Modelos Lineales Generalizados: MLG

Samuel Martínez

August 19, 2021

1 Introducción

Los modelos lineales generalizados son una extensión de la regresión clásica con una serie de supuestos más flexibles para el modelamiento. [3] Por ejemplo, no se requiere el supuesto de normalidad de los residuos y de varianza constante. El nombre se debe a que generalizan los modelos lineales clásicos para la diferentes distribuciones de la variable respuesta. [2]

En [1] Nelder y Wedderburn (1972) acuñaron por primera vez el término, donde demostraron que la técnica iterativa de pesos para la regresión lineal puede ser usada para obtener estimadores de máxima verosimilitud para modelos con variable respuesta que pertenece a la familia exponencial y sistematizar linealmente las relaciones existentes usando una transformación lineal.

2 ¿Cuándo usar los modelos lineales generalizados?

Existen cuatro situaciones muy usuales en que el uso de MLG es una buena opción de modelado. La variable respuesta es:

1. una **proporción**.
2. **binaria**.
3. un **conteo**.
4. continua **positiva**.

el elemnto fundalental para que variable de intéres pueda ser modelado mediante MLG es que pertenezca a la familia exponencial.

3 La Familia Exponencial

Una variable aleatoria Y pertenece a la familia exponencial si la función de densidad de probabilidad se puede escribir de la forma:

$$f(y, \theta, \psi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\psi)} + c(y; \psi)\right\}$$

donde $a(\psi)$, $b(\theta)$ y $c(y; \psi)$ son funciones específicas. Donde:

- Si ψ es conocido entonces el modelo pertenece a la familia exponencial y se denomina al único parámetro canónico.
- Si ψ es desconocido, el modelo puede pertenecer a la familia exponencial de 2 parámetros.

3.1 Ejemplos:

1. Distribución normal.
2. Distribución Poisson.

3.2 Problemas:

1. Determine cuál de las siguientes distribuciones pertenece a la familia exponencial e identifique θ, $a(\phi)$, $b(\theta)$ y $c(y;\phi)$ en caso que sea de la familia exponencial.

   (a) La distribución arcoseno es usada para modelar datos de conteo:

   $$P(y;p) = A(y;1)\frac{p^y}{y!}\exp\left(-\arcsin(p)\right)$$

   para $y = 0, 1, ...$ y $0 < p < 1$, donde $A(y;1)$ es una función de normalización.

   (b) La distribución *Cauchy*:

   $$P(y;c,s) = \frac{1}{\pi s\left(1 + \left[\frac{y-c}{s}\right]^2\right)}$$

   parar $-\infty < y < \infty$, $-\infty < c < \infty$ y $s > 0$

2. Sea Y una variable aleatoria con distribución inversa-gaussiana, con parámetros ϕ y λ ambos positivos, la función de densidad es

   $$fy = \frac{\sqrt{\lambda}}{\sqrt{2\pi y^3}}\exp\left(\sqrt{\lambda\phi}\right)\exp\left(\frac{-1}{2}\left(\frac{\lambda}{y} + \phi y\right)\right)$$

   muestre que pertenece a la familia exponencial. Muestre las expresiones para $E[Y]$ y $Var[Y]$ en términos de ϕ y λ.

3. Compruebe que la distribución de probabilidad $f(y) = \theta y^{-\theta-1}$ pertenece a la familia exponencial.
4. Sea $Y \sim G(\alpha, \lambda)$ demuestre que pertenece a la familia exponencial.
5. Muestre que la distribución $t - student$ no pertenece a la familia exponencial.

3.3 Esperanza y Varianza de la familia exponencial

Para un vector que pertenece a la familia exponencial $E[Y] = b'(\theta)$ y $Var[Y] = b''(\theta)a(\psi)$.

3.3.1 Problemas:

En los problemas de la sección 3.2 aquellas distribuciones que pertenecen a la familia exponencial. Compruebe el cumplimiento de las expresiones anteriores derivando las funciones $b(\theta)$ y $a(\psi)$.

4 Componentes de un modelo lineal generalizado

En los modelos lineales la relación entre la media y la variable respuesta se supone lineal. Ahora, en los MLG la relación que existe entre la variable respuesta y la media es no-lineal los MLG usan usa serie de elementos que permiten linealizar la relación:

- *Componente aleatoria*: Esta específica la variable aleatoria y y su distribución de probabilidad.
- *El predictor lineal*: para un vector de parámetros y la matriz de variables independientes se tiene que el predictor lineal es $\eta = XB$
- *La función de enlace*: Una función g monótona y diferenciable que se aplica a cada componente de $E[Y]$ para relacionar con el predictor lineal mediante $g[E[y]] = XB$

References

[1] W. Nelder, J. & Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384, 1972.

[2] M. Tellez, C. Morales. *Modelos Estadísticos Lineales con aplicaciones en R*. Ediciones de la U, 2015.

[3] D. Yadolah. *The concise encyclopedia of statistics*. Springer, 2008.