



<Artículo metodológico>

Análisis de regresión lineal múltiple con SPSS: un ejemplo prácticoRuth Vilà Baños¹ , Mercedes Torrado Fonseca² , Mercedes Reguant Álvarez³ 

Enviado: 10/09/2018. Aceptado: 27/11/2018. Publicado en prensa: 28/02/2019. Publicado: 01/07/2019

//Resumen

INTRODUCCIÓN: El artículo presenta el análisis de regresión lineal múltiple mediante la exposición de un caso práctico y utilizando el SPSS como programa estadístico. Se trata de una técnica aplicada en una amplia variedad de situaciones para predecir fenómenos diversos, poniendo en juego más de dos variables que se exploran y cuantificando la relación entre la variable dependiente y las variables independientes.

MÉTODO: La aplicación de la técnica implica cinco pasos: 1. Seleccionar la variable dependiente; 2. Seleccionar las variables explicativas; 3. Verificar los supuestos necesarios para poder aplicar adecuadamente la técnica; 4. Interpretar el modelo resultante identificando las variables predictoras propuestas y su importancia explicativa; y 5. Establecer la bondad de ajuste del modelo y su capacidad predictiva.

RESULTADOS: En el ejemplo de estudio propuesto sobre violencias de género entre adolescentes en los entornos virtuales, la aplicación de esta técnica ha permitido identificar factores asociados con el comportamiento ciberagresor vinculados a una serie de variables de tipo tecnológico y personal.

//Palabras clave

Regresión lineal múltiple; Técnicas multivariantes; Técnicas predictoras; Ciberviolencias de género; SPSS.

//Datos de las autoras

¹ Profesora. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE), Universidad de Barcelona, España.

² Profesora. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE), Universidad de Barcelona, España.

³ Profesora. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE), Universidad de Barcelona, España. Autora para la correspondencia: mreguant@ub.edu

//Referencia recomendada

Vilà Baños, R., Torrado Fonseca, M., y Reguant Álvarez, M. (2019). Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico. *REIRE Revista d'Innovació i Recerca en Educació*, 12(2), 1–10. <http://doi.org/10.1344/reire2019.12.222704>

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





//Títol

Anàlisi de regressió lineal múltiple amb SPSS: un exemple pràctic

//Resum

INTRODUCCIÓ: L'article presenta l'anàlisi de regressió lineal múltiple mitjançant l'exposició d'un cas pràctic utilitzant SPSS com a programa estadístic. Es tracta d'una tècnica aplicada a una àmplia varietat de situacions per a la predicció de fenòmens diversos. Es posen en joc més de dues variables que s'exploren, i es quantifica la relació entre la variable dependent i les variables independents.

MÈTODE: L'aplicació de la tècnica implica cinc passos: 1. Seleccionar la variable dependent; 2. Seleccionar les variables explicatives; 3. Verificar els supòsits necessaris per poder aplicar adequadament la tècnica; 4. Interpretar el model resultant, identificant les variables predictorres proposades i la seva importància explicativa; 5. Establir la bondat d'ajust del model i la seva capacitat predictiva.

RESULTATS: En l'exemple d'estudi proposat sobre violències de gènere entre adolescents en els entorns virtuals, l'aplicació d'aquesta tècnica ha permès identificar factors associats al comportament ciberagressor vinculats a una sèrie de variables de tipus tecnològic i personal.

//Paraules clau

Regressió lineal múltiple; Tècniques multivariables; Tècniques predictorres; Ciberviolències de gènere; SPSS.

//Title

Multiple regression analysis using SPSS Statistics: a practical example

//Abstract

INTRODUCTION: The article explains how to perform the technique of multiple regression using SPSS Statistics to analyse data in a practical case. This technique is used in a wide variety of situations to predict different phenomena, quantifying the relationship between a dependent variable and two or more independent variables.

METHOD: The technique involves five steps: (1) Select your dependent variable; (2) Select your independent variables; (3) Verify that your data passes the assumptions required for the technique to give you a valid result; (4) Interpret the resulting regression model, identifying the proposed predictor variables and their statistical significance; and (5) Establish whether the model is a good fit for the data and the independent variables can predict the values of the dependent variable.

RESULTS: This technique was used in a study of gender violence among adolescents in virtual environments to predict identifying factors associated with cyber-based aggressive behaviour from a series of variables of a technological and personal nature.

