajando con un 95% de confianza, valores iguales o menores a 0,05 en la significación corroboran que hay asociación entre las variables. Sin embargo, si la significación es mayor al error alfa o nivel de significación establecido, no podemos señalar que existe asociación entre las variables, por más que el estadístico así lo indique, porque la probabilidad de estarnos equivocando al señalar que hay asociación es muy alta o mayor al nivel de confianza establecido.

Al analizarla entonces para este caso, vemos que esta es de 0,000, es decir, está bajo el 0,05 (y está marcada con dos asteriscos por ser bajo 0,01), por lo que estaría indicando que existe una fuerte correlación y que es verdadera la correlación encontrada por el estadístico de Pearson.

Correlación de Spearman en SPSS

El coeficiente de correlación de Spearman, al igual que el de Pearson, muestra una asociación entre variables. A diferencia del anterior, permite obtener un coeficiente de asociación entre variables que no se comportan normalmente, entre variables ordinales. Se calcula en base a una serie de rangos asignados. Tal como en Pearson, los valores van de -

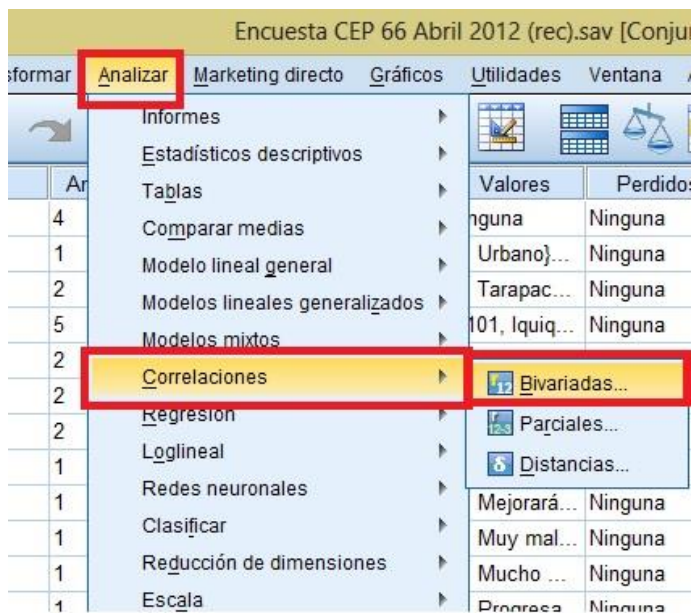
1 a 1, siendo 0 el valor que indica no correlación, y los signos indican correlación directa e inversa.

Para calcular el coeficiente de Spearman seleccionaremos dos variables ordinales que nos permitan realizar el análisis, a las que no les haremos ninguna prueba ya que no necesita cumplir supuestos:

Por un lado la variable MBP02.- ¿Cómo calificaría Ud. la actual situación del país?, cuyas alternativas de respuesta son: 1=muy mala; 2=mala; 3= ni buena ni mala; 4= buena; 5= muy buena; 8 = no sabe; 9= no contesta. En este caso se dan por perdidos los valores 8 y 9. La otra variable a correlacionar es la MBP18_29.- ¿Cuál de estás frases describe mejor su opinión sobre Camila Vallejo?, cutas alternativas de respuesta son: 1= muy negativa; 2= negativa; 3= ni negativa ni positiva; 4= positiva; 5= muy positiva; 6= no conoce a la persona; 8= no sabe; 9= no contesta. Se dan por perdidos los valores 6, 8 y 9.

Se escogieron estas dos variables ya que en el momento de la encuesta (2012), el conflicto estudiantil estaba en un momento álgido de enfrentamiento al gobierno de Sebastián Piñera, y Camila Vallejo como presidenta de la Fech y vocera del Confech, era la figura con mayor número de apariciones en medios de comunicación y con mayor peso simbólico con respecto al movimiento estudiantil. Por lo tanto, es posible suponer que a mayores niveles de aprobación de su figura, peor se calificará la situación del país (aunque existen otros factores mediando, claro).

Para llevar a cabo el análisis es muy similar el procedimiento. Vamos a analizar > correlaciones > bivariadas:



Luego en el cuadro que se nos abre, seleccionamos la casilla 'Spearman', y el resto lo dejamos tal como en Pearson.



Damos clic en el botón 'pegar', y seleccionamos la sintaxis de la ventana de sintaxis, y damos clic en el botón 'ejecutar'. La sintaxis debiese ser la siguiente:

```
NONPAR CORR
/VARIABLES=MBP02 MBP18_29
/PRINT=SPEARMAN TWOTAIL NOSIG
/MISSING=PAIRWISE.
```

A continuación se presenta la tabla de resultados:

Correlaciones

			¿Cómo calificaría Ud. la actual situación del país?	¿Cuál de estas frases describe mejor su opinión sobre Camila Vallejo?
Rho Spearman	de	de	¿Cómo calificaría Ud. la actual situación del país?	¿Cuál de estas frases describe mejor su opinión sobre Camila Vallejo?
			Coefficiente de correlación Sig. (bilateral)	Coefficiente de correlación Sig. (bilateral)
			N	N
			1,000	-,137**
			.	,000
			1541	1300
			-,137**	1,000
			,000	.
			1300	1304

** . La correlación es significativa al nivel 0,01 (bilateral).

Al analizar la tabla que nos entrega el SPSS vemos que se establece un nivel de correlación mínimo (-0,137), siendo que está mucho más cercano a 0 que a -1. Por otro lado, se establece que es una correlación negativa, es decir, inversa, por lo que a mayor nivel de aprobación de Camila Vallejo, peor será la calificación que se le otorga a la situación del país, y viceversa, por lo tanto se cumple lo que se planteó al escoger las variables. Al analizar la significación, vemos que es de 0,000 nuevamente, por lo que se cumple que sea menor a 0,05, e incluso es menor a 0,01, lo que indica que la correlación que se ha establecido (mínima) es muy probablemente cierta.