UNIVERSIDAD NACIONAL DE SAN LUIS

FACULTAD DE PSICOLOGÍA

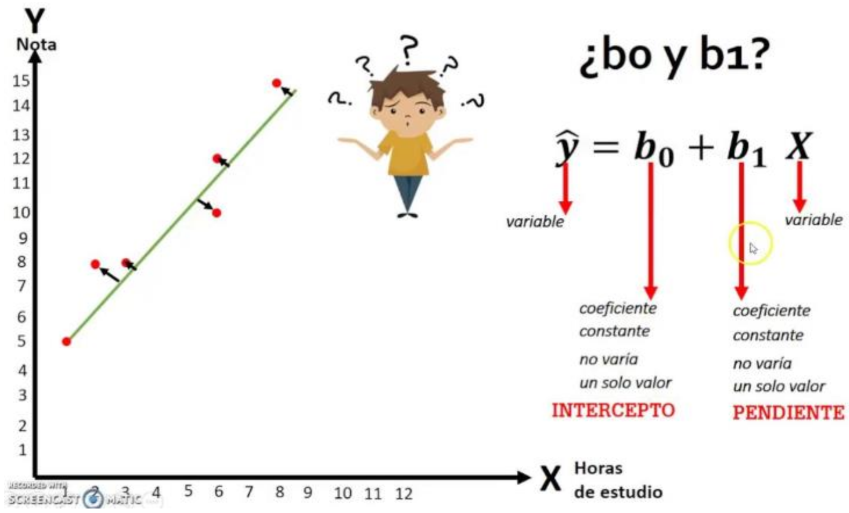
METODOLOGÍA DE LA INVESTIGACIÓN I

Licenciatura en Psicología

Licenciatura en Psicomotricidad

Prof. Responsable: Dr. HORACIO DANIEL GARCIA

Trabajos prácticos: Lic. DANIEL PITONI, Lic. MAXIMILIANO SAPINO y Mg. ELIANA ZÁRATE



Unidad 7: Análisis de Relaciones
Concepto de correlación. Relaciones entre variables cuantitativas: introducción al análisis de correlación lineal. Coeficientes de correlación. Análisis correlacional para variables numéricas: Análisis paramétrico (Coeficiente de Correlación R de Pearson) y Análisis no paramétrico (Coeficiente de Correlación Rho de Spearman). Análisis Correlacional para variables cualitativas: Tablas de contingencia. Prueba ji-cuadrado. Corrección de Yates. Introducción al análisis de regresión lineal simple.

Autor: HORACIO DANIEL GARCIA

Año 2023

Concepto de correlación

En esta oportunidad continuamos con uno de los capítulos más interesantes de la estadística.

Pensemos por el momento lo importante que es verificar si un grupo de variables se relacionan entre sí; por ejemplo: ¿hay relación entre la duración del tratamiento y la disminución de los síntomas? ¿la agresividad se encuentra asociada a algún factor de personalidad? … como verás, encontrarles respuestas a interrogantes de este tipo, resulta esencial en el progreso de nuestras disciplinas; más aún cuando estas deducciones aportarán información que conducirá al rechazo o la aceptación de hipótesis que posiblemente se consolidarán en teorías.

Por otra parte, este tipo de estrategias estadísticas están en la base de análisis más complejos, como cuando buscamos determinar si existe alguna relación de causalidad entre una o varias variables independientes y la variable dependiente, o cuando deseamos identificar la solidez y validez de un instrumento de medición, etc.

Para entender mejor el concepto de correlación, te anticipamos que las palabras asociación y relación son buenos sinónimos que se suelen emplear al momento de describir los resultados que nos proporcionan los estadísticos correspondientes.

Pero, es momento de hacer algunas precisiones ¿Qué es una correlación?

Entendemos que existe relación o correlación entre dos variables cuando observamos que, considerando un conjunto de unidades de análisis, el valor de una variable acompaña proporcionalmente al de otra variable. Por el contrario, la ausencia de correlación nos indica que los valores de ambas variables son completamente independientes.

Existen pruebas que permiten analizar relaciones entre variables cuantitativas, entre variables cualitativas y entre variables cuantitativas y cualitativas. Sin embargo, como las relaciones entre variables cuantitativas suele ser más intuitivas y más fáciles de comprender, comenzaremos desarrollando éstas.

Pensemos, por un instante. Propongámonos por el momento trabajar con variables poco complejas; tomemos la variable altura y, por el otro lado, la variable peso. Si tomamos una muestra, es muy probable que las pruebas estadísticas identifiquen una correlación; lo lógico sería pensar que, en una muestra aleatoria encontraremos que las personas, a medida que son más altas ↑, tienden a pesar más ↑. Por el contrario, si pensamos en las variables número de teléfono y largo de cabello, lo más probable es que digamos que no existirá relación.

Veamos el siguiente ejemplo analizando una pequeña matriz de datos

Unidad de análisis	Depresión	Angustia
001	20	30
002	15	25
003	25	35
004	35	45
005	22	32
006	38	48
007	14	24

Si observas con detenimiento, en todas las unidades de análisis, la variable Angustia tiene 10 puntos más que la variable Depresión. También podemos decir que, en todos los casos, a medida que aumenta los valores de depresión aumentan los valores de Angustia.

Claro está que las cifras no son reales y sólo sirven a modo explicativo de lo que sucede cuando se presenta una relación.

Entonces, podemos resumir que se presenta una correlación cuando se cumplen los siguientes aspectos:

- Los valores de ambas variables presentan cierta proporcionalidad considerando cada unidad de análisis.

- En la mayoría de las unidades de análisis se observa que esta proporcionalidad se da en una misma dirección. En el caso del ejemplo, Angustia toma constantemente valores más altos que Depresión. No obstante, existen otras correlaciones donde a mayores valores de una variable, menores valores se encontrarán en la otra.

Aunque en unos párrafos más lo retomaremos, te anticipamos que las correlaciones pueden ser:

Positivas: Cuando a medida que la variable A aumenta ↑ y la variable B también lo hace ↑, o cuando a medida que la variable A disminuye ↓, la variable B disminuye ↓.

Negativas: Cuando mientras que la variable A aumenta ↑, la variable B disminuye ↓, o a la inversa.

Relaciones entre variables cuantitativas: introducción al análisis de correlación lineal

¿Por qué se introduce la idea de correlación lineal?... bueno, en principio te comentamos que uno de los estadísticos que veremos más adelante (Pearson), lo que en realidad nos ofrece es un coeficiente que describe la linealidad expresada en una correlación. Así, es necesario que conozcas previamente a qué nos estamos refiriendo, para poder interpretar adecuadamente la información que nos ofrece, ya que las variables a veces pueden estar relacionadas, pero no de forma lineal. Este punto resulta importante, ya que frecuentemente se cae en el error de asumir la no asociación entre dos variables sólo porque observamos los coeficientes y no profundizamos en el análisis.

Cuando abordamos la relación entre dos variables numéricas, necesitamos realizar un análisis descriptivo previo, que consiste fundamentalmente en representarlas gráficamente, para recién luego realizar los cálculos (con el estadístico apropiado) a los efectos de obtener la información necesaria que conduzca al Rechazo o a la Aceptación de la hipótesis que nos hemos planteado.

Diagramas de dispersión

Este tipo de gráfico resulta adecuado para representar las puntuaciones relacionadas, consiste en identificar para cada unidad de análisis (participante) el lugar que le corresponde en un plano de dos coordenadas. El punto estará dado por los valores que tengan las variables ubicadas en las coordenadas X e Y.

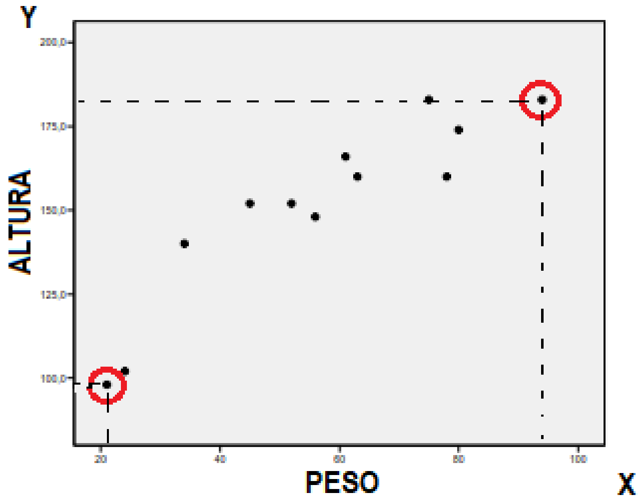


Gráfico 1. Diagrama de dispersión para las variables Altura y Peso.

Si observamos el gráfico 1 veremos que un participante pesa un poco más de 20 kg. y mide poco menos de 1metro de altura, en el caso extremo vemos una persona que pesa cerca de 90 kg. y mide poco más de 1,75 metros… si puedes interpretar cada uno de los puntos comprenderás cuanto pesó y midió cada individuo de la muestra. Además, observarás que se representa una relación positiva, es decir, que un incremento en los valores de una variable se asocia al incremento de la otra.

