

## Regresión logística

- Puede considerarse un caso especial de l análisis de regresión en donde la variable dependiente es dicotómica («Sí» [1] o «No» [0])
  - ➡ Se trata de calcular la probabilidad en la que una de las opciones de la variable dicotómica dependiente sucederá en función de cómo puntúa en una serie de variables dependientes que pueden estar en diferentes escalas de medida.
- Es por tanto, una técnica multivariante de dependencia ya que trata de estimar la probabilidad de que ocurra un suceso en función de la dependencia de otras variables.
  - ➡ Si tenemos un conjunto de variables independientes  $X_1, X_2, \dots, X_p$  que nos clasificas a los n sujetostrataremos de saber a cual de las dos categorías de la variable  $Y$  pertenece. La probabilidad de que un sujeto « $i$ » pertenezca a una de ellas será la combinación lineal  $Z = b_1 x_1 + b_2 x_2 + \dots + b_p x_p + b_0$  y será igual a:

Por lo que para el sujeto « $i$ »:

$$p_i = \frac{e^z}{1 + e^z}$$

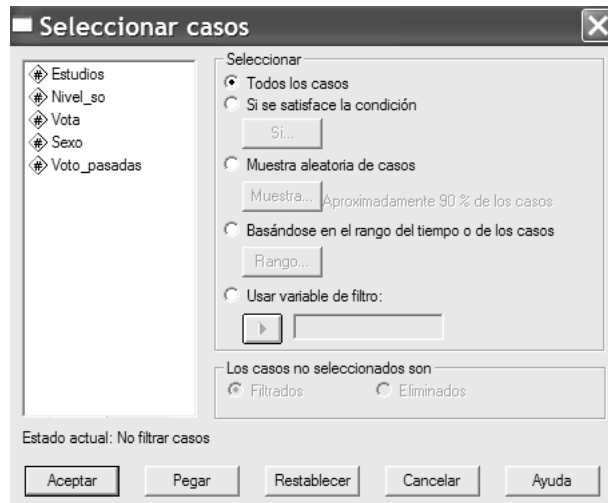
Que es igual a:

$$p_i = \frac{1}{1 + e^{-z}}$$

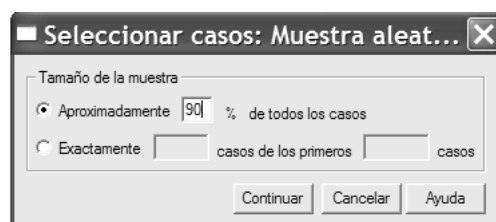
$$p_i = \frac{1}{1 + e^{-(b_1 x_1 + \dots + b_p x_p + b_0)}}$$

- Si la probabilidad  $p_i$  de que el sujeto este encuadrado en esa categoría es mayor que 0,5 se le asigna, si es menor se le asignara la otra categoría
  - ➡ Se trata de predecir esta probabilidad ya no sólo para los sujetos observados (los de la muestra que están en los datos) sino para la población en general. Para ello se calcula la función  $Z$  considerando una muestra aleatoria de nuestra muestra, después se realiza sobre la totalidad y se mide si el número de aciertos es suficientemente elevado. Esto constituye una prueba de bondad de su capacidad de predicción.

- La mecánica con SPSS comienza por ir a «**Datos** → **Seleccionar**» y activar «**Muestra**»

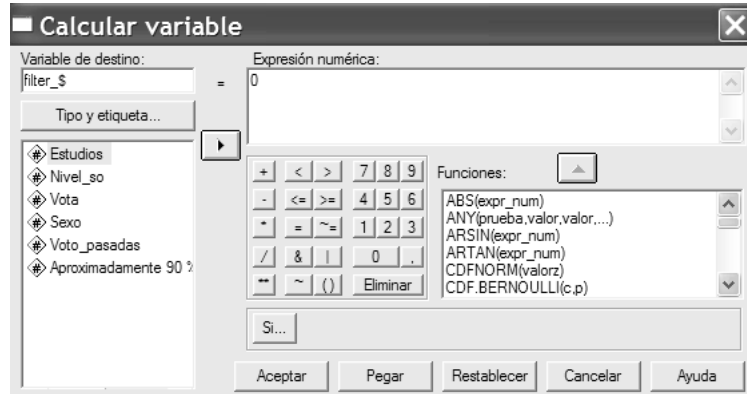


- Una vez allí seleccionamos «**Aproximadamente**» y escribimos el 90% de todos los casos como muestra el gráfico siguiente:



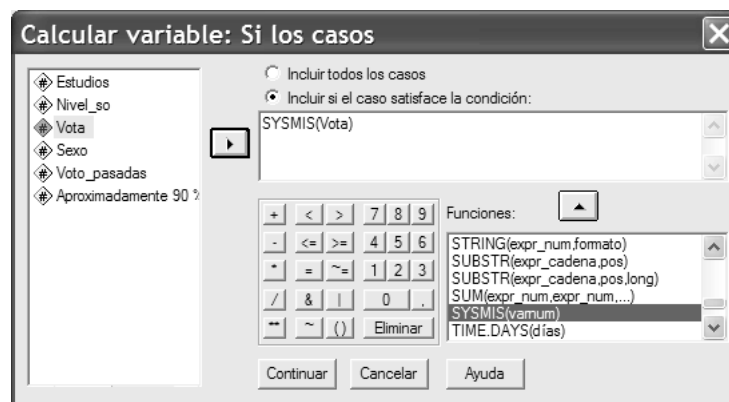
- El programa crea una variable nueva llamada «**filter\_\$**» que rellenará con los valores (1 y 0) según seleccione o no cada caso
- Una vez que tenemos el filtro volvemos a seleccionar todos los datos llenando a «**Datos** → **Seleccionar** → **Todos los datos**» .

- Como para algunos pueden haber valores perdidos en la variable dependiente, conviene eliminarlos mediante una operación que ponga a cero los valores perdidos en la variable «*filter\_\$*», para ello vamos a «*Transformar → Calcular*»

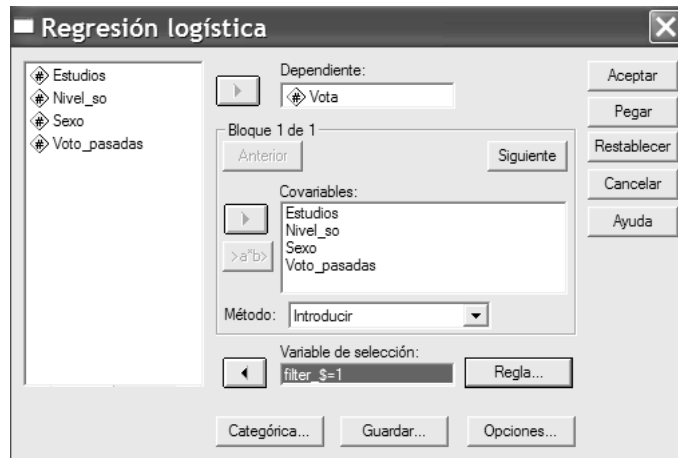


- Allí escribimos como variable destino la «*filter\_\$*» y ponemos en «*Expresión numérica*» un 0 activamos la condición «*Si...*»

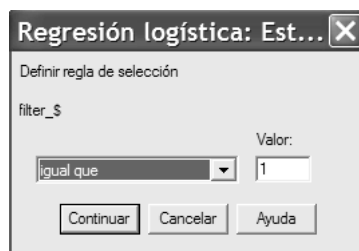
- Buscamos la función «*SYSMIS (Vota)*» que constituirá la condición *Si...* y damos a «*Continuar*» y después a «*Aceptar*» con ello se eliminan de la selección los casos en blanco que pretendíamos quitar del filtro



- Ahora estamos en condiciones de hacer la Regresión Logística llenando a «**Analizar**→**Regresión** →**Logística**» e introducimos la variable «**Dependiente**» (*Vota*) y las Covariables (*Estudios* y *Nivel so*) así como el filtro «**Variable de selección:**» el (*filter\_\$*)

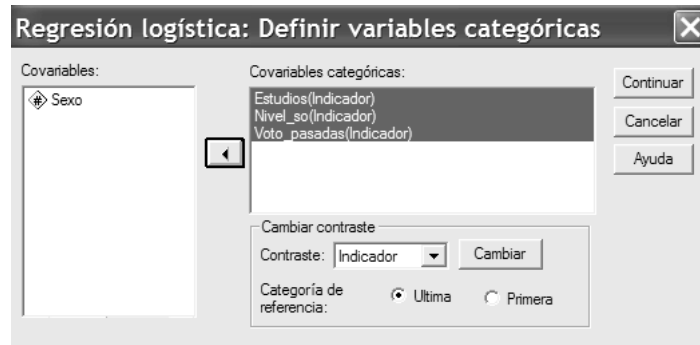


- Para activar la condición de *Filter\_\$=1* pulsamos sobre la opción «**Regla...**»



