

ConfoundReg

Modelling confounding in Linear, Logistic and Cox Regression

Tabla de contenidos


Presentación	1
Instalar el comando de extensión	2
Comprobación de la instalación	3
Desinstalar un comando de extensión	3
Regresión lineal múltiple	4
Ejemplo 1	4
Ejemplo 2	6
Ejemplo 3	7
Ejemplo 4	8
Regresión logística	11
Regresión de Cox	13
Variables dependientes del tiempo (VDT)	14
Tratamiento de variables binarias	17
Referencia de sintaxis de UAB_ConfoundReg	19

Presentación

En modelos de regresión con **finalidad explicativa**, para evaluar la magnitud de la confusión debida a una variable confundidora (Z) se debe **comparar** el efecto de la variable explicativa (X) estimado con el modelo que incluye la variable confundidora (referencia), con el estimado con el modelo (reducido) que no la incluye. Se prefiere el modelo de referencia cuando la magnitud del cambio es prácticamente importante. El estudio de simulación de Maldonado y Greenland (1993) para respuestas binarias sugiere que el ajuste es necesario para corregir el sesgo de confusión si el cambio entre el efecto ajustado y sin ajustar es superior al 10%.

Cuando el número de potenciales variables confundidoras es elevado resulta muy laborioso realizar todas las comparaciones entre el modelo de referencia y los modelos reducidos. La complejidad es aún mayor si la variable explicativa es categórica y se ha debido descomponer en ficticias, o si el modelo de referencia incluye una variable modificadora, pues en este caso las comparaciones se deben realizar para los diferentes valores de dicha variable.

El comando de extensión UAB_ConfoundReg simplifica notablemente el proceso al realizar de forma automática todas las comparaciones entre modelos y presentar los resultados en tablas de fácil lectura.

 La versión actual del comando de extensión (v0.0.7) ha sido comprobada con las versiones de SPSS 18.0.3, 19.0 y 20.0. Para SPSS 18.0.3 es necesario instalar el comando desde los archivos UAB_CONFOUNDREG_ML_SPSS18.spe y UAB_CONFOUNDREG_COX_SPSS18.spe. El comando de extensión no funciona en versiones anteriores a la 18.

Instalar el comando de extensión

Para poder utilizar comandos de extensión debe instalar en su ordenador, en primer lugar, los complementos de integración de Python y R en SPSS Statistics. Esta instalación la realizará sólo una vez y servirá para todos los comandos de extensión que posteriormente instale.

Para ayudarle en el proceso hemos creado un documento titulado *Guía de instalación de los complementos de integración de Python y R en SPSS Statistics*, que debe descargar del campus virtual, en concreto se encuentra en el enlace "Material", pestaña "SPSS Statistics", en el enlace "Guía de instalación de los complementos de integración de Python y R".

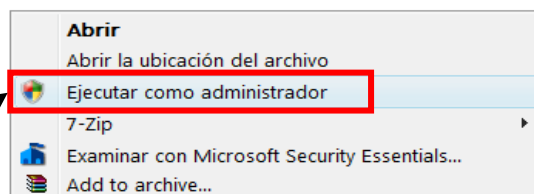
❗ **Para evitar problemas de instalación es fundamental que siga estrictamente los pasos que se indican en la guía de instalación.**

❗ **Recuerde que en SPSS 18 debe instalar desde los archivos UAB_CONFOUNDRREG_ML_SPSS18.spe y UAB_CONFOUNDRREG_ML_SPSS18.spe.**

Una vez instalados con éxito los complementos de integración de Python y R puede proceder a instalar los comandos de extensión.

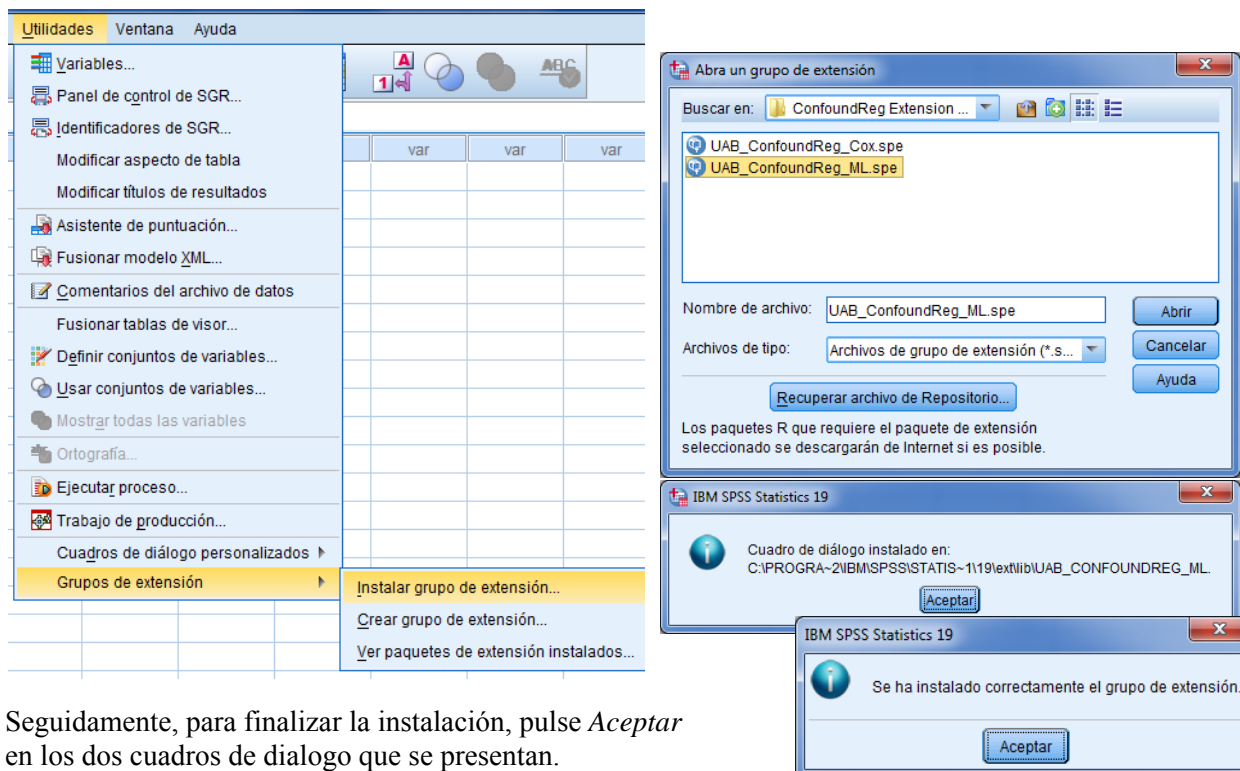
La instalación de un comando de extensión modifica la carpeta de instalación de SPSS y otros componentes protegidos del sistema operativo, por lo que **debe realizarse sobre una sesión de SPSS Statistics abierta con derechos de Administrador**. Además el ordenador **debe estar conectado a Internet** para poder los descargar paquetes de R que necesite el comando.

Para ello debe situarse sobre el icono de SPSS, hacer clic con el botón derecho y en el menú contextual escoger la opción *Ejecutar como administrador*.



Tras lanzar *SPSS Statistics* con derechos de Administrador puede instalar (o reinstalar) un comando de extensión con la opción de menú *Utilidades / Grupos de extensión / Instalar grupo de extensión...*

Las siguientes imágenes muestran la instalación del procedimiento de evaluación de variables confundidoras para regresión lineal y logística (UAB_ConfoundReg_ML). Se abre una ventana para indicar donde se encuentra el archivo .spe con el comando de extensión:



Seguidamente, para finalizar la instalación, pulse *Aceptar* en los dos cuadros de dialogo que se presentan.

Si un comando de extensión utiliza un paquete de R que no está en el ordenador la instalación intentará descargar e instalar el paquete, por lo que es necesario **tener una conexión a Internet activa** cuando se instala un comando de extensión. Si la instalación automática falla, será necesario descargar e instalar manualmente los paquetes de R requeridos por el comando de extensión siguiendo las instrucciones del apéndice de la guía *Cómo instalar manualmente un paquete de R*.

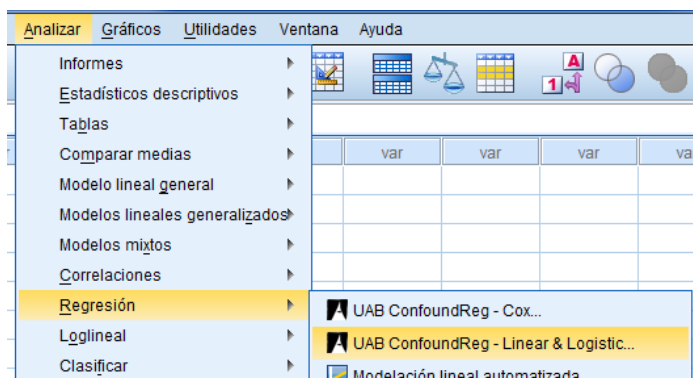
Para que la operativa del nuevo comando de extensión esté totalmente disponible, hay que cerrar y volver a abrir *SPSS Statistics* tras la instalación.