//Keywords

Multiple regression; Multivariate techniques; Predictive techniques; Gender-based cyber violence; SPSS Statistics.

1. Características generales de la prueba

El Análisis de regresión lineal múltiple es una técnica estadística utilizada para estudiar la relación entre variables en una amplia variedad de situaciones y predecir fenómenos diversos. La función de regresión más simple es la *lineal*, donde cada variable participa de forma aditiva y constante para todo el fenómeno observado (Hernández, Ramírez y Ferri, 2005). La regresión lineal múltiple pone en juego más de dos variables que se exploran, y se cuantifica la relación entre la variable dependiente y las variables independientes.

La técnica analiza la relación entre una variable dependiente métrica y varias variables independientes también métricas, con el objetivo principal de predecir la única variable dependiente seleccionada por el investigador, utilizando las variables independientes (Pérez, 2008). El modelo lineal viene dado de la siguiente forma (Pérez y Santín, 2008):

$$Y=b_0+b_1X_1+b_2X_2+\dots+b_kX_k+u$$

Los coeficientes b_1, b_2, \dots, b_k denotan la magnitud del efecto que las variables independientes (explicativas) tienen sobre la dependiente Y . El coeficiente b_0 es un término constante del modelo y u es el error del modelo.

Supuestos del modelo de regresión lineal múltiple

Para poder crear un modelo de regresión lineal es necesario que se cumpla con los siguientes supuestos:

- Linealidad: Que la relación entre las variables sea lineal.
- Independencia: Que los errores en la medición de las variables explicativas sean independientes entre sí.
- Homocedasticidad: Que los errores tengan varianza constante.
- Normalidad: Que las variables sigan la Ley Normal.
- No colinealidad: Que las variables independientes no estén correlacionadas entre ellas.

2. Un ejemplo práctico de aplicación de la prueba

Para aplicar la técnica se deben seguir los siguientes pasos:

- Paso 1. Seleccionar la variable dependiente.
- Paso 2. Seleccionar las variables explicativas.
- Paso 3. Verificar los supuestos del modelo.
- Paso 4. Interpretar el modelo.
- Paso 5. Bondad de ajuste del modelo.



Paso 1. Seleccionar la variable dependiente

En un estudio que analiza las experiencias en violencias de género en los entornos virtuales de chicos y chicas adolescentes se pretenden identificar factores asociados con el comportamiento ciberagresor en este colectivo. Se tiene información sobre los siguientes datos:

Tabla 1. Dimensiones del análisis

	Dimensiones de análisis	Nº de ítems
Caracterización	Datos sociodemográficos	7
	Perfil de uso tecnológico	8
	Experiencia en ciberbullyng (escala Likert -3)	9
Escala Likert-5	Escala concepto de VG 2.0	15
Escala Likert-5	Escala ciberagresión en VG2.0	20
Escala Likert-5	Escala cibervictimización en VG2.0	20
Escala Likert-5	Escala ciberobservación en VG2.0	20

La variable cuyos valores se pretenden explicar mediante el resto de las variables, es decir, la variable dependiente, es la puntuación en la escala de ciberagresión en VG2.0.

Paso 2. Seleccionar las variables explicativas

Ante un número grande de posibles variables predictoras la mejor solución es seleccionar aquellas que mejor puedan explicar la variable dependiente, en este caso, la ciberagresión en violencia de género 2.0.

Esta selección de variables independientes debe estar orientada por su pertinencia según la fundamentación teórica del estudio, así como por la significación empírica que se obtenga con el modelo. En el ejemplo, se seleccionan las siguientes variables explicativas:

Tabla 2. Selección de variables independientes

	Variables explicativas	Nº de ítems
Caracterización	Perfil de uso tecnológico	8
	Experiencia en ciberbullyng (escala Likert -3)	9
Escala Likert-5	Escala concepto de violencia de género	15

