Clase 8: Modelos Lineales

Justo Andrés Manrique Urbina

22 de octubre de 2019

1. Modelos Lineales Generalizados

Permite mayor flexibilidad:

- Variable respuesta (dependiente) puede asumir otras distribuciones no normales.
- Relación entre una función de $E(Y)$ y X "más"flexible.

$$Y_i = X_i\beta + \varepsilon_i.$$

En dónde $X_i\beta$ es el componente sistemático. ε_i es el componente aleatorio.

1.1. Componentes de MLG

1.1.1. Componente aleatorio:

- Y_i : variable aleatoria.
- Y_i ∼Familia Exponencial.
- $f_{Y_i}(y_i) = exp\frac{y\theta_i - b(\theta_i)}{a(\phi)} - c(y_i, \phi); y_i \in \mathbb{R}_y$

Asimismo, también se tiene que:

- θ_i : Parámetro asociado a la media.
- $a(\phi)$:Parámetro de escala (dispersión).
- $a(\cdot), b(\cdot), c(\cdot)$:funciones específicas.
- R_{Y_i} no depende de θ_i

Sea $(Y_1, Y_2, \ldots, Y_n)$ muestra aleatoria de población de la variable aleatoria Y, entonces:

- Y_i es independiente de $Y_j, \forall i \neq j$.
- $E(Y_i) = \mu_i$

Cada $Y_i \sim FE(\theta_i, \phi)$. Cada Y_i es independiente de la otra pero no tienen la misma media.

1.1.2. Componente Sistemático

El predictor lineal (n_i) es una función de las variables explicativas.

1.1.3. Función de Enlace

La función de enlace es una función monótona (no decreciente) diferenciable que asocia ambos componentes. Es decir:

$$g(E(Y_i)) = N_i.$$

$$g(\mu_i) = N_i.$$

Ejemplo: $g(\mu_i) = N_i = \beta_0 + \beta_1 X_{i1}$

Cuando $N_i = \theta_i$, la función de enlace se llama canónica. Esto tiene algunas propiedades (revisar).

2. Resumen:

Los modelos lineales generalizados se resumen en:

$$Y_i \sim FE(\theta_i, \phi).$$

$$E(Y_i) = \mu_i.$$

$$g(\mu_i) = N_i = \phi(\beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}).$$

Los parámetros por estimar en el modelo son: β, θ.

Propiedades:

- $E(Y) = b'(\theta)$
- $V(Y) = a(\phi)b''(\theta) = a(\theta)V(\mu)$, en dónde $V(\mu)$ es la varianza de la media.

3. Modelos para datos binarios

$$Y_i \sim Bernoulli(p_i).$$

en dónde p_i : probabilidad de éxito. Es decir, probabilidad de que el i-ésimo cliente realice un fraude.

3.1. Componente aleatorio:

$$Y_1 \sim Bernoulli(p_i).$$

Pertenece a la familia exponencial:

$$f_{Y_i} = p^{Y_i}(1-p)^{(1-Y_i)}.$$
$$f_{Y_i} = expy_i log(\frac{p_i}{1-p_i}) + log(1-p_i).$$

En dónde:

- $a(\phi) = 1$
- $c(y_i, \phi) = 0$
- $\theta_i = log(\frac{p_i}{1-p_i})$
- $b(\theta_i) = -log(1-p_i) \rightarrow$ en función de θ_i
- $\theta_i = log(\frac{p_i}{1-p_i}) \rightarrow e^{\theta_i} = \frac{p_i}{1-p_i} \rightarrow e^{\theta_i} - e^{\theta_i} p_i = p_i \rightarrow p_i = \frac{e^{\theta_i}}{1+e^{\theta_i}}$
- $b(\theta_i) = -log(1-p_i) \rightarrow log(\frac{1}{1+e^{\theta_1}})^{-1} \rightarrow b(\theta_i) = log(1+e^{\theta_i})$

3.2. Componente sistemático

$$N_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}.$$

3.3. Función de Enlace

Sea p_i : probabilidad de éxito, está entre 0 y 1.

$$E(Y_i) = p_i = \mu_i \rightarrow 0 \leq u_i \leq 1.$$

$$g(\mu_i) = N_i.$$
$$p_i = \mu_i = g^{-1}(N_i).$$

$g(\cdot)$ tiene que asegurar que las estimaciones de p_i sean en $[0, 1]$. En dónde $p_i \in [0, 1]$, una elección natural de $g(\cdot)$ es la función de distribución acumulada para cualquier variable Ztal que:

$$F_Z(Z) = P(Z < z) \in [0, 1].$$

Por ejemplo: $Z \sim$ logística$(\mu = 0, s = 1)$

$$f_Z(z) = \frac{exp(-z)}{(1 + exp(-z))^2}.$$

$$F_Z(z) = \frac{e^z}{1+e^z} = \text{logit}^{-1}(z); F_z(z) = p(Z < z) \in [0,1].$$

Se utiliza esta función de distribución acumulada para el enlace logit:

$$\eta_i = log(\frac{p_i}{1-p_i}) = logit(p_i).$$

Entonces $g(\cdot) = \text{logit}$.

