

Gráficos interactivos con redes lineales y logísticas.

Modesto Escobar y Luis Martínez

(modesto@usal.es & lmartinez@march.es)

Universidad de Salamanca
Fundación Juan March

XI JORNADAS DE USUARIOS DE R

Madrid, 15 de noviembre de 2019

Contenido

1 Introducción

- Grafos
- Los grafos en la investigación social

2 CARIG

- Datos
- Coincidencias
- Grados
- Regresiones

3 Programas

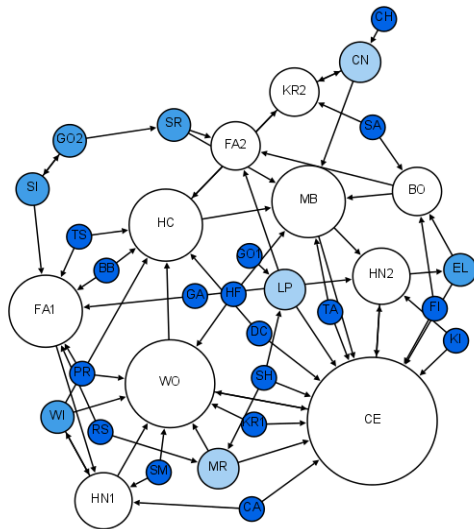
4 Propuestas

Grafos en la investigación social

Clasificación

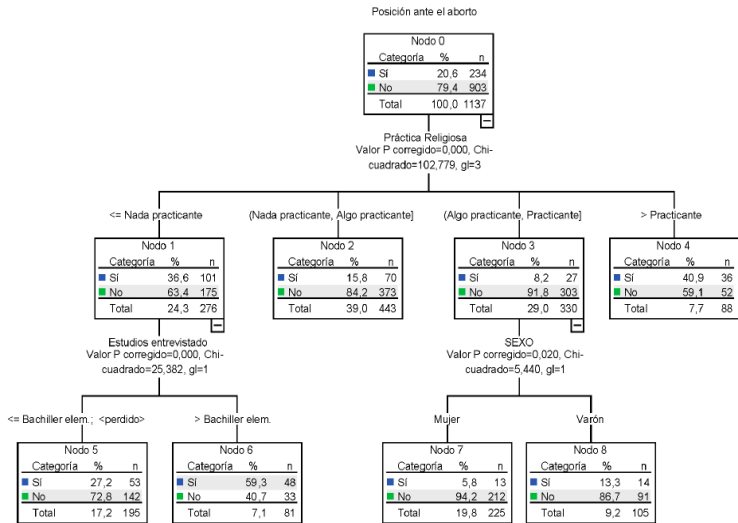
- Grafos (Análisis de redes sociales)
- Árboles (Árboles de clasificación y regresión)
- Senderos (Análisis de senderos)
- Modelos causales y de medida (Modelos de ecuaciones estructurales)
- Grafos reticulares de clasificación y regresión. (NCARG)

Sociograma de Moreno en una clase de 2º grado. (Grandjean 2015)



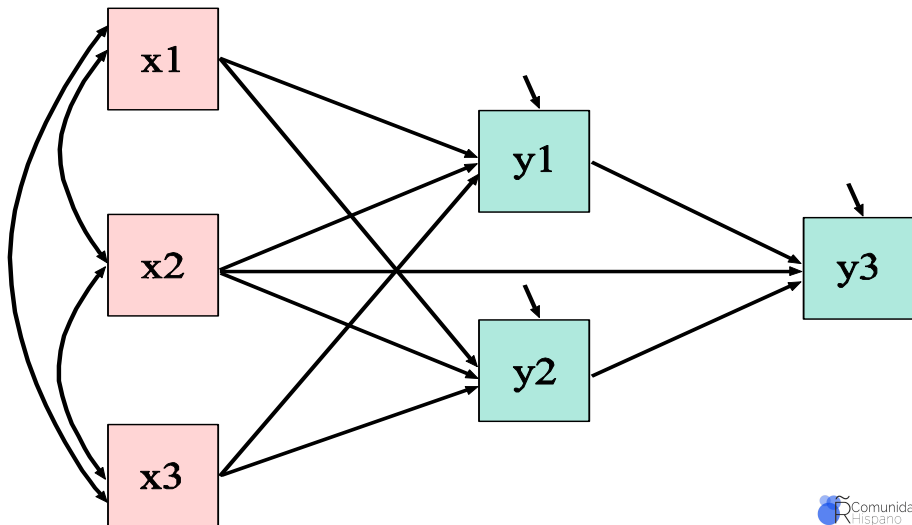
CART

Necesidad percibida de inmigrantes (Fuente: CIS-2511)



