Clase 5: Modelos Lineales 2

Justo Andrés Manrique Urbina

17 de septiembre de 2019

1. Modelos Lineales Generalizados

- $Y_i = X_i^T \beta + \varepsilon_i; \varepsilon_i \sim N(0, \sigma^2)$
- $Y_i \sim N(\mu_i, \sigma^2)$, se le conoce como componente aleatorio.
- $N_i = X_i^T \beta$, se le conoce como componente sistemático.
- $\mu_i = N_i$, se le conoce como la función de enlace.

Lo que haremos va a ser extender el modelo. Para ello, admitiremos que el modelo **no es necesariamente normal, sino que pertenece a la familia exponencial.** Pensaremos que el componente sistemático se mantendrá igual. Sin embargo, $g(\mu_i) = N_i$. En general, el $g(\cdot)$ tiene que ser una función monótona y doblemente diferenciable.

Definition 1. *Una variable aleatoria y pertenece a la familia exponencial si su función de densidad o de probabilidad es de la forma:*

$$f(y, \theta, \phi) = \exp\{\phi(y\theta - b(\theta)) + c(y, \phi)\}, \theta\mathbb{R}, \phi > 0.$$

dónde $b(\cdot)$ y $c(\cdot, \cdot)$ son funciones conocidas. Bajo ciertas condiciones de regularidad, se cumple que:

$$E(\frac{\partial}{\partial\theta} \log f(y, \theta, \phi)) = 0.$$

$$E(\frac{\partial^2}{\partial\theta^2} \log f(y, \theta, \phi)) = -E(\{\frac{\partial}{\partial\theta} \log f(y, \theta, \phi)\}^2).$$

Dadas las propiedades anteriores, se cumple lo siguiente:

$$E(Y) = \mu = b^{'}(\theta).$$

$$Var(y) = \frac{1}{\phi} b^{''}(\theta) = \frac{1}{\theta} V(\mu).$$

en dónde $V(\cdot)$ es una función.

Tarea: *Demostrar que las expresiones de $E(Y)$ y $V(Y)$ son correctas utilizando las propiedades dadas anteriormente.*

1.1. Distribución de Poisson

Sea:

$$Y \sim P(\mu) \rightarrow f(y) = \frac{e^{-\mu} u^y}{y!}, y = 0, 1, 2 \ldots.$$

$$f(y) = \exp\{y \log \mu - \mu - \log(y!)\}.$$

$$f(y) = \exp\{\phi(y\theta - b(\theta)) - c(y, \theta)\}; b(\theta) = e^\theta, \phi = 1, c(y, \phi) = -\log(y!).$$

Por lo tanto, la distribución de Poisson pertenece a la familia exponencial. Asimismo, se tiene que:

$$E(y) = b^{'}(\theta) = e^\theta = \mu.$$

$$Var(y) = \frac{1}{\phi} b^{''}(\theta) = e\theta = \mu.$$

$$Var(y) = \frac{1}{\phi} V(\mu); V(\mu) = \mu.$$

1.2. Distribución Normal

Sea:

$$Y \sim N(\mu, \sigma^2).$$

$$f(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}.$$

$$f(y) = \exp\{-\frac{1}{2}\frac{y^2}{\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{\sigma^2} - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi)\}.$$

$$f(y) = \exp\{\phi(y\theta - b(\theta)) + c(y, \theta)\}.$$

$$\phi = \frac{1}{\sigma^2}; \theta = \mu; b(\theta) = \frac{\theta^2}{2}.$$

La distribución normal pertenece a la familia exponencial.

$$E(y) = b^{'}(\theta) = \theta = \mu.$$

$$Var(y) = \frac{1}{\phi} b^{''}(\theta) = \frac{1}{\phi} = \sigma^2.$$

$$V(\mu) = 1.$$

Se define el parámetro θ como el parámetro de regresión.

