Clase 6: Modelos Lineales

Justo Andrés Manrique Urbina

24 de septiembre de 2019

1. Score

Definamos el score de la forma $U_\beta = \phi X^T w^{0,5} v^{-0,5}(Y - \mu)$. La matriz de información de Fisher se define como:

$$I_{\beta\beta} = \phi X^T w X.$$

Tenemos como observaciones lo siguiente:

$$\eta_i = X_i^T \beta.$$

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

$$V_i = V(\mu_i).$$

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

$$W = diag(w_1, w_2, \ldots, w_n).$$

$$V = diag(v_1, v_2, \ldots, v_n).$$

Para hallar el estimador de máxima verosimilitud, utilizaremos el algoritmos de Scoring de Fisher, este es:

$$\beta^{m+1} = \beta^m + I_{\beta\beta}(\beta^m)^{-1} U_\beta(\beta^m).$$

$$\beta^{m+1} = \beta^m + (\phi X^T w^m X)^{-1}(\phi X^T w^{m^{0,5}} v^{m^{-0,5}}(y - \mu^m)).$$

$$\beta^{m+1} = \beta^m + (X^T w^m X)^{-1} X^T w^{m^{0,5}} V^{m^{-0,5}}(y - \mu^m).$$

Aplicando algunas identidades, se obtiene que:

$$\beta^{m+1} = (X^T + w^m + X)^{-1} X^T w^m (X\beta^m + w^{m^{-0,5}} v^{m^{-0,5}}(Y - \mu^m)).$$

Se tiene la forma parecida a unos mínimos cuadrados ponderados, la cual es:

$$\beta^{m+1} = (X^T w^m X)^{-1} X^T w^m Z^m.$$

$$Z^m = \eta^m + w^{m^{-0,5}} v^{m^{-0,5}} (y - \mu^m).$$

Sin embargo, el estimador se le conoce como el estímador de mínimos cuadrados iterativamente reponderados (*iteratively reweighted least squares*).

Por teoría de máxima verosimilitud, se tiene que:

$$\hat{\beta} \sim N(\beta, \frac{1}{\phi}(X^T w X)^{-1}).$$

Dado que no tenemos el parámetro ϕ entonces utilizamos el estimado:

$$\hat{\beta} \sim N(\beta, \frac{1}{\hat{\phi}}(X^T \hat{w} X)^{-1}).$$

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

$$se(\beta_j) = (\hat{Var}(\hat{\beta}_j))^{0,5} = \frac{1}{\hat{\phi}^{0,5}} \sqrt{C_{jj}}.$$

Si deseamos hacer una prueba de hipótesis en donde $\beta_j = 0$. $\frac{\hat{\beta}_j}{se(\beta_j)} \sim N(0,1)$ si H_0 es verdadera.

El intervalo de confianza se definiría como:

$$IC_{100(1-\alpha)\,\%}(\beta_j) = \hat{\beta}_j + - Z_{1-\frac{\alpha}{2}} se(\beta_j).$$

Para hacer la prueba de hipótesis de forma general, definamos lo siguiente:

$$\hat{\beta}, \hat{\phi} \text{ EMV del modelo sin restricciones.}$$

$$L(\hat{\beta}, \phi) = \text{ log-verosimilitud evaluada en el EMV.}$$

$\hat{\beta}^0, \hat{\phi}^0$ el EMV del modelo si $H_0 : CB = d$ es verdadera. Definamos $L(\hat{\beta}, \hat{\phi}^0) \to$ log-verosimilitud evaluada en $\hat{\beta}^0$ y $\hat{\phi}^0$. Entonces se define la razón de verosimilitud como:

$$\varepsilon_{RV} = 2(L(\hat{\beta}, \hat{\sigma}) - L(\hat{\beta}^0, \hat{\phi}^0)) \sim X^2_{(r)}.$$

Se rechaza si $\varepsilon_{RV} > X^2_{1-\alpha,r}$.