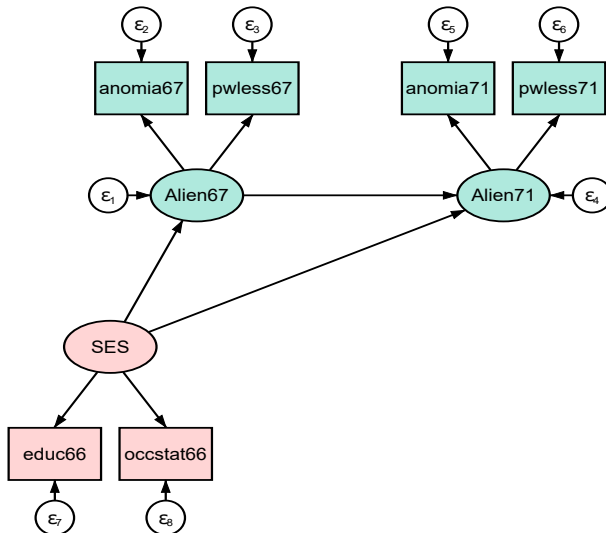
Análisis de senderos

Análisis de senderos con variables continuas (Muthén & Muthén 2017,32)



SEM

Ejemplo clásico de la alienación (Wheaton et al. 1977,84-136)



Contenido

1 Introducción

- Grafos
- Los grafos en la investigación social

2 CARIG

- Datos
- Coincidencias
- Grados
- Regresiones

3 Programas

4 Propuestas

Grafos interactivos de coincidencias y regresión

CARIG

- Mi **propuesta** es emplear grafos para la representación gráfica, exploratoria e interactiva de un conjunto de regresiones al estilo de los modelos de ecuaciones estructurales.
- El principal **escollo** es la representación de las variable categóricas, que paso a reseñar brevemente a continuación.
- En primer lugar, hablaré de la representación de las **coincidencias** (estadística bivariada).
- Pera pasar, a continuación, a hablar de la representación de modelos de **regresión lineal generalizada** (estadística multivariada).

Datos

Fuente: CIS-3.244 (2019)

- Vamos a ver los grafos reticulares de coincidencias y regresión con datos de un superbarómetro del CIS.
- Por simplicidad se usarán solo 3 variables: voto, ideología y religión.
- El voto se considerará una variable endógena, mientras que ideología y religión serán tratadas como exógenas.

Distribución de frecuencias

Voto, ideología y religión (sin ponderar)

Distribución de frecuencias

Categoría	%
Ideología:1-3	26.09
Ideología:4	14.05
Ideología:5	20.33
Ideología:6	9.38
Ideología:7	6.99
Ideología:8-10	9.46
Ideología:NC	4.35
Ideología:NS	9.35
Religión:Católico/a no practicante	48.62
Religión:Católico/a practicante	22.74
Religión:No creyente	24.41

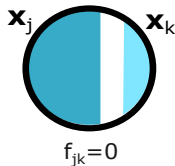
Categoría	%
Voto:VOX	4.42
Voto:PP	10.65
Voto:Ciudadanos	7.84
Voto:PSOE	21.20
Voto:Podemos	7.08
Voto:Otro	10.05
Voto:Indeciso	29.58
Voto:No votará	9.18