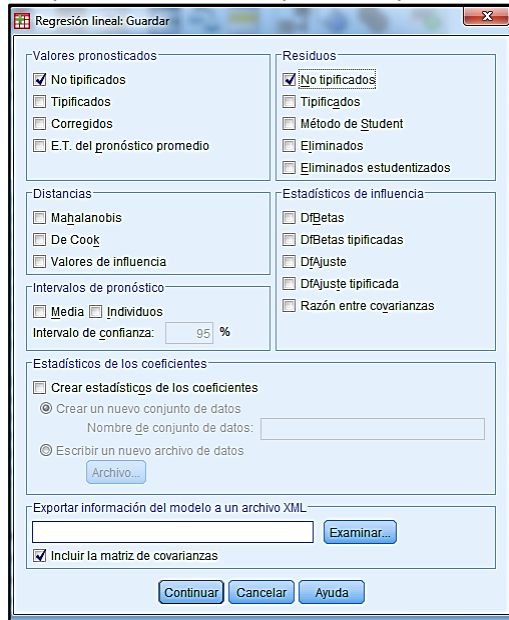
Paso 3. Verificar los supuestos del modelo

Para aplicar la regresión lineal múltiple que nos estamos proponiendo, los datos deben cumplir con los 5 supuestos ya mencionados: linealidad, independencia, homocedasticidad, normalidad y no colinealidad.

Estos supuestos pueden estudiarse con las puntuaciones residuales, analizando la diferencia entre las puntuaciones observadas y las predichas por el modelo de regresión. El procedimiento con SPSS se encuentra en la opción "guardar" (tercer cuadro de la derecha dentro de la opción Analizar –menú principal de SPSS– → Regresión → Lineales) y en el cuadro de diálogo correspondiente marcaremos

en Valores pronosticados No tipificados y Residuos No tipificados (figura 1). Se generan así, dos variables *PRE_1* y *RES_1*. Con ellas procederemos a iniciar el estudio de los supuestos del modelo. Además, en los gráficos se puede detectar alguna anomalía (Camacho, 2006).

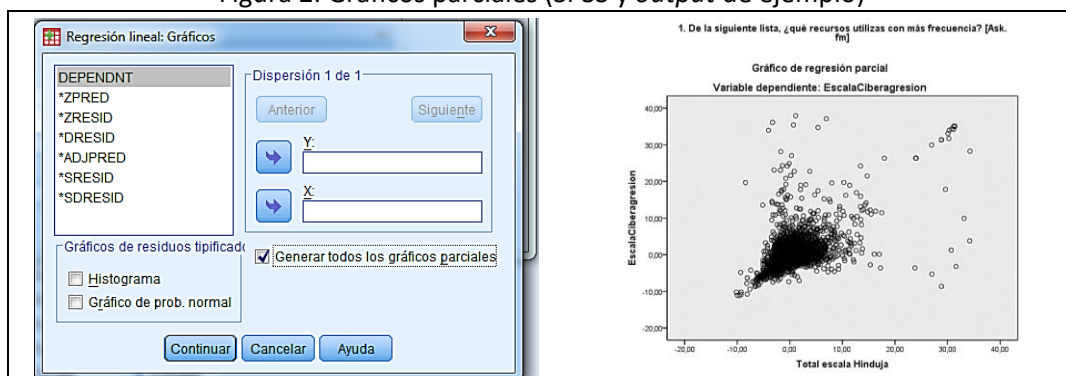
Figura 1. Valores pronosticados no tipificados y residuos no tipificados



Supuesto 1. Linealidad

El supuesto de la linealidad implica que la relación entre las variables sea lineal. Este supuesto lo verificamos en la opción de SPSS V. 22 Analizar → Regresión Lineal → Gráficos → Generar todos los gráficos parciales (figura 2). En los diagramas de dispersión parcial de cada variable independiente se ha eliminado el efecto proveniente de las otras variables. De esta manera, se muestra la relación neta entre las variables representadas (Martín Martín, 2008).

Figura 2. Gráficos parciales (SPSS y output de ejemplo)