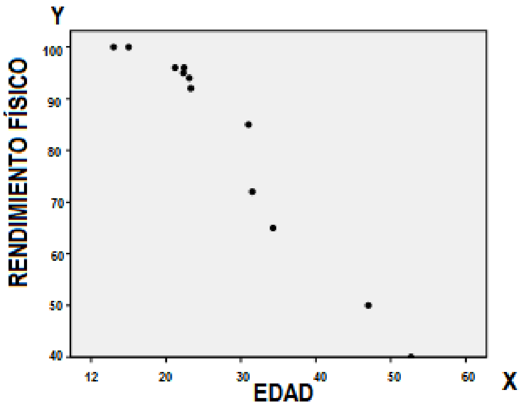
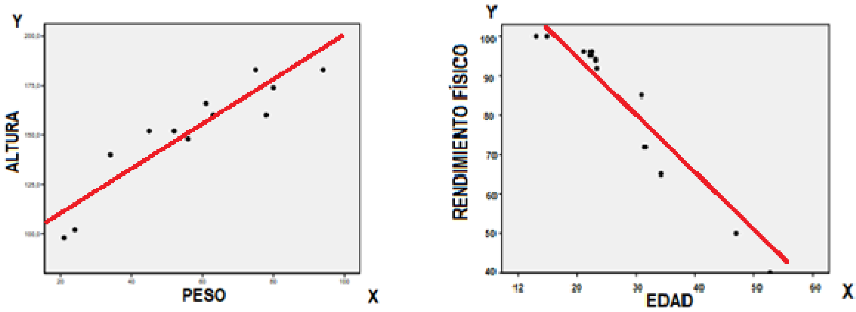


Gráfico 2. Diagrama de dispersión para las variables Rendimiento Físico y Edad.

En el Gráfico 2 observamos una relación negativa, que se sintetiza de este modo: a medida que aumenta la edad de los participantes de esta muestra, disminuye el rendimiento físico.

En ambos gráficos, las nubes de puntos describen una tendencia en torno a una línea recta imaginaria que puede ser creciente o decreciente, según el signo de la relación (positiva o negativa).



Gráficos 3. Diagrama de dispersión con sus respectivas líneas de relación.

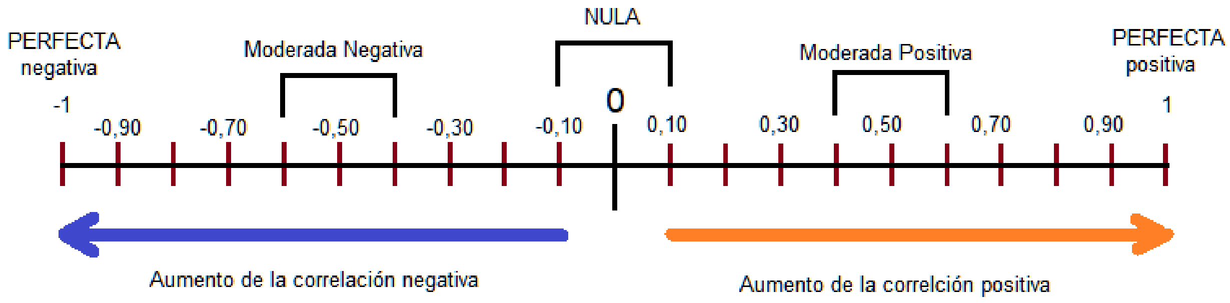
Las líneas son trazadas de modo de encontrar un equilibrio entre todos los puntos del diagrama, procurando un ajuste de las distancias que conservan los valores respecto de esa línea.

Ahora vamos a anticiparnos un poco a lo que veremos más adelante; sin embargo, es necesario hacerlo para que podamos comprender los conceptos de relación y de linealidad.

Coeficientes de correlación

A la medida que representa la fuerza de la correlación estadística se la denomina coeficiente de correlación (*r*); y asume un valor numérico que varía de -1,0 a +1,0. A medida que la cifra del coeficiente se acerca a los valores extremos, nos indica una relación más fuerte; en tanto que si toma valores próximos a 0 nos niega la relación. Por otro lado, el signo del coeficiente señala la dirección de la asociación: un coeficiente + nos dirá que estamos frente a una relación positiva (**A**↑ **B**↑ o **A**↓ **B**↓); en tanto que un valor **–** nos describe una asociación negativa (**A**↑ **B**↓ o **A**↓ **B**↑)

Los gráficos ayudan mucho cuando se trata de analizar los coeficientes de correlación, por ejemplo, a veces un coeficiente de correlación de Pearson bajo no significa que no exista relación entre las variables, sino que estas pueden tener una *relación no lineal*.



La ponderación de los coeficientes de correlación se la suele caracterizar del siguiente modo:

Negativas		Positivas
r= -1	Correlación perfecta	r= 1
Entre r= -0,80 y r= -1	Correlación muy fuerte	Entre r= 0,80 y r= 1
Entre r= -0,60 y r= -0,80	Correlación fuerte	Entre r= 0,60 y r= 0,80
Entre r= -0,60 y r= -0,40	Correlación moderada	Entre r= 0,60 y r= 0,40
Entre r= -0,40 y r= -0,20	Correlación baja	Entre r= 0,40 y r= 0,20
Entre r= -0,20 y r= -0,10	Correlación muy baja	Entre r= 0,20 y r= 0,10
> r= -0,10	Correlación nula o muy débil	< r= 0,10

Un aspecto que no tienes que olvidar es que, confirmada una correlación, ***no implica que necesariamente deba existir una relación causal directa***; es decir que, si bien podemos decir que las variables A y B están relacionadas, esto ***no significa necesariamente que A sea la causa de B***. Sin embargo, suele suceder frecuentemente que las correlaciones altas sirvan para hipotetizar relaciones de causa-efecto, la cuál debe ser comprobada con otras pruebas estadísticas específicas para tales efectos, y que veremos al finalizar este documento.

¿Qué pasará cuando no exista relación? En ese caso, verán una nube de puntos muy dispersa o una nube de puntos en la que la línea no representa una pendiente (horizontal).

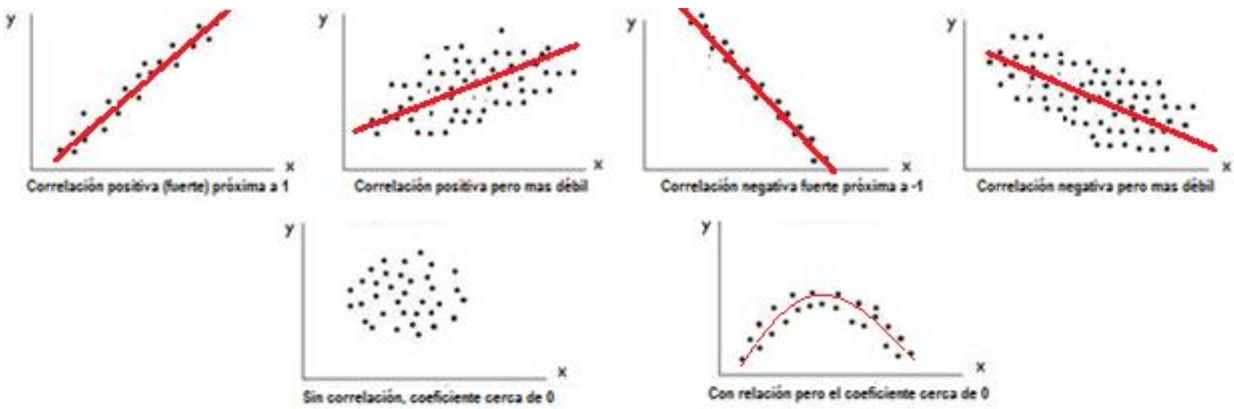


Figura 1. Diagramas de dispersión más representativos

En el primer caso de la figura 1 vemos una relación donde las dos variables aumentan simultáneamente a un ritmo constante, por lo que decimos que existe una relación lineal positiva y si calculáramos el Coeficiente de Pearson es muy posible que éste esté en torno al +1 (probablemente +0,95); ¿por qué deducimos esto?, bueno, lo que hicimos fue prestarles atención a dos elementos: a) el tipo de pendiente que traza la línea y, b) la proximidad de los puntos respecto de la línea. Observa ahora el segundo diagrama de dispersión: veras una pendiente similar a la descripción anterior, pero notarás una mayor distancia (mayor dispersión) de los

puntos respecto de la recta. En este caso decimos que mantiene una correlación positiva, pero más débil que el caso anterior, donde el coeficiente de correlación de Pearson probablemente se encuentre en un valor aproximado de +0,42.
En los diagramas de dispersión de abajo, mostrados en la figura 1, vemos dos situaciones donde característicamente tendremos Coeficientes de correlación de Pearson cercanos a 0. El diagrama de dispersión de abajo a la izquierda muestra una nube de puntos que parece distribuirse aleatoriamente, por lo que es muy difícil trazar una línea que refleje estrictamente la tendencia de los datos. En ese caso hablamos de ausencia de correlación. Pero si nos fijamos en el de su derecha veremos una línea que describe una curva. En este caso la tasa de aumento o descenso de una de las variables cambia a medida que la otra variable aumenta, por lo que tenemos una relación, pero ésta no es lineal; por consiguiente, el coeficiente de Pearson no la va a poder detectar adecuadamente.

Con esto demostramos lo que anticipábamos: la importancia de graficar los datos para poder explorar las relaciones que pudieran existir.

Análisis correlacional para variables numéricas

Propósito del análisis
Como hemos visto, mediante un análisis correlacional podremos establecer si dos variables numéricas se encuentran asociadas o no. Obviamente, dependiendo la naturaleza de los datos escogeremos la estrategia apropiada
Partamos de la idea de que un investigador en su estudio plantea la siguiente hipótesis:

- Existe relación entre las variables Ansiedad y Depresión en la muestra de estudio

Lo que pretende confirmar con esta hipótesis es si ambas variables están relacionadas, sin aclarar qué tipo de relación pudiera existir (positiva o negativa). Es decir, podría suceder que, los participantes, a medida que experimentan mayores niveles de Ansiedad, presentan mayores niveles de Depresión (correlación positiva); o bien, que a medida que posean mayores niveles de Ansiedad, evidencien menores niveles de Depresión (correlación negativa).
En este caso puntual, donde no se aclara si lo que se espera en una correlación positiva o una correlación negativa, la prueba de hipótesis será bilateral. En cambio, cuando se plantea una de las mencionadas posibilidades, la prueba será unilateral derecha o izquierda, según corresponda.