Representación de las coincidencias

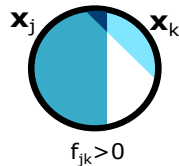
Nula, mera, probable, subtotal, total

Coincidencias

Nula



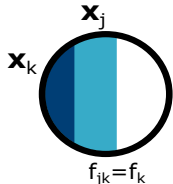
Mera



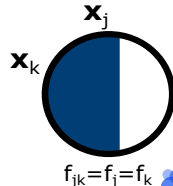
Probable



Subtotal



Total



Leyenda

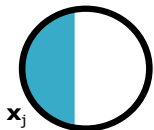


Representación de las coincidencias

Independencia y coincidencia condicional

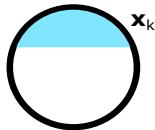
Coincidencias condicionales

Marginal de \mathbf{x}_j



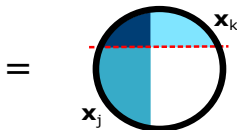
$$f_{jj} < 50\%$$

Marginal de \mathbf{x}_k



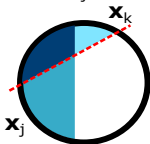
$$f_{kk} < 50\%$$

Independencia de \mathbf{x}_j y \mathbf{x}_k



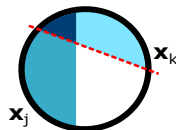
$$f_{jk} = f_{jj}f_{kk}/n$$

Coincidencia condicional
de \mathbf{x}_j y \mathbf{x}_k



$$f_{jk} > f_{jj}f_{kk}/n$$

No coincidencia condicional
de \mathbf{x}_j y \mathbf{x}_k



$$f_{jk} < f_{jj}f_{kk}/n$$

Leyenda



De las coincidencias a las adyacencias

Ejemplos de distintos grados de coincidencias

De las coincidencias a las adyacencias

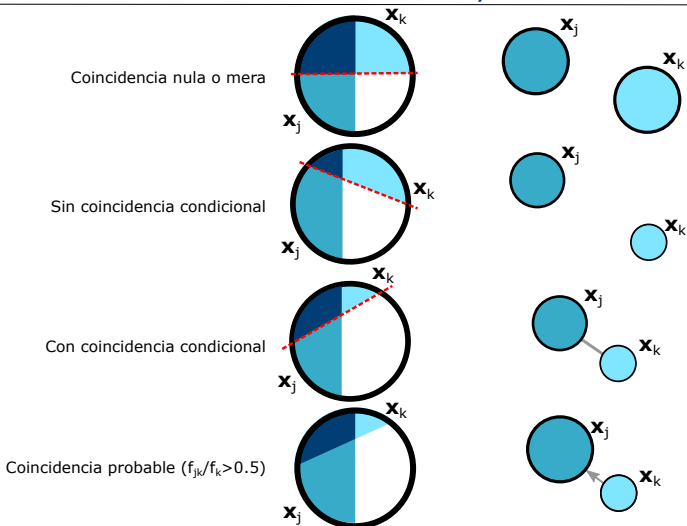
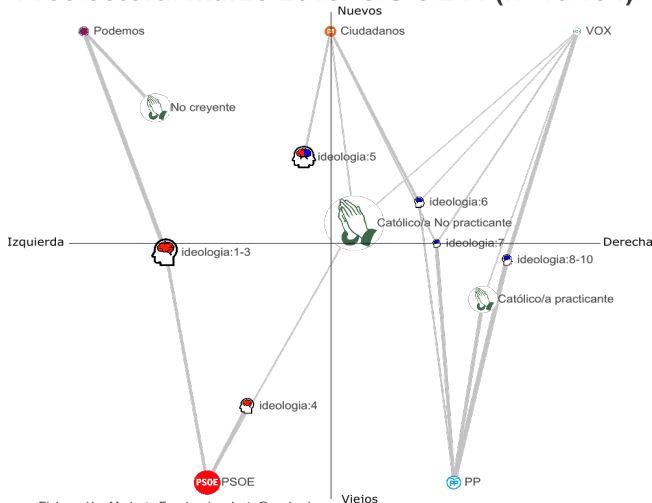


Ilustración de coincidencias

Coincidencias condicionales significativas ($p \leq 0.05$)

Preelectoral marzo 2019. CIS-3.244 (n=16.194).



De las coincidencias a las regresiones

Con el mismo ejemplo

- Ahora se distinguirá entre variables dependientes e independientes.
- Como en este ejemplo solo se está trabajando con variables cualitativas, las categorías endógenas se convierten en variables dicotómicas para ser incluidas en regresiones logísticas como variables dependientes.
- Tantas como categorías dependientes consideremos de relevancia. En este caso, solo se emplean los cinco partidos más votados.
- Como variables independientes pueden introducirse variables cuantitativas o cualitativas(éstas como indicadores dicotómicos)
 - Religión
 - Ideología
- De este modo, se trabajaría con cinco regresiones logísticas, cada una de ellas con dos variables independientes de 4 y 8 categorías respectivamente.

Regresión logística del voto al PP

Sobre religión e ideología

Regresión del voto al PP sobre religión e ideología

	Estimador	Error típ.	valor z	$\Pr(> z)$
(Constante)	-5.4423	0.4123	-13.20	0.0000
Católico/a no practicante	-0.8156	0.0670	-12.17	0.0000
No creyente	-1.5612	0.1392	-11.21	0.0000
Otra religión	-1.2133	0.2091	-5.80	0.0000
Ideología: 4	1.5425	0.4728	3.26	0.0011
Ideología: 5	2.9734	0.4202	7.08	0.0000
Ideología: 6	4.6218	0.4157	11.12	0.0000
Ideología: 7	5.4992	0.4152	13.25	0.0000
Ideología: >7	5.6725	0.4138	13.71	0.0000
Ideología: NS	3.6774	0.4218	8.72	0.0000
Ideología: NC	2.7713	0.4688	5.91	0.0000