- En el cuadro que se abre buscamos la condición «**igual que**» e introducimos en «**Valor**» (*1*). Damos a «**Continuar**» y después a «**Aceptar**». Produciéndome a continuación la salida SPSS correspondiente a este análisis

- Como las variables *Nivel\_so* , *Estudios* y *Voto\_pasadas* son categóricas, no están en una escala numérica realmente, podemos pedir que nos recodifique cada categoría de éstas en una variable binaria seleccionando la opción «*Catagórica*» y damos «*indicador*» y «*cambiar*»



- Finalmente estas variables las analizará como *Nivel\_so ( )* , *Nivel\_so (1)* y *Nivel\_so (2)*. De forma parecida hará *con Estudios ( )* , *Estudios(1)* , *Estudios(2)* y *Estudios (3)* y *Voto\_pasadas( )* , *Voto\_pasadas(1)* , *Voto\_pasadas(2)*



- La salida de SPSS nos informa de los casos ponderados y procesados, de los perdidos así como de la codificación de la variable dependiente

→ **Regresión logística**

Resumen del procesamiento de los casos

| Casos no ponderados <sup>a</sup> |                          | N   | Porcentaje |
|----------------------------------|--------------------------|-----|------------|
| Casos seleccionados              | Incluidos en el análisis | 284 | 92,2       |
|                                  | Casos perdidos           | 0   | ,0         |
|                                  | Total                    | 284 | 92,2       |
| Casos no seleccionados           |                          | 24  | 7,8        |
| Total                            |                          | 308 | 100,0      |

a. Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

Codificación de la variable dependiente

| Valor original | Valor interno |
|----------------|---------------|
| No             | 0             |
| Sí             | 1             |

- A continuación nos muestra la tabla con las codificaciones que se han hecho en las variables categóricas y las frecuencias de cada una de las categorías

Codificaciones de variables categóricas

|              |                      | Frecuencia | Codificación de parámetros |       |       |
|--------------|----------------------|------------|----------------------------|-------|-------|
|              |                      |            | (1)                        | (2)   | (3)   |
| Estudios     | < Que básicos        | 103        | 1,000                      | ,000  | ,000  |
|              | Básicos obligatorios | 92         | ,000                       | 1,000 | ,000  |
|              | Post-básicos         | 65         | ,000                       | ,000  | 1,000 |
|              | universitarios       | 24         | ,000                       | ,000  | ,000  |
| Voto_pasadas | 1                    | 73         | 1,000                      | ,000  |       |
|              | 2                    | 134        | ,000                       | 1,000 |       |
|              | 3                    | 77         | ,000                       | ,000  |       |
| Nivel_so     | baja                 | 54         | 1,000                      | ,000  |       |
|              | Media                | 176        | ,000                       | 1,000 |       |
|              | Alta                 | 54         | ,000                       | ,000  |       |

- Pasa a continuación a mostrar los resultados de la primera etapa «*Bloque inicial*» donde incorporará las variable que mayor información aporte.

- En el bloque inicial nos muestra, en primer lugar una tabla de probabilidad de que ocurra el suceso tanto en los casos seleccionados como en los no seleccionados. Analizando seguidamente las variables que entran en la ecuación. El estadístico «*Wald*», nos muestra que es significativo ( $<0,05$ )