Comprobación de la instalación

Es necesario cerrar SPSS y volverlo a abrir para que aparezca el comando de extensión en los menús.

Abra el menú *Analizar / Regresión y* además de los procedimientos propios de SPSS encontrará el procedimiento *UAB ConfoundReg - Linear & Logistic...* que acaba de instalar.

La mejor prueba para asegurarse de que ha instalado correctamente un comando de extensión es realizar un análisis con dicho comando.



Desinstalar un comando de extensión

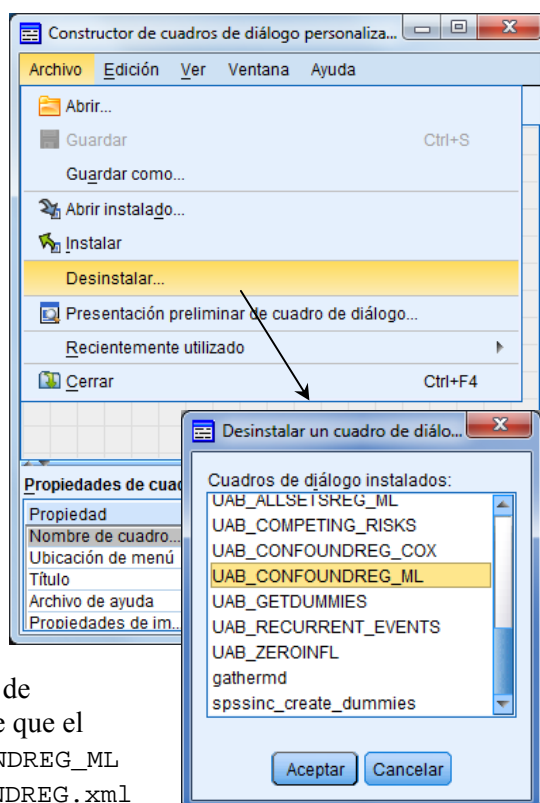
Es posible desinstalar el cuadro de diálogo (y la opción de menú asociada) desde la opción de menú *Utilidades | Cuadros de diálogo personalizados | Constructor de cuadros de diálogo personalizados*.

En la ventana del constructor de cuadros de diálogo escoja la opción de menú *Archivo / Desinstalar*, y en la lista que se abre a continuación escoja el comando de extensión que desea desinstalar.

Cierre SPSS y vuelva a abrirlo para que la desinstalación del cuadro de diálogo tenga efecto.

El proceso indicado desinstala la opción de menú y el cuadro de diálogo, pero el comando de extensión **sigue estando disponible en sintaxis**. Para desinstalar completamente un comando de extensión, de forma que no sea posible utilizarlo en sintaxis, es necesario localizar la carpeta *extensions* dentro de la carpeta de instalación de *SPSS Statistics* (*C:\Program Files\IBM\SPSS\Statistics\19\extensions* para SPSS 19 en Windows 7) y borrar el contenido asociado al comando de extensión: una carpeta y los archivos con el mismo nombre que el comando de extensión. Por ejemplo, para el *UAB_CONFFOUNDREG_ML* hay que borrar el *UAB_CONFFOUNDREG.py*, *UAB_CONFFOUNDREG.xml* y la carpeta *UAB_CONFFOUNDREG_ML*.

Esta operación hay que realizarla con *SPSS Statistics* cerrado.



En los próximos apartados se explica detalladamente el funcionamiento del comando de extensión para la regresión lineal múltiple, y a continuación se detallan las particularidades de funcionamiento para los modelos de regresión logística y de Cox.

Regresión lineal múltiple

Para ilustrar el funcionamiento del comando de extensión en regresión lineal múltiple se emplea el archivo de datos de prueba **DataTest_ConfoundReg_MR.sav**, que contiene 243 casos y las variables:

- X1 (cuantitativa) y X2 (categórica con los valores 1, 2 y 3): actuarán como variables de exposición; la variable X2 sólo se utiliza en el ejemplo 2.
- Z1 (binaria), Z2 y Z3 (cuantitativas) y Z4 (categórica con los valores 1, 2, 3, 4 y 5): actuarán como modificadoras o potenciales confundidoras.
- Y (cuantitativa): actuará como variable de respuesta.

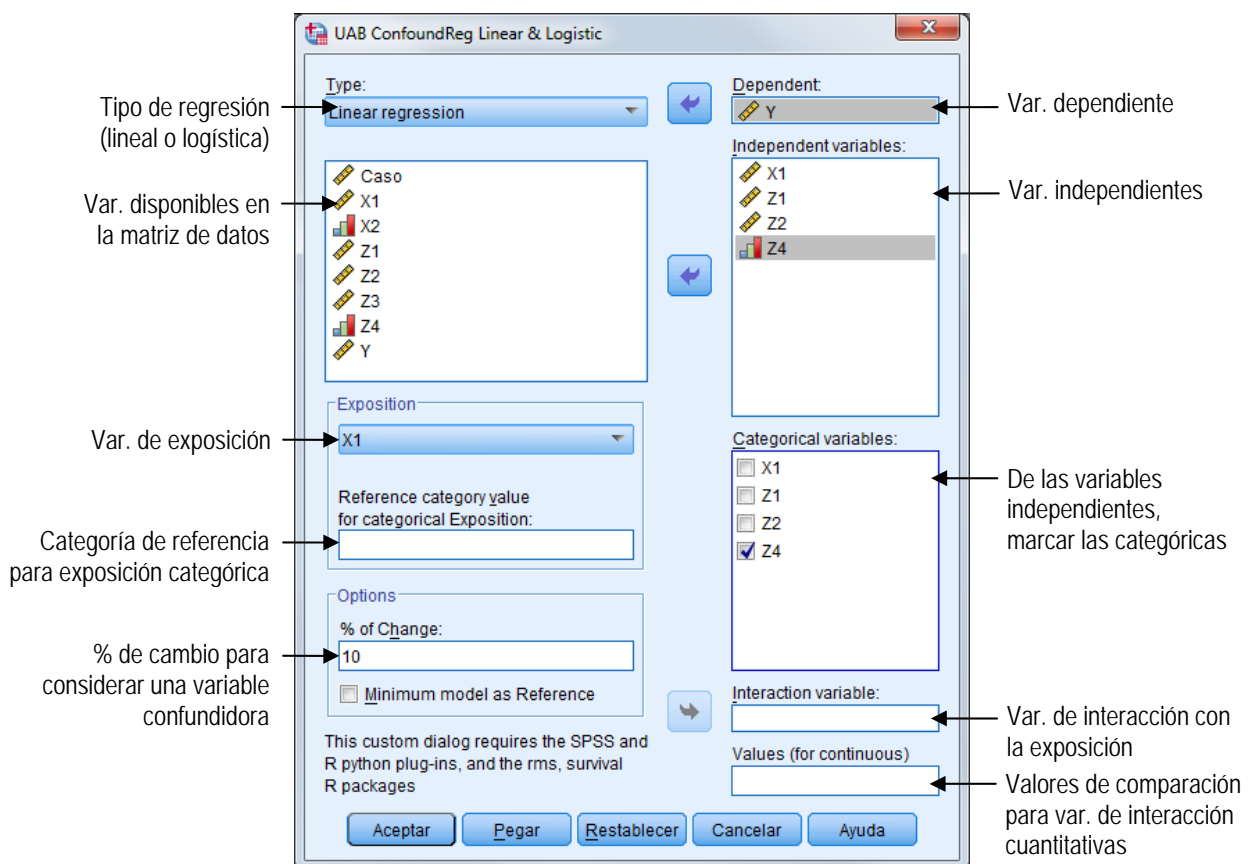
Abra el archivo de datos y ejecute la opción de menú *Analizar | Regresión | UAB ConfoundReg - Linear & Logistic...* Si ha realizado correctamente todo el proceso de instalación aparecerá un cuadro de diálogo con la lista de todas las variables de la ventana de datos.

Ejemplo 1

El primer ejemplo consiste en establecer un modelo de referencia con la exposición cuantitativa (X1) y con los potenciales confundidores Z1 (binario), Z2 (cuantitativo) y Z4 (categórico). No se incluyen términos de interacción.

Puesto que la respuesta modelada es cuantitativa seleccione **Linear regression** como modelo de regresión. Pase la variable dependiente Y a **Dependent**. Todas las variables independientes, incluida la exposición y las variables de interacción, se arrastran al cuadro **Independent Variables** (X1, Z1, Z2 y Z4), y para los predictores que sean categóricos (Z4) se marca su casilla de verificación en el cuadro **Categorical variables**. En la lista **Exposition** se selecciona la variable de exposición (X1) y, en caso de ser categórica, se indica en **Reference category value** el valor que actuará como categoría de referencia al codificar la variable en ficticias. El porcentaje de cambio para considerar una variable confundidora se deja en el 10%.

La siguiente imagen muestra la configuración del cuadro de diálogo.



