

Grafos interactivos de regresión con modelos lineales generalizados

Modesto Escobar y Cristina Calvo

Universidad de Salamanca

modesto@usal.es cristinacalvolopez@usal.es

Este trabajo introduce una metodología innovadora en el análisis de datos dentro de la investigación social, destacando la aplicación de grafos en la representación de resultados de diferentes modelos de regresión. La propuesta central se centra en el uso de gráficos reticulares para una representación de múltiples relaciones entre variables, tanto cuantitativas como cualitativas. Este enfoque se complementa con un análisis crítico sobre métodos tradicionales, especialmente en lo que se refiere a la categoría base-contraste y a la relevancia de los márgenes y efectos marginales en los modelos estadísticos.

Palabras clave: grafos integrativos, análisis de coincidencias, análisis de regresión

1. INTRODUCCIÓN

En el campo de la investigación social, la aplicación de métodos cuantitativos y analíticos es crucial para la indagación de fenómenos complejos. Entre estos métodos, el análisis de regresión y los modelos de ecuaciones estructurales ocupan un lugar destacado. Este artículo propone una innovadora forma de representación gráfica de los resultados de una regresión mediante gráficos reticulares, con el objetivo de mejorar el análisis y la interpretación y de los datos en estudios sociológicos.

Los grafos no solo han sido empleados para solucionar problemas topográficos y para representar estructuras sociales, sino también para estudiar relaciones entre variables. Son bien conocidos los análisis de senderos (path analysis) y los modelos de ecuaciones estructurales. Ambos, sin embargo, estaban inicialmente restringidos al uso de variables cuantitativas. En este artículo, se aborda cómo pueden también representarse las relaciones entre variables cualitativas, tal como ya lo hace el análisis de correspondencias, pero empleando los recursos técnicos del análisis de redes y de otras conocidas técnicas multivariantes como la regresión lineal y la logística.

El núcleo de esta presentación consiste en dos propuestas concretas: primero, recordar un modo alternativo de mostrar los coeficientes de las regresiones, que permite una interpretación más intuitiva sobre todo cuando se trabaja con modelos no lineales. Segundo, enseñar unos gráficos reticulares abiertos e interactivos que representen al mismo tiempo varias regresiones con modelos lineales generalizados.

2. METODOLOGÍA

Como aplicación práctica de esta metodología, se expondrá el resultado de aplicar un análisis de contenido a casi 90.000 respuestas emitidas por unas 5.800 personas de cinco países diferentes, quienes respondieron a la pregunta "¿quién soy yo?" mediante el TST (Kuhn y McPartland, 1954). De sus respuestas se entresacaron todas las referencias explícitas en ellas clasificadas en Este caso de estudio no solo ilustra la aplicabilidad y eficacia de los gráficos reticulares propuestos, sino que también destaca la capacidad de estos métodos para manejar y analizar simultáneamente datos complejos,

La metodología empleada en este trabajo integra varios componentes claves para abordar la complejidad inherente a la investigación social, utilizando una combinación de técnicas

cuantitativas avanzadas. La aplicación de grafos en la investigación social, el uso de regresiones, y la integración de ambos en una representación gráfica innovadora, constituyen el núcleo de este enfoque metodológico.

3. RESULTADOS

Los modelos de regresión lineal múltiple examinados en la Tabla 1 evidencian una estructura compleja de relaciones entre un conjunto de variables independientes –género, edad, nivel educativo, ocupación y hábitat– y una serie de variables dependientes que son el número de referencias mencionadas clasificadas en áreas como genéricas, biosociales, físicas, entre otras. (Escobar, 2019). Los coeficientes de regresión, acompañados de sus respectivos errores típicos y niveles de significación estadística, son cruciales para interpretar la dinámica de estas relaciones.