Análisis paramétrico: Coeficiente de Correlación R de Pearson
Para aplicar un análisis correlación de tipo paramétrico se deben cumplir ciertos supuestos; de lo contrario recuerda que podemos apelar a una estrategia no paramétrica, de la cual hablaremos más adelante.
La correlación de Pearson ofrece un coeficiente comprendido entre los valores -1 y 1, tal como lo hemos explicado en los puntos anteriores. Adicionalmente, y sobre este coeficiente, los programas estadísticos calculan el valor de probabilidad asociado, dato determinante para confirmar o rechazar la hipótesis del investigador.
Es un tipo de estrategia óptima cuando las variables son numéricas, la muestra es grande y se ha comprobado que la asociación puede ser lineal, caso contrario el análisis puede ser infructuoso u ofrecer datos que conduzcan a conclusiones erróneas.

Comprobación de supuestos

- <u>Linealidad de la asociación</u>: Pearson busca establecer el grado de ajuste de los datos combinados de ambas variables, a una línea recta teórica. Por tal motivo, no se puede aplicar a variables cuyo gráfico de dispersión ofrezca nubes de puntos que no garanticen la posibilidad de describir una relación lineal (ver figura 1).

- Normalidad de los datos: Como toda prueba paramétrica, los resultados son inferidos de la población real, por lo que el requerimiento es que la distribución de los datos, de ambas variables, se aproximen a una distribución normal.
- Nivel de medición: Este tipo de análisis es apropiado para variables cuantitativas continuas, con una escala de intervalo (por lo menos).

Pasos

Formular las hipótesis nula y alternativa

- H_0= No existe relación entre las variables Ansiedad y Depresión en la muestra de estudio
- H_1= Existe relación entre las variables Ansiedad y Depresión en la muestra de estudio

Establecer el estadístico de prueba adecuado: Prueba Coeficiente de Correlación R de Pearson

Seleccionar un nivel de significación: Alfa de 0,05

Establecer la regla de decisión

$$\begin{cases} \mathbf{H_0: R_{xy} = 0} \text{ (El coeficiente procede de una población cuya correlación es cero } (\rho = 0)). \\ \mathbf{H_1: R_{xy} \neq 0} \text{ (El coeficiente procede de una población cuya correlación no es cero } (\rho \neq 0)). \end{cases}$$

Calcular el valor observado del estadístico de prueba (Procedimiento mediante Jamovi)

Inicialmente debemos corroborar que se cumplan los tres supuestos de este análisis.

- ✓ En cuanto al nivel de medición, ambas variables son continuas con un nivel de medición intervalar, por lo que se corrobora este requisito.
- ✓ Linealidad

Comprobaremos si la nube de puntos descrita por ambas variables se ajusta a una línea recta. Para ello iremos a Análisis> Regresión> Matriz de correlaciones

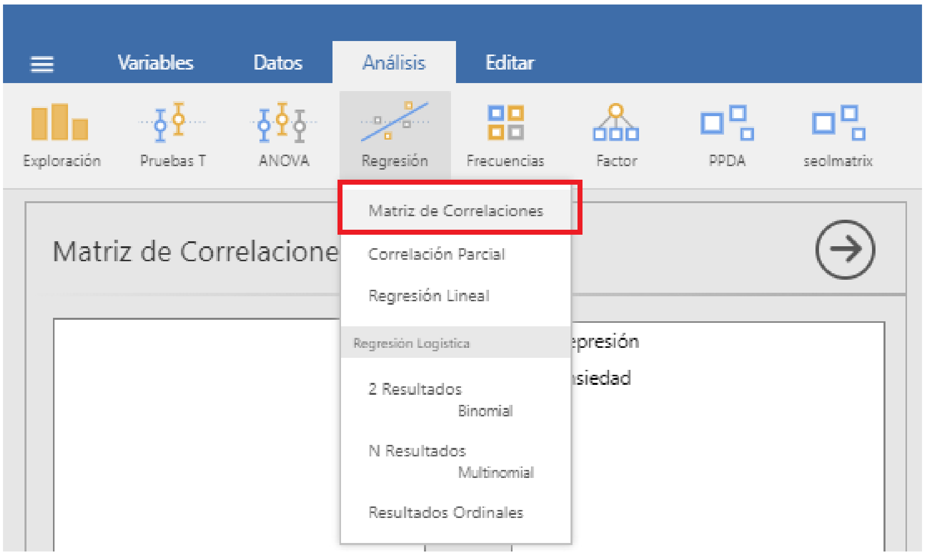


Figura 1: Captura de pantalla del procedimiento para cálculo de relaciones mediante el programa estadístico Jamovi

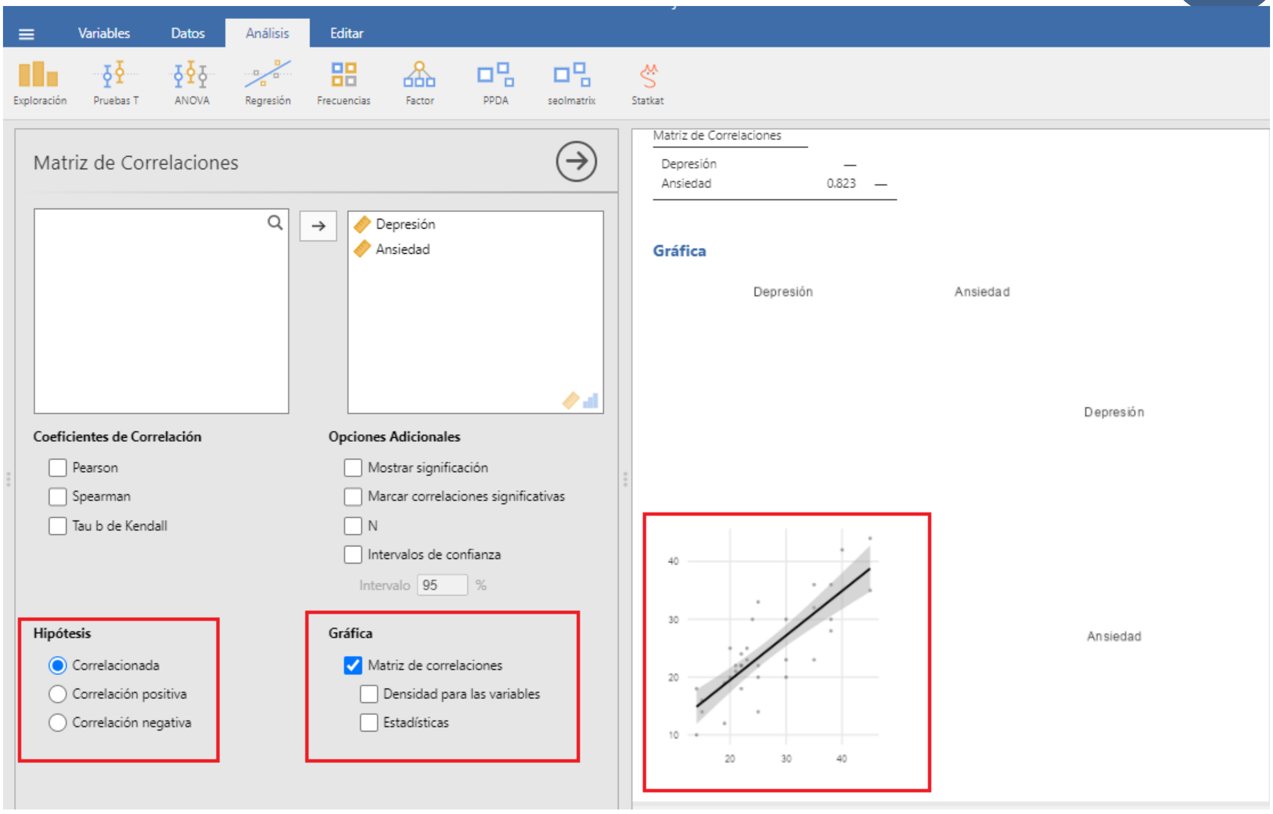


Figura 2: Captura de pantalla de las opciones para obtener el Diagrama de dispersión, realizado mediante Jamovi

Tal como se puede apreciar en la figura 2, los datos describen una nube de puntos en la cual es admisible proyectar una línea recta, lo que permite satisfacer el requisito de linealidad.

- ✓ Normalidad de los datos: para indagar este aspecto podemos ejecutar un análisis mediante Shapiro-Wilk. Para ello iremos a Análisis> Exploración> Normalidad y marcaremos la opción correspondiente para luego seleccionar las variables de interés

Tabla 1.

Análisis de normalidad

Descriptivas

	Depresión	Ansiedad
N	33	33
W de Shapiro-Wilk	0.924	0.962
Valor p de Shapiro-Wilk	0.024	0.296

La tabla 1 nos informa el valor de probabilidad asociado a la prueba de Shapiro-Wilk. Como hemos visto con anterioridad, las pruebas de normalidad parten de la premisa de que, si rechaza la hipótesis nula, se puede afirmar que los datos provienen una distribución no normal. Por lo tanto, si los comparamos con el Alfa propuesto por el investigador (generalmente 0,05) rechazaremos H_0 en la variable Depresión (variable que no cumple con el criterio de normalidad) y aceptaremos H_0 en la variable Ansiedad (variable con distribución normal).

