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:

$$Y_i = \beta_i + \beta_2 x_{2i} + ... + \beta_p x_{pi} + U_i \quad (1)$$
$$E(U_i) = 0 \qquad i = 1, ..., n \quad (2)$$
$$(3)$$

Por lo que tomando la esperanza de Y_i obtendríamos:

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

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.

2. Normalidad

Si los datos no tienen una estructura normal, habitualmente se realiza una transformación de la variable respuesta o utilizar métodos no paramétricos. Otra posible solución es utilizar modelos lineales generalizados. Estos nos permiten especificar otros tipos de distribución de errores.

- Poisson: muy útiles para conteo de acontecimientos, por ejemplo: número de heridos por accidentes de tráfico, número de hogares asegurados que dan parte de siniestro al día,..,etc.

- Binomial: de gran utilidad para proporciones y datos de presencia o ausencia, por ejemplo: tasas de mortalidad, tasas de infección, porcentaje de siniestros mortales.
- Gamma muy útiles con datos que muestran un coeficiente de variación constante, esto se da cuando la varianza aumenta conforme aumenta la media de la muestra, por ejemplo: número de heridos en función del número de siniestros.
-

Además en los modelos lineales habituales, se asume que la variable dependiente así como también los errores del modelo siguen una distribución normal. Por ejemplo, supongamos que un investigador està interesado en predecir cuantos accidentes se producen al día en un lugar determinado, en este caso es razonable asumir que la variable dependiente seguirá una distribución de tipo Poisson y no una normal como algunas veces se utiliza por comodidad.