

UAB ZeroInfl

Regresión de Poisson y binomial negativa con ceros inflados

Comando de extensión (*Extension Command*) para calcular un modelo de regresión con ceros inflados (*Zero-inflated Regression*).

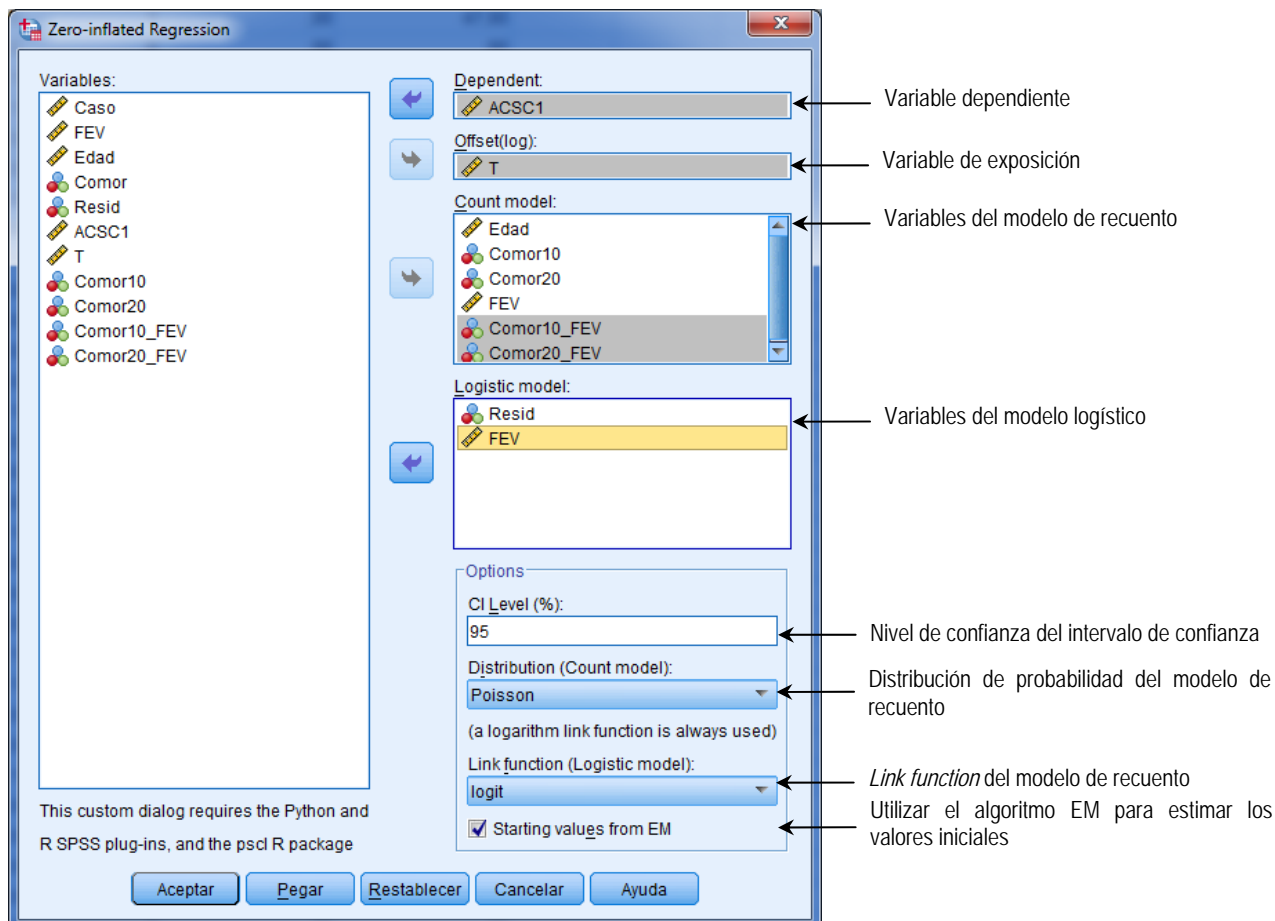
Instalación

Los comandos de extensión requieren como mínimo la versión 18 de SPSS. Antes de instalar *UAB ZeroInfl* es necesario instalar los complementos de integración de *Python (Essentials for Python)* y de *R (R Essentials)*. Una vez se cumplen los requisitos previos, instalar *UAB ZeroInfl* es muy fácil. Ejecutando SPSS como administrador, se ejecuta la opción de menú "Utilidades | Grupos de extensión | Instalar grupo de extensión..." de la ventana de datos de SPSS. En el cuadro de diálogo que se abre debe indicar la dirección en disco del archivo .spe correspondiente. Es conveniente cerrar y reiniciar SPSS para tener la operativa completa.

Atención: En el enlace "Material" del campus virtual hay un documento (*Guia instalación Python y R.pdf*) que explica el proceso de instalación de los complementos de integración y de los comandos de extensión. Siga detalladamente las instrucciones para asegurar que realiza correctamente la instalación.