*NOTA: tal como puedes apreciar, si bien se han corroborado los supuestos de nivel de medición de las variables y de linealidad, no ha sido posible confirmar que ambas variables sometidas a análisis posean una distribución normal. Este solo hecho, así como el incumplimiento de cualquiera de los otros supuestos, invalidan la posibilidad de continuar con la estrategia paramétrica, debiendo optar por su equivalente no paramétrico.
Habiendo aclarado esto, continuaremos sólo de manera didáctica con el análisis mediante el Coeficiente de correlación de Pearson.

Análisis de correlación mediante el Coeficiente de correlación de Pearson

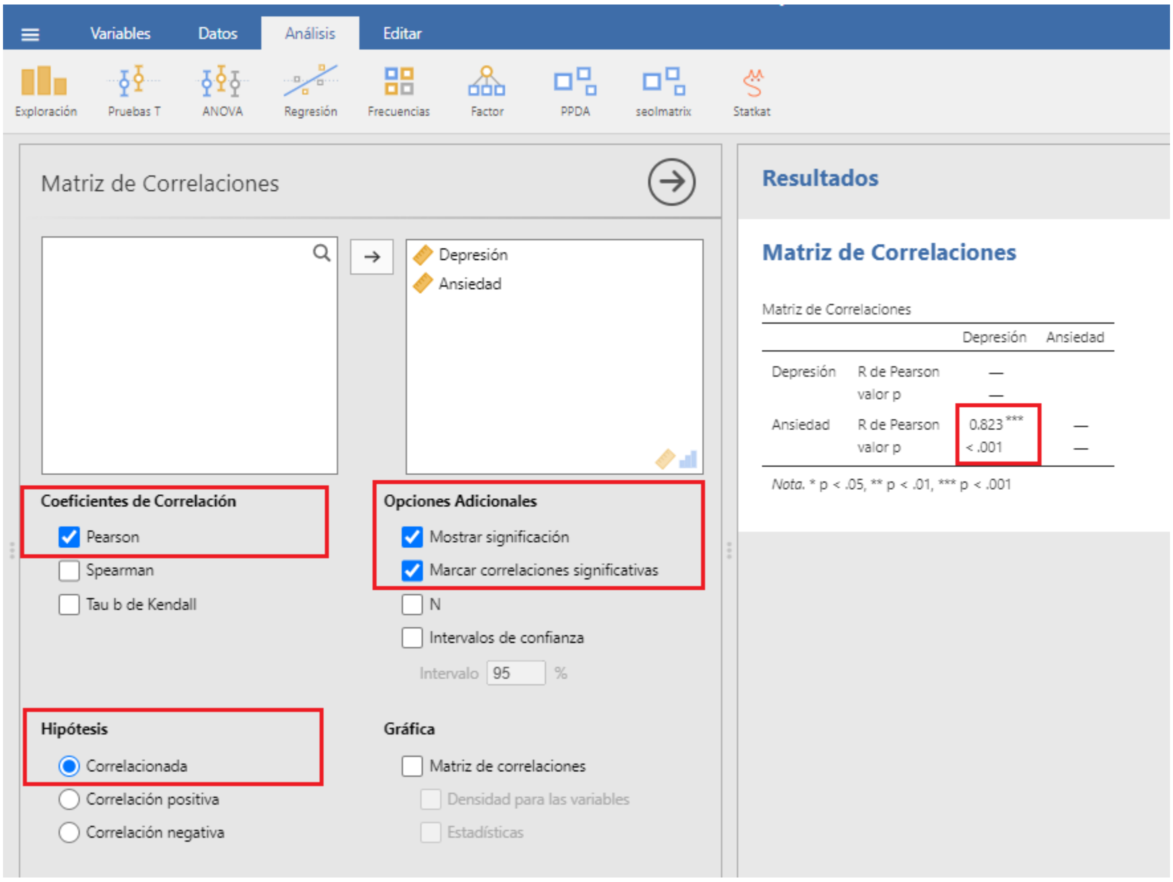


Figura 3: Captura de pantalla opciones para el análisis de Prueba de Correlación de Pearson y resultados de la matriz de correlaciones realizado mediante Jamovi

Resultados

En la figura 3 encontrarás marcadas las opciones para realizar el cálculo y, adicionalmente la tabla con los resultados respectivos. En lo que respecta a los resultados debes concentrarte inicialmente en el valor p, ya que de él depende si se rechazará la hipótesis nula o no. Claro está que, de no ser posible el rechazo de H_0, no tiene sentido continuar con el análisis del Coeficiente de correlación, ya que en este caso simplemente diremos que no existe asociación entre las variables en estudio.

Interpretación para las conclusiones

Interpretación de los resultados

El valor p del coeficiente de correlación es un dato muy importante. Recuerda que esta cifra indica en qué medida el análisis realizado (en este caso la relación) pudo haberse debido al simple efecto del azar.

Muy bien, veamos: los resultados nos muestran un valor de probabilidad menor a 0,001, cifra claramente menor al Alfa (nivel de significación tomado por el investigador), que era de 0,05. Con esta información podemos decir que estamos en condiciones de rechazar H_0 y, por lo tanto, aceptar H_1 que afirmaba la existencia de relación entre las variables Ansiedad y Depresión en la muestra de estudio.

Muy bien, habiendo podido rechazar H_0, estamos en condiciones de describir cómo es dicha asociación entre las variables. El coeficiente R de Pearson fue de 0,823; de esa información analizaremos la dirección (en este caso prestaremos atención si ese valor es positivo o negativo) y la intensidad. En cuanto a la dirección describe una relación positiva y, en lo que respecta a la intensidad, podemos decir que la correlación es muy fuerte. Por tal motivo, podríamos concluir diciendo:

El valor de probabilidad (valor p=<0,001), derivado del análisis correlacional realizado, permite rechazar la hipótesis nula y confirmar la existencia de relación entre las variables Depresión y Ansiedad. El Coeficiente R de Pearson describe una relación positiva de muy fuerte intensidad, por lo que podemos afirmar que a medida que los participantes del estudio presentan valores más elevados en Depresión, han tendido a puntuar más alto también en Ansiedad.

*NOTA: recuerda que este análisis no sería apropiado, dado que la distribución de una de sus variables no cumple con el supuesto de normalidad exigido; por lo que tampoco sería válida la conclusión arribada. A continuación, subsanaremos esta situación ofreciendo un análisis realizado mediante la estrategia no paramétrica.

Análisis no paramétrico: Coeficiente de Correlación Rho de Spearman

Como hemos señalado con anterioridad, cuando la relación entre las variables no es lineal, debemos utilizar otro estadístico que sea apropiado. En este caso, resulta especialmente indicado el Coeficiente de correlación de Spearman (también conocido como Rho de Spearman). En términos generales presenta muchas similitudes con el coeficiente de Pearson; sin embargo, no exige normalidad en sus datos y se puede implementar con variables que posean un nivel de medición ordinal.

Por lo demás, se interpreta igual que el Coeficiente de Pearson tanto en la intensidad de la correlación (valores próximos a 1 indican una relación más fuerte) como por la dirección dada por los signos + o -.

Pasos

Formular las hipótesis nula y alternativa

- H_0= No existe relación entre las variables Ansiedad y Depresión en la muestra de estudio
- H_1= Existe relación entre las variables Ansiedad y Depresión en la muestra de estudio

Establecer el estadístico de prueba adecuado: Prueba Coeficiente de Correlación Rho de Spearman

Seleccionar un nivel de significación: Alfa de 0,05

Establecer la regla de decisión

$$\begin{cases} \mathbf{H_0: Rho_{xy} = 0} \text{ (El coeficiente procede de una población cuya correlación es cero } (\rho = 0)\text{).} \\ \mathbf{H_1: Rho_{xy} \neq 0} \text{ (El coeficiente procede de una población cuya correlación no es cero } (\rho \neq 0)\text{).} \end{cases}$$

Calcular el valor observado del estadístico de prueba (Procedimiento mediante Jamovi)

En este caso, recuerda que no es necesario corroborar los criterios de normalidad y linealidad que exigía Pearson. Por lo que directamente presentaremos el cálculo sólo teniendo en cuenta que las variables tengan un nivel de medición ordinal, o superior.

Para ello repetiremos el procedimiento: Análisis> Regresión> Matriz de correlaciones

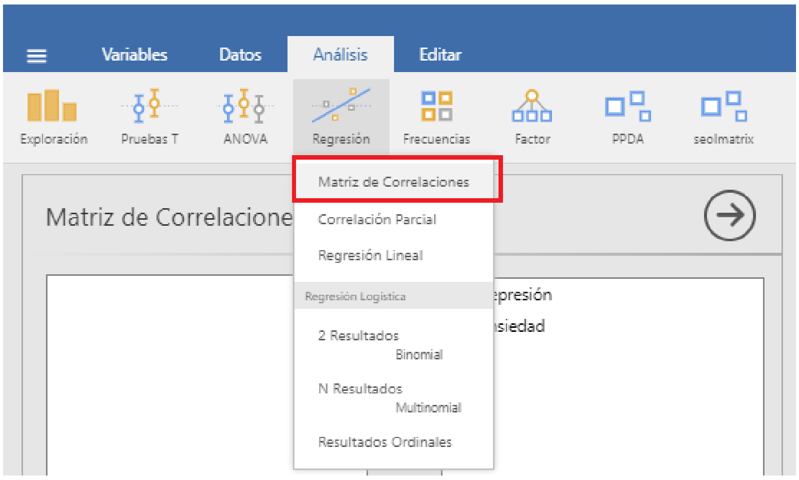


Figura 4: Captura de pantalla del procedimiento para cálculo de relaciones mediante el programa estadístico Jamovi