Existe la función de enlace canónica (ver imagen, el θ). Con la función de enlace canónica se garantiza la unicidad del estimador de máxima verosimilitud. Si se desea explicar un modelo de explicar, se utiliza un enlace logit (binomial), Poisson, o Normal.

2. Sección 2

Sea:

$$f(y_i) = \exp\{\phi(y_i\theta_i - b(\theta_i)) + c(y_i, \phi)\}.$$

La función de log-verosimilitud se define como:

$$L = \sum_{i=1}^{n} \log f(y_i).$$

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

$$L(\mu, Y) = \sum_{i=1}^{n} L(\mu_i, Y_i).$$

3. Tipos de modelos

- **Modelo nulo:** No hay covariables.
- **Modelo bajo estudio:** Se tiene $p < n$
  - En este modelo bajo estudio, se tiene que $L(\hat{\mu}, Y)$.
- **Modelo saturado:** Se tiene $p = n$
  - Dado que es un modelo saturado, se tiene que $L(Y, Y)$.

En base a esto, se define la *devianza*.

Definition 2. *La devianza se define como:*

$$D^*(y, \hat{\mu}) = 2(L(Y, Y) - L(\hat{\mu}, Y))$$

Se tiene también que:

$$\hat{\theta}_i = \theta(\hat{\mu}_i) \leftarrow \text{Modelo bajo estudio.}$$

$$\tilde{\theta}_i = \theta(Y_i) \leftarrow \text{modelo saturado.}$$

Se asume que ϕ es fijo.

Bajo esta definición, se tiene que:

$$D^*(Y, \hat{\mu}) = 2\{L(Y, Y) - L(\hat{\mu}, Y)\}.$$

$$= 2\sum_{i=1}^{n}\{\phi(Y_i\tilde{\theta}_i - b(\tilde{\theta}_i) + c(Y_i, \phi) - \phi(Y_i\hat{\theta}_i - b(\hat{\theta}_i)) - c(y_i, \phi)\}.$$

$$= \phi 2\sum_{i=1}^{n}\{(Y_i(\tilde{\theta}_i - \hat{\theta}_i)) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}.$$

$$\phi 2\sum_{i=1}^{n} d^2(Y_i,\hat{\mu}_1).$$

$$= \phi D(Y,\hat{\mu}).$$

$$D^*(y,\hat{\mu}) = \phi D(Y,\hat{\mu}).$$

en dónde el primer término es la devianza escalada y el segundo es la devianza. $D(Y,\hat{\mu}) = \sum_{i=1}^{n} d^2(Y_i,\hat{\mu}_i)$

3.1. Poisson

Definamos $\theta = \log(\mu)$, $b(\theta) = e^{\theta}$, $\hat{\theta}_i = \log(\hat{\mu}_i)$, $\tilde{\theta}_i = \log(Y_i)$.

$$D(Y,\hat{\mu}) = \sum_{i=1}^{n} 2\{Y_i(\tilde{\theta}_i - \hat{\theta}_i + b(\tilde{\theta}_i))\}.$$

$$= \sum_{i_1}^{n} 2\{Y_i \log(Y_i) - \log(\hat{\mu}_i) + \hat{\mu}_i - Y_i\}.$$

$$\sum_{i=1}^{n} 2\{Y_i \log(\frac{Y_i}{\hat{\mu}_i}) - (Y_i - \hat{\mu}_i)\}.$$

En la normal sale $\sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2$.
Si ϕ es constante: $\beta = (\beta_1,\beta_2)^T$ con:

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 \neq 0.$$

$\rightarrow$ Devianza del modelo completo:$D(y,\hat{\mu})$.

$\rightarrow$ Devianza del modelo reducido:$D(y,\hat{\mu}^o$.

Podemos definir que la razón de verosimilitud, definida como $\varepsilon_{RV} = \phi D(y,\hat{\mu}^o) - D(y,\hat{\mu}) \sim_{aprox} X_q^2$, dónde q es igual a la cantidad de parámetros. Se rechaza H_0 si $\varepsilon_{RV} > X^2_{1-\alpha,q}$.