Problemas con las regresiones

- Los contrastes son una herramienta flexible para comprobar los efectos de los predictores nominales en los análisis de varianza y regresión.
- En las regresiones habituales los contrastes presentan estas características
 - Se contrasta en relación con otra categoría.
 - Generalmente se realizan con la primera categoría (la que arbitrariamente presenta el valor menor).
- Sin embargo, hay otro tipo de contrastes
 - Contrastes adyacentes (invertidos o no).
 - Diferencias en relación con la media de niveles previos o posteriores.
 - **Diferencias en relación con la media global (ponderada o sin ponderar)**
 - Contrates polinómicos ortogonales.

Regresión alternativa del voto al PP (1^{er} paso)

Cambiar la base de los predictores nominales: menor coeficiente

Regresión alternativa del voto al PP -cambio de base-

	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-7.0036	0.4209	-16.64	0.0000
Católico/a practicante	1.5612	0.1392	11.21	0.0000
Católico/a no practicante	0.7456	0.1387	5.37	0.0000
Otra respuesta	0.3479	0.2417	1.44	0.1500
Ideología: 4	1.5425	0.4728	3.26	0.0011
Ideología: 5	2.9734	0.4202	7.08	0.0000
Ideología: 6	4.6218	0.4157	11.12	0.0000
Ideología: 7	5.4992	0.4152	13.25	0.0000
Ideología: 8-10	5.6725	0.4138	13.71	0.0000
Ideología: NS	3.6774	0.4218	8.72	0.0000
Ideología: NC	2.7713	0.4688	5.91	0.0000

Regresión alternativa del voto al PP (2° paso)

Aplicar el contraste de la media ponderada

Regresión alternativa del voto al PP -cambio de contraste-

	Estimador	Error típ.	valor z	$\Pr(> z)$
(Constante)	-2.9951	0.0857	-34.94	0.0000
Católico/a practicante	0.8290	0.0547	15.17	0.0000
Católico/a no practicante	0.0134	0.0423	0.32	0.7510
Otra respuesta	-0.3843	0.1991	-1.93	0.0537
Ideología: 4	-1.0975	0.2314	-4.74	0.0000
Ideología: 5	0.3334	0.1359	2.45	0.0142
Ideología: 6	1.9818	0.1315	15.07	0.0000
Ideología: 7	2.8592	0.1310	21.82	0.0000
Ideología: 8-10	3.0325	0.1262	24.02	0.0000
Ideología: NS	1.0374	0.1465	7.08	0.0000
Ideología: NC	0.1313	0.2467	0.53	0.5945

Regresión alternativa del voto al PP (3^{er} paso)

Eliminar las categorías con coeficiente negativo

Regresión alternativa del voto al PP -sin estimadores no contribuyentes-

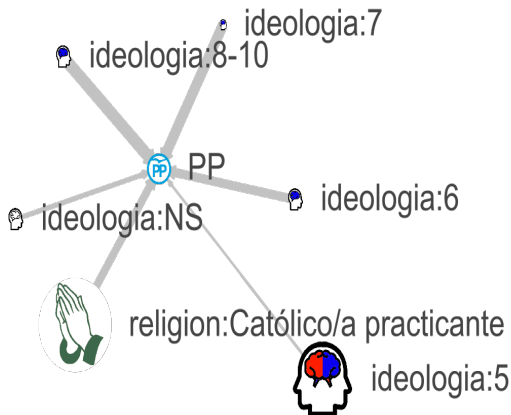
	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-2.9951	0.0857	-34.94	0.0000
Católico/a practicante	0.8290	0.0547	15.17	0.0000

Ideología: 5	0.3334	0.1359	2.45	0.0142
Ideología: 6	1.9818	0.1315	15.07	0.0000
Ideología: 7	2.8592	0.1310	21.82	0.0000
Ideología: 8-10	3.0325	0.1262	24.02	0.0000
Ideología: NS	1.0374	0.1465	7.08	0.0000

Grafo de regresión

Regresión del voto al PP

Intención de voto en Elecciones Generales España (Barómetro del CIS de enero-2019)