Análisis de correlación mediante el Coeficiente de Correlación Rho de Spearman

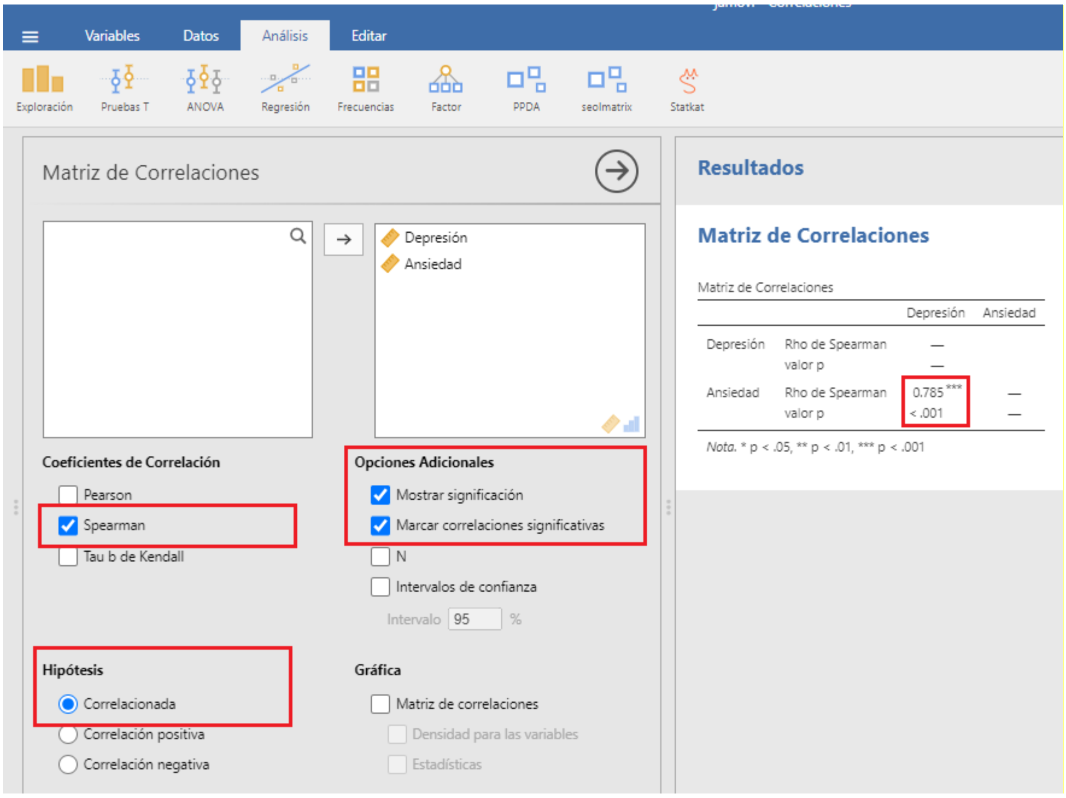


Figura 5: Captura de pantalla opciones para el análisis de Prueba de Correlación Rho de Spearman y resultados de la matriz de correlaciones realizado mediante Jamovi

Resultados

En la figura 5 podrás observar las opciones correspondientes para realizar el cálculo mediante este coeficiente y, adicionalmente, la tabla con los resultados. Tal como sugerimos para Pearson, en los resultados debes prestarle atención inicialmente al valor p y luego, si fue posible rechazar H0, al coeficiente Rho.

Interpretación de los resultados

Los resultados exhiben que el análisis realizado tiene una probabilidad de error menor a 0,001; este valor, por ser inferior al Alfa (nivel de significación tomado por el investigador), nos conduce a rechazar H_0 y, por lo tanto, a confirmar la existencia de relación entre las variables Ansiedad y Depresión en la muestra de estudio.

Habiendo rechazado H_0, ahora podremos describir cómo es dicha asociación. El coeficiente Rho de Spearman fue de 0,785, lo que describe una relación positiva y, en lo que respecta a la intensidad, podemos decir que la correlación es fuerte. Concluyendo podríamos decir:

> El valor de probabilidad (valor p=<0,001), derivado del análisis correlacional realizado, permite rechazar la hipótesis nula y confirmar la existencia de relación entre las variables Depresión y Ansiedad. El Coeficiente Rho de Spearman describe una relación positiva de fuerte intensidad, por lo que podemos afirmar que a medida que los participantes del estudio presentan valores más elevados en Depresión, han tendido a puntuar más alto también en Ansiedad.

NOTA: como observas, la interpretación es similar a la que mostramos para Pearson, solo ha variado la intensidad de la relación, ya que este procedimiento, si bien es más generoso en cuanto a que no presenta las restricciones de Pearson, es más exigente a nivel estadístico para lograr el rechazo de H_0.

Análisis Correlacional para variables cualitativas

Cómo puedes ver, hasta el momento hemos abordado las relaciones entre variables cuantitativas. En este punto comenzaremos a desarrollar las posibilidades de análisis que existen para estudiar la relación entre variables cualitativas. Intentaremos señalar las particularidades descriptivas, además de ver cómo se cuantifica la relación y la significación de las relaciones.

Recordemos que una variable es cualitativa categórica o nominal, cuando los valores que toma la variable son cualidades o categorías o nombres. Por ejemplo: la variable Posee o No posee una determinada patología, Lugar de residencia, Sexo, Género, etc. En tanto, una variable cualitativa ordinal es aquella cuyas categorías se pueden ordenar. Por ejemplo, la variable Nivel de formación (primario, secundario, universitario).

También existe la posibilidad de transformar una variable cuantitativa y convertirla en cualitativa si la categorizamos. Imaginemos que estamos trabajando con la variable Extroversión y tenemos valores comprendidos entre los 45 y 110 puntos. Claramente por el momento esta variable es cuantitativa (numérica) pero, sí establecemos categorías, la convertiremos en cualitativa, por ejemplo: a) *Baja extroversión* (puntuaciones menores a 55); b) *Extroversión media* (puntajes entre 56 y 87) y; c) *Extroversión alta* (valores mayores a 88).

Generalmente, dada la naturaleza de las variables cualitativas, se suele hacer un conteo de los casos que tenemos en cada categoría, en este caso estamos hablando de las frecuencias absolutas que describimos al inicio en los capítulos iniciales. Seguramente recuerdas también que podemos calcular frecuencias relativas y porcentajes, pudiéndolos representar por medio de gráficas como por ejemplo Diagramas de frecuencia o Gráficos de torta.

A continuación, nos centraremos en variables dicotómicas (con dos valores posibles,) para que la comprensión resulte más fácil; Sin embargo, debes saber que los conceptos son fácilmente aplicables a cualquier variable cualitativa con más de dos valores.

Tablas de contingencia. Prueba ji-cuadrado. Corrección de Yates.

Antes de llegar al análisis de una relación entre variables cualitativas, necesitamos conocer una de las formas más eficientes de resumir datos combinados: Las tablas de contingencia, ya que brindan la base sobre la cual se analizan las asociaciones.

Una tabla de contingencia no es otra cosa que una tabla de doble entrada, donde en las columnas se ubican las categorías de una variable (X), y en las filas las categorías de otra variable (Y). En las intersecciones de las filas y las columnas se ubican las

frecuencias absolutas, que reflejan las cantidades de individuos que poseen las características combinadas de alguna de las categorías de la variable ubicada en las filas y de alguna de las categorías de la variable ubicada en las columnas.
Veamos un ejemplo:

Tabla 2.
Tabla de contingencia para las variables Sexo y Agresivo-No agresivo.

	No agresivo	**Agresivo**	**Total**
Hombre	28	20	48
Mujer	39	13	52
Total	67	33	100

Como puedes ver, en la tabla 2 se hace un conteo de hombres y de mujeres según sean agresivos o no. Las intersecciones entre filas y columnas nos informan la cantidad de elementos (individuos) que comparten esas características. Por ejemplo, la celda de color verde refleja que en la muestra existen 20 hombres agresivos, en tanto que la celda de color naranja anuncia 39 mujeres no agresivas. En la columna derecha verás el total de cada fila (48 hombres y 52 mujeres); y en la fila inferior encontrarás el total de cada columna (67 no agresivos y 33 agresivos), y finalmente el total de la muestra (100).
Algunas veces, cuando pretendemos hacer sólo un análisis descriptivo, y tenemos tamaños muestrales distintos entre los grupos, conviene representar en la tabla de contingencia, no las frecuencias, sino los porcentajes. Por ejemplo, si nos interesa comparar los porcentajes de agresividad según sexo podríamos presentar la siguiente tabla

Tabla 3.
Tabla de contingencia con porcentajes de la variable Agresivo-No agresivo según Sexo.

	No agresivo	**Agresivo**	**Total**
Hombre	58,33%	41,66%	100%
Mujer	75%	25%	100%

¿Cómo obtuvimos esos valores? Fácil, sabíamos que el N de la muestra de Hombres era de 48, por lo que para obtener el porcentaje de hombres No agresivos tomamos la frecuencia (28), ese valor lo multiplicamos por 100 (ese dato representa el 100% de varones) y lo dividimos por el N (48); es decir: 28*100/48= 58,33%. Con el mismo razonamiento hicimos el resto de los cálculos.
Ahora estamos en condiciones de analizar que, en términos de proporción, si consideramos las muestras de hombres y mujeres por separado, la de mujeres tiene una proporción mayor de no agresividad que la muestra de varones.
Como podrás darte cuenta, las tablas de continencia tienen un poder revelador en el análisis descriptivo gracias a su capacidad de resumir datos y representarlos de manera clara y ordenada. No obstante, guardan un poder adicional; son la base con la cual se pueden calcular pruebas de hipótesis para las diferencias entre proporciones y pruebas de hipótesis para relaciones.
Como veníamos proponiendo, en esta oportunidad nos enfocaremos en la posibilidad de someter a prueba la hipótesis de que dos variables cualitativas están relacionadas y que los resultados observados no sean producto del azar.