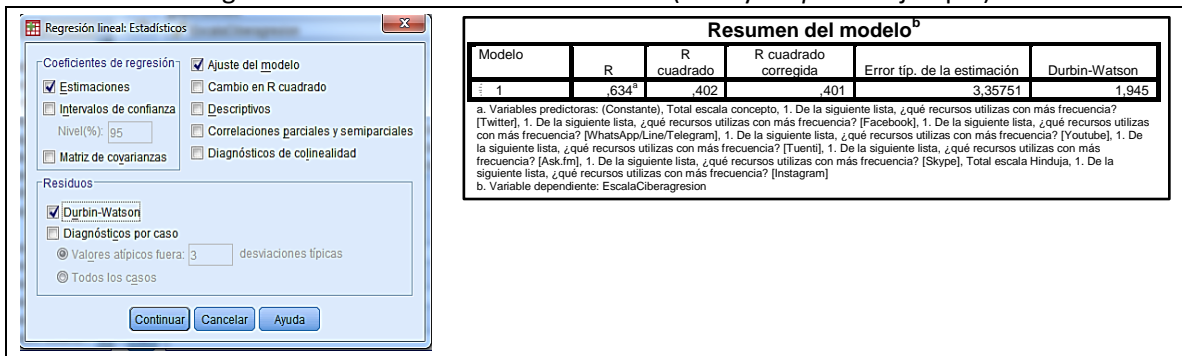
A modo de ejemplo, la figura 2 representa uno de los gráficos donde se puede interpretar la linealidad de relación entre las dos variables seleccionadas, verificando el supuesto de linealidad.

Supuesto 2. Independencia de los errores

El supuesto de la independencia de los errores implica que los errores en la medición de las variables explicativas sean independientes entre sí. Verificar este aspecto es especialmente importante en los estudios longitudinales. Ante diferentes observaciones de una misma variable en el tiempo, puede darse una cierta inercia en el sistema que hace que los valores sucesivos se aproximen entre sí (Camacho, 2006).

Este supuesto lo verificamos mediante el estadístico de Durbin-Watson, en la opción de SPSS, Analizar → Regresión Lineal → Estadísticos → Durbin-Watson (figura 3). Se consideran independientes si el estadístico se encuentra entre 1,5 y 2,5.

Figura 3. Estadístico de Durbin-Watson (SPSS y output de ejemplo)

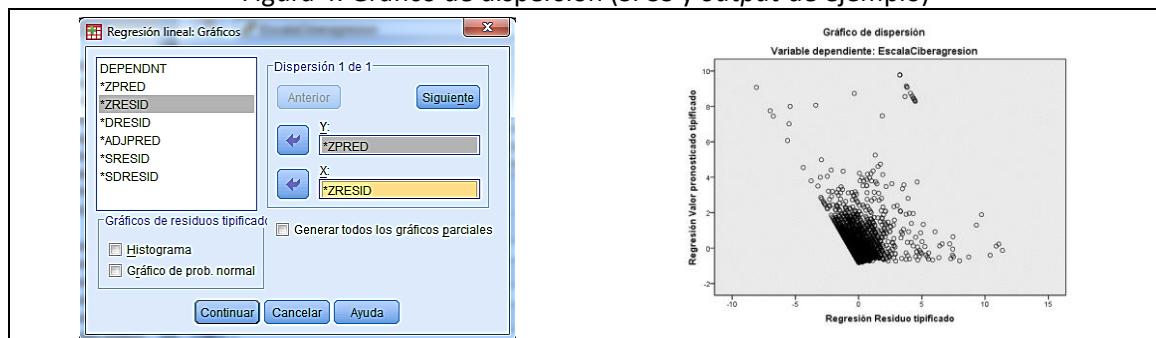


En la tabla de la figura 3 se representa el estadístico Durbin-Watson del *output* que ofrece SPSS en el ejemplo propuesto, verificando la independencia de los errores, puesto que 1,945 se encuentra entre 1,5 y 2,5.

Supuesto 3. Homocedasticidad

El supuesto de la homocedasticidad implica que los errores tengan varianza constante. Este supuesto lo verificamos mediante el estadístico de Levene o un gráfico de dispersión (ZPRED = pronósticos tipificados y ZRESID = residuos tipificados), que se obtiene siguiendo los pasos: Analizar → Regresión → Lineal-Gráficos (figura 4). El supuesto de homocedasticidad supone que la variación de los residuos es uniforme, es decir, que en el gráfico no se visualizan pautas de asociación (Martín Martín, 2008).