La prueba F, que no depende de ϕ, se define como la siguiente:

$$F = \frac{(D(Y,\hat{\mu}^o - D(y,\hat{\mu})))/q}{(D(Y,\hat{\mu}))/(n-p)} \sim_{aprox} F(q,n-p).$$

Rechazar H_0 si $F > F_{1-\alpha,q,n-p}$.

4. Log-verosimilitud

$$L(\beta,\phi)=\sum_{i=1}^{n}\phi(Y_i\theta_i-b(\theta_i))+c(Y_i,\phi).$$

$$\frac{\partial L}{\partial B_j}=\sum_{i=1}^{n}\phi(Y_i\frac{\partial\theta_i}{\partial u_i}\frac{\partial\mu_i}{\partial N_i}\frac{\partial N_i}{\partial B_j}-\frac{\partial b(\theta_i)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial N_i}\frac{\partial N_i}{B_j}).$$

Definamos como

$$\frac{\partial\theta_i}{\partial\mu_i}=\frac{1}{\frac{\partial\mu_i}{\partial\theta_i}}=\frac{1}{V_i};\text{dónde } V_i=v(\mu_i).$$

Asimismo, definamos que:

$$w_i=(\frac{\partial\mu_i}{\partial N_i})^2*\frac{1}{V_i}.$$

$$w_i^{\frac{1}{2}}=\frac{\partial\mu_i}{\partial N_i}*\frac{1}{V_i^{\frac{1}{2}}}.$$

$$V_i^{\frac{1}{2}}w_i^{\frac{1}{2}}=\frac{\partial\mu_i}{\partial N_i}.$$

Por lo tanto, se tiene que:

$$\frac{\partial L}{\partial\beta_j}=\sum_{i=1}^{n}\phi\{\sqrt{\frac{w_i}{V_i}}(Y_i-\mu_i)x_{ij}\}.$$

Posteriormente, sacamos segunda derivada para hallar la varianza del estimador, entonces se tiene que: Ver imagen.

Posteriormente, se saca la información de Fisher. Los primeros términos se hacen 0, por lo que se tiene que:

$$-E(\frac{\partial L}{\partial\beta_j\partial\beta_p})=\phi\sum_{i=1}^{n}\frac{\partial\theta_i}{\partial\mu_i}(\frac{\partial\mu_i}{\partial N_i})^2x_{ij}x_{ip}.$$

$$=\phi\sum_{i=1}^{n}\frac{1}{V_i}(\frac{\partial\mu_i}{\partial N_i})^2x_{ij}x_{ip}.$$

$$\phi\sum_{i=1}^{n}w_ix_{ij}x_{ip}.$$

$$I_{\beta\beta}=X^TWX.$$

Definamos el Score como:

$$U=(U_\beta,U_\theta)^T.$$

La matriz de información de fisher es:

$$I = (I_{\beta\beta}, I_{\beta\theta} | I_{\sigma\beta}, I_{\theta\theta}), \text{ matriz 2x2.}$$

$$U_\theta = \frac{\partial L}{\partial \phi} = \sum_{i=1}^{n} (Y_i\theta_i - b(\theta_i)) + c^{'}(Y_i, \theta).$$

$$\frac{\partial^2 L}{\partial \phi^2} = \sum c^{''}(y_i, \phi).$$

$$I_{\theta\theta} = -E(\frac{\partial^2 L}{\partial \phi^2}) = -\sum_{i=1}^{n} E(c^{''}(Y_i, \phi)).$$

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_l} = \phi su.$$

$$I_{\beta\phi} = -E(\frac{\partial L}{\partial \beta \partial \theta}) = 0.$$

$$I = (I_{\beta\beta}, 0 | 0, I_{\theta\theta}).$$

β y θ son ortogonales asintóticamente.

El algoritmo de scoring de Fisher entonces es:

$$\beta^{j+1} = \beta^j + I_{\beta\beta}^{-1} U_\beta.$$