2. Métodos de diagnóstico

Consideremos el algoritmo iterativo de mínimos cuadrados.

$$\hat{\beta} = (X^T \hat{w} X)^{-1} X^T \hat{w} \hat{Z}.$$

$$\hat{\beta} = (X^T \hat{w}^{0,5} \hat{w}^{0,5} X)^{-1} X^T \hat{w}^{0,5} \hat{w}^{0,5} \hat{Z}.$$

$$\hat{\beta} = ({X^*}^T X^*)^{-1} {X^*}^T Z^*.$$

Definamos la matriz hat como lo siguiente:

$$H = X^* ({X^*}^T X^*)^{-1} {X^*}^T.$$

$$\hat{H} = \hat{w}^{0,5} X (X^T \hat{w} X)^{-1} X^T \hat{w}^{0,5}.$$

3. Tipos de residuales

- **Residuales de Pearson:** Definamos:

$$\hat{\beta} = (X^T \hat{w} X)^{-1} X^T \hat{w} \hat{Z}.$$

$$\hat{Z} = \hat{\eta} + \hat{w}^{-0,5} \hat{v}^{-0,5} (y - \hat{\mu}).$$

$$E(\hat{Z}) =_{aprox} \hat{\eta}.$$

$$Cov(\hat{Z}) =_{aprox} \hat{w}^{-1}.$$

El residual de Pearson es:

$$r = \hat{w}^{0,5} (\hat{z} - \hat{\eta}).$$

$$r = \hat{w}^{0,5} (\hat{\eta} + \hat{w}^{-0,5} \hat{v}^{-0,5} (y - \hat{\mu}) - \hat{\eta}).$$

$$r = \hat{v}^{0,5} (y - \mu).$$

$$cov(r) =_{aprox} \frac{1}{\phi} (I - \hat{H}).$$

Definimos el residual como:

$$t_{s-i} = \frac{y_i - \hat{\mu}_i}{\hat{v}_i^{0,5} * \frac{1}{\hat{\phi}^{0,5}} * (1 - \hat{h}_{ii})^{0,6}}.$$

$$t_{s_i} = \frac{\hat{\phi}^{0,5} (y_i - \hat{\mu}_i)}{\sqrt{\hat{v}_i (1 - \hat{h}_{ii})}}.$$

Sin embargo esto no tiene simetría, por lo tanto no es tan usado.

- **Residual de devianza:** Se define como el siguiente residual:

$$t_{D_i} = \frac{\hat{\phi}^{0,5} d(y_i, \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}}.$$

4. Influencia

Definimos la influencia como:

$$D_i = 2(L(\hat{\beta}) - L(\hat{\beta})_{(-i)}).$$

Expandimos por Taylor y se obtiene:

$$D(\beta) = 2(L(\hat{\beta}) - (L(\hat{\beta}) + (\beta - \hat{\beta}) L'(\hat{\beta}) + \frac{1}{2} (\beta - \hat{\beta})^T L''(\hat{\beta}) (\beta - \hat{\beta})).$$

$$D(\beta) = (\beta - \hat{\beta})^T L''(\hat{\beta}) (\beta - \hat{\beta}).$$

$$D(\beta) = (\hat{\beta} - \hat{\beta}_{(i)})^T L^{''}(\hat{\beta})(\hat{\beta} - \hat{\beta}_{(i)}).$$

Cambiamos el $L^{''}(\hat{\beta})$ por la matriz de información de Fisher y finalmente la Influencia se puede obtener realizando lo siguiente:

$$D_i = (\frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}})t_{s_i}.$$

5. Banda de confianza simulada para los residuales

Estimar los parámetros del modelo por máxima verosimilitud: $\hat{\beta}, \hat{\phi}$. Simulaciones consideran estas estimaciones como los valores reales de la población. Entonces se hace el siguiente algoritmo:

- Simular $Y^k \ FE(\hat{\mu}, \hat{\phi})$
- Estimar $\hat{\beta}^k$ y $\hat{\phi}^k$
- Calcular residuales de devianza.
- Ordenar los residuales de devianza y se guardan.
- Volver al paso 1 y repetir la simulación.