Figura 4. Gráfico de dispersión (SPSS y output de ejemplo)

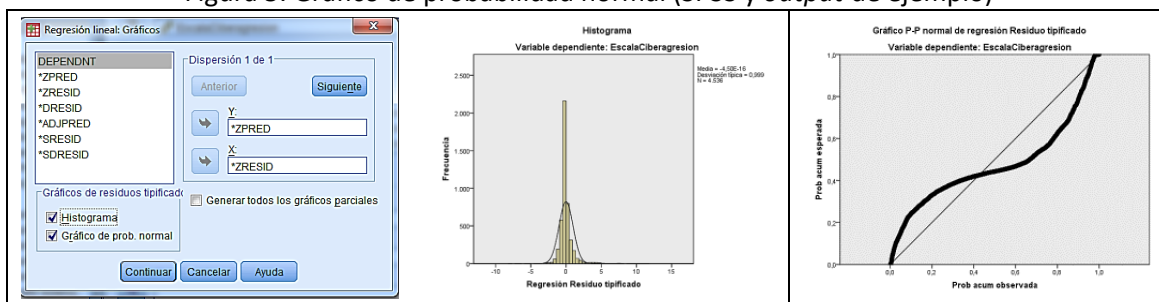




Supuesto 4. Normalidad

El supuesto de la normalidad implica que las variables siguen la ley normal. Este supuesto lo verificamos mediante la prueba de Kolmogorov-Smirnov o gráficamente en Analizar → Regresión → Lineal → Gráficos (figura 5). La opción Histograma añade una curva N (0,1) y el Gráfico de Probabilidad Normal representa las proporciones acumuladas de la variable esperada respecto a las proporciones acumuladas de la variable observada.

Figura 5. Gráfico de probabilidad normal (SPSS y output de ejemplo)



En los gráficos presentados en la figura 5 se intuye la normalidad de las variables del ejemplo. Para una aproximación más analítica podemos optar por la prueba de K-S (tabla 3).

Tabla 3. Prueba de Kolmogorov- Smirnov (output de ejemplo)

Prueba de Kolmogorov-Smirnov para una muestra		
		Unstandardized Residual
N		4536
Parámetros normales ^{a,b}	Media	,0000000
	Desviación típica	3,35380354
Diferencias más extremas	Absoluta	,135
	Positiva	,135
	Negativa	-,076
Z de Kolmogorov-Smirnov		,687
Sig. asintót. (bilateral)		,733

a. La distribución de contraste es la Normal.
b. Se han calculado a partir de los datos.

Como puede observarse, la significación asintótica bilateral es 0,733 > 0,05, se asume que la variable sigue la Ley Normal y que por lo tanto podemos continuar con el análisis de regresión lineal.

Supuesto 5. No colinealidad

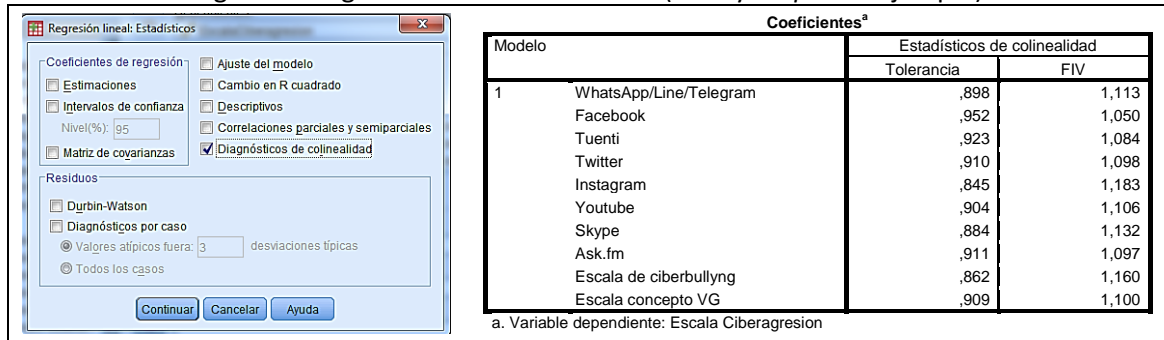
El supuesto de la no colinealidad implica que las variables independientes no estén correlacionadas entre ellas. Existe multicolinealidad entre las variables explicativas cuando existe algún tipo de dependencia lineal entre ellas, o lo que es lo mismo, si existe una fuerte correlación entre las mismas. La correlación no solamente se refiere a las distintas variables dos a dos, sino a cualquiera de ellas con cualquier grupo de las restantes.

El SPSS adopta varios procedimientos para detectar multicolinealidad entre las variables independientes (figura 6), como por ejemplo, la tolerancia y el factor de inflación de la varianza (FIV). La tolerancia está basada en la correlación múltiple de una determinada variable independiente con las restantes. Es un indicador de la variabilidad de cada variable independiente que no está explicada por las otras variables independientes en el modelo.