Ejecución

El comando de extensión funciona como cualquier otro procedimiento SPSS. Es accesible a través de la opción de menú "Analizar | Regresión | UAB Zero-Inflated Regression...", que abre un cuadro de diálogo que, una vez configurado, puede ejecutar directamente o pegar la sintaxis. Con los datos del archivo de ejemplo *ZeroInfl_Test.sav*, que incluye las ficticias que representan *Comor* y su interacción con *FEV*:



Como resultado se obtiene el siguiente listado, que comienza con un resumen del modelo estimado, que incluye el logaritmo de su verosimilitud, los términos del submodelo de recuento y los términos del submodelo logístico. La primera tabla de resultados presenta un resumen del número de casos válidos y con valor *missing* para cada variable. En las dos tablas siguientes se presenta, para cada coeficiente de regresión de cada submodelo, su valor estimado, su error estándar, la prueba de Wald con sus grados de libertad y su grado de significación, la exponencial del coeficiente y su intervalo de confianza.

```
UAB ZEROINFL DEP=ACSC1 OFFSET= T
```

```
  COUNT=Edad Comor10 Comor20 FEV Comor10_FEV Comor20_FEV
```

```
  LOGISTIC=Resid FEV
```

```
  /OPTIONS CL=95 DIST=POISSON LINK= LOGIT EM= TRUE.
```

UAB ZEROINFL

Extension Command UAB ZeroInfl v0.1.0 (2012.06.07)

(c) JM Domenech & JB Navarro

Programmer: R Sesma

Laboratori d'Estadística Aplicada - Universitat Autònoma de Barcelona

This Extension Command uses the `pscl.zeroinfl` R function:

(c) Achim Zeileis, <http://rss.acs.unt.edu/Rdoc/library/pscl/html/zeroinfl.html>.

Dependent variable: ACSC1

Offset variable: T

Count model (poisson with log link): Edad Comor10 Comor20 FEV Comor10_FEV Comor20_FEV

Logistic model (binomial with logit link): Resid FEV

Starting values from EM

CI Level(%): 95

Log likelihood: -128.4652

Residual degree of freedom: 65

Case Processing Summary

| | ACSC1 | T | Edad | Comor10 | Comor20 | FEV | Comor10_FEV | Comor20_FEV | Resid |
|---------|-------|----|------|---------|---------|-----|-------------|-------------|-------|
| Valid | 75 | 76 | 76 | 76 | 76 | 76 | 76 | 76 | 76 |
| Missing | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Valid number of cases (listwise): 75

Count model coefficients

| | B | SE(B) | Wald | df | Sig. | Exp(B) | 95% CI low | 95% CI upp |
|-------------|--------|--------|--------|----|------|--------|------------|------------|
| Constant | -7,705 | 1,5615 | 24,351 | 1 | ,000 | ,000 | 2,111E-5 | ,010 |
| Edad | ,047 | ,0138 | 11,474 | 1 | ,001 | 1,048 | 1,020 | 1,076 |
| Comor10 | -,474 | ,7741 | ,375 | 1 | ,540 | ,623 | ,137 | 2,838 |
| Comor20 | 1,083 | ,9785 | 1,224 | 1 | ,268 | 2,953 | ,434 | 20,095 |
| FEV | -,009 | ,0110 | ,624 | 1 | ,429 | ,991 | ,970 | 1,013 |
| Comor10_FEV | ,014 | ,0144 | ,896 | 1 | ,344 | 1,014 | ,986 | 1,043 |
| Comor20_FEV | -,017 | ,0176 | ,900 | 1 | ,343 | ,983 | ,950 | 1,018 |

Logistic model coefficients


| | B | SE(B) | Wald | df | Sig. | Exp(B) | 95% CI low | 95% CI upp |
|----------|--------|--------|-------|----|------|--------|------------|------------|
| Constant | -5,742 | 2,6698 | 4,625 | 1 | ,032 | ,003 | 1,714E-5 | ,601 |
| Resid | ,728 | 1,0834 | ,451 | 1 | ,502 | 2,070 | ,248 | 17,306 |
| FEV | ,066 | ,0311 | 4,467 | 1 | ,035 | 1,068 | 1,005 | 1,135 |

Si se estima un modelo de regresión binomial negativa para ceros inflados, se obtiene también la estimación del logaritmo del parámetro de dispersión (log theta).

Es posible estimar un modelo sin variable de OFFSET y sin variables predictoras (sólo la constante) en el submodelo de recuento y en el logístico. La sintaxis para este modelo, el más simple que se puede plantear, sería:

```
UAB ZEROINFL DEP=ACSC1.
```

Referencia de sintaxis

| |
|--|
|  Analizar Regresión UAB Zero-inflated regression... |
| <pre>UAB ZEROINFL DEP=var [OFFSET=var] [COUNT=list_var] [LOGISTIC=list_var] /OPTIONS [CL=95**] [DIST= {POISSON**}] [LINK={LOGIT**}] [EM={TRUE**}] {NEGBIN } {PROBIT } {FALSE } {GEOMETRIC} {CLOGLOG} {CAUCHIT} {LOG } [HELP] ** Valor por defecto</pre> |

- DEP: Variable **dependiente**.
- OFFSET: Variable de **exposición** (se utilizará el log de la variable).
- COUNT: Variable o lista de variables del modelo de **recuento**.
- LOGISTIC: Variable o lista de variables del modelo **logístico**.
Las variables de COUNT y LOGISTIC pueden coincidir. Las variables DEP y OFFSET no pueden aparecer en COUNT y LOGISTIC. Ningún nombre de variable puede contener los caracteres @ o \$, ya que son caracteres reservados en la sintaxis R. Se supone que las variables ficticias (correspondientes a categóricas o interacciones) necesarias para evaluar el modelo ya se han generado (por ejemplo, con el *extension command* UAB GETDUMMIES).
- CL: **Nivel de confianza** del intervalo de confianza (en porcentaje). Por defecto, 95%.
- DIST: **Distribución** de probabilidad del modelo de recuento. Puede ser POISSON (**defecto**), NEGBIN, o GEOMETRIC. Se utilizará una *link function* logarítmica.
- LINK: **Link function** del modelo logístico. Puede ser LOGIT (**defecto**), PROBIT, CLOGLOC, CAUCHIT o LOG.
- EM: Si EM = TRUE (**defecto**) se utilizará el algoritmo EM para estimar los valores iniciales de los parámetros. Si EM = FALSE los parámetros se estimarán por mínimos cuadrados.