

El Análisis Multinivel

con SPSS y R¹²

Mattia Vacchiano

Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (QUIT)

Institut d'Estudis del Treball (IET)

mattia.vacchiano@uab.cat

Jordi Merino i Noé

Unitat de Sociologia. Departament d'Empresa.

Universitat de Girona.

jordi.merino@udg.edu

Este taller propone una introducción al Análisis Multinivel, facilitando a los participantes los principios teóricos y metodológicos para el uso de esta técnica de análisis en la investigación sociológica. La potencialidad del Análisis Multinivel es que su lógica considera la importancia del contexto en el que se generan las observaciones, midiendo la influencia e interacción entre los distintos niveles en el que se articulan los datos. Los modelos multinivel constituyen la metodología de análisis más adecuada para tratar datos estructurados de forma jerárquica, convirtiendo esta técnica en un recurso fundamental para los investigadores sociales.

A lo largo del taller se realizará un ejercicio práctico con Spss y R, analizando el proceso de modelaje con un ejemplo real y concluyendo con una reflexión sobre las aportaciones, las ventajas y las desventajas de esta metodología de análisis.

El taller requiere un conocimiento básico de análisis de regresión logística y de los softwares Spss y R.

¹ Presentación oral en el marco del III Working Doctoral Group. Universitat Autònoma de Barcelona, 6 de Junio 2017.

² This article was elaborated in the context of INCASI Network, a European project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (Ref. 691004) and coordinated by Dr. Pedro López-Roldán. This article reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

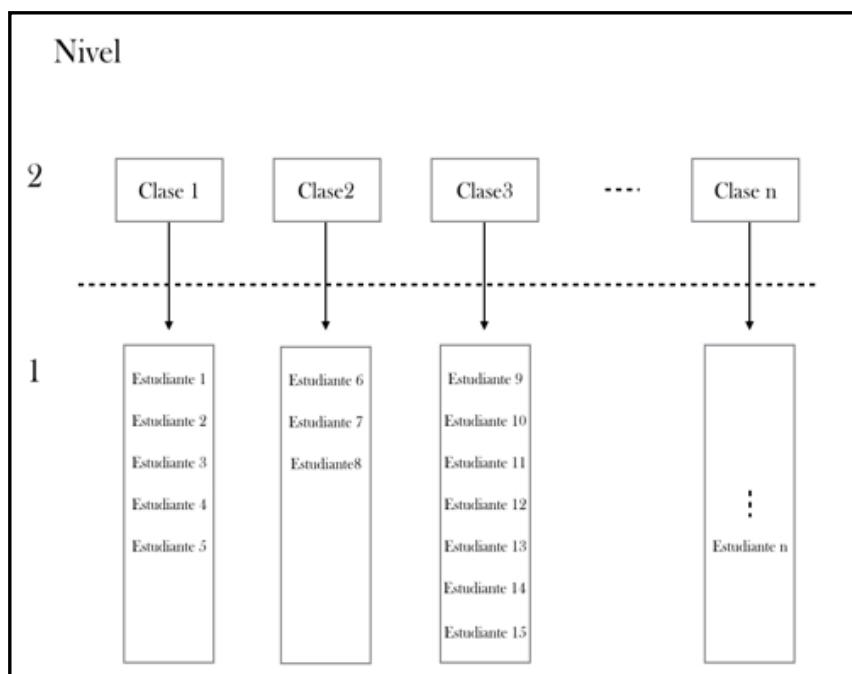
1. La lógica multinivel

1.1. Datos jerarquizados: la importancia del contexto

Una de las preocupaciones de la sociología, en tanto que disciplina, ha sido intentar medir las regularidades sociales y procurar entender qué factores contextuales inciden y ayudan a mejorar la comprensión de un fenómeno. Con ello, tradicionalmente, las ciencias sociales utilizan técnicas de análisis cuantitativas para analizar conjuntamente las características de una *unidad de análisis* (personas, organizaciones, países) con otras variables de contexto, si bien estas últimas no se configuran únicamente como un atributo de la unidad misma. Los análisis de este tipo resultan inexactos, ya que tratan los datos como si estos se articulasen en un único nivel cayendo en la llamada «falacia ecológica»: es decir, atribuyen erróneamente las características de un contexto a las unidades de análisis.

Para obviar este problema cabe considerar que las observaciones que investigamos se articulan en múltiples niveles. Significa que algunas variables están agrupadas dentro otras variables y, en este sentido, los datos que estamos analizando se organizan en un orden jerárquico. Es el caso, por ejemplo, de una investigación que quiera analizar la nota media en historia del arte de los estudiantes de secundaria en Cataluña: (i) los estudiantes están agrupados en aulas o cursos; (ii) las aulas están agrupadas en escuelas o centros educativos; (iii) las escuelas están agrupadas en distritos o provincias. Aulas, escuelas o distritos no son simples atributos de un estudiante. En un mismo aula se comparten experiencias, métodos de enseñanza y docentes que pueden determinar o influir en los resultados escolares de un grupo de estudiantes. Es decir, estos factores contextuales pueden incidir en la variabilidad de la característica que queremos investigar: la nota en historia del arte.

Figura 1.
Ejemplo de datos estructurados en dos niveles.
Estudiantes (nivel 1) y clases (nivel 2).



1.2. Los Modelos Multinivel: conceptos clave

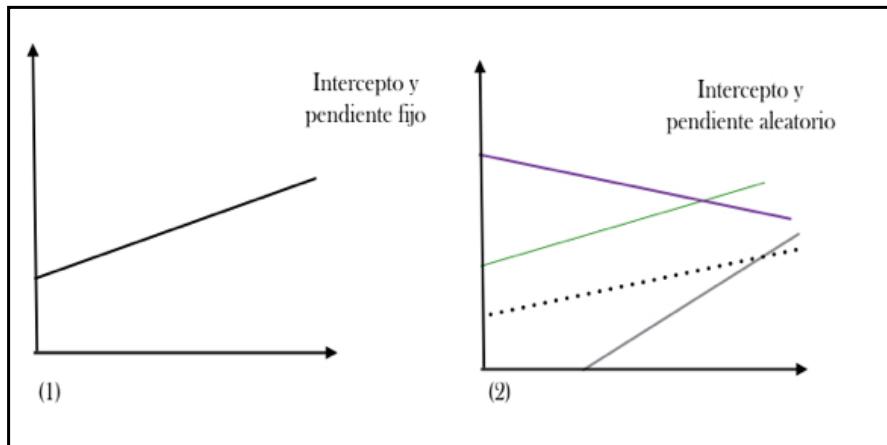
Una vez aclarada la estructuración jerárquica de los datos, la estrategia de análisis multinivel permite trabajar con dos (o tres, o más) unidades de análisis, dependiendo del número de niveles en el que se articulan nuestros datos. Es el investigador/a el que establece la clasificación de los niveles según sus objetivos, modelo de análisis y bagaje teórico. En definitiva, los modelos multinivel³ representan ampliaciones de los modelos de regresión lineal y logística, ya que nos permiten elaborar varios modelos de regresión para cada nivel de análisis, permitiéndonos entender la interacción en cada uno de ellos (Murillo Torrecilla, F.J., 2008). Antes de proponer el proceso de modelaje de un modelo multinivel, es necesario definir dos conceptos fundamentales para comprender la lógica y potencialidad de esta técnica de análisis: (1) la diferencia entre *random effect* y *fixed effect*, (2) la correlación intraclass:

1. Para entender la distinción entre *random effect* (coeficiente aleatorio) y *fixed effect* (coeficiente fijo), cabe considerar que en los modelos de regresión lineal clásicos los parámetros, el intercepto y la pendiente, son comunes a todos los sujetos: es decir, son estimados a partir de todas las observaciones y se asumen como coeficientes fijos, representativo de toda la muestra. En una estructura multinivel, en cambio, estos coeficientes son aleatorios: es decir, cada grupo tiene su propio intercepto y pendiente, como representación específica de cada contexto. Con ello, los coeficientes de regresión del nivel 1 se regresan en el nivel superior, y así sucesivamente hasta obtener un modelo general. Esta distinción representa la gran aportación del análisis multinivel, ya que la composición de los modelos multinivel en dos partes (una parte general, fija, y una variable, que varía en cada grupo) implica que las variables interactúan en distintos niveles (*interacción internivel*). Así pues, básicamente se puede estimar el efecto de las variables — ó factores — explicativas del fenómeno a través de los efectos fijos y la estimación de qué parte de la variabilidad podría ser imputable a cada nivel a través de los efectos aleatorios.
2. La *correlación intraclass* es la medida del grado de dependencia de los individuos (Murillo Torrecilla, F.J., 2008). Es una estimación de cuanto las unidades de análisis sean similares en un mismo grupo y, en este sentido, nos ayuda a comprender cuanto los resultados de nuestro modelo general dependan de la homogeneidad del contexto. Si su valor se acerca a cero, significa que los grupos no son homogéneos internamente y por lo tanto este nivel grupal (este contexto) no influye en los resultados, ya que ser o no ser parte de un grupo es prácticamente irrelevante en el análisis. Es decir, si pertenecer a un grupo o contexto es importante y hay puntuaciones en los datos bastante similares a nivel interno la diferencia entre grupos será más acentuada al existir una mayor variabilidad entre ellos, por tanto una ICC mayor. El coeficiente de correlación intraclass (ICC) es uno de los aspectos más importantes de tener en cuenta a la hora de evaluar un modelo ya que facilita una medida de la similitud que hay en los grupos de un nivel individual, indicando qué parte de la variabilidad total es debida a la variabilidad de las observaciones en los niveles.

³ A nivel general se habla de modelos multinivel, pero también son denominados modelos jerárquicos (por la estructuración de los datos), modelos mixtos (porque utilizan parámetros fijos y parámetros aleatorios) y también de regresión logística multinivel.

Figura 2.

Comparación entre un modelo de regresión lineal (1) y un modelo de regresión lineal multinivel (2), con el modelo general (línea discontinua) y los modelos de cada grupo (líneas coloreadas).



1.3. ¿Cuándo se puede aplicar el análisis multinivel?

Como se ha mencionado, considerar el orden jerárquico en el que se articulan los datos implica afrontar el problema principal de la dependencia entre las observaciones, en razón de la pertenencia a un determinado contexto o grupo (o nivel, como descrito). El análisis multinivel, entonces, se aplica para todo aquellos conjuntos de datos en el que las observaciones no son independientes, y que además pueden estar agrupadas en diferentes niveles. Esta atención hacia el tipo de estructuración de los datos soluciona el problema de la dependencia entre los datos, y permite obtener resultados más robustos y fiables (Cebolla Boado, 2013).

El análisis multinivel se puede utilizar en todos aquellos estudios en el que:

1. los datos se estructuren de forma jerárquica;
2. se quiera trabajar en un mismo modelo diferentes predictores tanto individuales como contextuales sobre una variable dependiente (interacción internivel);
3. se quiera observar por separado la incidencia y varianza dentro del grupo como con los otros niveles de agrupación (correlación intraclass).

2. Aplicación práctica con Spss y R

2.1 El modelo multínivel: estructura, objetivos e hipótesis

Planteamos un ejercicio práctico de aplicación de la lógica multínivel utilizando el programa R. Con este fin, utilizaremos un conjunto de datos longitudinales concernientes las trayectorias laborales de 250 jóvenes entre 20 y 34 años en Cataluña.⁴

Con este fin utilizaremos la matriz de datos: **Ejercicio.sav.**

a. Organización de los datos

Como primer paso, cabe identificar la estructura jerárquica en el que se organizan nuestros datos. Nótense que en las primeras dos columnas se señalan los niveles que componen nuestra matriz: las filas representan los empleos (o eventos laborales) que componen la carrera laboral, y por lo tanto representan el nivel más bajo de la estructura (ID_evento_Nivel1). En la primera columna, en cambio, la variable (ID_caso_Nivel2) representa el número de identificación de los 250 jóvenes de la muestra, lo cual nos indica la trayectoria laboral de pertenencia de los empleos. En este caso consideramos:

1. los empleos (eventos laborales) como nuestra unidad de análisis (**nivel 1**)
2. los jóvenes como el contexto en el que se generan las observaciones (**nivel 2**)

En otras palabras, cada fila representa un empleo (nivel 1) parte de una trayectoria laboral de un joven (nivel 2). Por lo tanto, cabe considerar como la variabilidad de las características de los empleos que componen una trayectoria es dependiente de las características del joven, que en este caso representaría nuestro contexto.

Como se ha mencionado, aplicamos un análisis multínivel ya que:

1. los casos no son independientes y se estructuran de forma jerárquica. Los eventos laborales tienen algo en común, la pertenencia a la trayectoria laboral de un joven.
2. se quiere trabajar en un mismo modelo diferentes predictores, relacionados tanto con las características de los eventos laborales como con las características de los jóvenes sobre una variable dependiente (interacción internivel).
3. se quiere observar por separado la incidencia y varianza de los eventos laborales, a nivel grupal, como al nivel de los jóvenes (correlación intraclasa).

b. Objetivos, hipótesis y modelo

El planteamiento de un modelo multínivel, como toda técnica de análisis, responde de los objetivos de investigación y las hipótesis específicas relacionadas con el objeto de estudio. En este caso, para el planteamiento de este ejercicio hemos decidido centrar nuestro interés en una variable: el canal de acceso al empleo (V7.Metodo.de.inserción)

⁴ Los datos utilizados pertenecen al proyecto de investigación: «Las redes sociales, en sus diferentes modalidades, como recursos y mecanismos de búsqueda laboral e inserción laboral en el empleo y de apoyo social en los jóvenes» (Ministerio de Economía y competitividad. Ref. CSO2012-36055)

Con ello, la variable dicotómica **V7.Metodo.de.inserción** nos indica como ha sido obtenido el trabajo, distinguiendo entre: mecanismos formales (anuncios, convocatorias u ofertas públicas de empleo) o mecanismos informales (contactos personales). A partir de este interés planteamos entonces la siguiente hipótesis:

H1. Las dificultades del mercado del trabajo, en tiempo de crisis, convierten las características de las redes personales, con su informalidad y rapidez, en un canal informal de acceso al empleo prioritario por los jóvenes. Con ello, nos esperamos que los eventos laborales obtenidos a través los contactos personales han aumentado en los años de crisis. Además, nos esperamos que los empleos obtenidos a través los contactos se asocien con mayor probabilidad a empleos precarios, irregulares o complementares.

Con esta hipótesis, planteamos un modelo de regresión logística binomial multínivel. La variable dependiente (VD) es el *mecanismo de acceso al empleo* (formales o informales) mientras las variables independientes (VI) son relacionadas con las características del evento y las características de los y las jóvenes.

En este caso, incluimos en nuestro modelo cuatro variables dicotómicas:

V5. Importancia.actividad. <i>Nivel 1</i>	Indica si el empleo es una actividad laboral principal [1] o un actividad secundaria y complementar [0]
V9.1. Tipodecontrato_illegal. <i>Nivel 1</i>	Identifica la relación laboral del empleo: con contrato [0] o sin contrato [1]
V10.Crisis. <i>Nivel 1</i>	Señala si el empleo se ha obtenido antes de la crisis (1996 – 2008)[0] o durante la crisis (2008-2014) [1]
V13.Sexo. <i>Nivel 2</i>	El sexo del entrevistado, mujer [0] o hombre [1]

Nótense que nuestro el modelo, como condición esencial, plantea una variable dependiente relacionada con el Nivel 1 (característica del empleo) y agrega variables de ambos niveles como variables independientes:

(VD) Mecanismo formal o informal → Importancia.actividad + Tipo de contrato_illegal + Crisis + Sexo

2.2 El proceso de modelaje en R

a. Instalación de RStudio

Presentamos el procedimiento para construir el modelo multínivel con R. Sabemos que este programa es un lenguaje de programación orientado a objetos y que es un software libre y abierto, lo cual representa una primera ventaja en comparación al uso del programa Spss.