R. Vilà Baños, M. Torrado Fonseca, M. Reguant Álvarez. *Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico*

Para que no haya multicolinealidad la tolerancia debe ser alta: Una tolerancia menor de 0,10 diagnostica graves problemas de colinealidad. Por otra parte, el FIV es un indicador recíproco de la tolerancia. Cuanto menor sea este valor, menor será la multicolinealidad. Se estima que un valor FIV mayor de 10 diagnostica graves problemas de colinealidad.

Figura 6. Diagnóstico de la colinealidad (SPSS y *output* de ejemplo)

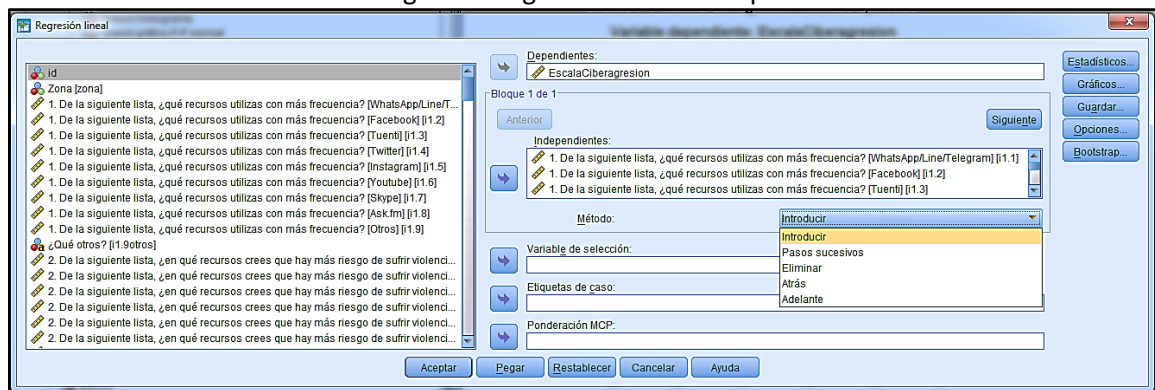


En nuestro ejemplo, la tabla representa el *output* que verifica el supuesto de no colinealidad entre las variables, como puede apreciarse en la columna de tolerancia, esta es > 0,10 en todos los casos. En la columna FIV los valores son < 10 en todos los casos.

Paso 4. Interpretar el modelo

Un modelo de regresión lineal múltiple se obtiene en SPSS desde el menú analizar → Regresión → Lineal (figura 7). Tal como se resume en la figura, en SPSS se ofrecen diferentes métodos para el análisis: introducir, pasos sucesivos, eliminar, atrás y adelante. Utilizaremos el método por defecto.

Figura 7. Regresión lineal múltiple



La tabla de coeficientes de regresión parcial contiene toda la información necesaria para construir la ecuación de regresión.

En la columna de coeficientes no estandarizados se encuentran los coeficientes B que forman parte de la ecuación en puntuaciones directas. Estos coeficientes no son independientes entre ellos, están calculados considerando la presencia del resto de variables independientes. Los coeficientes tipificados Beta se basan en puntuaciones típicas, así que pueden ser comparables entre ellos. Cuantifican el cambio que se produce en la variable dependiente por cada cambio en la variable independiente que corresponda, cuando el resto de las variables son constantes. Identifica la

importancia de cada variable independiente en la ecuación (De la Fuente Fernández, 2011). Así, las variables con un mayor coeficiente son las que tienen más peso.

Las columnas de pruebas *t* y sus niveles de significación sirven para identificar las variables significativas que contribuyen a explicar la variable dependiente (significación menor de 0,05).