**Bloque 0: Bloque inicial**

Tabla de clasificación<sup>c,d</sup>

| Observado         |      | Pronosticado                     |     |                     |                                     |       |                     |    |       |
|-------------------|------|----------------------------------|-----|---------------------|-------------------------------------|-------|---------------------|----|-------|
|                   |      | Casos seleccionados <sup>a</sup> |     |                     | Casos no seleccionados <sup>b</sup> |       |                     |    |       |
|                   |      | Vota                             |     | Porcentaje correcto | Vota                                |       | Porcentaje correcto |    |       |
| No                | Sí   | No                               | Sí  |                     |                                     |       |                     |    |       |
| Paso 0            | Vota | No                               | 171 | Sí                  | 0                                   | 100,0 | 18                  | 0  | 100,0 |
|                   |      | Sí                               | 113 | 0                   | ,0                                  | 6     | 0                   | ,0 |       |
| Porcentaje global |      |                                  |     |                     | 60,2                                |       |                     |    | 75,0  |

a. Casos seleccionados Aproximadamente 90 % de los casos (MUESTRA) EQ 1

b. Casos no seleccionados Aproximadamente 90 % de los casos (MUESTRA) NE 1

c. En el modelo se incluye una constante.

d. El valor de corte es ,500

Variables en la ecuación

|                  | B     | E.T. | Wald   | gl | Sig. | Exp(B) |
|------------------|-------|------|--------|----|------|--------|
| Paso 0 Constante | -.414 | ,121 | 11,677 | 1  | ,001 | ,661   |

- Seguidamente nos muestra las puntuaciones Rao así como los grados de libertad de todas las variables y su nivel de significación. Por tanto la más significativa como candidata a entrar en la ecuación es la variable *Estudios*, seguida del *Nivel\_so*

Variables que no están en la ecuación

| Paso                  | Variables       | Puntuación | gl | Sig. |
|-----------------------|-----------------|------------|----|------|
| 0                     | Estudios        | 140,665    | 3  | ,000 |
|                       | Estudios(1)     | 128,863    | 1  | ,000 |
|                       | Estudios(2)     | 9,039      | 1  | ,003 |
|                       | Estudios(3)     | 47,420     | 1  | ,000 |
|                       | Nivel_so        | 47,084     | 2  | ,000 |
|                       | Nivel_so(1)     | 10,551     | 1  | ,001 |
|                       | Nivel_so(2)     | 7,508      | 1  | ,006 |
|                       | Sexo            | 4,557      | 1  | ,033 |
|                       | Voto_pasadas    | 5,984      | 2  | ,050 |
|                       | Voto_pasadas(1) | 4,982      | 1  | ,026 |
|                       | Voto_pasadas(2) | ,167       | 1  | ,683 |
| Estadísticos globales |                 | 164,150    | 8  | ,000 |

- En el paso siguiente incorpora la variable más significativa al modelo y nos muestra los coeficientes de eficiencia del mismo  
**Bloque 1: Método = Por pasos hacia adelante (Wald)**

En este caso las pruebas y coeficientes determinan que no es necesario seguir desarrollando nuevas etapas donde ya que al incorporar nuevos elementos el modelo no gana en valorar la probabilidad

Pruebas omnibus sobre los coeficientes del modelo

|        |        | Chi-cuadrado | gl | Sig. |
|--------|--------|--------------|----|------|
| Paso 1 | Paso   | 164,002      | 3  | ,000 |
|        | Bloque | 164,002      | 3  | ,000 |
|        | Modelo | 164,002      | 3  | ,000 |
| Paso 2 | Paso   | 76,815       | 2  | ,000 |
|        | Bloque | 240,817      | 5  | ,000 |
|        | Modelo | 240,817      | 5  | ,000 |

Resumen de los modelos

| Paso | -2 log de la verosimilitud | R cuadrado de Cox y Snell | R cuadrado de Nagelkerke |
|------|----------------------------|---------------------------|--------------------------|
| 1    | 217,776 <sup>a</sup>       | ,439                      | ,593                     |
| 2    | 140,962 <sup>a</sup>       | ,572                      | ,773                     |

a. La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

- A continuación nos muestra la tabla de probabilidad en donde la capacidad de predicción de la variable dependiente es del 90,5% en los casos seleccionados y del 79,2% para los no seleccionados, teniendo en cuenta las variables que entran en el modelo que presenta a continuación