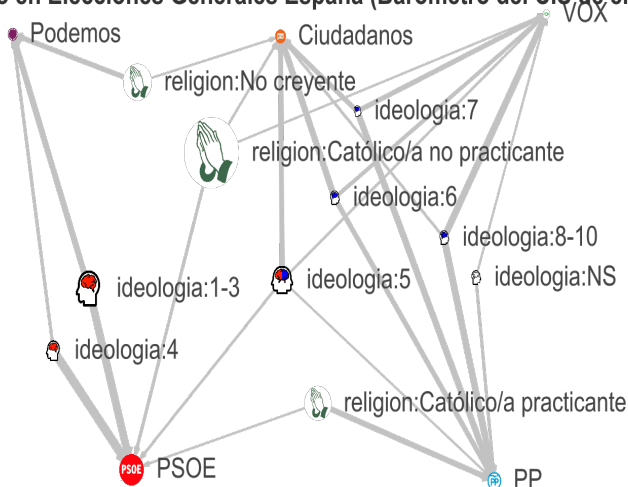


Fuente: CIS (3226) . Analisis: Modesto Escobar. Gráfico elaborado con netCoin. Escobar, M. et al. (2019). netCoin: Interactive analytical networks. Retrieved from CRAN.R packages.

De una a varias regresiones

PP, PSOE, Ciudadanos, U. Podemos, VOX

Intención de voto en Elecciones Generales España (Barómetro del CIS de enero-2019)



Fuente: CIS (3226) . Analisis: Modesto Escobar. Gráfico elaborado con netCoin. Escobar, M. et al. (2019). netCoin: Interactive analytical networks. Retrieved from CRAN.R packages.

Contenido

- 1 Introducción
 - Grafos
 - Los grafos en la investigación social
- 2 CARIG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 Propuestas

¿Cómo hacer estos gráficos

netCoin

- netCoin es un paquete gratuito para R (ubicado en el repositorio CRAN).
 - netCoin calcula
 - Una matriz de coincidencias a partir de una matriz de incidencias (coin).
 - Matrices de semejanzas (distancias) a partir de una matriz de coincidencias (sim).
 - netCoin representa gráficamente
 - Desde una lista de vínculos (netCoin)
 - Desde un conjunto de variables cuantitativas (netCorr)
 - Desde un conjunto de variables dicotómicas (allNet)
 - Desde un conjunto de datos en SPSS o Stata (surCoin and glmCoin)
 - Desde un objeto lavaan (pathCoin)
 - netCoin convierte
 - Una matriz de incidencias/coincidencias/semejanzas/distancias en una lista de vínculos.
 - Un gráfico netCoin en otro igrph o Pajek.

Contenido

- 1 Introducción
 - Grafos
 - Los grafos en la investigación social
- 2 CARIG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 **Propuestas**

Algunas propuestas

- Proponemos los grafos reticulares de coincidencias y regresión como un marco analítico visual.
 - Representar los nodos por su importancia (frecuencia) en lugar de por su grado.
 - Representar los enlaces en función de su significación en lugar de por sus frecuencias.
 - Para variables categóricas (factores) expresar solo asociaciones positivas en relación con las medias ajustadas ponderadas.

Última diapositiva

Agradecimiento

Muchas gracias.
modesto@usal.es & lmartinez@march.es

Bibliografía

Análisis reticular de coincidencias

Bibliografía sobre análisis reticular de coincidencias

- Escobar, M. (2009). Redes semánticas en textos periodísticos: propuestas técnicas para su representación. *Empiria*, 17, 13-39.
- Escobar, M., y Gómez Isla, J. (2015). "La expresión de la identidad a través de la imagen: los archivos fotográficos de Miguel de Unamuno y Joaquín Turina". *Revista Española de Investigaciones Sociológicas*, 152, 23-46.
- Escobar, M. (2015). "Studying Coincidences with Network Analysis and Other Multivariate Tools". *The Stata Journal*, 15(4), 1118-1156.
- Escobar, M. (2016). "Ensayo sobre las coincidencias". En A. Almarcha, P. González, y L. Román (Eds.), *Donde la Sociología te lleve*. A Coruña: Universidad de A Coruña.
- Escobar, M., y C. Tejero (2018). "El análisis reticular de coincidencias". *Empiria*, 39, 129-148.
- Escobar, M. y L. Martínez (under review) "Network Coincidence Analysis: the netCoin R Package".

netCoin

Otros ejemplos interactivos

- Bases de datos
 - M. de Unamuno's album
 - British Library
- Encuestas
 - CIS
 - Series de TV
 - Otras iniciativas
- Coincidencias históricas
 - Historia de la ciencia
 - History de la cultura
- Análisis de contenido
 - TST
 - Blogs de mujeres
- Modelos de ecuaciones estructurales
 - Innovacion empresarial (CIE)