TABLA 1. REGRESIÓN LINEAL MÚLTIPLE SOBRE EL USO DE REFERENCIAS

| | Genéricas | Biosociales | Físicas | Familiares | Grupales | Activas | Sociales | Primarias | Secundarias | Laborales | Educativas | Lúdicas |
|---------------------------------|----------------------|-----------------------|-----------------------|----------------------|-----------------------|----------------------|-----------------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|
| Género | | | | | | | | | | | | |
| Mujer | -0,110 (0,065) | 0,243 *** (0,052) | 0,076 * (0,031) | 0,771 *** (0,053) | 0,127 *** (0,028) | -0,121 (0,078) | -0,176 *** (0,041) | 0,918 *** (0,061) | -0,080 (0,080) | 0,084 (0,048) | 0,031 * (0,014) | -0,236 *** (0,047) |
| Edad | | | | | | | | | | | | |
| 30-39 | -0,167 (0,091) | -0,319 *** (0,073) | -0,185 *** (0,044) | 0,462 *** (0,074) | 0,188 *** (0,039) | -0,158 (0,110) | -0,011 (0,057) | 0,581 *** (0,086) | -0,023 (0,113) | -0,004 (0,068) | -0,076 *** (0,020) | -0,078 (0,066) |
| 40-49 | -0,156 (0,100) | -0,477 *** (0,080) | -0,286 *** (0,048) | 0,680 *** (0,081) | 0,227 *** (0,043) | -0,061 (0,121) | 0,072 (0,062) | 0,822 *** (0,094) | 0,084 (0,123) | 0,040 (0,074) | -0,113 *** (0,021) | 0,012 (0,072) |
| 50-59 | -0,241 * (0,110) | -0,310 *** (0,088) | -0,215 *** (0,053) | 0,860 *** (0,090) | 0,297 *** (0,047) | 0,145 (0,133) | 0,283 *** (0,069) | 1,042 (0,104) | 0,507 *** (0,136) | 0,209 * (0,082) | -0,100 (0,024) | 0,036 (0,080) |
| 60-70 | -0,182 (0,142) | -0,205 (0,113) | -0,288 *** (0,069) | 1,175 *** (0,115) | 0,383 *** (0,061) | 0,952 *** (0,172) | 0,287 ** (0,089) | 1,401 *** (0,133) | 1,260 *** (0,176) | 0,899 *** (0,105) | -0,081 ** (0,031) | 0,134 (0,103) |
| Estudios | | | | | | | | | | | | |
| Secundarios | 0,200 * (0,089) | 0,132 (0,071) | 0,054 (0,043) | 0,196 ** (0,073) | 0,067 (0,038) | 0,401 *** (0,108) | 0,126 * (0,056) | 0,232 ** (0,084) | 0,549 *** (0,111) | 0,338 *** (0,066) | 0,054 ** (0,019) | 0,009 (0,065) |
| Universitarios | 0,152 (0,092) | 0,049 (0,074) | -0,048 (0,045) | 0,154 * (0,075) | 0,091 * (0,040) | 0,688 *** (0,111) | 0,247 *** (0,058) | 0,171 * (0,087) | 0,910 *** (0,114) | 0,446 *** (0,068) | 0,145 *** (0,020) | 0,097 (0,067) |
| Ocupación | | | | | | | | | | | | |
| Empresario, profesional y otros | 0,135 (0,119) | 0,087 (0,095) | -0,031 (0,058) | 0,032 (0,097) | 0,005 (0,051) | 0,330 * (0,144) | 0,114 (0,075) | 0,003 (0,112) | 0,289 * (0,148) | 0,168 (0,089) | -0,027 (0,026) | 0,189 * (0,087) |
| Desempleado | -0,063 (0,095) | 0,070 (0,076) | 0,001 (0,046) | 0,079 (0,077) | 0,038 (0,041) | 0,026 (0,115) | 0,042 (0,059) | 0,104 (0,089) | 0,043 (0,118) | -0,068 (0,071) | 0,055 ** (0,020) | 0,038 (0,069) |
| Estudiante | 0,469 *** (0,118) | 0,215 * (0,094) | 0,039 (0,057) | -0,156 (0,096) | 0,188 *** (0,051) | 0,627 *** (0,143) | 0,315 *** (0,074) | -0,113 (0,111) | 0,709 *** (0,146) | -0,212 * (0,088) | 0,462 *** (0,025) | 0,377 *** (0,086) |
| Trabajo en casa | -0,353 ** (0,126) | -0,014 (0,100) | 0,042 (0,061) | 0,660 *** (0,102) | -0,073 (0,054) | -0,065 (0,152) | -0,225 ** (0,079) | 0,641 *** (0,118) | -0,313 * (0,156) | -0,003 (0,093) | -0,032 (0,027) | -0,030 (0,091) |
| Jubilado | -0,078 (0,130) | 0,317 (0,104) | 0,021 (0,063) | -0,086 (0,106) | -0,096 (0,056) | 0,154 (0,157) | 0,167 * (0,081) | -0,206 (0,122) | 0,273 (0,161) | 0,061 (0,097) | 0,021 (0,028) | 0,072 (0,094) |
| Hábitat | | | | | | | | | | | | |
| Pequeña ciudad | -0,109 (0,071) | 0,105 (0,057) | 0,157 *** (0,035) | -0,096 (0,058) | -0,113 *** (0,031) | -0,178 * (0,086) | 0,047 (0,045) | -0,156 * (0,067) | 0,109 (0,088) | 0,135 * (0,053) | -0,020 (0,015) | -0,293 *** (0,052) |
| Gran ciudad | -0,047 (0,085) | 0,168 * (0,068) | 0,120 ** (0,041) | 0,021 (0,069) | 0,051 (0,037) | 0,045 (0,103) | 0,198 *** (0,053) | 0,011 (0,080) | 0,413 *** (0,106) | 0,156 * (0,063) | -0,003 (0,018) | -0,108 (0,062) |
| Constante | 1,144 *** (0,118) | 1,614 *** (0,094) | 0,743 *** (0,057) | 0,680 *** (0,096) | 0,494 *** (0,051) | 2,272 *** (0,143) | 0,820 *** (0,074) | 0,993 *** (0,111) | 1,994 *** (0,146) | 0,854 *** (0,088) | 0,141 *** (0,025) | 1,278 *** (0,086) |
| n | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 |
| R2 ajustado | 0,01 | 0,02 | 0,02 | 0,09 | 0,02 | 0,02 | 0,02 | 0,08 | 0,04 | 0,04 | 0,11 | 0,01 |