Tabla de clasificación<sup>c</sup>

| Observado |                   | Pronosticado                     |     |                     |                                     |    |                     |      |
|-----------|-------------------|----------------------------------|-----|---------------------|-------------------------------------|----|---------------------|------|
|           |                   | Casos seleccionados <sup>a</sup> |     |                     | Casos no seleccionados <sup>b</sup> |    |                     |      |
|           |                   | Vota                             |     | Porcentaje correcto | Vota                                |    | Porcentaje correcto |      |
| No        | Sí                | No                               | Sí  |                     |                                     |    |                     |      |
| Paso 1    | Vota              | No                               | 154 | 17                  | 90,1                                | 14 | 4                   | 77,8 |
|           |                   | Sí                               | 27  | 86                  | 76,1                                | 1  | 5                   | 83,3 |
|           | Porcentaje global |                                  |     |                     | 84,5                                |    |                     | 79,2 |
| Paso 2    | Vota              | No                               | 163 | 8                   | 95,3                                | 15 | 3                   | 83,3 |
|           |                   | Sí                               | 19  | 94                  | 83,2                                | 2  | 4                   | 66,7 |
|           | Porcentaje global |                                  |     |                     | 90,5                                |    |                     | 79,2 |

a. Casos seleccionados Aproximadamente 90 % de los casos (MUESTRA) EQ 1

b. Casos no seleccionados Aproximadamente 90 % de los casos (MUESTRA) NE 1

c. El valor de corte es ,500



- Las variables que entran en el modelo final son los *Estudios* y el *Nivel\_so*, mostrando los coeficientes  $b_0, b_1, \dots, b_p$  así como los Walds, grados de libertad y significación

Variables en la ecuación

|        |             | B       | E.T.     | Wald   | gl | Sig. | Exp(B)  |
|--------|-------------|---------|----------|--------|----|------|---------|
| Paso 1 | Estudios    |         |          | 77,993 | 3  | ,000 |         |
|        | Estudios(1) | 22,824  | 8204,331 | ,000   | 1  | ,998 | 8,2E+09 |
|        | Estudios(2) | 20,217  | 8204,331 | ,000   | 1  | ,998 | 6,0E+08 |
|        | Estudios(3) | 17,753  | 8204,331 | ,000   | 1  | ,998 | 5,1E+07 |
|        | Constante   | -21,203 | 8204,331 | ,000   | 1  | ,998 | ,000    |
| Paso 2 | Estudios    |         |          | 62,343 | 3  | ,000 |         |
|        | Estudios(1) | 25,336  | 6932,795 | ,000   | 1  | ,997 | 1,0E+11 |
|        | Estudios(2) | 21,997  | 6932,795 | ,000   | 1  | ,997 | 3,6E+09 |
|        | Estudios(3) | 17,679  | 6932,795 | ,000   | 1  | ,998 | 4,8E+07 |
|        | Nivel_so    |         |          | 10,272 | 2  | ,006 |         |
|        | Nivel_so(1) | 24,816  | 4477,795 | ,000   | 1  | ,996 | 6,0E+10 |
|        | Nivel_so(2) | 21,395  | 4477,795 | ,000   | 1  | ,996 | 2,0E+09 |
|        | Constante   | -44,505 | 8253,133 | ,000   | 1  | ,996 | ,000    |

a. Variable(s) introducida(s) en el paso 1: Estudios.

b. Variable(s) introducida(s) en el paso 2: Nivel\_so.

- Finalmente nos muestra aquellas variables que no entran en el modelo ya que no contribuyen a explicar el comportamiento de la variable dependiente

Variables que no están en la ecuación

|        |           |                       | Puntuación | gl | Sig. |
|--------|-----------|-----------------------|------------|----|------|
| Paso 1 | Variables | Nivel_so              | 57,798     | 2  | ,000 |
|        |           | Nivel_so(1)           | 26,281     | 1  | ,000 |
|        |           | Nivel_so(2)           | ,871       | 1  | ,351 |
|        |           | Sexo                  | ,018       | 1  | ,894 |
|        |           | Voto_pasadas          | 4,270      | 2  | ,118 |
|        |           | Voto_pasadas(1)       | ,680       | 1  | ,410 |
|        |           | Voto_pasadas(2)       | 1,337      | 1  | ,247 |
|        |           | Estadísticos globales | 59,620     | 5  | ,000 |
| Paso 2 | Variables | Sexo                  | ,649       | 1  | ,421 |
|        |           | Voto_pasadas          | 2,960      | 2  | ,228 |
|        |           | Voto_pasadas(1)       | ,026       | 1  | ,873 |
|        |           | Voto_pasadas(2)       | 1,924      | 1  | ,165 |
|        |           | Estadísticos globales | 3,764      | 3  | ,288 |