Para conseguir directamente los resultados puede hacer click en **Aceptar**, si bien recomendamos que haga uso de la sintaxis, que consigue con el botón **Pegar**:

```
UAB CONFOUNDRREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z1 Z2 Z4 EXPOSITION=X1 CATEGORICAL=Z4
/OPTIONS CHANGE=10.
```

Observe que aunque Z1 es binaria no se marca como categórica, ya que las variables binarias se analizan como las cuantitativas.

Tras ejecutar la anterior sintaxis obtendrá un resumen del proceso realizado en la ventana de resultados, indicando los términos del modelo de referencia y el tiempo de ejecución (9 segundos). Puesto que el comando de extensión trabaja con la opción LISTWISE, se genera una tabla con los valores válidos y desconocidos de cada variable y se indica el número de sujetos (con valores válidos en todas las variables) que se han utilizado para efectuar los análisis, en el ejemplo 242, ya que se elimina el caso 10.

```
Extension Command UAB ConfoundReg v0.0.7 (2013.01.18)
(c) JM Domenech & JB Navarro
Programmer: R Sesma
Laboratori d'Estadística Aplicada - Universitat Autònoma de Barcelona

This Extension Command uses functions of the rms and survival R packages.
rms R package (c) Frank E Harrell Jr, http://CRAN.R-project.org/package=rms.
survival R package (c) Terry Therneau and original R port by Thomas Lumley,
http://CRAN.R-project.org/package=survival.
-----
Response: Y
Exposition: X1
Potential confounders: Z1, Z2, Z4
Important change: >= 10%
Type: linear
A new dataset named @4773Results has been created with the results.
Total time: 9 seconds
```

Case Processing Summary

	Y	X1	Z1	Z2	Z4
Valid	243	242	243	243	243
Missing	0	1	0	0	0

Valid number of cases (listwise): 242

Los resultados de la comparación de modelos se presentan en una nueva matriz de datos:

	Nvar	Variables	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	4	X1, Z1, Z2, Z4	43,820	0.0%	28,149	59,490	31,341	,000	.
2	2	X1, Z4	44,220	0.9%	28,664	59,777	31,114	-,228	1
3	3	X1, Z1, Z4	44,585	1.7%	28,931	60,238	31,308	-,034	1
4	3	X1, Z2, Z4	43,422	0.9%	27,841	59,004	31,163	-,178	1
5	1	X1	51,707	18.0%	36,440	66,973	30,533	-,808	0
6	2	X1, Z1	52,137	19.0%	36,819	67,455	30,636	-,705	0
7	2	X1, Z2	50,948	16.3%	35,702	66,194	30,492	-,849	0
8	3	X1, Z1, Z2	51,403	17.3%	36,113	66,693	30,580	-,762	0


La primera fila presenta, para el modelo de referencia que contiene 4 variables (Nvar=4; X1,Z1,Z2,Z4) la estimación del coeficiente B de la exposición (B=43.820), su intervalo de confianza del 95% (lbCI=28.149 a ubCI=59.777) y la amplitud de dicho intervalo (Range=31.341).

Las siguientes filas contienen los modelos reducidos. En primer lugar se encuentran los modelos con cambios en el coeficiente de regresión de X1 inferiores al 10%. En el ejemplo analizado se trata de 3 modelos, que incluyen X1,Z4 ; X1,Z1,Z4 y X1,Z2,Z4. En la nueva matriz de datos estos modelos se identifican porque tienen el valor 1 en la última columna (Select), mientras que los modelos reducidos con cambios en B mayores al 10% se identifican con el valor 0 en la última columna.

Aparte de la misma información que para el modelo de referencia, para cada modelo reducido se añade el porcentaje de cambio del coeficiente B respecto al del modelo de referencia (Change), y en la penúltima columna (Range Diff) la diferencia entre la amplitud del IC95% del modelo reducido y el de referencia.

El primer modelo reducido que proporciona un cambio prácticamente no importante es el modelo que incluye X1,Z4 (cambio en B del 0.9%). Además la estimación de B es más precisa (diferencia de -0.228 respecto al modelo referencia) y es el más parsimonioso, por lo que este modelo reducido se convierte en candidato a ser elegido como mejor modelo explicativo de Y.

También podrían considerarse los modelos X1,Z1,Z4 y X1,Z2,Z4 aunque son menos parsimoniosos y la estimación es ligeramente menos precisa que con el modelo X1, Z4.

 Puesto que en este ejemplo la exposición es cuantitativa y no se han definido interacciones, para cada modelo reducido sólo se debe evaluar el cambio en un coeficiente de regresión. En ejemplos posteriores, cuando la exposición es categórica o se definen interacciones será necesario comparar varios coeficientes en cada modelo reducido.

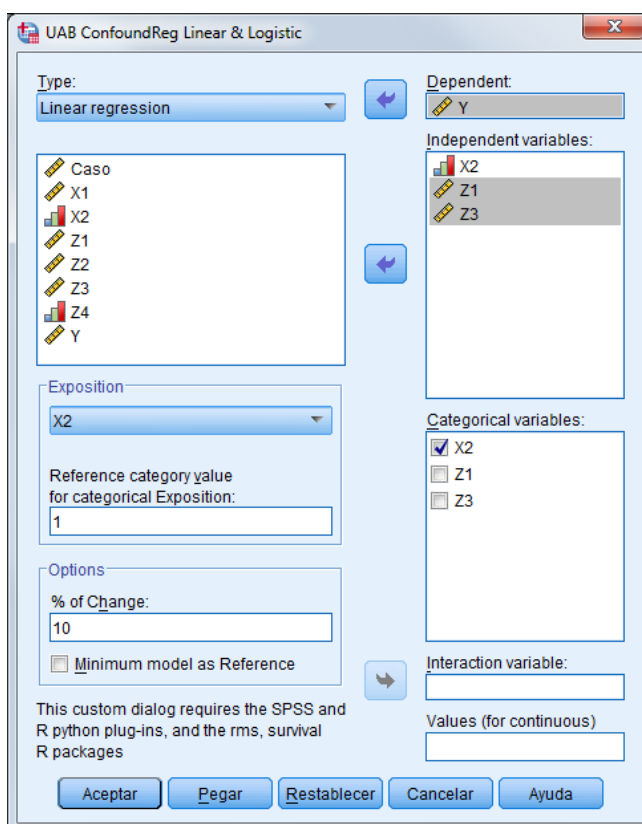
Ejemplo 2

En este segundo ejemplo se establece un modelo de referencia con una exposición con 3 categorías (X2) y con los potenciales confundidores Z1 (binario) y Z3 (cuantitativo). No se incluyen términos de interacción. La imagen muestra la configuración del cuadro de diálogo y la sintaxis que genera el botón **Pegar**.

```
UAB CONFOUNDRREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=X2 Z1 Z3
EXPOSITION=X2 CATEGORICAL=X2
/OPTIONS CHANGE=10 REFCAT=1.
```

La nueva matriz de datos con los resultados contiene un total de 8 líneas.

Puesto que la exposición X2 tiene 3 categorías se generan automáticamente 2 variables ficticias, tomando como categoría de referencia el valor 1 indicado por el usuario. Los nombres de las variables ficticias siguen la nomenclatura propuesta en el texto del curso: X2_21 y X2_31, y se encuentran en la columna Label.



	Model	Nvar	Variables	Label	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	0	3	X2, Z1, Z3	X2_21	-1,139	0.0%	-22,805	20,528	43,333	,000	.
2	0	3	X2, Z1, Z3	X2_31	4,259	0.0%	-26,868	35,386	62,255	,000	.
3	1	1	X2	X2_21	-5,632	394.6%	-27,961	16,697	44,658	1,325	0
4	1	1	X2	X2_31	4,895	14.9%	-27,353	37,142	64,495	2,240	0
5	2	2	X2, Z1	X2_21	-5,898	418.0%	-28,304	16,508	44,812	1,479	0
6	2	2	X2, Z1	X2_31	4,576	7.4%	-27,766	36,918	64,684	2,429	0
7	3	2	X2, Z3	X2_21	-,918	19.4%	-22,506	20,669	43,175	-,158	0
8	3	2	X2, Z3	X2_31	4,514	6.0%	-26,519	35,547	62,066	-,188	0

En este ejemplo, para evaluar la confusión se debe evaluar el cambio en cada modelo reducido respecto al de referencia en los coeficientes de regresión de las dos variables ficticias. Cada modelo ocupa por tanto dos líneas y se identifica en la primera columna (Model), con el valor 0 para el modelo de referencia y números secuenciales a partir de 1 para los modelos reducidos.

Los resultados obtenidos indican que en los 3 modelos reducidos se obtienen cambios superiores al 10% en el coeficiente de la variable ficticia X2_21, por lo que se debe seleccionar el modelo de referencia como mejor modelo explicativo de Y, ya que Z1 y Z3 son necesarias para ajustar la estimación del efecto de X2, aunque sólo sea para una de las dos variables ficticias que la representan.