Fuente: Elaboración propia. *** $p < 0,001$, ** $p < 0,01$, * $p < 1$.

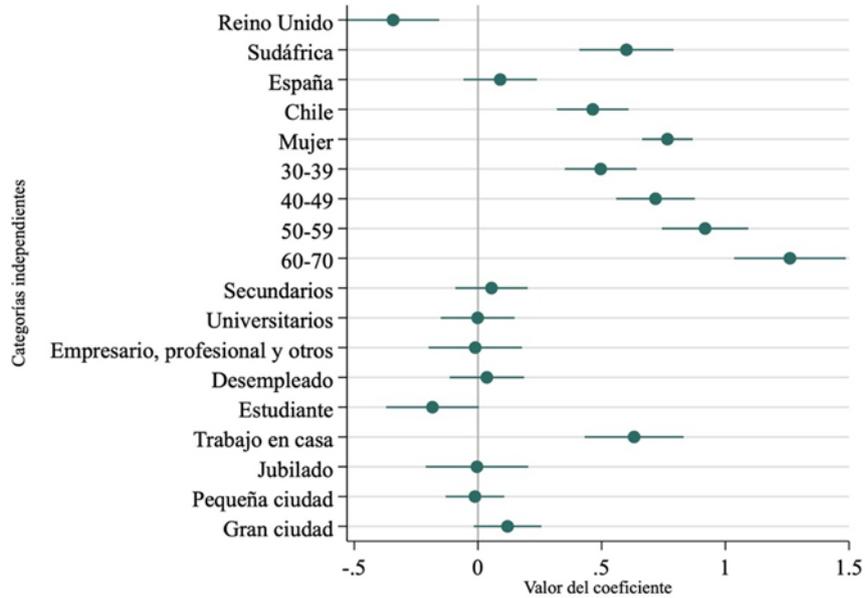
Se observa que la categoría de género, específicamente ser mujer, presenta asociaciones tanto positivas como negativas significativas en diversas categorías dependientes. El efecto de la edad, especialmente en el grupo de 50-59 años, muestra una relación positiva fuerte en la categoría física (0,86), lo que podría alinearse con estudios que indican un cambio en las prioridades y comportamientos con la edad. También se detecta relación entre nivel educativo y ciertas variables dependientes. Asimismo la ocupación y el hábitat también muestran relaciones notables. Entre ellas, el hecho de ser estudiante se asocia positivamente con la mayor parte de referencias (Revilla, 2018).