Prueba ji-cuadrado o X^2
El estadístico ji-cuadrado o X^2 (también conocido como chi-cuadrado), es útil para someter a prueba de hipótesis las distribuciones de frecuencias de datos cualitativos, los que suelen representarse convenientemente en las tablas de contingencia. Lo que busca este estadístico es, de acuerdo a la hipótesis nula que se plantee, hacer un contraste entre las frecuencias

observadas (aquellas que surgen de los datos recogidos por el investigador) y las frecuencias esperadas (cantidad de datos que se esperaría encontrar en las celdas si las variables fuesen independientes).

Así, como con los estadísticos "z" y "t" hemos sometido a prueba hipótesis diferencias de valores medios y relaciones entre variables cuantitativas, con el estadístico X^2 (o ji-cuadrado) lograremos poner a prueba hipótesis si las frecuencias denotan asociación de las variables, o no. Sin embargo, este coeficiente no informa en qué dirección y cuál es la intensidad de la relación. Por este motivo suelen implementarse algunos coeficientes adicionales a modo de ampliar el análisis.

- ✓ Análisis complementarios para pruebas de hipótesis de relación entre dos variables cualitativas nominales

  En estos casos, los programas estadísticos suelen ofrecer: el Coeficiente de Contingencia y Phi y V de Cramer

  Coeficiente de contingencia: se trata de una medida, basada en ji-cuadrado, que ofrece información acerca de la intensidad de la asociación. Este coeficiente toma valores entre 0 y 1. Cifras cercanas a 0 indican que no hay asociación entre las variables, o que es de muy baja intensidad, en cambio un valor cercano a 1 es indicativo de una gran intensidad de relación entre las variables. Como el resultado es sensible al número de filas y columnas de la tabla de contingencia se suele complementar con otros coeficientes.

  Phi y V de Cramer: Ambos estadísticos, al igual que el Coeficiente de contingencia, son una medida de asociación basada en ji-cuadrado. Se interpreta de manera similar a este, en cuanto también ofrece valores comprendidos entre 0 y 1. Sin embargo, tienen una mayor precisión en la valoración de la intensidad al dividir el estadístico de ji-cuadrado por el tamaño de la muestra y extraer la raíz cuadrada del resultado.

- ✓ Análisis complementarios para pruebas de hipótesis de relación entre dos variables cualitativas ordinales

  Cuando en las tablas de contingencia en las filas y las columnas tenemos variables ordinales se pueden realizar los siguientes análisis: *Gamma* y *Tau-b de Kendall*. Ambos coeficientes toman valores comprendidos entre -1 y 1. El signo indica la dirección de la relación, en tanto que el valor señala la fuerza de la relación.

- ✓ Análisis complementarios para pruebas de hipótesis de relación entre una variable nominal y otra de intervalo

  Cuando se dan estos casos se puede implementar Eta, aunque frecuentemente la variable categórica debe codificarse numéricamente con anterioridad. Por el momento esta opción no está disponible en Jamovi.

Cálculo y análisis

Veamos un ejemplo. Supongamos que deseamos estudiar con un nivel de significación de 0,05 la posible relación (asociación) entre: haber realizado algún tratamiento psicoterapéutico o psicomotriz y lugar de residencia (urbana o rural). Partimos de la suposición de que quienes residen en zonas urbanas tienden a realizar tratamientos psicoterapéuticos o psicomotrices

Bien, ahora procedemos a realizar la prueba de hipótesis

Pasos

Formular las hipótesis nula y alternativa

H0: Haber realizado tratamientos psicológicos no está relacionado con el lugar de residencia de la persona
H1: Haber realizado tratamientos psicológicos está relacionado con el lugar de residencia de la persona

Establecer el estadístico de prueba adecuado: <u>Prueba ji-cuadrado (X^2) con análisis complementario mediante Coeficiente de contingencia, Phi y V de Cramer</u>

Seleccionar un nivel de significación: Alfa de 0,05

Establecer la regla de decisión

H₀: X^2_{xy} = 0 (Los datos proceden de una población cuya correlación es cero (ρ =0)).

H₁: X^2_{xy} ≠ 0 (Los datos proceden de una población cuya correlación no es cero (ρ ≠ 0)).

Calcular el valor observado del estadístico de prueba (Procedimiento mediante Jamovi)

Para ello aplicaremos el procedimiento: Análisis> Frecuencias> Muestras independientes (Prueba de asociación de X^2)

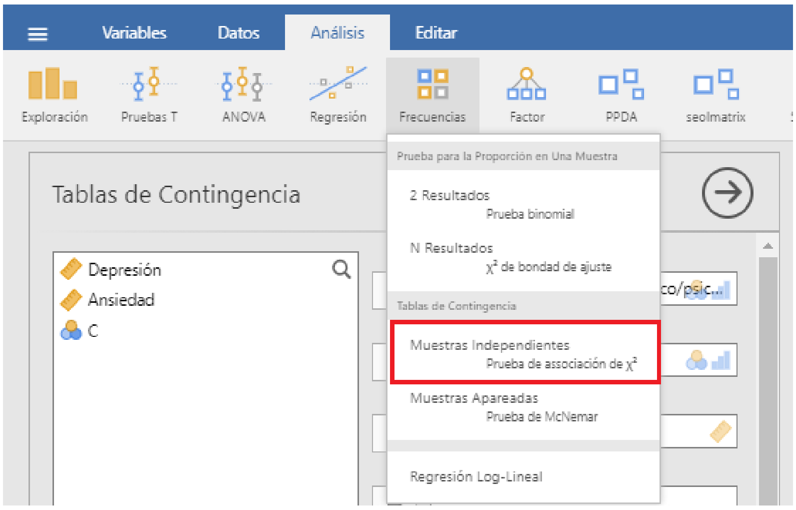


Figura 6: Captura de pantalla del procedimiento para cálculo de relaciones cualitativas mediante el programa estadístico Jamovi

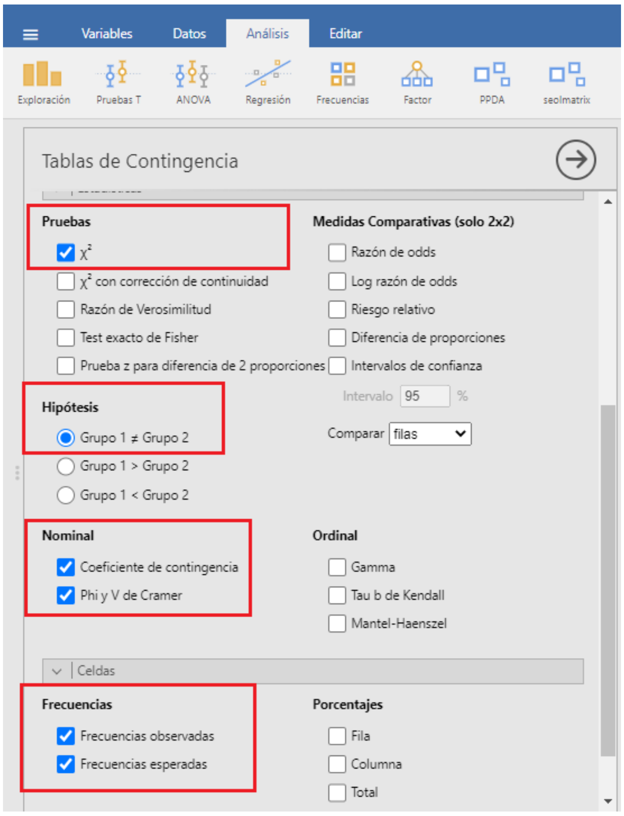


Figura 7: Captura de pantalla opciones para el análisis de correlación mediante X2 (chi-cuadrado) con análisis complementario mediante Coeficiente de contingencia, Phi y V de Cramer realizado mediante Jamovi.

Los recuadros rojos de la figura 7 señalan las opciones usualmente solicitadas cuando se busca corroborar la hipótesis de relación entre variables nominales y obtener información complementaria de utilidad para dar cuenta de las características de tal posible relación.

Resultados

Para el análisis solicitado, Jamovi ofrece tres tablas con información. A continuación, interpretaremos cada una de ellas.

Tabla 4.

Tabla de contingencia con frecuencias observadas y esperadas para las variables Lugar de residencia y Tratamiento psicoterapéutico/psicomotriz.

Tratamiento psicoterapéutico/psicomotriz		Lugar de residencia: Rural	Urbana	Total
Si	Observado	8	14	22
	Esperado	11.27	10.73	22.0
No	Observado	13	6	19
	Esperado	9.73	9.27	19.0
Total	Observado	21	20	41
	Esperado	21.00	20.00	41.0

La tabla de contingencia mostrada arriba, si momentáneamente tomamos las frecuencias observadas (recuadro verde), sólo ofrece información de carácter descriptivo. Como nuestro interés es comprobar si la hipótesis que planteamos es válida continuaremos con la tabla siguiente, para luego, siempre que podamos rechazar H_0, volver y utilizar toda la información de la tabla 4 a los efectos de lograr una mayor comprensión del análisis.

Tabla 5.

Análisis correlacional mediante Ji-Cuadrado (X^2) para las variables Lugar de residencia y Tratamiento psicoterapéutico/psicomotriz.

	Valor	gl	p
χ^2	4.19	1	0.041
N	41		

La tabla 5 nos informa el valor de probabilidad asociado al cálculo de X^2 (valor p=0,041). Dado que el mismo es menor al Alfa del Investigador (0,05), podemos rechazar la hipótesis nula y confirmar que las variables Lugar de residencia y Tratamiento psicoterapéutico/psicomotriz se encuentran relacionadas.