Sólo en el caso de incrementar el porcentaje de cambio prácticamente importante al 20% se obtendría otra conclusión, ya que entonces se podría seleccionar el modelo reducido 3, con X2,Z3 (excluyendo Z1) porque da estimaciones diferentes al modelo de referencia en 19.4% y 6.0%.

Ejemplo 3

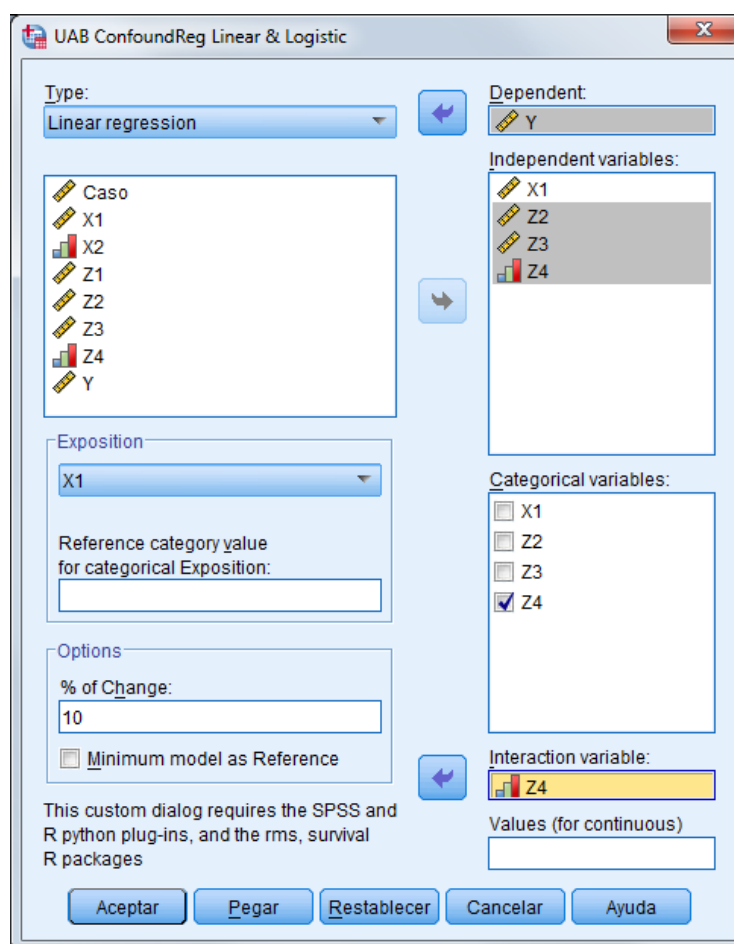
En este tercer ejemplo se establece un modelo de referencia con una exposición cuantitativa (X1), con los potenciales confundidores Z2 y Z3 (cuantitativos) y con la variable de interacción Z4 (categórica).

En primer lugar se deben especificar en el cuadro de diálogo principal las variables implicadas, incluyendo la variable modificadora Z4 como una variable independiente más, marcándola como categórica y pasándola después al control **Interaction variable**.

La imagen muestra la configuración del cuadro de diálogo y la sintaxis que genera el botón **Pegar**.

```
UAB CONFOUNDREG TYPE=LINEAR
DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z2 Z3 Z4
  EXPOSITION=X1          CATEGORICAL=Z4
/INTERACTIONS Z4
/OPTIONS CHANGE=10.
```

En presencia de interacción, la comparación del modelo de referencia y de los modelos reducidos se debe realizar para cada valor de la variable modificadora (Z4).



Al ejecutar el comando de extensión se obtiene en la ventana de resultados una advertencia indicando un problema de multicolinealidad con la variable Z4.

```
Response: Y
Exposition: X1
Potential confounders: Z2, Z3
Interactions: Z4 (5 categories)
Important change: >= 10%
Type: linear

WARNING!!! Variable Z4 must be removed to avoid collinearity.
There are non estimable models.
```

Si aún así intentamos estimar el modelo máximo observamos que el efecto de la exposición no puede ser estimado para los valores 4 y 5 de la variable modificadora Z4. Ante esta situación se puede optar por agrupar las categorías 4 y 5 de Z4. Para ello se crea la nueva variable Z4R con esta agrupación y se ejecuta de nuevo el comando de extensión:

```
RECODE Z4 (4,5=4) (ELSE=COPY) INTO Z4R.
FORMATS Z4R (F1.0).
UAB CONFOUNDREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z2 Z3 Z4R
  EXPOSITION=X1      CATEGORICAL=Z4R
/INTERACTIONS Z4R
/OPTIONS CHANGE=10.
```

Se obtienen los siguientes resultados:

	Model	Nvar	Variables	Label	B	Change	lbCl	ubCl	Range	RangeDiff	Select
1	0	5	X1 Z4R X1*Z4R, Z2, Z3	Z4R = 1	44,138	0.0%	18,916	69,359	50,443	,000	.
2	0	5	X1 Z4R X1*Z4R, Z2, Z3	Z4R = 2	32,543	0.0%	1,269	63,817	62,549	,000	.
3	0	5	X1 Z4R X1*Z4R, Z2, Z3	Z4R = 3	40,163	0.0%	4,942	75,385	70,443	,000	.
4	0	5	X1 Z4R X1*Z4R, Z2, Z3	Z4R = 4	10,205	0.0%	-25,433	45,843	71,276	,000	.
5	3	4	X1 Z4R X1*Z4R, Z3	Z4R = 1	44,866	1.6%	19,615	70,116	50,501	,059	1
6	3	4	X1 Z4R X1*Z4R, Z3	Z4R = 2	32,567	0.1%	1,230	63,903	62,674	,125	1
7	3	4	X1 Z4R X1*Z4R, Z3	Z4R = 3	41,039	2.2%	5,769	76,309	70,540	,097	1
8	3	4	X1 Z4R X1*Z4R, Z3	Z4R = 4	10,661	4.5%	-25,042	46,364	71,407	,131	1
9	1	3	X1 Z4R X1*Z4R	Z4R = 1	51,028	15.6%	25,486	76,569	51,083	,640	0
10	1	3	X1 Z4R X1*Z4R	Z4R = 2	45,609	40.1%	14,546	76,671	62,125	-,424	0
11	1	3	X1 Z4R X1*Z4R	Z4R = 3	47,264	17.7%	11,412	83,116	71,704	1,261	0
12	1	3	X1 Z4R X1*Z4R	Z4R = 4	16,521	61.9%	-19,798	52,840	72,638	1,362	0
13	2	4	X1 Z4R X1*Z4R, Z2	Z4R = 1	50,150	13.6%	24,645	75,655	51,010	,567	0
14	2	4	X1 Z4R X1*Z4R, Z2	Z4R = 2	45,389	39.5%	14,402	76,376	61,974	-,575	0
15	2	4	X1 Z4R X1*Z4R, Z2	Z4R = 3	46,226	15.1%	10,436	82,016	71,581	1,138	0
16	2	4	X1 Z4R X1*Z4R, Z2	Z4R = 4	15,942	56.2%	-20,296	52,179	72,475	1,199	0

Sólo se da un cambio menor al 10% en los 4 coeficientes comparados en el modelo reducido que excluye Z2 y contiene Z3 (indicado con un 1 en la última columna para cada coeficiente comparado). Observe que el modelo reducido con menos términos contiene los 3 términos fijos del modelo: la exposición X1, la interacción X1*Z4R y la modificadora Z4R que debe permanecer en el modelo por el principio jerárquico.

Por tanto, en presencia de interacción con Z4R, el mejor modelo para estudiar la influencia de X1 sobre Y incluye X1,Z4R,X1*Z4R y la variable de ajuste Z3.

Ejemplo 4

En este ejemplo se establece un modelo de referencia con una exposición cuantitativa (X1), con los potenciales confundidores Z1 y Z3 (cuantitativos) y con la variable de interacción Z2 (cuantitativa).

En presencia de interacción con una variable cuantitativa, la comparación del modelo de referencia con los modelos reducidos se realiza, por defecto, para 3 valores de la variable modificadora que corresponden a los percentiles 5, 50 y 95 y que el comando de extensión calcula de forma automática.

En este ejemplo, los percentiles de la variable modificadora Z2 que se usarán para calcular los efectos de la exposición X1 vienen dados en la tabla de la derecha.

Además de estos 3 valores prefijados, al definir la interacción se pueden introducir valores adicionales que se encuentren dentro del rango de valores observados.

Mire dos ejemplos que puede escribir en el recuadro **Values (for continuous)**:

Percentiles		
5	50	95
53,0	68,0	84,0

- "P5 P50 P95 60 75": compara modelos para 5 valores de la variable de interacción, los percentiles 5, 50 y 95, y los valores 60 y 75.
- "60 70 80": compara modelos para 3 valores de la variable de interacción, 60, 70 y 80.