Existen múltiples formas de representar gráficamente los resultados de una regresión. Una de las opciones más habitualmente usadas son los gráficos de coeficientes, como el que se presenta en la Figura 1. Un gráfico de coeficientes, comúnmente conocido como *coefplot* (Jann, 2014), es una herramienta visual utilizada para representar los efectos estimados de diferentes variables independientes sobre una variable dependiente, basada en un modelo de regresión. Cada punto en el gráfico corresponde al coeficiente estimado de una variable, y las líneas horizontales representan los intervalos de confianza al 95% para cada coeficiente.

Como no es posible combinar más de una regresión en un gráfico de coeficientes habitual, se ha seleccionado uno de los tipos de referencias posibles: las familiares, que se emiten en aquellas ocasiones en las que los sujetos mencionan a su familia para definirse a sí mismos (“soy

madre”, “tengo dos nietos”). En el gráfico, las variables independientes se presentan en el eje vertical y los coeficientes en el eje horizontal.

FIGURA 1. GRÁFICO DE COEFICIENTES SOBRE LAS REFERENCIAS FAMILIARES



Fuente: Elaboración propia.

Otro modo de presentar los modelos de regresión es mediante los efectos marginales, que son la diferencia en la predicción media de la variable dependiente que resulta de un cambio unitario en la variable independiente, manteniendo constantes las demás variables en el modelo. Estos efectos ofrecen una interpretación directa del impacto que tiene cambiar una variable independiente sobre la variable dependiente. La utilidad de los efectos marginales radica en su capacidad para proporcionar información sobre la influencia que ejerce una variable en la presencia de otras variables en el modelo (Bartus, 2005).

Una de las ventajas que tiene el uso de los efectos es el no estar anclados a una constante o categoría de referencia, lo que permite que cada categoría de una variable se evalúe en términos de su efecto absoluto en lugar de su efecto relativo a otra categoría. Esto simplifica la comparación entre categorías al eliminar la necesidad de interpretar los coeficientes en relación con una categoría base, que puede no ser intuitiva para todos los usuarios, especialmente para aquellos menos familiarizados con el análisis estadístico (Williams, 2012).

Además, al comparar siempre con la media de la muestra, los efectos marginales proporcionan una comprensión más clara del efecto de las variables independientes. Esto contrasta con las tablas de regresión convencionales, donde los coeficientes de las variables categóricas se interpretan en relación con una categoría omitida, lo que puede dificultar la interpretación cuando el interés reside en la magnitud del efecto de cada categoría en particular.

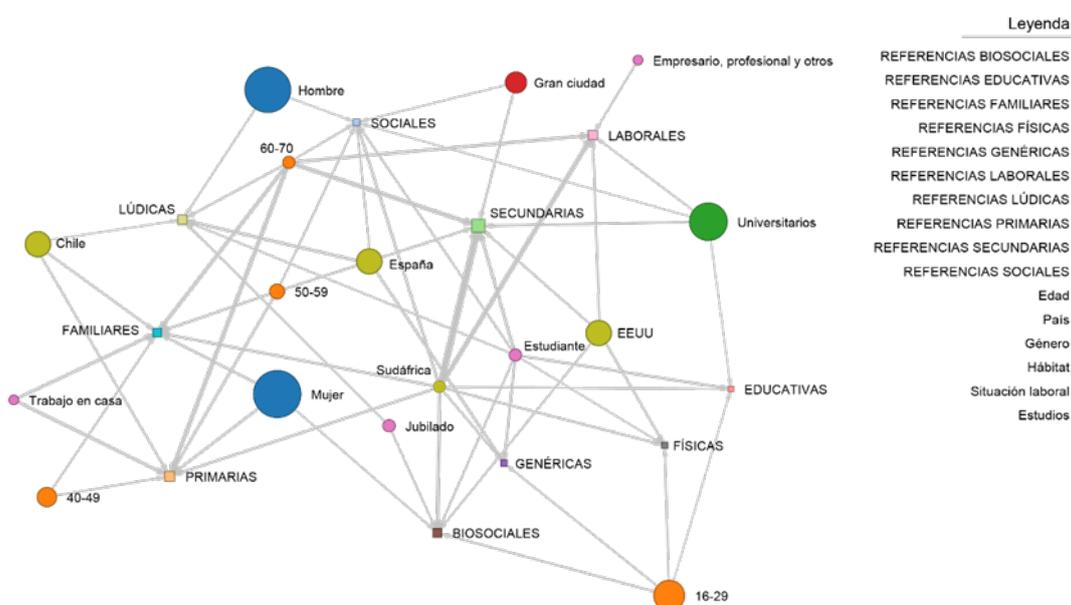
TABLA 2. EFECTOS MARGINALES EN LAS REFERENCIAS