En primer lugar, necesitamos instalar el programa RStudio a través del enlace: <http://rstudio.org/>. Este programa permite la visualización simultánea de los entornos de trabajo de R, y mejora la eficiencia y usabilidad del programa.

RStudio se organiza en cuatro áreas de trabajo diferenciadas⁵:

1. En el área en alto a la izquierda podemos abrir y editar ficheros con código R, denominados Scripts. Este código representa el lenguaje de R.
2. En la parte inferior izquierda podemos observar la *console* de R, a través la cual ejecutamos las operaciones en R y visualizamos los resultados.
3. En alto a la derecha encontramos las ventanas del History (el historial de las operaciones en R, con las líneas de código ejecutada) y Environment (donde visualizamos el listado de objetos creados en R).
4. Bajo a la derecha, finalmente, encontramos cuatro ventanas:
 - Files: para acceder a los files del disco duro;
 - Plots: para visualizar los gráficos;
 - Packages: para la gestión de los paquetes instalados.
 - Help: para acceder a la ayuda en línea.

b. Ejecutar el modelo

Procedemos paso a paso para ejecutar el modelo. En primer lugar, cabe controlar nuestra working directory. La working directory es la carpeta donde guardamos todos los archivos con los cuales trabajamos en R. Para saber cuál es nuestra working directory ejecutamos el comando: `getwd()`, en la consola de R (ventana baja a la izquierda).

Con todo, podemos elegir la carpeta donde trabajar a través del comando `setwd("Nombre carpeta")` o bien a través del Menú, a través Session → Set Working directory → Choose Directory...

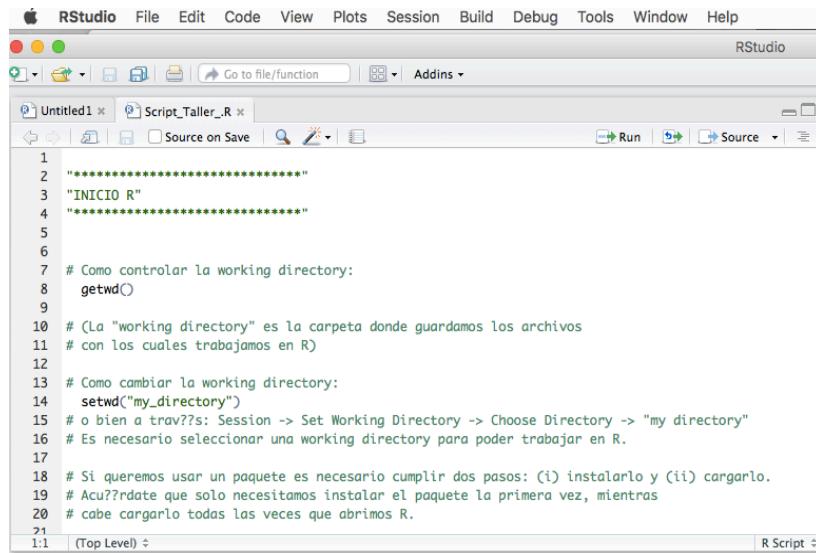
Elegiremos la carpeta donde tenemos guardado nuestra matriz de datos Ejercicio1.sav y el file Script_1.R.

Una vez seleccionada la working directory, abrimos el file: Script_1, a través el menú contextual File → Open File...

Nótese que en alto a la izquierda aparecerá nuestro Script con el código necesario para ejecutar todos los pasos para obtener el modelo:

⁵ Esta información se ha extraído del material de la asignatura “Mètodes Quantitatius de Recerca en Criminologia” del Dr. Pedro López-Roldán y Dr. Sandra Fachelli.

Figura 3.
Visualización del código



```

 1
 2 "*****"
 3 "INICIO R"
 4 "*****"
 5
 6
 7 # Como controlar la working directory:
 8 getwd()
 9
10 # (La "working directory" es la carpeta donde guardamos los archivos
11 # con los cuales trabajamos en R)
12
13 # Como cambiar la working directory:
14 setwd("my_directory")
15 # o bien a trav??s: Session -> Set Working Directory -> Choose Directory -> "my directory"
16 # Es necesario seleccionar una working directory para poder trabajar en R.
17
18 # Si queremos usar un paquete es necesario cumplir dos pasos: (i) instalarlo y (ii) cargarlo.
19 # Acu??date que solo necesitamos instalar el paquete la primera vez, mientras
20 # cabe cargarlo todas las veces que abrimos R.
21
 1:1 (Top Level) R Script

```

Después de haber seleccionado nuestra working directory, si queremos ejecutar una función es necesario cumplir dos pasos: (i) instalar los paquetes necesarios y (ii) cargarlos. En el lenguaje de R, un paquete es una recolección de funciones, datos y códigos que nos permiten operar en el entorno de R. Solo necesitamos instalar el paquete la primera vez, mientras cabe cargarlo todas las veces que abrimos R.

Por lo tanto, el paso siguiente es instalar los paquetes necesarios para poder ejecutar la función que nos permite calcular nuestro modelo multinivel. Con ello, tenemos que ejecutar en la consola el siguiente código:

```

install.packages("lme4")
install.packages("optimx")
install.packages("RVAideMemoire")
install.packages("sjPlot")
install.packages("arm")
install.packages("VGAM")
install.packages("MASS")

```

Sucesivamente, el comando “Library” nos permite cargar los paquetes:

```

library(foreign)
library(lme4)
library(optimx)
library(RVAideMemoire)
library(sjPlot)
library(MASS)
library(arm)
library(VGAM)

```

Podemos comprobar que los paquetes hayan sido instalados y cargados correctamente a través la ventana “Packages” en la parte baja a la derecha de Rstudio.

Figura 4.
La ventana “Packages”

The screenshot shows the RStudio interface with the 'Packages' tab selected in the top menu bar. The window displays a table of packages, with columns for Name, Description, and Version. A search bar and a refresh button are at the top right. The 'lme4' package is highlighted with a blue checkmark in the 'Install' column.

Name	Description	Version
labeling	Generation in R	0.3
latentnet	Latent Position and Cluster Models for Statistical Networks	2.7.1
lattice	Lattice Graphics	0.20-31
latticeExtra	Extra Graphical Utilities Based on Lattice	0.6-26
lazyeval	Lazy (Non-Standard) Evaluation	0.2.0
leaps	regression subset selection	2.9
lme4	Linear Mixed-Effects Models using 'Eigen' and S4	1.1-12
lmeSplines	Add smoothing spline modelling capability to nlme.	1.1-10
lmm	Linear Mixed Models	1.0
lmtest	Testing Linear Regression Models	0.9-34
locfit	Local Regression, Likelihood and Density Estimation.	1.5-9.1
logistf	Firth's bias reduced logistic regression	1.21
logmult	Log-Multiplicative Models, Including Association Models	0.6.2
lpSolve	Interface to 'lp_solve' v. 5.5 to Solve Linear/Integer Programs	5.6.12
lsmeans	Least-Squares Means	2.20-23
magrittr	A Forward-Pipe Operator for R	1.5

Una vez instalados y cargados los paquetes, el paso siguiente es cargar en R nuestra matriz de datos, a través los comandos:

```
NUEVO_OBJETO_R <- read.spss("NOMBRE_MATRIZ.sav",
to.data.frame=TRUE)

attach(NUEVO_OBJETO_R)
```