Tabla 6.

Coeficiente de contingencia, Phi y V de Cramer para las variables Lugar de residencia y Tratamiento psicoterapéutico/psicomotriz.

	Valor
Coeficiente de contingencia	0.305
Coeficiente Phi	0.320
V de Cramer	0.320

Tal como lo explicáramos en párrafos anteriores, los tres coeficientes nos ofrecen información de la intensidad de la relación, no de la dirección, y en cuanto a ello podemos decir que es baja. Una vez que hemos confirmado la relación de las variables mediante X2 y la fuerza de las relaciones, por medio de los coeficientes de Contingencia, Phi y V de Cramer, queda interpretar la forma en que se da la relación. Para ello, volveremos a la tabla 4 e interpretaremos, en cada celda, la diferencia que se da entre las frecuencias observadas y las esperadas.

Tomando los valores correspondientes de las personas que han realizado tratamiento, encontramos que las frecuencias observadas de quiénes viven en zona rural son menores a las esperadas, en tanto que, para quienes viven en zonas urbanas son mayores las observadas que las esperadas. Una situación inversa se presenta en las personas que no han realizado tratamiento psicoterapéutico o psicomotriz. Esto nos lleva a comprender que vivir en zonas urbanas se relaciona con una mayor tendencia a realizar tratamientos de estos tipos.

Conclusiones del análisis

Dados los valores de probabilidad asociados a X^2 (ji-cuadrado) y su comparación con el nivel de significación propuesto estamos en condiciones de confirmar la relación entre las variables. La intensidad de la asociación es baja según lo observado en los coeficientes de Contingencia, Phi y V de Cramer, en tanto que el análisis de las diferencias entre las frecuencias observadas y esperadas de la tabla de contingencia sugieren que residir en zonas urbanas se encuentra relacionado con una mayor tendencia a realizar tratamientos psicoterapéuticos o psicomotrices.

Corrección de Yates

Cuando aplicamos la distribución ji cuadrado existe una pequeña distorsión respecto a la realidad, se produce porque estamos utilizando una distribución continua, para representar un fenómeno discreto. Cuando se trabaja con números grandes de frecuencias esperadas esta desviación es muy pequeña y la podemos descartar. Sin embargo, cuando tenemos valores esperados en las celdas menores a 5, la desviación puede ser importante por lo que se debe corregir.

Yates en 1934 presentó un procedimiento muy sencillo para corregir los métodos empleados para hallar el ji cuadrado, logrando aumentar la concordancia entre los resultados del cálculo y la distribución teórica.

Esta corrección se puede usar siempre y cuando la muestra sea mayor de 20 elementos y que el valor esperado sea menor que 5 pero igual o mayor que 3; en el caso de que tengamos muestras más pequeñas de 20 casos o tengamos valores esperados menores que 5 se recomienda la prueba de Fisher.

La corrección se hace restando 0,5 a cada diferencia entre el valor observado y el valor esperado. De esta manera es más difícil rechazar la hipótesis nula, ya que el valor de ji cuadrado calculado va a ser menor.

Con la corrección de Yates el estadístico de prueba pasa a ser el siguiente:

$$\chi^2 = \sum \frac{(|\text{F. Observada} - \text{F. Esperada}| - 0.5)^2}{\text{F. Esperada}}$$

Introducción al análisis de regresión lineal simple

Hasta acá hemos visto cómo podemos indagar si dos o más variables están relacionadas, o si por el contrario son independientes. También hemos dicho que encontrar una relación no implica afirmar que estamos seguros que esa relación es de causalidad (causa → efecto).

Pero, ¡imaginemos por un momento lo interesante que puede ser descubrir una relación de este tipo! Supongamos que podemos afirmar que la asertividad es causa de un mejor desempeño académico! ;) … o que determinada técnica ayuda a mejorar a nuestros pacientes!!! Bueno, para estos propósitos tenemos las técnicas de regresión.

Los análisis de regresión son estrategias estadísticas utilizadas para determinar el tipo de relación que puede existir entre una variable llamada dependiente, o criterio, (Y) y una o más variables llamadas independientes, o predictoras, (X1, X2, ..., Xk). Se suele implementar para identificar relaciones potencialmente causales o, en el caso de que no existan dudas sobre esta relación causal, para predecir los valores de una variable a partir de otra.
Hay fenómenos que son claramente deterministas; por ejemplo, la presión ejercida en un cuerpo y la deformación que se genera en él. Sin embargo, en nuestras disciplinas las relaciones deterministas son más difíciles de identificar, ya que suelen intervenir un amplio conjunto de variables, a veces desconocidas, que influyen sobre la relación que existe entre las que son sometidas a estudio.

Hay distintos tipos de regresiones y se las puede clasificar del siguiente modo:
Según la cantidad de variables independientes:

- **Regresión simple**: Cuando la variable *Y* depende únicamente de una única variable *X*.
- **Regresión múltiple**: Cuando la variable *Y* depende de varias variables (X_1, X_2, ..., X_r)

Según el tipo de función *f*(*X*):

- **Regresión lineal**: Cuando el modelo propone una función lineal.
- **Regresión no lineal**: Cuando el modelo asume una función no lineal.

En esta oportunidad nos ocuparemos de explicar las regresiones lineales simples, intentaremos exponer los fundamentos esenciales y, al igual que en los capítulos anteriores, trataremos no ceñirnos tanto a aspectos más técnicos del análisis, sino que buscaremos que comprendas cuándo, cómo utilízalo y cómo interpretar sus resultados.

Concepto de regresión lineal entre dos variables

La regresión es un modelo estadístico que tiene por objeto estimar el efecto de una variable sobre otra. Es una prueba que tiene mucha relación con el coeficiente r de Pearson, pero a diferencia de este, mediante su implementación se puede predecir los valores de una variable, considerando los valores de la otra variable. Como deducirás más adelante, mientras mayor sea la correlación entre las variables, mayor será la fuerza de la predicción.
Si bien el cálculo es complejo, actualmente con los programas informatizados se pueden obtener sus resultados en un segundo; por lo que nos centraremos en los rudimentos necesarios para su interpretación.
La prueba requiere de dos variables cuantitativas con un nivel de medición de intervalos o razón; a una de ellas se la considera como independiente (X) y otra como dependiente (Y). Pero recuerda, tiene sentido hacerlo sólo cuando exista un sólido sustento teórico, de lo contrario podríamos afirmar algo que en la realidad no existe o es producto de la casualidad.

Comprobación de supuestos

Como la regresión lineal simple es una extensión del coeficiente de Correlación de Pearson, se mantienen los requisitos para su implementación y se incorpora uno adicional.

- *Nivel de medición:* variables cuantitativas continuas, con una escala de intervalo (por lo menos).
- *Linealidad de la asociación*
- *Normalidad de los datos*
- *Homogeneidad de las varianzas*: distribución de los datos de las variables relativamente constantes.

Procedimiento e interpretación

Como hemos afirmado con anterioridad, los análisis de regresión (cualquiera de ellos) se ejecutan proponiendo de antemano un modelo con una función matemática determinada. En el caso de la regresión lineal, la función incluida en el modelo propuesto es, obviamente, lineal. Luego, lo que se pone a prueba, en definitiva, es si el modelo es válido para explicar el fenómeno.

Partiremos del siguiente ejemplo:

Hemos tomado una muestra probabilística de 44 universitarios extraídos de la población de estudiantes de Metodología de la investigación I. Nos interesa determinar si la cantidad de horas de estudio (variable independiente) es la que predice las Calificaciones que obtienen (variable dependiente).

Pasos

Formular las hipótesis nula y alternativa

- H_0= La variable Horas de estudio (independiente) NO predice las Calificaciones de los estudiantes universitarios (variable dependiente).
- H_1= La variable Horas de estudio (independiente) predice las Calificaciones de los estudiantes universitarios (variable dependiente).

Establecer el estadístico de prueba adecuado: Regresión lineal simple

Seleccionar un nivel de significación: Alfa de 0,05

Establecer la regla de decisión

H_0: Y≠ A+B de X. El valor de la variable dependiente Y (Calificaciones) NO está en función de los valores A y B de la variable independiente X (Horas de estudio).

H_1: Y= A+B de X. El valor de la variable dependiente Y (Calificaciones) está en función de los valores A y B de la variable independiente X (Horas de estudio).

Calcular el valor observado del estadístico de prueba (Procedimiento mediante Jamovi)

Por ser una prueba de tipo paramétrica nos encontramos, inicialmente, con la necesidad de comprobar los supuestos antes de dar comienzo a la prueba en sí. No obstante, la mayoría de los softwares estadísticos ofrecen el análisis de manera simultánea. Para ejecutar todo el análisis en Jamovi debes ir a **Análisis > Regresión > Regresión lineal**

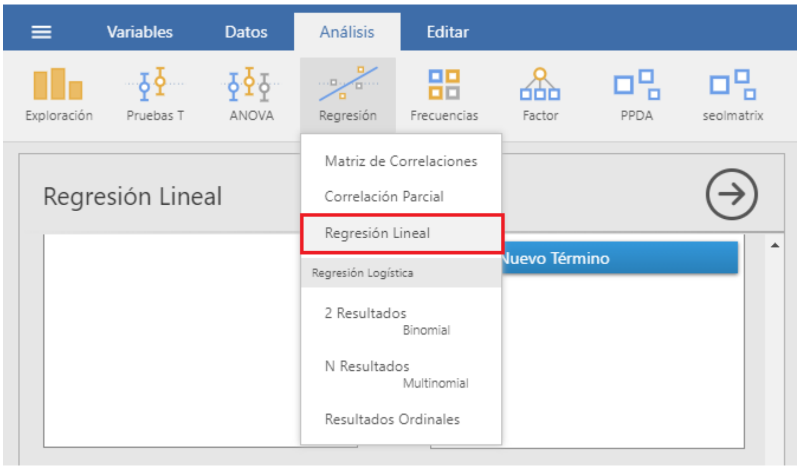


Figura 8: Captura de pantalla del procedimiento para cálculo de Relaciones lineales mediante el programa estadístico Jamovi

Una vez ingresado en el análisis (figura 8) se deben ubicar las variables a analizar en el casillero correspondiente (variable dependiente, y la variable independiente como covariable) según la manera en la que ha sido planteado el modelo.