| | Genéricas | Biosociales | Físicas | Familiares | Grupales | Activas | Sociales | Primarias | Secundarias | Laborales | Educativas | Lúdicas |
|---------------------------------|-----------|-------------|----------|------------|----------|----------|----------|-----------|-------------|-----------|------------|----------|
| Pais | | | | | | | | | | | | |
| EEUU | -0,4 *** | 0,2 *** | 0,3 *** | -0,2 ** | -0,2 *** | -0,4 *** | 0,0 | -0,3 *** | 0,2 * | 0,2 *** | 0,0 | -0,6 *** |
| Reino Unido | -0,4 *** | -0,1 | 0,0 | -0,5 *** | -0,2 *** | -0,5 *** | -0,2 ** | -0,6 *** | -0,3 * | 0,0 | -0,1 * | -0,5 *** |
| Sudáfrica | 0,3 ** | 0,6 *** | 0,3 *** | 0,4 *** | 0,2 *** | 1,2 *** | 0,5 *** | 0,4 *** | 1,8 *** | 1,0 *** | 0,1 *** | 0,1 |
| España | 0,5 *** | -0,3 *** | -0,3 *** | -0,1 | 0,3 *** | 0,3 *** | 0,2 *** | 0,1 | 0,1 | -0,3 *** | 0,0 | 0,6 *** |
| Chile | -0,1 | -0,1 * | -0,2 *** | 0,3 *** | -0,1 *** | -0,3 *** | -0,3 *** | 0,3 *** | -0,8 *** | -0,3 *** | -0,1 *** | 0,1 * |
| Género | | | | | | | | | | | | |
| Hombre | 0,1 | -0,1 ** | 0,0 | -0,4 *** | -0,1 *** | 0,1 | 0,1 ** | -0,5 *** | 0,0 | 0,0 | 0,0 | 0,1 *** |
| Mujer | 0,0 | 0,1 ** | 0,0 | 0,4 *** | 0,1 *** | -0,1 | -0,1 ** | 0,5 *** | 0,0 | 0,0 | 0,0 | -0,1 *** |
| Edad | | | | | | | | | | | | |
| 16-29 | 0,1 * | 0,2 *** | 0,2 *** | -0,5 *** | -0,2 *** | -0,1 | -0,1 | -0,6 *** | -0,2 * | -0,1 ** | 0,1 *** | 0,0 |
| 30-39 | -0,1 | -0,1 | 0,0 | 0,0 | 0,0 | -0,3 *** | -0,1 ** | 0,0 | -0,3 *** | -0,1 ** | 0,0 | -0,1 * |
| 40-49 | -0,1 | -0,2 ** | -0,1 | 0,2 *** | 0,0 | -0,2 | 0,0 | 0,2 ** | -0,1 | 0,0 | 0,0 ** | -0,1 * |
| 50-59 | -0,1 | -0,1 | -0,1 * | 0,4 *** | 0,1 *** | 0,1 | 0,2 *** | 0,5 *** | 0,3 ** | 0,1 | 0,0 * | 0,1 |
| 60-70 | 0,0 | 0,0 | -0,2 *** | 0,7 *** | 0,2 *** | 0,9 *** | 0,2 * | 0,9 *** | 0,9 *** | 0,7 *** | 0,0 | 0,3 ** |
| Estudios | | | | | | | | | | | | |
| Primarios | -0,1 | 0,0 | 0,0 | 0,0 | -0,1 | -0,3 *** | -0,1 * | 0,0 | -0,5 *** | -0,2 *** | -0,1 *** | 0,0 |
| Secundarios | 0,1 | 0,0 | 0,0 | 0,0 | 0,0 | -0,1 | 0,0 | 0,1 | -0,1 | 0,0 | 0,0 ** | 0,0 |
| Universitarios | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,2 *** | 0,1 ** | 0,0 | 0,3 *** | 0,1 ** | 0,1 *** | 0,0 |
| Ocupación | | | | | | | | | | | | |
| Empleado | 0,0 | -0,1 * | 0,0 | 0,0 | 0,0 | -0,1 * | -0,1 | 0,0 | -0,1 * | 0,0 | -0,1 *** | -0,1 |
| Empresario, profesional y otros | 0,1 | 0,0 | 0,0 | 0,0 | 0,0 | 0,2 | 0,1 | -0,1 | 0,2 | 0,2 * | -0,1 *** | 0,1 |
| Desempleado | -0,2 | 0,0 | 0,0 | 0,0 | 0,0 | -0,2 | 0,0 | 0,0 | -0,1 | 0,0 | 0,0 | -0,1 * |
| Estudiante | 0,3 ** | 0,2 * | 0,1 * | -0,2 * | 0,1 * | 0,4 ** | 0,2 *** | -0,2 * | 0,6 *** | -0,1 | 0,4 *** | 0,2 * |
| Trabajo en casa | -0,2 * | -0,1 | 0,0 | 0,6 *** | 0,0 | 0,0 | -0,2 ** | 0,6 *** | -0,3 | 0,0 | -0,1 ** | 0,0 |
| Jubilado | 0,0 | 0,2 * | -0,1 | 0,0 | 0,0 | 0,2 | 0,1 | -0,1 | 0,1 | 0,0 | 0,0 | 0,2 * |
| Hábitat | | | | | | | | | | | | |
| Pueblo | -0,1 | -0,1 | 0,0 | 0,0 | 0,0 | -0,1 | -0,1 ** | 0,0 | -0,2 ** | -0,1 | 0,0 | 0,0 |
| Pequeña ciudad | 0,1 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| Gran ciudad | 0,0 | 0,1 | 0,0 | 0,1 | 0,1 ** | 0,1 | 0,1 ** | 0,1 | 0,3 ** | 0,1 | 0,0 | 0,0 |
| n | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 | 5808 |