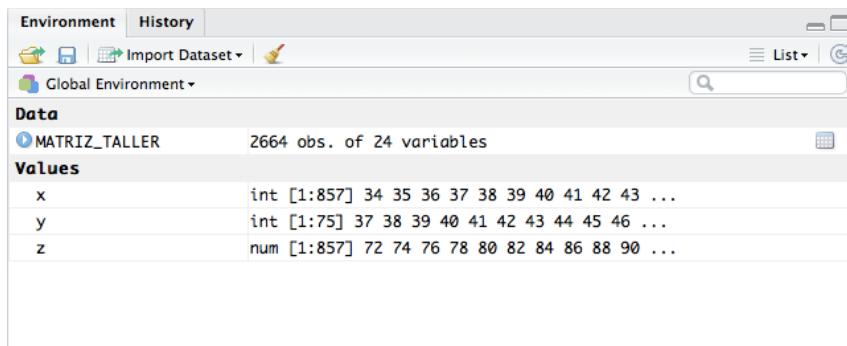
Con este código, creamos un objeto en R: en este caso, una matriz de datos. Podemos decidir el nombre para nuestro objeto. En cambio de NUEVO_OBJETO_R podemos etiquetar la matriz como: MATRIZ_TALLER y escribir el nombre de la matriz en la parte correspondiente:

```
MATRIZ_TALLER <- read.spss("Ejercicio1.sav", to.data.frame=TRUE)

attach(MATRIZ_TALLER)
```

Podemos averiguar que la matriz ha sido cargada correctamente a través la ventana “Environment”.

Figura 5.
La ventana “environment”



Visualizamos la matriz en la ventana del Script clicando dos veces sobre el objeto MATRIZ_TALLER a través la misma ventana.

Figura 6.
Visualización de la matriz de datos

ID_caso_Nivel2	ID_evento_Nivel1	V1.Total.Numero.Eventos	V2.Crono_evento	V3.1.Datainici	V3.2.Dat
ID_Entrevistado	Eventos de la trayectoria	Nombre d'esdeveniments en la trajectoria	Cronología de los eventos		
1	1	101	13	1	13376448000 13375
2	1	102	13	2	13386988800 13428
3	1	103	13	3	13431657600 13436
4	1	104	13	4	13442198400 13450
5	1	105	13	5	13463193600 13468
6	1	106	13	6	13471142400 13476
7	1	107	13	7	13479004800 13486
8	1	108	13	8	13520995200 13528
9	1	109	13	9	13531536000 13563
0	1	110	13	10	13565836800 13585
1	1	111	13	11	13599964800 13610
2	1	112	13	12	13592016000 13613
3	1	113	13	13	13276396800 13613

Una vez cumplidos todos estos pasos podemos efectivamente ejecutar el código de la función multinivel **glmer** (*Generalized Linear Mixed-Effects Models*). La formula separa la variable dependiente, a la izquierda, con el operador ~ y las variables independientes con el operador +, a la derecha. Los efectos aleatorios se distinguen a través la barra vertical (“|”).

```
MODELO <- glmer(VD ~ VI + VI + VI + VI + (1 | "VARIABLE_NIVEL2"), family = binomial(logit), control=glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
```

Creamos el objeto “MODELO1” e insertamos las variables del modelo en las partes correspondientes:

```
MODELO1<- glmer(V7.Metodo.de.insercion ~ V13.Sex + V5.Importancia.actividad
+ V9.1.TipodeContrato_Illegal + V10.Crisis + (1 | ID_caso_Nivel2), family = binomial(logit), control=glmerControl(optimizer="bobyqa", optCtrl = list(maxfun =
100000)))
```

Es importante saber que todos estos pasos pueden ser ejecutados directamente a través del código presente en el Script. En este caso podemos subrayar la línea de código del Modelo y clicar el botón RUN, en alto a la derecha de la ventana.

Figura 7.
El código del modelo

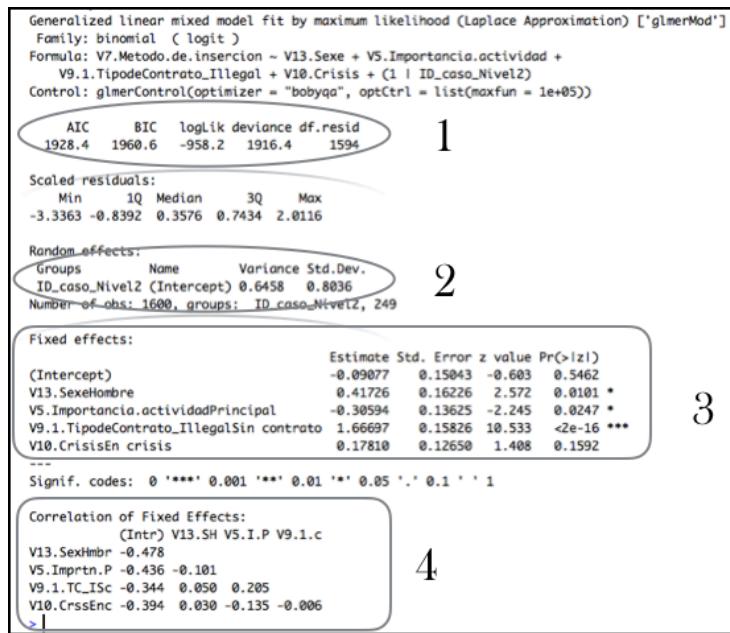
```

70 NOMBRE_DEL_MODELO <- glmmPQL(VD ~ VI1 + VI2 + VI3, random = ~ 1 | EgoID,
71                               family = binomial)
72
73
74 "*****"
75 "EJEMPLO MODELO"
76 "*****"
77
78
79 MODELO1<- glmer(V7.Metodo.de.insercion ~ V13.Sex + V5.Importancia.actividad + V9.1.TipodeContrato_Illegal + V10.Crisis + (1 | ID_caso_Nivel2), family = binomial(logit), control=glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
80
81 MODELO2 <- glmer(V7.Metodo.de.insercion ~ V13.Sex + V12.Origen + V11.1.Edat_Evento + V5.Importancia.actividad + V9.1.TipodeContrato_Illegal + V10.Crisis + (1 | ID_caso_Nivel2), family = binomial(logit), control=glmerControl(optimizer="bobyqa", optCtrl = list(maxfun = 100000)))
82
83
84 # Visualizar los resultados del modelo
85
86 summary (MODELO1)
87
88 # Cancelar el modelo
89
90 rm (MODELO1)
91
```

Una vez ejecutado este comando notaremos haber creado el objeto “MODELO1” en la ventana “Environment”. Significa que ya disponemos de nuestros resultados y solo necesitamos visualizarlo a través el comando:

summary (MODELO1)

Figura 8.
Visualización de los resultados



Comentamos punto por punto los resultados obtenidos:

(1) Como en un regresión logística para conocer la bondad de ajuste disponemos de distintos test⁶. Dos son los principales:

- Akaike's information criterion (AIC)
- Schwarz's Bayesian criterion (BIC)

Ninguno de los dos test puede ser intrínsecamente interpretable. Esencialmente valores más bajos señalan una mayor bondad de ajuste, aunque algunos autores señalan que estos indicadores son más adecuados en el caso de regresiones lineales.

(2) Como sabemos los modelos multinivel se componen de dos partes: una parte fija y una variable. El apartado “Random effect” señala la desviación estándar de la parte variable, permitiéndonos calcular el coeficiente de correlación intraclase (ICC), es decir el porcentaje (%) de varianza de la VD explicada por el factor aleatorio (en este caso, los jóvenes). Para calcular el ICC usamos la formula:

$$(STdev)^2 / (3.29 + (STdev)^2).$$

En nuestro modelo, esto es: $ICC = 0.6458 / (3.29 + 0.6458) = 0.164$. El 16,4 % de la probabilidad que un empleo se consiga por contacto personal depende de las características del joven.

(3) En este apartado visualizamos efectivamente nuestro modelo general.