La siguiente imagen reproduce el cuadro de diálogo del comando de extensión para este ejemplo, junto a la sintaxis obtenida. Observe que al definir la interacción $X1*Z2$, además de los tres percentiles, se han añadido los valores 60 y 75.

```
UAB CONFOUNDRREG TYPE=LINEAR
DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z1 Z2 Z3
EXPOSITION=X1
/INTERACTIONS Z2 (P5 P50 P95 60 75)
/OPTIONS CHANGE=10.
```

	Model	Nvar	Variables	Label	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	0	5	X1 Z2 X1*Z2, Z1, Z3	Z2 = 53	39,342	0.0%	11,065	67,619	56,553	,000	.
2	0	5	X1 Z2 X1*Z2, Z1, Z3	Z2 = 60	41,717	0.0%	21,581	61,852	40,271	,000	.
3	0	5	X1 Z2 X1*Z2, Z1, Z3	Z2 = 68	44,430	0.0%	28,691	60,169	31,478	,000	.
4	0	5	X1 Z2 X1*Z2, Z1, Z3	Z2 = 75	46,805	0.0%	27,678	65,931	38,253	,000	.
5	0	5	X1 Z2 X1*Z2, Z1, Z3	Z2 = 84	49,858	0.0%	20,380	79,336	58,956	,000	.
6	3	4	X1 Z2 X1*Z2, Z3	Z2 = 53	38,980	0.9%	10,738	67,221	56,483	-,071	1
7	3	4	X1 Z2 X1*Z2, Z3	Z2 = 60	41,307	1.0%	21,212	61,402	40,191	-,080	1
8	3	4	X1 Z2 X1*Z2, Z3	Z2 = 68	43,967	1.0%	28,281	59,654	31,373	-,105	1
9	3	4	X1 Z2 X1*Z2, Z3	Z2 = 75	46,295	1.1%	27,223	65,366	38,143	-,110	1
10	3	4	X1 Z2 X1*Z2, Z3	Z2 = 84	49,288	1.1%	19,865	78,710	58,845	-,111	1
11	1	3	X1 Z2 X1*Z2	Z2 = 53	47,828	21.6%	19,702	75,955	56,253	-,301	0
12	1	3	X1 Z2 X1*Z2	Z2 = 60	49,157	17.8%	29,384	68,930	39,546	-,725	0
13	1	3	X1 Z2 X1*Z2	Z2 = 68	50,676	14.1%	35,347	66,005	30,658	-,820	0
14	1	3	X1 Z2 X1*Z2	Z2 = 75	52,004	11.1%	32,975	71,034	38,059	-,194	0
15	1	3	X1 Z2 X1*Z2	Z2 = 84	53,713	7.7%	23,928	83,498	59,570	,614	0
16	2	4	X1 Z2 X1*Z2, Z1	Z2 = 53	48,143	22.4%	19,996	76,290	56,293	-,260	0
17	2	4	X1 Z2 X1*Z2, Z1	Z2 = 60	49,536	18.7%	29,737	69,334	39,597	-,674	0
18	2	4	X1 Z2 X1*Z2, Z1	Z2 = 68	51,128	15.1%	35,760	66,495	30,735	-,743	0
19	2	4	X1 Z2 X1*Z2, Z1	Z2 = 75	52,521	12.2%	33,450	71,592	38,142	-,111	0
20	2	4	X1 Z2 X1*Z2, Z1	Z2 = 84	54,312	8.9%	24,485	84,138	59,653	,697	0
21											

Para los 5 valores de interacción analizados los resultados reflejan cambios menores al 10% al excluir del modelo Z1, como indican los valores 1 de la última columna. Por tanto, el modelo de elección sería el que contiene X1, Z2, $X1*Z2$ y la variable de confusión Z3.

Si el modelo de referencia contiene más de una variable modificadora, puede incluirlas todas escribiéndolas en la sintaxis (el cuadro de diálogo sólo permite definir una interacción). Por ejemplo, para incluir las variables de interacción Z2 y Z3, y comparar el modelo de referencia con los modelos reducidos en los percentiles 5, 50 y 95 de Z2, y en los valores 80, 100 y 120 de Z3, debe escribir la siguiente sintaxis:

```
UAB CONFOUNDRREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z1 Z2 Z3
  EXPOSITION=X1
/INTERACTIONS Z2 (P5 P50 P95) Z3 (80 100 120)
/OPTIONS CHANGE=10.
```

La ventana de resultados refleja que cada comparación de modelos se realiza para 9 valores, resultantes de combinar los tres de Z2 con los tres de Z3:

	Model	Nvar	Variables	Label	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 53 & Z3 = 80	18,916	0.0%	-20,386	58,218	78,605	,000	.
2	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 53 & Z3 = 100	43,790	0.0%	14,960	72,620	57,661	,000	.
3	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 53 & Z3 = 120	68,664	0.0%	20,299	117,029	96,730	,000	.
4	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 68 & Z3 = 80	22,688	0.0%	-10,406	55,781	66,187	,000	.
5	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 68 & Z3 = 100	47,562	0.0%	31,310	63,813	32,503	,000	.
6	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 68 & Z3 = 120	72,436	0.0%	31,760	113,112	81,351	,000	.
7	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 84 & Z3 = 80	26,711	0.0%	-16,027	69,449	85,476	,000	.
8	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 84 & Z3 = 100	51,585	0.0%	22,088	81,082	58,994	,000	.
9	0	6	X1 Z2 Z3 X1*Z2 X1*Z3, Z1	Z2 = 84 & Z3 = 120	76,459	0.0%	30,252	122,666	92,413	,000	.
10	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 53 & Z3 = 80	18,075	4.4%	-21,123	57,273	78,396	-,209	1
11	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 53 & Z3 = 100	43,573	0.5%	14,778	72,369	57,591	-,070	1
12	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 53 & Z3 = 120	69,072	0.6%	20,768	117,375	96,608	-,122	1
13	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 68 & Z3 = 80	21,723	4.3%	-11,236	54,681	65,917	-,270	1
14	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 68 & Z3 = 100	47,221	0.7%	31,012	63,430	32,419	-,084	1
15	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 68 & Z3 = 120	72,719	0.4%	32,092	113,347	81,255	-,096	1
16	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 84 & Z3 = 80	25,613	4.1%	-16,980	68,207	85,187	-,289	1
17	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 84 & Z3 = 100	51,112	0.9%	21,672	80,552	58,880	-,114	1
18	1	5	X1 Z2 Z3 X1*Z2 X1*Z3	Z2 = 84 & Z3 = 120	76,610	0.2%	30,451	122,768	92,317	-,096	1

Una última consideración es que el modelo de referencia puede ser muy complejo, dando lugar a un gran número de modelos reducidos y a muchos coeficientes que comparar. Por ejemplo, si se estudia una exposición categórica con 4 categorías (que debe ser codificada en 3 variables ficticias), que interacciona con una variable categórica con 3 categorías y con una variable cuantitativa en la que se definen 3 valores de interacción adicionales a los tres percentiles... ¡cada modelo reducido deberá compararse con el de referencia en un total de 54 coeficientes!

Regresión logística

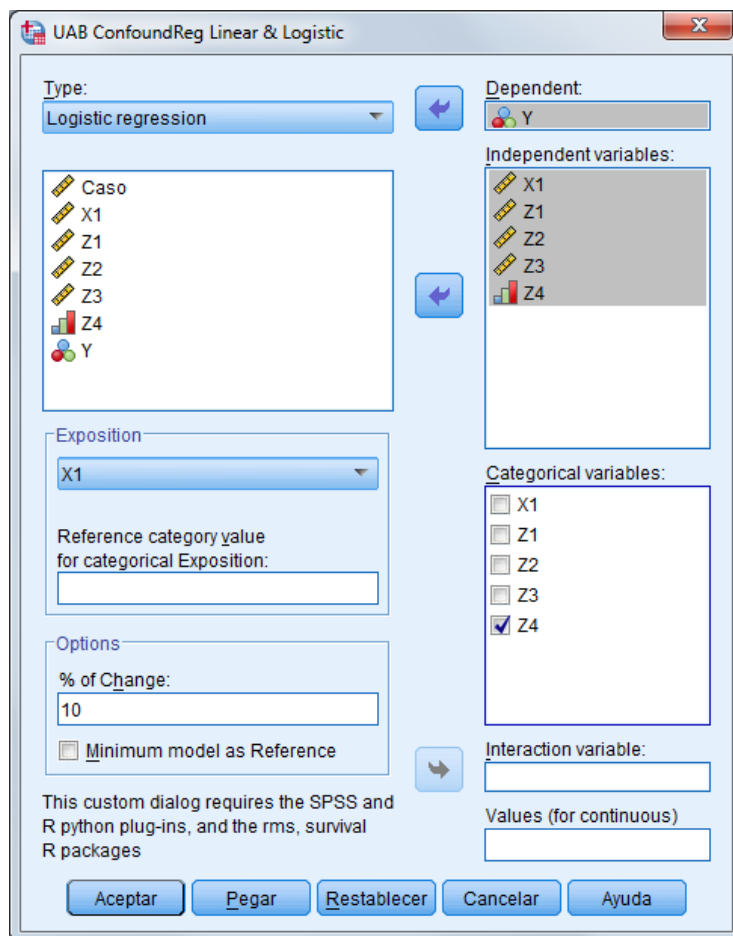
El funcionamiento del comando de extensión en **regresión logística** es muy similar a regresión lineal. Para ilustrarlo usaremos el archivo de datos de prueba **DataTest_ConfoundReg_LR.sav**, que contiene las variables:

- X1 (cuantitativa): actuará como exposición.
- Z1 (binaria), Z2 y Z3 (cuantitativas) y Z4 (categórica con los valores 1, 2 y 3): actuarán como modificadoras/potenciales confundidoras.
- Y (binaria): actuará como respuesta.

