

Ejercicio modelo de frecuencia

Boris Polanco

13 de julio de 2015

1. Introducción

Los modelos de regresión lineal, se basan en los siguientes supuestos:

- Los errores se distribuyen normalmente.
- La varianza es constante.
- La variable dependiente se relaciona linealmente con las variables independientes

De este modo tendríamos:

$$Y_i = \beta_i + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + U_i \quad (1)$$

$$E(U_i) = 0 \quad i = 1, \dots, n \quad (2)$$

$$(3)$$

Por lo que tomando la esperanza de Y_i obtendríamos:

$$E(Y_i) = \beta_i + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

Sin embargo suele pasar que algunos de estos supuestos no se cumplen por la naturaleza de la información. Por lo que se utiliza los modelos lineales generalizados. Estos modelos son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (binomiales, Poisson, gamma, etc) y varianzas no constantes.

Generalmente utilizamos modelos GLM cuando la variable dependiente es:

- Una variable de conteo de casos, como por ejemplo: número de colisiones, viviendas, accidentes, ..., etc.
- Una variable de conteo de casos expresados como proporciones, por ejemplo: porcentaje de heridos en un accidente, porcentaje de personas con empleo, ..., etc.
- Una variable binaria, por ejemplo: vivo o muerto, hombre o mujer, mayor de edad o no, ..., etc.