⁶ Para los modelos multinivel se usan «chi-square maximum likelihood ratio test».

En el modelo se evidencian:

- Estimate: los coeficientes estimados por cada variable.
- St. Error: la desviación estándar
- Z value: los residuos
- P value: la significación

Notamos que nuestra hipótesis se confirmada solo parcialmente. Los coeficientes señalan una asociación positiva de las variables “sexo”, “tipo de contrato” y “crisis” (aunque solo las primeras dos son significativas) y una asociación negativa en la variable “importancia de la actividad”. En este caso:

- el modelo destaca una mayor probabilidad que los empleos obtenidos por contacto personal sean **trabajos sin contrato**, obtenido por **hombres**, y que no sean una actividad principal, sino **secundaria o complementar** a otra actividad de estudio o de trabajo, como planteado en nuestro hipótesis.
- Sin embargo, la variable crisis, si bien se asocia positivamente, no es significativa, por lo tanto no podemos afirmar que obtener empleos por contacto personal sea más probable en una época de crisis, sino más bien este mecanismo de acceso al empleo se configura como una característica estructural del mercado laboral en el contexto investigado.

(4) Finalmente, en el último apartado visualizamos las correlaciones entre las variables. Esta matriz nos permite identificar problemas de multicolinealidad en el modelo general.

2.3 El proceso de modelaje en SPSS

El procesador IBM SPSS es un paquete estadístico históricamente utilizado en las universidades, en combinación con otros paquetes, y con una implantación considerable en el mercado privado. A partir de su versión 19.2 incorpora la posibilidad de poder realizar modelos multínivel con variables respuesta de tipo categórico a través de los denominados modelos lineales mixtos generalizados (GLMM). Con las versiones posteriores se han ido mejorando básicamente los procesos y el manejar diferentes tipos de variables y funciones, como es el caso de la versión 22.0.

Es central antes de proceder a la explotación estadística se tenga claro la estructura del modelo a realizar. En primer lugar, hay que tener en cuenta que la composición de las variables disponga de la estructura necesaria para proceder a su análisis. La variable respuesta ha de ser dicotómica y las variables predictores sean categóricas (nominales u ordinales) o también variables de tipo lineal.

Este modelo responderá a la pregunta de investigación que el investigador/a quiera hacer y sustentando en algún bagaje teórico que facilite la posterior interpretación de los resultados. Siguiendo el modelo anterior y con la misma hipótesis en R realizaremos el modelo paso a paso con SPSS.

a. Visualización de los datos

Los datos se estructuran en diferentes ventanas: una de visualización de la información (dónde se puede apreciar la matriz de datos y el listado de variables), una ventana de sintaxis (que recoge los procedimientos utilizados) y una ventana de resultados.

Figura 9.

Distribución de la información de las variables en la pestaña “vista de variables”.

Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Role	
4	V2.Crono e...	Numérico	4	0	Cronología de ...	Ninguna	Ninguna	13	Derecha	Nominal	Entrada
5	V3.1 Datainici	Fecha	8	0		Ninguna	Ninguna	15	Derecha	Escala	Entrada
6	V3.2 Datafinal	Fecha	10	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
7	V4.Durada	Numérico	8	0	Durada del es...	Ninguna	Ninguna	10	Derecha	Escala	Entrada
8	V5.Importan...	Numérico	8	0	Es la actividad ...	[0, Secundaria...]	99	26	Derecha	Nominal	Entrada
9	V6.Categor...	Numérico	8	0	No cualificado ...	Ninguna	99	31	Derecha	Nominal	Entrada
10	V7.Metodo...	Numérico	8	0	MECANISMO...	[0, Otros...]	99	29	Derecha	Nominal	Entrada
11	V8.Primer e...	Numérico	8	0	La actividad e...	Ninguna	Ninguna	19	Derecha	Nominal	Entrada
12	V9.Relacion...	Numérico	8	0	TipodeContrato ...	[1, Autonom...]	99	13	Derecha	Nominal	Entrada
13	V9.1.Tipode...	Numérico	8	0	TipodeContrato ...	[0, Con cont...]	99	25	Derecha	Nominal	Entrada
14	V9.2.Tipode...	Numérico	8	0	TipodeContrato ...	Ninguna	99	27	Derecha	Nominal	Entrada
15	V9.3.Tipode...	Numérico	8	0	TipodeContrato ...	Ninguna	99	25	Derecha	Nominal	Entrada
16	V9.4.Tipode...	Numérico	8	0	TipodeContrato ...	Ninguna	99	25	Derecha	Nominal	Entrada
17	V10.Crisis	Numérico	8	0	Evento durante ...	[0, Antes de...]	99	28	Derecha	Nominal	Entrada
18	Caracterist...	Numérico	8	2		Ninguna	Ninguna	118	Derecha	Desconocido	Entrada
19	V11.Edad	Numérico	8	0	Edad del entrev.	Ninguna	Ninguna	5	Derecha	Escala	Entrada
20	V11.1.Edat...	Numérico	8	0	Edad del entrev.	Ninguna	Ninguna	14	Derecha	Escala	Entrada
21	V12.Origen	Numérico	8	0	Origen del entr...	[0, Immigr...]	Ninguna	21	Derecha	Nominal	Entrada
22	V13.Sexo	Numérico	1	0	[0, Mujer...]	Ninguna	12	Derecha	Nominal	Entrada	
23	V14.Nivel e...	Numérico	8	0	Estudiante Uni...	[0, No Univ...]	9	33	Derecha	Nominal	Entrada
24	V15.Cat Pro...	Numérico	8	0	La categoría pr...	Ninguna	99	27	Derecha	Nominal	Entrada
25											
26											
27											
...											

a. Ejecución del modelo.

Realizamos las operaciones a partir de la secuencia de comandos de la barra superior, o bien a través la ventana de *syntax* si previamente se ha realizado el ejercicio, pulsando icono ejecutar.

Cabe remarcar que en este ejercicio no se pide la información diferenciada en los modelos como puede ser el sexo, las región de los individuos u otras variables. Si fuere así cabría seleccionar los casos previamente desde la opción de datos de la barra de comandos y hacer un ejercicio para cada grupo, ya que en modelos mixtos aún no se ha incorporado esta segmentación por grupos en los resultados.

Existen diferentes rutas para realizar modelos multínivel, pero en el caso de variables dependientes categóricas (dicotómica) la ruta es la siguiente:

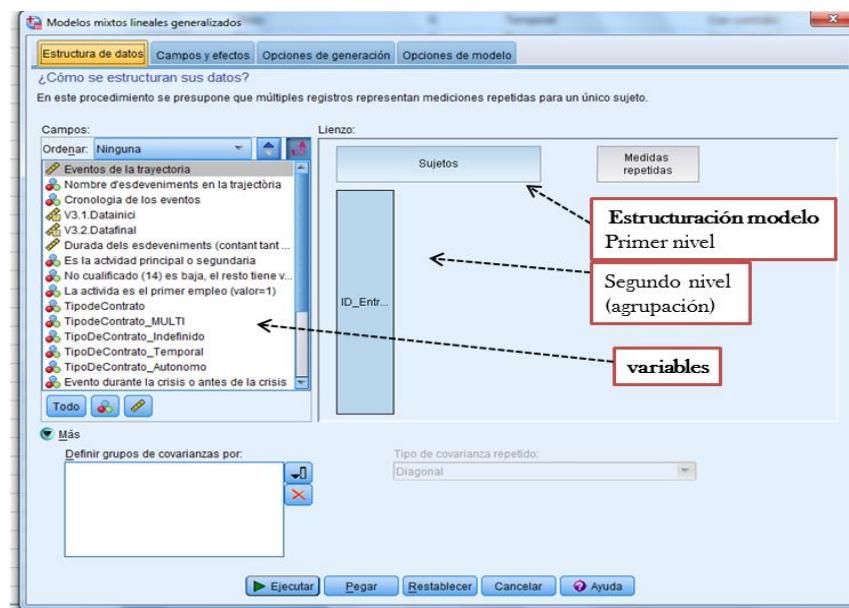
Analizar > Modelos Mixtos > Lineales generalizados (ó modelos mixtos según versión)

Si es la primera vez que se realiza en un procesador SPSS es posible aparezca una ventana de diálogo que recordará que el nivel de medición cabe realizarlo manualmente. Se aconseja indicar “asignar manualmente”, ya que se realizará el ejercicio. Después de la asignación aparecerá una ventana de diálogos donde hay diferentes pestañas para proceder al análisis paso a paso. En los cuadros de las pestañas hay diferentes opciones, pero solo se hará mención a las que tienen a ver con el modelo que se está realizando.