Fuente: Elaboración propia. *** p<,0001, ** p<,0,01, * p<,1.

Usando los efectos marginales positivos puede construirse un gráfico reticular de regresión como el presentado en la Figura 2, donde se visualiza los efectos marginales presentados en la Tabla 2 y derivados de la regresión múltiple lineal de la Tabla 1. Este grafo proporciona una representación visual de los efectos marginales asociados a las variables independientes utilizadas. En, los nodos simbolizan las variables dependientes– representadas con los nodos cuadrados y la leyenda en mayúscula- y las categorías de cada una de las variables independientes representadas con nodos redondos. La presencia de un enlace entre dos nodos denota un efecto positivo y significativo de una variable o categoría sobre otra, cuya dirección marca la flecha. Además, los colores de los nodos corresponden a las diferentes variables, según se indica en la leyenda del grafo.

FIGURA 2. GRÁFICO RETICULAR DE REGRESIÓN



Fuente: Elaboración propia.

Una característica significativa del grafo es su capacidad para mostrar las relaciones positivas significativas entre las variables dentro de un solo marco visual. La utilización de enlaces solo cuando existen valores positivos simplifica la interpretación al facilitar la identificación de relaciones de influencia directa, sin la distracción de efectos negativos o neutros. El grosor de los enlaces sirve como un indicador visual intuitivo del tamaño del efecto: los enlaces más gruesos corresponden a efectos más elevados. Esta característica permite a los observadores discernir rápidamente la importancia relativa de las diferentes influencias en más de un modelo al mismo tiempo.

4. BIBLIOGRAFÍA

Bartus, T. (2005): Estimation of marginal effects using `margeff`. *The Stata Journal*, 5(3), 309-329.

Escobar, M. (2018): "Tesoro de la Identidad". En *La interacción social. Escritos en homenaje a José Ramón Torregrosa*. Madrid: Centro de Investigaciones Sociales (CIS).

Jann, B. (2014). Plotting regression coefficients and other estimates. *The Stata Journal* 14(4): 708-737

Kuhn, M. H. Y T. McPartland (1954): "An Empirical Investigation of Self-Attitudes". En *American Sociological Review*, 19, 68-76.

Revilla, J. C. (1998): *La identidad personal de los jóvenes: pluralidad y autenticidad*. Madrid: Entinema.

Williams, R. (2012): Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects. *Stata Journal*, 12(2), 308-331.