Ilustraremos la regresión logística con un único ejemplo en que el modelo de referencia contiene la exposición cuantitativa (X1) y los potenciales confundidores Z1, Z2 y Z3 (cuantitativos) y Z4 (categórico).

Ejecute la opción de menú *Analizar | Regresión | UAB ConfoundReg - Linear & Logistic...*. La imagen muestra la configuración del cuadro de diálogo de acuerdo al ejemplo (no olvide seleccionar **Logistic regression** como tipo de regresión).

```
UAB CONFOUNDRREG TYPE=LOGISTIC
DEPENDENT=Y
/VARIABLES INDEPENDENT=X1 Z1 Z2 Z3 Z4
EXPOSITION=X1
CATEGORICAL=Z4
/OPTIONS CHANGE=10.
```



Al ejecutar la sintaxis se obtiene en la ventana de resultados el habitual resumen del proceso realizado:

```
Extension Command UAB ConfoundReg v0.0.7 (2013.01.18)
(c) JM Domenech & JB Navarro
Programmer: R Sesma
Laboratori d'Estadística Aplicada - Universitat Autònoma de Barcelona

This Extension Command uses functions of the rms and survival R packages.
rms R package (c) Frank E Harrell Jr, http://CRAN.R-project.org/package=rms.
survival R package (c) Terry Therneau and original R port by Thomas Lumley,
http://CRAN.R-project.org/package=survival.
```

```
Response: Y
Exposition: X1
Potential confounders: Z1, Z2, Z3, Z4
Important change: >= 10%
Type: logistic
A new dataset named @9776Results has been created with the results.
Total time: 8 seconds
```

Case Processing Summary

	Y	X1	Z1	Z2	Z3	Z4
Valid	243	242	243	243	242	243
Missing	0	1	0	0	1	0

Valid number of cases (listwise): 241

La nueva matriz de datos con los resultados obtenidos presenta el modelo de referencia y los 15 modelos reducidos resultantes de eliminar:

- Las 4 confundidoras (1 modelo reducido)
- Sólo 3 de las 4 confundidoras (4 modelos reducidos)
- Sólo 2 de las 4 confundidoras (6 modelos reducidos)
- Sólo 1 de las 4 confundidoras (4 modelos reducidos)

	Nvar	Variables	ExpB	Change	lbCI	ubCI	Range	RangeDiff	p_fit_HL	p_fit_CH	Select
1	5	X1, Z1, Z2, Z3, Z4	2,108	0.0%	1,246	3,564	2,318	,000	,965	,016	.
2	3	X1, Z3, Z4	2,117	0.4%	1,257	3,566	2,309	-,008	,503	,005	1
3	4	X1, Z1, Z3, Z4	2,116	0.4%	1,253	3,571	2,318	,000	,423	,005	1
4	4	X1, Z2, Z3, Z4	2,102	0.3%	1,246	3,545	2,299	-,019	,908	,018	1
5	1	X1	3,254	54.4%	2,021	5,238	3,216	,899	,726	,010	0
6	2	X1, Z1	3,294	56.3%	2,042	5,314	3,271	,954	,854	,006	0
7	2	X1, Z2	3,190	51.3%	1,984	5,128	3,144	,826	,512	,238	0
8	2	X1, Z3	2,807	33.2%	1,727	4,564	2,837	,520	,715	,090	0
9	2	X1, Z4	2,563	21.6%	1,540	4,265	2,726	,408	,527	,026	0
10	3	X1, Z1, Z2	3,238	53.6%	2,009	5,217	3,207	,890	,587	,156	0
11	3	X1, Z1, Z3	2,843	34.9%	1,745	4,632	2,887	,569	,982	,064	0
12	3	X1, Z1, Z4	2,565	21.7%	1,538	4,277	2,739	,421	,734	,025	0
13	3	X1, Z2, Z3	2,770	31.4%	1,706	4,499	2,794	,476	,738	,215	0
14	3	X1, Z2, Z4	2,516	19.4%	1,510	4,191	2,681	,364	,900	,278	0
15	4	X1, Z1, Z2, Z3	2,815	33.6%	1,729	4,583	2,854	,536	,656	,157	0
16	4	X1, Z1, Z2, Z4	2,524	19.7%	1,511	4,214	2,703	,385	,840	,259	0

Para cada modelo se presenta el valor Exp(B) de la exposición X1, junto al intervalo de confianza del 95% de Exp(B), la amplitud de dicho IC95%, y el grado de significación de los índices de ajuste de Hosmer-Lemeshow (p_fit_HL) y de le Cessie-van Houwelingen (p_fit_CH).

Para los modelos reducidos se incluye además el porcentaje de cambio de Exp(B) respecto al modelo de referencia, la diferencia en la amplitud del IC95% y una última columna ("Select") que indica si los cambios porcentuales en ese modelo son inferiores al criterio fijado.

Analizando los resultados se observa que en 3 modelos reducidos los cambios en Exp(B) son inferiores al 10% (indicados con el valor 1 en la última columna), en concreto los modelos que incluyen las confundidoras Z3,Z4 ; Z1,Z3,Z4 y Z2,Z3,Z4. Además, los cambios en la amplitud del IC95% de Exp(B) en los 3 modelos reducidos son pequeños en comparación con el modelo de referencia, como se aprecia en la columna "Range Diff". Por tanto, atendiendo a un criterio de parsimonia el modelo finalmente seleccionado es el que incluye la exposición X1 y las confundidoras Z3, Z4, aunque en su contra está el hecho de que la prueba de bondad de ajuste de le Cessie-van Houwelingen ("p_fit_CH") da un resultado estadísticamente significativo.

⚠ Los índices de ajuste de *Hosmer-Lemeshow* y de *le Cessie-van Houwelingen* pueden dar resultados muy diferentes en determinados modelos. Si para el modelo seleccionado uno de ellos es estadísticamente significativo y el otro no, es recomendable realizar un análisis de residuales para garantizar el cumplimiento de los supuestos del modelo.

Regresión de Cox

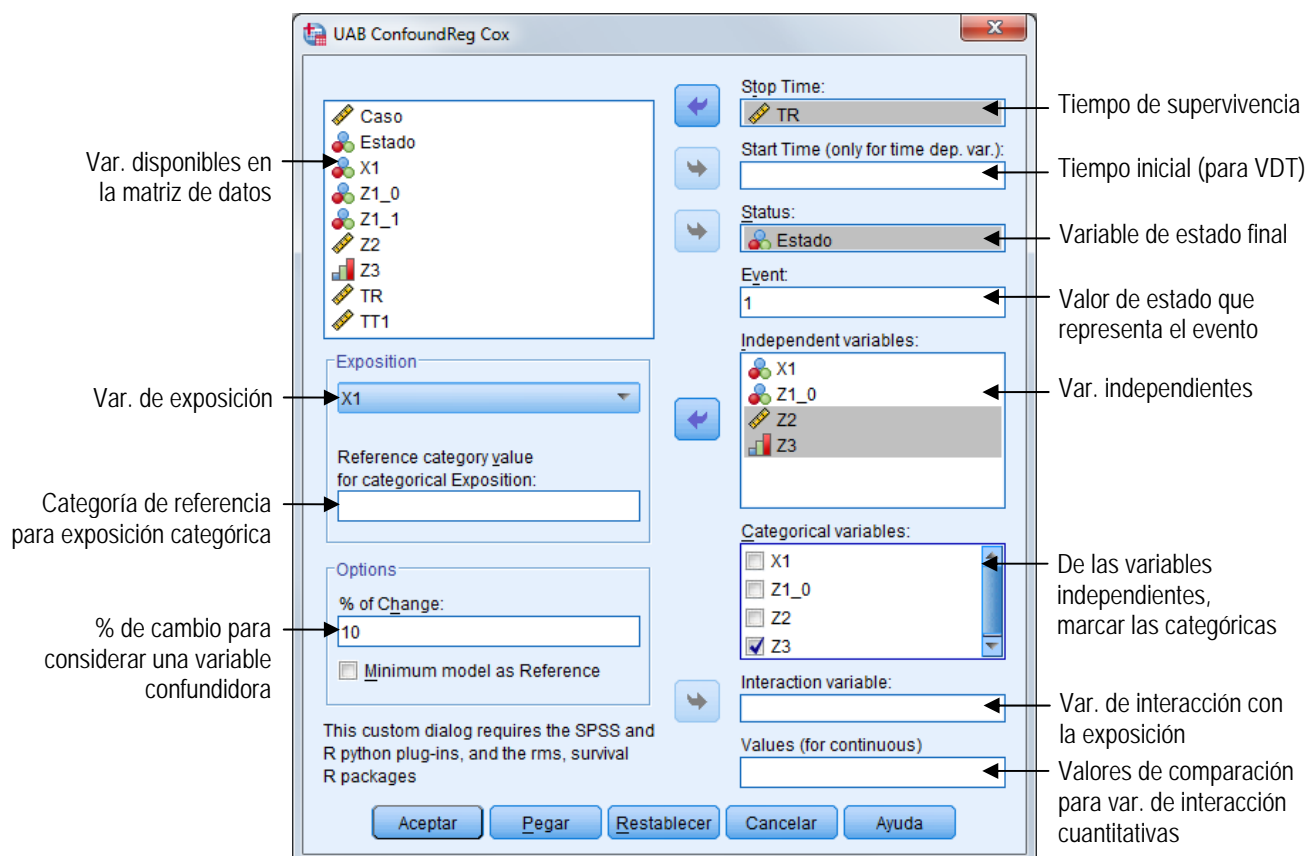
En primer lugar debe **instalar el comando de extensión para regresión de Cox mediante el archivo Uab_ConfoundReg_Cox.spe** que ha descargado junto a esta documentación (en el apartado 'Instalar el comando de extensión' de este mismo documento se explica el procedimiento).