La primera de ella es la de estructura de datos que presenta un lienzo en la parte derecha dónde se arrastran las variables en los niveles que se quiera agrupar. Se entiende que el primer nivel lo conforman los empleos (sujetos, en la ventana SPSS) por tanto

solo hay que incorporar el segundo nivel u otros sucesivos si hubieren. En este caso se arrastra la variable de segundo nivel, los jóvenes (ID_Entrevistado).

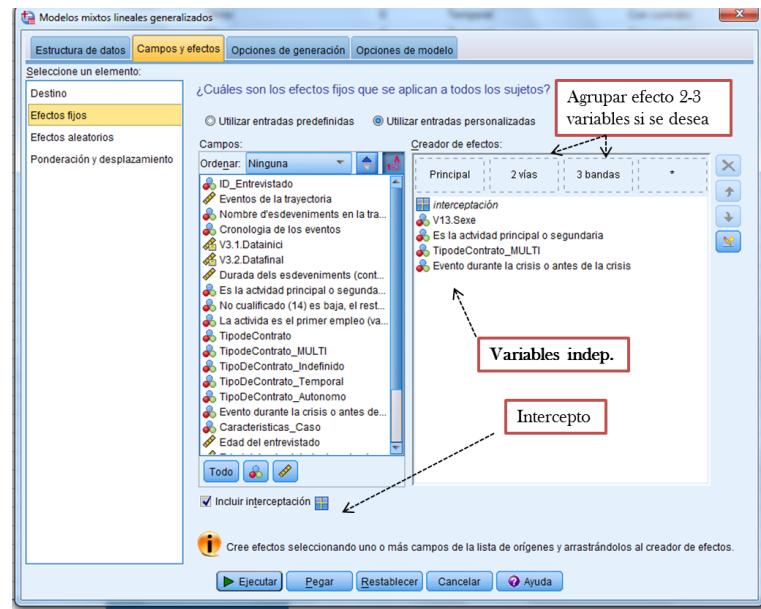
Figura 10.
Cuadro de dialogo de estructuración de los datos.



En la segunda pestaña se definen los campos y los efectos. Este está compuesto por cuatro acciones que aparecen en el cuadro de diálogo de la banda izquierda. En la primera acción, que es el destino, hay que definir la variable independiente y remarcar que la regresión que se ejecutará será logística binaria. La segunda acción será indicar que variables son los predictores en el modelo en los efectos fijos. Por defecto también aparece el intercepto y la opción de añadir otros efectos, o incluso agrupar dos variables en la opción 2 vías.

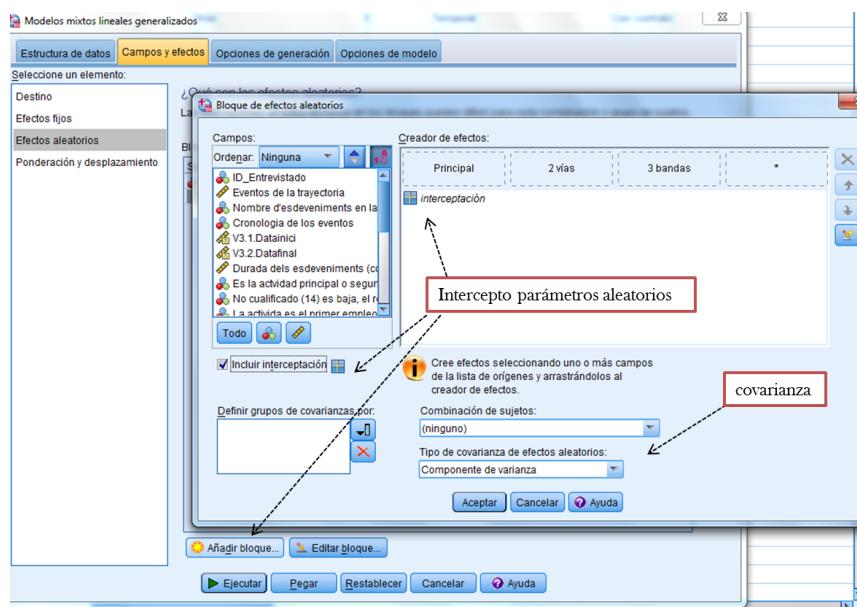
Si se desea hacer un análisis paso a paso, se aconseja no añadir ninguna variable y así se obtendría el modelo nulo, que indicaría información sobre la variabilidad del contexto sin los predictores y si procede continuar con el análisis. Para modelos más complejos con agrupaciones de factores y covariables (variable cualitativa y cuantitativa, respectivamente), en este apartado se podrían ir incorporando según los pasos que se quiera seguir.

Figura 11.
Cuadro de dialogo de diseño de las variables independientes en los efectos fijos.



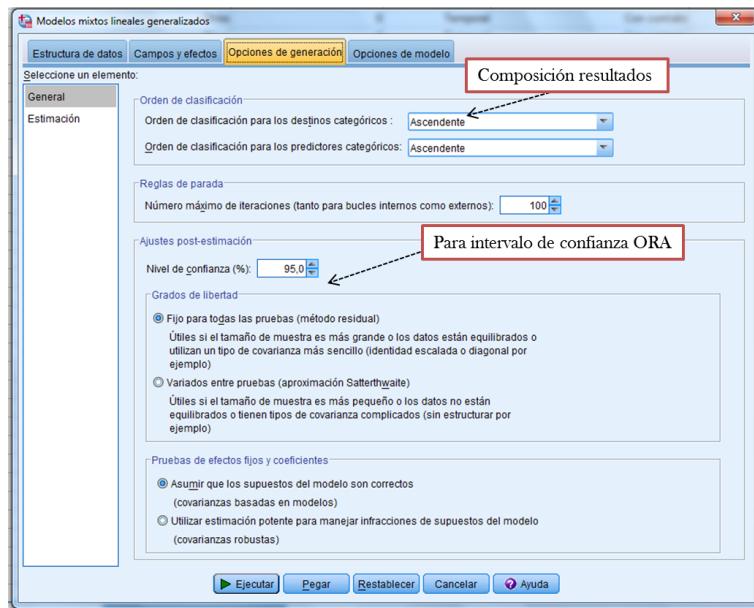
La acción tercera corresponde a la elección del intercepto del modelo en los parámetros aleatorios. Se aconseja en el tipo de componentes de la varianza en relación a los efectos aleatorios, aunque hay otras opciones. Para modelos más complejos se podrían incorporar otros bloques de efectos aleatorios, pero en el caso de modelos dicotómicos es suficiente el intercepto por su tipo de distribución en dos valores.

Figura 12.
Cuadro de dialogo de efectos aleatorios.



La cuarta opción es por si es una muestra que haya que ponderar, si es necesario y no se ha hecho por otras opciones de SPSS. Las opciones de generación, que son la última pestaña, indican cómo se desea ver los resultados. Es importante remarcar el intervalo de confianza al 95% (que nos ayudara por ejemplo a interpretar la ORA) y la composición de los resultados en función de la categoría de referencia (por ejemplo hombres = 1, mujeres = 0).

Figura 13.
Opciones de generación de los resultados.