Posteriormente hay que marcar las opciones necesarias para la comprobación de los supuestos: En *Constante* marcar Nivel de referencia; en *Comprobación de supuestos* seleccionar Prueba de normalidad; en *Ajuste del modelo* tildar R, R^2 y R^2 ajustada; y en *Coeficientes del modelo* seleccionar ANOVA.

Ahora estamos en condiciones de desglosar los resultados ofrecidos

Inicialmente debemos corroborar que se cumplan los tres supuestos de este análisis.

- ✓ *Nivel de medición:* ambas variables son continuas con un nivel de medición intervalar.
- ✓ *Linealidad:* Confirmada por gráfico 1 y tabla 7

Gráfico 1:

Diagrama de dispersión de las variables Calificaciones y Horas de estudio.

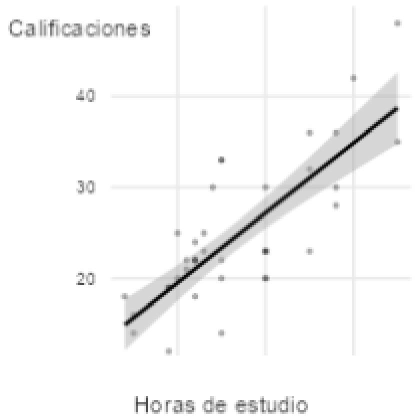


Tabla 7.

Prueba para determinar la relación lineal del modelo propuesto mediante ANOVA.

Prueba Omnibus ANOVA

	Suma de Cuadrados	gl	Media Cuadrática	F	p
Calificaciones	1661	1	1661.1	64.9	< .001
Residuos	998	39	25.6		

Nota. Suma de cuadrados tipo 3

En la tabla 7, mediante el procedimiento de ANOVA, se busca información sobre si efectivamente existe una relación lineal significativa entre la variable dependiente y la variable independiente, dando una pista más de la adecuación del modelo de regresión. Acá lo que tenemos que prestar mucha atención es al nivel de significación, ya que con valores <0,05 podemos decir que la variable dependiente (Calificaciones) efectivamente está influida linealmente por la variable independiente (Horas de estudio).

- ✓ *Normalidad de los datos:* los valores p observados en la tabla 8 sugieren que las distribuciones de ambas variables se ajustan a la normal

Tabla 8.

Prueba de normalidad de Shapiro-Wilk para el modelo propuesto.

Normality Tests

	Statistic	p
Shapiro-Wilk	0.944	0.083
Kolmogorov-Smirnov	0.153	0.291
Anderson-Darling	0.890	0.051

Nota. Additional results provided by *moretests*

- ✓ Homogeneidad de las varianzas: Se asume que las variables del modelo son homocedásticas (que tienen varianzas similares). En caso de sospecha del incumplimiento de este supuesto se debe calcular la prueba correspondiente: Levene

Ahora bien, habiéndose confirmado que el modelo cumple con los supuestos exigidos, pasamos a interpretar la información propia de él.

Tabla 9.

Medidas de ajuste del modelo

Modelo	R	R^2	R^2 Adjustada
1	0.790	0.625	0.615

En la tabla 9 podemos ver el coeficiente R (0,79), éste es el valor de correlación entre las variables, lo que nos indica que tienen una asociación fuerte (dato deseable cuando se quiere realizar este tipo de análisis). Este valor es importante porque si no existe relación entre las variables o ésta es muy débil, no tiene sentido proceder con el análisis. La R^2 es el coeficiente de determinación, nos dice que los puntos se ajustan muy bien a la recta propuesta; mientras más cercano a 1 sea este valor, mejor es (en este caso, con R^2=0,625, podemos decir que es aceptable). La R^2 ajustada nos dice la bondad de ajuste de la recta del modelo. En nuestro ejemplo, es 0,615, lo que indica que la variable *Horas de estudio* aporta capacidad de predicción de la variable dependiente (*Calificaciones*) en un 61,5%.

Tabla 10.

Coeficientes del modelo

Predictor	Estimador	EE	t	p
Constante	6.430	2.572	2.50	0.017
Calificaciones	0.813	0.101	8.06	< .001

En la tabla 10 podremos ver qué dirección y qué impacto estimado tiene la recta de regresión que propone el modelo. Antes de comenzar a interpretar esta tabla debemos corroborar si los valores p son menores al nivel de significación escogido, ya que de lo contrario no podríamos rechazar la hipótesis nula y tendríamos que desestimar el modelo. Si observamos el recuadro rojo encontraremos que los valores de probabilidad, por ser más pequeños que 0,05, confirman la robustez del modelo y, por lo tanto, la relación causal: Horas de estudio determina la variable Calificación.

En la primera columna tenemos los valores de la constante (encuadrado en color azul) y el valor que está delimitado en verde es la inclinación o pendiente de la línea de regresión. Al primer valor se lo denomina A = 6,43 (Ordenada de origen o Intercepto) y nos dice el punto donde la recta corta el eje vertical de Y. El valor de abajo se lo denomina B = 0,813 (Pendiente o Slope) y refleja que por cada punto de aumento de Horas de estudio se espera un cambio de 6,43 puntos en Calificaciones.

Entonces, la Recta de regresión calculada para nuestros datos es Y= (A+B).X lo que sería: → Y= 6,43+0,813 * X (ten en cuenta que luego del valor de B hay un asterisco que simboliza una multiplicación con el valor que nos interese de la variable X (Horas de estudio).

Muy bien, pero: ¿Cuál es la utilidad práctica de estos datos?

Y= 6,43+0,813 * X es lo que ha sido confirmado, y esto determina la mejor recta de regresión posible para los datos recogidos. Con este modelo podremos calcular qué valor podemos esperar en la variable dependiente si proponemos un determinado valor para la variable independiente.

Veamos: Queremos predecir qué valor puede asumir Calificaciones si un individuo invierte 20 Horas de estudio. Esto se calcula del siguiente modo:

$$Y= 6{,}43 + 0{,}813 * 20$$

Lo que nos da el siguiente resultado Y=22,69

Es decir que se espera que la persona que obtenga una Calificación de 22,69 puntos

Realicemos un cálculo más: ¿qué Calificación podrá tener un estudiante si invierte 50 Horas de estudio?

$$Y= 6{,}43 + 0{,}813 * 50$$

Lo que nos da el siguiente resultado Y= 47,08

Se espera que el estudiante que destine 50 Horas de estudio a la asignatura obtendrá una Calificación de 47,08 puntos

Recuerda que la expresión derivada del modelo no deja de ser una *aproximación estadística* a la *verdadera relación que existe entre* X e Y. Obviamente, mientras mejor sea el modelo, mejor será la predicción.

Bibliografía

Bonmatí, A.N & Vasallo, J.M. (2016). Estadística básica en Ciencias de la Salud. Universidad de Alicante. Recuperado el 17/07/2019 de https://rua.ua.es/dspace/bitstream/10045/60526/1/Estad%c3%adstica%20b%c3%a1sica%20en%20Ciencias%20de%20la%20Salud.pdf

Cabrera, J. (2017). Regresión Lineal en SPSS. Recuperado el 17/07/2019 de https://investigayanaliza.blogspot.com/2017/12/regresion-lineal-en-spss.html

Carollo-Limeres, M.C. (2012). *Regresión lineal simple.* Universidad de Santiago Compostela. Recuperado el 17/07/2019 de http://eio.usc.es/eipc1/BASE/BASEMASTER/FORMULARIOS-PHP-DPTO/MATERIALES/Mat_50140116_Regr_%20simple_2011_12.pdf

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2010). *Metodología de la Investigación* (5ta edición). México D.F.: McGraw Hill.

Laguna, C. (2016). *Correlación y regresión lineal*. Instituto Aragonés de Ciencias de la Salud. Recuperado el 17/07/2019 de http://www.ics-aragon.com/cursos/salud-publica/2014/pdf/M2T04.pdf.

Orellana, L. (2008). *Análisis de regresión. Regresión Lineal Simple.* Recuperado el 17/07/2019 de http://www.dm.uba.ar/materias/estadistica_Q/2011/1/clase%20regresion%20simple.pdf

Palacios-Cruz, L.; Pérez, M.; Rivas-Ruiz, R. & Talaverab, J. (2013). Investigación clínica XVIII. Del juicio clínico al modelo de regresión lineal. *Rev Med Inst Mex Seguro Soc., 51*(6), 656-661.

Recursos On Line

- Haz diagramas de dispersión y calcula la recta de regresión http://www.alcula.com/calculators/statistics/scatter-plot/
- Calculadora de Regresión Lineal http://www.alcula.com/calculators/statistics/linear-regression/
- Calculadora de Coeficiente de Correlación de Pearson http://www.alcula.com/calculators/statistics/correlation-coefficient/
- Evalúa tu conocimiento y aprendé más!!! https://es.khanacademy.org/math/statistics-probability