Para practicar emplearemos el archivo de datos de prueba **DataTest_ConfoundReg_Cox.sav** con:

- FE; FU: Fecha de entrada y Fecha de último seguimiento.
- Estado: Estado en el último seguimiento (0:Sin recaída, 1:Recaída, 2:Muerto). El evento de interés es la recaída, mientras que los sujetos sin recaída y los muertos se consideran datos censurados, ya que en el momento del último seguimiento no habían presentado el evento de interés.
- X1: Predictor binario que actuará como variable de exposición.
- Z1_0: Variable binaria dependiente del tiempo. Actuará como potencial confundidora. Esta variable registra el valor inicial.
- Z1_1: Variable binaria dependiente del tiempo. Esta variable registra los cambios que se han producido respecto al valor inicial, que se dan en los casos 2, 3, 8, 12, 17, 27, 32 y 34.
- FZ1_1: Fecha en que se produce un cambio en el predictor binario dependiente del tiempo Z1.
- Z2: Variable binaria. Actuará como potencial confundidora.
- Z3: Variable cuantitativa. Actuará como potencial confundidora.
- TR: Tiempo de supervivencia. Se ha generado con la instrucción:
COMPUTE TR= CTIME.DAYS(FU-FE).
- TT1: Tiempo desde FE hasta FZ1_1 (sólo tiene valor para los 8 casos que cambian en Z1). Se ha generado con la instrucción:
COMPUTE TT1= CTIME.DAYS(FZ1_1 - FE).

En primer lugar practicaremos la selección del mejor modelo explicativo **sin tener en cuenta la variable dependiente del tiempo (VDT) Z1**.

El modelo de referencia incluirá la exposición X1 y los potenciales confundidores Z1_0 (al considerar sólo el valor inicial de Z1 no se tratará como una VDT) Z2 y Z3. Elija la opción de menú *Analizar | Regresión | UAB ConfoundReg - Cox...* Si ha realizado correctamente el proceso de instalación aparecerá un cuadro de diálogo que deberá configurar como muestra la imagen:



Al **Pegar** obtenemos la sintaxis del comando de extensión:

```
UAB CONFOUNDRREG TYPE=COX StopTime=TR Status=Estado Event=1
/VARIABLES INDEPENDENT=X1 Z1_0 Z2 Z3
          EXPOSITION=X1 CATEGORICAL=Z3
/OPTIONS CHANGE=10.
```

En la ventana de resultados se muestra el resumen del proceso:

```
Extension Command UAB ConfoundReg v0.0.7 (2013.01.18)
(c) JM Domenech & JB Navarro
Programmer: R Sesma
Laboratori d'Estadística Aplicada - Universitat Autònoma de Barcelona

This Extension Command uses functions of the rms and survival R packages.
rms R package (c) Frank E Harrell Jr, http://CRAN.R-project.org/package=rms.
survival R package (c) Terry Therneau and original R port by Thomas Lumley,
http://CRAN.R-project.org/package=survival.
-----
Stop Time: TR
Status: Estado (Event value= 1)
Exposition: X1
Potential confounders: Z1_0, Z2, Z3
Important change: >= 10%
Type: cox

A new dataset named @2346Results has been created with the results.
Total time: 6 seconds
```

Case Processing Summary

	TR	Estado	X1	Z1_0	Z2	Z3
Valid	35	35	35	35	35	35
Missing	0	0	0	0	0	0

Valid number of cases (listwise): 35

En la nueva ventana de datos que se crea aparece el resultado de la comparación de modelos. Los tres modelos reducidos que contienen tres variables, es decir que excluyen un predictor, proporcionan cambios inferiores al 10% en la estimación del exponencial de B de X1, por lo que son candidatos a ser elegidos como mejor modelo explicativo. Puesto que en este ejemplo la parsimonia no sirve como criterio para seleccionar entre los tres modelos, nos fijamos en la precisión de la estimación, viendo que el modelo 3, que incluye la exposición X1 y las variables de ajuste Z1_0 y Z3, es el que tiene un **Range** menor (2.485), que a su vez es inferior al del modelo de referencia, y por tanto es el seleccionado.

	Nvar	Variables	ExpB	Change	lbCI	ubCI	Range	RangeDiff	Select
1	4	X1, Z1_0, Z2, Z3	1,146	0.0%	,409	3,210	2,801	,000	.
2	3	X1, Z1_0, Z2	1,138	0.7%	,414	3,124	2,709	-,092	1
3	3	X1, Z1_0, Z3	1,078	5.9%	,403	2,887	2,485	-,317	1
4	3	X1, Z2, Z3	1,057	7.7%	,384	2,915	2,531	-,270	1
5	1	X1	,625	45.5%	,261	1,493	1,232	-1,570	0
6	2	X1, Z1_0	,950	17.1%	,364	2,478	2,114	-,687	0
7	2	X1, Z2	,994	13.2%	,382	2,584	2,201	-,600	0
8	2	X1, Z3	,877	23.4%	,345	2,231	1,886	-,916	0

Variables dependientes del tiempo (VDT)

Para introducir en el modelo de referencia un predictor dependiente del tiempo, es decir cuyos valores puedan cambiar durante el seguimiento, **es necesario cambiar la estructura de la matriz de datos de manera que contenga un registro para cada cambio de valor en la VDT.**

En la matriz de datos de ejemplo hay un total de 35 registros, con 8 sujetos que cambian su valor en el predictor Z1. La matriz de datos reestructurada tendrá 43 registros, con los identificadores de los 8 sujetos duplicados.



Esta forma de estructurar la matriz de datos es la empleada por el programa de análisis estadístico R, con el que internamente se realizan los cálculos de selección del mejor modelo predictivo. Además es la estructura que se obtiene cuando la entrada de datos se ha realizado con un gestor de bases de datos relacional como Microsoft Access.

Vemos con detenimiento los cambios realizados a partir del caso 2, el primero que cambia su valor en Z1. Eliminando las 3 fechas que sirvieron para el cálculo de los tiempos TR y TT1, los datos originales de este caso son:

Caso	Estado	X1	Z1_0	Z1_1	Z2	Z3	TR	TT1
2	1	1	0	1	21	1	960	191

Por tanto se trata de un sujeto que desde el momento inicial (día 0) hasta el día 190 (TT1-1) tiene el valor Z1=0, y en ese intervalo de tiempo no se ha producido el evento. El día 191 cambia al valor Z1=1, permaneciendo en el estudio hasta el día TR=960 en el que se produce el evento (Estado=1).

En la matriz reestructurada el caso 2 se convertirá en dos registros con los siguientes valores:

Caso	Estado	X1	Z1	Z2	Z3	Start	Stop
2	0	1	0	21	1	0	190
2	1	1	1	21	1	191	960

Observe que las dos variables que registraban los valores de Z1_0 y Z1_1 se han agrupado en una única variable Z1, y que los tiempos de supervivencia se indican con las variables Start y Stop. El primer registro contempla desde el día 0 al día 190, con Z1=0 y Estado=0 porque no se produce el evento. El segundo registro contempla desde el día 191 al día 960, con Z1=1 y Estado=1 porque se produce el evento. Los valores en el resto de predictores que no cambian (X1, Z2, Z3) se mantienen idénticos en los dos nuevos registros.

De la misma manera se opera con los otros 7 casos cuyo valor en Z1 cambia. Observe el caso 3, es similar al caso 2 con la diferencia de que el cambio en Z1 es pasar del valor 1 al 0, y de que el Estado en los dos registros reestructurados tiene valor 0 porque en este sujeto no se produce el evento de interés.