Las opciones del modelo en este caso no son importantes, aunque en esa ventana el investigador/a puede pedir estimaciones de medias, tipos de contraste para estimar los destinos. También aparece la opción de guardas en campos nuevos las estimaciones en una variable de destino nueva.

Una vez realizado el modelo se puede proceder a indicar “ejecutar” para realizar los resultados o “pegar” para trasladarlo a lenguaje de syntax.

Figura 14.
Ventana de Sintaxy.

```

*Syntax1 - IBM SPSS Statistics Editor de sintaxis
Archivo Editar Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ejecutar Herramientas Ventana Ayuda
Ejecutar selección info_de_datos1
DATASET ACTIVATE Conjunto_de_datos1.
*Generalized Linear Mixed Models
GENLINMIXED
  /ANALYSIS SUBJECTS=id_casa_Nivel2
  /FIELDS TARGET=V7 Método_de_investigación=NONE OFFSET=NONE
  /TARGET_OPTIONS REFERENCE=1 DISTRIBUTION=BINOMIAL LINK=LOGIT
  /FIXED EFFECTS=V13_Sexo V5_Importancia_actividad V9_1_TipoDeContrato V10_Crisis
  /USE_INTERCEPT=TRUE
  /RANDOM SUBJECTS=id_casa_Nivel2 COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /CROWDER USE_INTERCEPT=TRUE COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERATIONS=100
  /CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COV_B=MODEL PCONVERGE=0.000001(ABSOLUTE) SCORING=0
  /SINGULAR=0.000000000001
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=SEQSDAK.
  
```

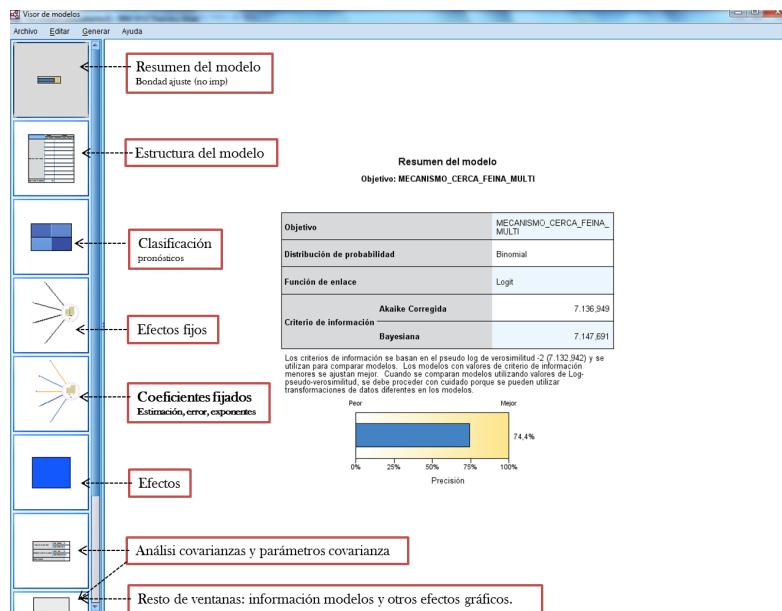
b. Visor de resultados.

Una vez ejecutado el análisis en la ventana de resultados aparece la información de los casos incorporados y la información genérica del modelo. Hay que clicar dos veces en el cuadro de resultados que aparece ya que SPSS construye un visor dinámico de resultados en formato flash.

En la parte izquierda del visor aparece el diálogo estructurado por parámetros (Figura 9), y permite ir directamente a la consulta de los datos. Para las primeras veces se aconseja ir a la última ventana donde hay un resumen de todos los pasos que se han realizado en el modelo y las opciones de construcción, así como las variables que se han utilizado, además de una información básica sobre el método y los modelos mixtos (no se ofrece figura).

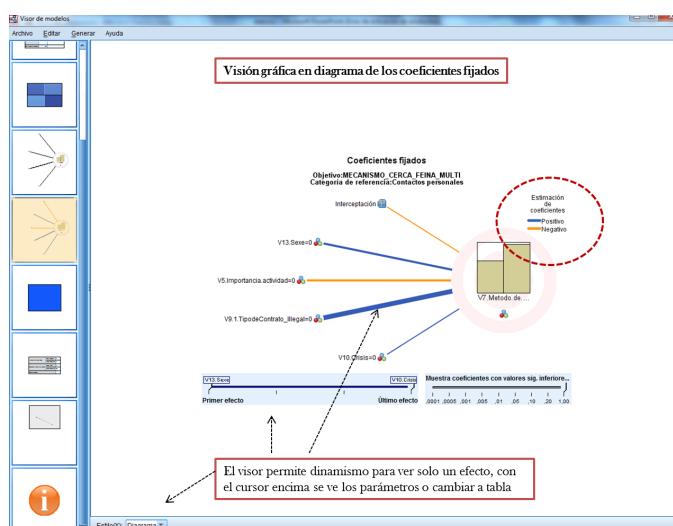
En la ventana de resumen del modelo se facilita poder consultar la bondad de ajuste del modelo a través de los criterios de información. Nótense que a diferencia del ejercicio con R, los valores de AIC y BIC son diferentes. Algunos autores no dan importancia a estos ajustes al ser la variable respuesta dicotómica y por tanto no tiene más interés que ver como varia con la incorporación de variables contextuales. Aun así se suele utilizar el logaritmo de la verosimilitud que aparece debajo del recuadro en la opción -2LL.

Figura 15.
Clasificación del visor de resultados según parámetros.



En la ventana de los coeficientes fijados la visualización de datos aparece de forma gráfica y coloreada según la asociación sea positiva o negativa y con un grosor proporcional a su peso sobre la variable dependiente. Si se pone el cursor encima de las líneas aparecerá la información y datos de ese valor. En la parte inferior aparece la opción de pasar a diagrama y ver así en tablas la información, a la que habrá que pedir que pase a exponente las estimaciones y observar los intervalos de confianza.

Figura 16.
Visión gráfica de los coeficientes fijados.

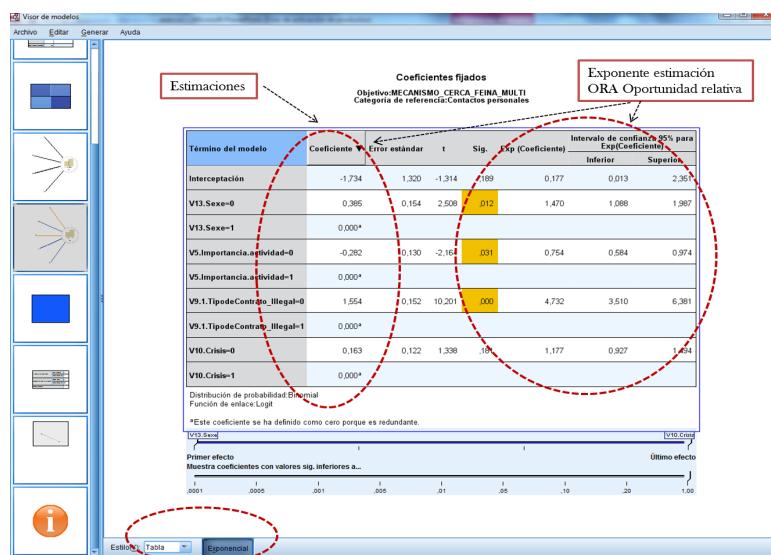


Según la tabla de resultados (Figura 11) se pueden interpretar las estimaciones o las Oportunidades Relativas Asociadas (ORA), que son la elevación exponencial de las

estimaciones y poder así interpretarlo dentro del intervalo de confianza y apreciar así su proporción si se desea. A nivel inicial tanto el sexo como la actividad principal o el tipo de contrato se asocian con encontrar trabajo de forma informal siendo ésta estadísticamente significativa:

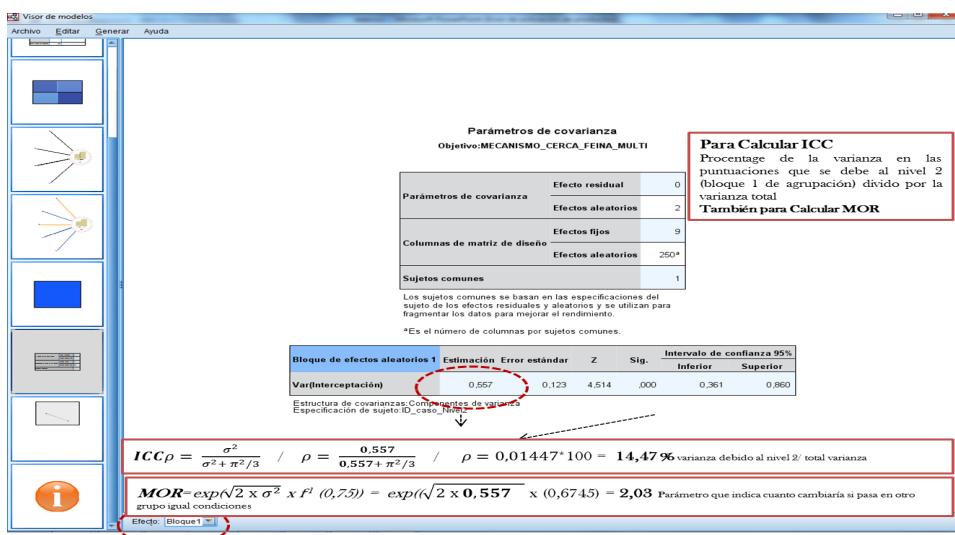
1. Se aprecia que el sexo determina el mecanismo informal de encontrar empleo, siendo la proporción de ORA significativa para los hombres con un valor de 1,47 (IC95% = 1,08-1,98) sobre la referencia de mujeres.
2. Se observa también que cuando el empleo no es una actividad laboral principal su asociación con la variable dependiente es preservadora. Lo que quiere decir que si es una actividad laboral principal no facilita el evento de búsqueda informal de empleo. Una lectura que también se podría hacer es que si el trabajo es la actividad laboral principal la proporción de no encontrar trabajo de manera informal sea de un 68% menor que si es una actividad laboral secundaria.
3. El mayor de los efectos con una ORA más elevada se contempla en el caso de encontrar trabajo de manera informal en jóvenes que tienen una relación laboral sin contrato, lo que también se puede leer que es 4,7 veces superior o que es un asociación facilitadora.
4. El momento en que se ha obtenido el empleo (evento) de manera informal (si ha sido precedente o durante la crisis) no es significativo. Aquí el investigador/a podría discutir estos resultados en función de su conocimiento, otras estadísticas o estudios, como por ejemplo para expresar que el uso de los contactos personales es una características estructural del mercado del trabajo, y no depende de la coyuntura económica. En este sentido, el empleo juvenil ya tenía algunas dinámicas previas a la crisis que se han mantenido o agraviado durante ese período. Pero como se dice, eso formaría parte de la discusión de los resultados.

Figura 17.
Distribución de las estimaciones y las ORA de los efectos fijados.



Otra ventana interesante es la que facilita la información de los parámetros de covarianza y los efectos aleatorios. Gracias al valor del intercepto, si es significativo, se puede proceder a hacer la ecuación para calcular la ICC i la MOR, que SPSS no ejecuta. Las ecuaciones se ofrecen encuadradas en la Figura 18.

Figura 18.
Visor de los parámetros de covarianza de los efectos aleatorios.



El parámetro MOR es discutible en algunas disciplinas de las Ciencias Sociales, pero en epidemiología social y sociología de la salud se considera un buen indicador gráfico de la variabilidad entre grupos, ya que es una mediana de la ORA y muestra cuánto varia la proporción de un grupo a otro si el individuo (o estructura del primer nivel) son similares. Es decir ayuda a explicar la varianza del segundo nivel.

Como se aprecia en los resultados, al utilizar un enlace logit diferente a R, hay alguna pequeña diferencia en los resultados. En el caso de la ICC nos indica que la varianza debida al nivel segundo en relación a la varianza total es del 14,47%. Además, la proporción que ofrece la MOR es de 2,03. Esto podría interpretarse como la proporción de la varianza que existe entre un individuo que ha encontrado trabajo de manera informal comparado, considerando los factores antes descritos similares para los dos, con otro que no lo encontró de manera informal. Una proporción bastante elevada, lo que indica que los eventos diferencian de manera considerable según los individuos.

De manera más fácil se podría decir que este parámetro aleatorio nos daría pistas para evidenciar las diferencias del éxito de los eventos en función de un hombre que ha encontrado empleo a través mecanismos informales siendo la actividad laboral principal, sin contrato y en tiempo de crisis, si cambiara a otro individuo.

A partir de este modelo simple, según la pregunta de investigación y los modelos teóricos, el investigador/a podría incorporar otras variables de los eventos o de los individuos en tanto que contextuales para apreciar mayor concreción en los efectos.

3. Conclusiones

En este taller se ha realizado una breve introducción a la lógica multinivel y se ha rea-lizado un ejercicio con el procesador estadístico SPSS y con el software R en el caso de un modelo con una variable respuesta categórica dicotómica. El análisis multinivel proporciona diferentes técnicas para trabajar con modelos más complejos que den res-puesta a diferentes problemas de investigación tanto en variables categóricas y lineales.

Principalmente, el ejercicio ayuda a observar las diferencias entre un procesador y otro en un modelo simple de dos niveles.

1. El procesador estadístico SPSS permite al investigador trabajar poco a poco el análisis multinivel a partir de su pregunta de investigación. Desde este punto de vista facilita el proceso secuencial desde el modelo nulo (mo-delo inicial sin variables predictores) hasta el modelo final y observar cómo van interfiriendo factores o covariables. Diversamente el entorno de trabajo R, por sus mismas características, propone un proceso de modelaje más rá-
pido, centrándose únicamente en la definición de las variables independien-tes y la variables dicotómica dependiente.
2. En relación a la visualización de los resultados, SPSS ofrece al investi-gador centrar la atención en aquellos parámetros que desea observar, al mis-mo tiempo que facilita mayor dinamismo en la expresión gráfica y visual, si bien , el formato flash de visualización de los datos dificulta su posterior explotación en presentaciones u otras composiciones. En comparación, el entorno de trabajo R permite una visualización rápida de los resultados a travé-s la consola, permitiendo una menor inversión de tiempo una vez ad-quirida la alfabetización básica del lenguaje y de la lógica multinivel.

4. Referencias y recursos bibliográficos

A. Literatura básica

- Andréu Abela, J. (2011). El análisis multinivel: una revisión actualizada en el ámbito sociológico. *Metodología de Encuestas*, 13, 161–176.
- Murillo Torrecilla, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis, Revista Internacional de Investigación en Educación*, 1, 45–62.
- Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Larsen, K. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, 60(4), 290–297.

B. Manuales

- Catalán-Reyes, M., & Galindo-Villardón, M. (2003). Utilización de los modelos multinivel en investigación sanitaria. *Gaceta Sanitaria*.
- Cebolla Boado, H. (2013). Introducción al análisis multinivel, Madrid, Centro de Investigaciones Sociológicas (CIS).
- Field, A. (2009). Discovering statistics using SPSS. London: Sage
- Heck, R. H., & Thomas, S. L. (2009). An introduction to multilevel modeling techniques. New York, Routledge.
- Heck, R. H., et al. (2012). Multilevel modeling of categorical outcomes using IBM SPSS, New York, : Routledge.
- Pardo, A., Ángel, M., San, R. (2007). Cómo ajustar e interpretar modelos multinivel con SPSS. *Psicothema*, 19(2), 308–321.