Tabla 4. *Output* SPSS modelo de regresión lineal múltiple

Modelo	Coeficientes ^a				
	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error tip.	Beta		
1 (Constante)	14,523	,571		25,436	,000
WhatsApp/Line/Telegram	,082	,129	,008	,634	,526
Facebook	,140	,064	,026	2,191	,029
Tuenti	1,195	,201	,071	5,934	,000
Twitter	-,097	,066	-,018	-1,461	,144
Instagram	,141	,068	,026	2,088	,037
Youtube	,033	,096	,004	,348	,728
Skype	-,050	,081	-,008	-,614	,539
Ask.fm	,610	,088	,083	6,891	,000
Escala de cyberbullyng	,641	,014	,559	45,163	,000
Escala concepto VG	-,032	,004	-,087	-7,218	,000

a. Variable dependiente: Escala Ciberagresion

En el ejemplo presentado, el *output* de la tabla 4 indica las variables que se incluyen en el modelo, así como aquellas que son significativas (la significatividad es necesaria para que la variable sea considerada como predictora), en este caso, Facebook, Tuenti, Instagram, Ask.fm, escala de experiencia en cyberbullying, y escala de concepto de violencia de género. También nos ofrece información sobre qué variables tienen más peso en la ecuación (tercera columna de la tabla 4, coeficientes tipificados Beta), en nuestro ejemplo ordenadas de mayor a menor peso son: escala de cyberbullyng, escala de concepto de violencia de género, uso de Ask.fm, Tuenti, Facebook e Instagram. Finalmente, también nos indica la tendencia (positiva o negativa) de influencia en el modelo. Por ejemplo, la escala de concepto de violencia de género tiene una B negativa, que nos indica que los jóvenes con menor concepto sobre violencia de género tienen una mayor tendencia a ser ciberagresores.

En resumen, en el ejemplo podría interpretarse que los factores asociados a la ciberagresión en violencia de género son por orden de importancia: tener experiencia como ciberagresor en general, tener un concepto poco desarrollado de la violencia de género, y usar en gran medida plataformas como Ask.fm, Tuenti, Facebook e Instagram.

Paso 5. Bondad de ajuste del modelo

Desde el resumen del modelo, puede valorarse la varianza explicada desde la R^2 corregida. Así, un valor R^2 corregida = 0,4 significa que en el modelo queda explicado el 40% de la varianza. Como mayor sea este valor mejor se ajusta el modelo.

Por otra parte, desde el resumen de la ANOVA el estadístico F nos permite valorar si existe relación lineal significativa entre la variable dependiente y el conjunto de variables independientes del modelo. Especialmente, con la significación, se puede valorar si esta relación es significativa (menor de 0,05).

Tabla 5. *Outputs* SPSS resumen del modelo y ANOVA

Resumen del modelo ^b				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,634 ^a	,402	,401	3,35751

a. Variables predictoras: (Constante), Total escala concepto, Twitter, FacebookWhatsApp/Line/Telegram, Youtube, Tuenti, Ask.fm, Skype, escala de ciberbullyng, Instagram.
b. Variable dependiente: Escala Ciberagresión

ANOVA ^b						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	34318,231	10	3431,823	304,432	,000 ^a
	Residual	51009,672	4525	11,273		
	Total	85327,903	4535			

a. Variables predictoras: (Constante), Total escala concepto, Twitter, FacebookWhatsApp/Line/Telegram, Youtube, Tuenti, Ask.fm, Skype, escala de ciberbullyng, Instagram.
b. Variable dependiente: Escala Ciberagresión

En el ejemplo, los *outputs* de la tabla 5 nos indican que en el modelo definido sobre la ciberagresión explica el 40% de la varianza de la variable dependiente, de forma significativa, tal como puede observarse en la significación de la ANOVA.

<Referencias bibliográficas>

Camacho, C. (2006). *Regresión lineal simple*. Documento inédito. Recuperado de <http://personal.us.es/vararey/adatos2/Regsimple.pdf>

De la Fuente Fernández, S. (2011). *Regresión múltiple*. Documento inédito. Madrid: Universidad Autónoma de Madrid. Recuperado de http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/REGRE_MULTIPLE/regresion-multiple.pdf

Hernández, J., Ramírez, M. J., y Ferri, C. (2005). *Introducción a la minería de datos*. Madrid: Pearson.

Martín Martín, R. (2008). *Supuestos del modelo de regresión lineal*. Documento inédito. Ciudad Real: Universidad de Castilla-La Mancha. Recuperado de https://previa.uclm.es/profesorado/raulmartin/Estadistica/PracticasSPSS/Supuestos_del_modelo_de_regresion_lineal.pdf

Pérez, C. (2008). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson.

Pérez, C., y Santín, D. (2008). *Minería de datos. Técnicas y herramientas*. Madrid: Paraninfo.