Caso	Estado	X1	Z1_0	Z1_1	Z2	Z3	TR	TT1
3	0	1	1	0	26	0	1033	658



Caso	Estado	X1	Z1	Z2	Z3	Start	Stop
3	0	1	1	26	0	0	657
3	0	1	0	26	0	658	1033

En los casos que no cambian de valor en Z1 se debe realizar un pequeño cambio, poniendo el valor Start=0 y el valor Stop en el del tiempo de supervivencia. Por ejemplo, para el caso 1:

Caso	Estado	X1	Z1_0	Z1_1	Z2	Z3	TR	TT1
1	2	0	0	.	18	1	745	.



Caso	Estado	X1	Z1	Z2	Z3	Start	Stop
1	2	0	0	18	1	0	745

En el .zip que ha descargado del campus, la matriz de datos reestructurada se encuentra almacenada en el archivo **DataTest_ConfoundReg_CoxVDT.sav**, ábralo y ejecute la opción de menú *Analizar | Regresión | UAB ConfoundReg - Cox...*

Estimaremos el mejor modelo explicativo de la tasa de riesgo del evento a partir de un modelo explicativo que incluya la exposición X1 y los potenciales confundidores Z1 (tratado como una VDT) Z2 y Z3.

```

UAB CONFOUNDREG TYPE=COX
StopTime=Stop StartTime=Start
Status=Estado Event=1
/VARIABLES INDEPENDENT=X1 Z1 Z2 Z3
            EXPOSITION=X1
            CATEGORICAL=Z3
/OPTIONS CHANGE=10.

```

	Nvar	Variables	ExpB	Change	lbCI	ubCI	Range	RangeDiff	Select
1	4	X1, Z1, Z2, Z3	1,035	0.0%	,370	2,893	2,523	,000	.
2	2	X1, Z2	,998	3.6%	,384	2,594	2,210	-,313	1
3	3	X1, Z1, Z2	,977	5.6%	,375	2,547	2,172	-,351	1
4	3	X1, Z1, Z3	,938	9.4%	,344	2,555	2,211	-,312	1
5	3	X1, Z2, Z3	1,060	2.4%	,385	2,921	2,537	,014	1
6	1	X1	,626	39.5%	,262	1,498	1,235	-1,287	0
7	2	X1, Z1	,694	32.9%	,283	1,702	1,419	-1,104	0
8	2	X1, Z3	,878	15.2%	,345	2,232	1,887	-,636	0

El modelo reducido que incluye la exposición X1 y la variable de ajuste Z2 produce cambios muy pequeños (3.6%) en la estimación de la tasa de riesgo respecto al modelo de referencia. Además su IC95% es menor en 0.313 puntos al de referencia, por lo que se trata del mejor modelo explicativo.

Tratamiento de variables binarias

En este documento se indica que las variables de exposición/confundidoras binarias (codificadas con 0/1) se traten como cuantitativas y por tanto no se marque su casilla de verificación en el cuadro **Categorical variables**, lo cual es correcto para cualquier modelo de regresión. No obstante, las variables binarias también se pueden tratar como categóricas y marcar por tanto la casilla de verificación

El resultado de ambas opciones es idéntico si la categoría de referencia escogida al ser tratadas como categóricas es la codificada con el valor 0, situación habitual en variables binarias. La ventaja de tratar una variable binaria como categórica es que se puede escoger como categoría de referencia la codificada con el valor 1, lo cual no es posible si se trata como cuantitativa.

Se puede comprobar en el siguiente ejemplo con los datos de prueba **DataTest_ConfoundReg_MR.sav**. La variable dependiente es Y, la exposición es Z1 (binaria) y las dos potenciales confundidoras son Z2 y Z3. Los resultados obtenidos con las dos siguientes configuraciones son idénticos:

La exposición Z1 se trata como cuantitativa

```
UAB CONFOUNDRREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=Z1 Z2 Z3
  EXPOSITION=Z1
/OPTIONS CHANGE=10.
```

La exposición Z1 se trata como categórica con el valor 0 como categoría de referencia.

```
UAB CONFOUNDRREG TYPE=LINEAR DEPENDENT=Y
/VARIABLES INDEPENDENT=Z1 Z2 Z3
  EXPOSITION=Z1 CATEGORICAL=Z1
/OPTIONS CHANGE=10 REFCAT=0.
```

Exposición Z1 categórica

	Model	Nvar	Variables	Label	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	0	3	Z1, Z2, Z3	Z1_10	-4,338	0.0%	-24,222	15,547	39,769	,000	.
2	1	1	Z1	Z1_10	-4,124	4.9%	-24,848	16,601	41,450	1,681	1
3	2	2	Z1, Z2	Z1_10	-5,096	17.5%	-25,747	15,554	41,302	1,533	0
4	3	2	Z1, Z3	Z1_10	-3,472	20.0%	-23,403	16,460	39,863	,094	0

Exposición Z1 →
cuantitativa

	Nvar	Variables	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	3	Z1, Z2, Z3	-4,338	0.0%	-24,222	15,547	39,769	,000	.
2	1	Z1	-4,124	4.9%	-24,848	16,601	41,450	1,681	1
3	2	Z1, Z2	-5,096	17.5%	-25,747	15,554	41,302	1,533	0
4	2	Z1, Z3	-3,472	20.0%	-23,403	16,460	39,863	,094	0

La ventaja de poder tratar la variable exposición Z1 como categórica es que permite especificar el valor 1 como categoría de referencia.

En este caso, como la exposición Z1 es binaria, el resultado obtenido es el mismo que con el valor 0, pero con los signos de los coeficientes de regresión cambiados:

Exposition

Z1

Reference category value for categorical Exposition:

1

	Model	Nvar	Variables	Label	B	Change	lbCI	ubCI	Range	RangeDiff	Select
1	0	3	Z1, Z2, Z3	Z1_01	4,338	0.0%	-15,547	24,222	39,769	,000	.
2	1	1	Z1	Z1_01	4,124	4.9%	-16,601	24,848	41,450	1,681	1
3	2	2	Z1, Z2	Z1_01	5,096	17.5%	-15,554	25,747	41,302	1,533	0
4	3	2	Z1, Z3	Z1_01	3,472	20.0%	-16,460	23,403	39,863	,094	0

Referencia de sintaxis de UAB_ConfoundReg

```
UAB CONFOUNDRREG  TYPE={LINEAR } DEPENDENT=varname
                   {LOGISTIC}
                   TYPE={COX      } StopTime=varname [StartTime=varname]
                                     Status=varname Event=integer

/VARIABLES  EXPOSITION=varname
            INDEPENDENT=varlist
            [CATEGORICAL=varlist]
            [/INTERACTIONS interactions list]

/OPTIONS    [CHANGE=10**]
            [REFCAT=int]
            [MINIMUM]

[HELP]
```

- **TYPE:** tipo de regresión. **LINEAR** realiza una regresión lineal, **LOGISTIC** realiza una regresión logística y **COX** realiza una regresión de Cox.
- **DEPENDENT:** variable dependiente (si **TYPE = LINEAR** o **LOGISTIC**). Si **TYPE = LOGISTIC**, esta variable debe ser binaria (0,1).
- **StopTime:** En modelos sin variables dependientes del tiempo es el tiempo de supervivencia. En modelos con variables dependientes del tiempo es el tiempo final de cada intervalo (si **TYPE = COX**).
- **StartTime:** En modelos sin variables dependientes del tiempo este parámetro debe quedar vacío. En modelos con variables dependientes del tiempo es el tiempo de inicio de cada intervalo (si **TYPE = COX**). Requiere que la matriz de datos haya sido preparada por el usuario de acuerdo a la estructura requerida para modelos con variables dependientes del tiempo.
- **Status:** variable de estado final (si **TYPE = COX**).
- **Event:** valor (entero ≥ 0) de la variable de estado que indica el evento (si **TYPE = COX**).
- **INDEPENDENT:** lista de variables independientes. Esta lista incluye la exposición, las variables categóricas y las variables de interacción con la exposición.
- **EXPOSITION:** variable de exposición. Debe ser una de las **INDEPENDENT**.
- **CATEGORICAL:** lista de variables categóricas entre las independientes.
- **INTERACTIONS:** variables de interacción. Las variables de interacción deben ser una de las variables **INDEPENDENT**. En el caso de variables categóricas, el cambio de los coeficientes se evalúa para cada categoría. En el caso de variables cuantitativas, el cambio se evalúa para los valores indicados por el usuario, siendo por defecto los valores correspondientes a los percentiles 5, 50 y 95. Para indicar manualmente los valores, es necesario indicarlo en sintaxis. Por ejemplo, **INTERACTIONS Edad (P5 P25 P50 34 56)** utiliza los valores correspondientes a los percentiles 5, 25 y 50, además de los valores 34 y 56 de la variable Edad.
- **CHANGE:** porcentaje de cambio del efecto que se considera prácticamente importante. Por defecto, 10%.
- **REFCAT:** valor de la categoría de referencia cuando la exposición es categórica.
- **MINIMUM:** si presente, indica que el modelo de referencia es el mínimo, el que sólo contiene la exposición. En este caso no se permiten interacciones ni exposiciones categóricas.

La variable **DEPENDENT** no puede aparecer en **INDEPENDENT**. Ningún nombre de variable puede contener los caracteres @, \$ o #.