4. Resumen: GLM para datos binarios.

Regresión logística: $(Y_1, Y_2, \ldots, Y_n)$ una muestra aleatoria de Y. En dónde $Y_i \sim \text{Bernoulli}(p_i)$. La función de enlace es $\text{logit}(p_i) = \eta_i = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_{ip}$. Entonces:

$$\text{logit}(p_i) = log(\frac{p_i}{1-p_i}).$$

4.1. Inferencia: EMV para MLG

Como existe independencia, entonces el estimador de máxima verosimilitud es:

$$L(\beta, \phi) = \prod_{n}^{i=1} f_{Y_i}(y_i).$$

$$= exp(\frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i\theta_i - b(\theta_i) + \sum_{i=1}^{n} c(Y_i, \phi))).$$

$$l(\beta, \phi) = \frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i\theta_i - b(\theta_i) + \sum_{i=1}^{n} c(Y_i, \phi)$$

Función Score para β:

$$S(\beta) = \frac{\partial l(\beta, \phi)}{\partial \beta} = \frac{1}{a(\phi)} X^T W \Delta (Y - \mu).$$

En dónde $W = diag(w_i)$, $\Delta = diag(g_u(\mu_i))$, $w_i = (V(\mu_i) g_\mu^2(\mu_i))^{-1}$. Asimismo se tiene que $V(\mu_i) = \frac{V(Y_i)}{a(\phi)}$ y $g_u(\mu_i) = \frac{\partial \eta_i}{\partial u_i}$. En general, los modelos de estimación de máxima verosimilitud no tienen solución analítica.

4.2. Ejemplo:

Para la regresión de Bernoulli:

$$v(\mu_i) = \frac{V(Y_i)}{a(\phi)} = V(Y_i) = \mu_i(1 - \mu_i).$$

$$g_u(\mu_i) = \frac{\partial \eta_i}{\partial \mu_i} = (\frac{\partial \mu_i}{\partial \eta_i})^{-1} = (\frac{\partial \frac{e^{\eta_i}}{1+e^{\eta_i}}}{\partial \eta_i})^{-1}.$$

$$= \left(\left(\frac{e^{\theta_i}}{1+e^{\theta_i}} \right) * \left(\frac{1}{1+e^{\theta_i}} \right) \right)^{1}.$$

$$= (p_i(1-p_i))^{-1}.$$

$$= (\mu_i(1-\mu_i))^{-1}.$$

$$g_u(\mu_i) = (V(\mu_1) g_\mu^2(\mu_i))^{-1} = \left(\frac{1}{(\mu_i)(1-\mu_i)} \right).$$

$$w = diag(\mu_i(1-\mu_i)).$$

$$\Delta = diag(\mu_i(1-\mu_i))^{-1}.$$

$$w\Delta = I_n.$$

Por lo tanto, la función de Score de β es:

$$S(\beta) = X^T(Y-\mu).$$

Dado que no tiene solución analítica, utilizamos métodos numéricos.

$$\beta^{m+1} = \beta^m + (I(\beta^m))^{-1} S(\beta^m).$$

en dónde $I(\beta) = -E(\frac{\partial^2}{\partial\beta\partial\beta^T} l(\beta))$ o $I(\beta) = \frac{1}{a(\phi)} X^T W X$.

Para datos binarios logísticos son:

$$\beta^{m+1} = \beta^m + (X^T W X)^{-1} X^T(Y-\mu).$$

4.3. Interpretación de parámetros de la salida de R

La razón de ODDS representa el incremento o reducción estimado en la chance de éxito por cada unidad en que se incrementa x_i.

$$log\left(\frac{p_i}{1-p_i}\right) = log(odds) = \eta_i.$$

El predictor lineal para $X_i = x_i + 1$ es $\eta(x_i+1) = \beta_0 + \beta_1 X_i + \beta_1$. Se realiza la diferencia en η_{X_i}, teniendo lo siguiente:

$$\eta(X_i+1) - \eta(X_i) = \beta_1.$$

$$log(odds(X_i+1)) - log(odds(X_i)) = \beta_1.$$

$$\frac{odds(X_i+1)}{odds(X_i)} = e^{\beta_1}.$$

Ejemplo: $e^{\hat{\beta_1}} = 1{,}005514$. Por cada unidad monetaria en que se incrementa el balance, la chance de realizar fraude se incrementa 1.005514 veces, es decir aumenta un 0.05514 %.

Bajo condiciones de regularidad, el EMV $\hat{\beta}$ del parámetro β asintóticamente cumple que:

$$\hat{\beta} \to N(\beta, (I(\beta))^{-1}).$$

$$cov(\hat{\beta}) = (\frac{1}{a(\phi)} X^T W X)^{-1}.$$

$$\widehat{cov(\hat{\beta})} = (\frac{1}{a(\phi)} X^T \hat{w} X)^{-1}.$$

5. Propiedad de invarianza del EMV

Sea $\hat{\theta}$ el estimador de máxima verosimilitud $\theta \to \widehat{g(\theta)} = g(\hat{\theta})$.