



PUC Minas
Virtual

Tipos de resíduos nos GLMs

Tipos de Resíduos nos GLMs

O **resíduo de Pearson** é definido como (r_i^P) é definido como:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}},$$

Componente da estatística de Pearson generalizada X_p^2 .

Para o modelo linear o **resíduo de Pearson** coincide com o **resíduo ordinário**. Matricialmente, o resíduo de Pearson é expresso da seguinte forma:

$$r^P = \widehat{V}^{1/2}(y - \hat{\mu}) = \widehat{W}^{1/2}(\hat{z} - \hat{\eta})$$

Em que $V = \text{diag}\{V(\mu_1), \dots, V(\mu_n)\}$ e $W = \text{diag}\{g'(\mu_i)^{-2} V(\mu_i)^{-1}\}$

Desvantagem: Muito assimétrico para modelos não normais

Resíduo de Pearson Padronizado

O resíduo de Pearson tem uma versão padronizada, com média 0 e variância aproximadamente 1, definido por:

$$r_i^{P'} = \frac{y_i - \hat{\mu}_i}{\sqrt{\phi V(\hat{\mu}_i)(1 - \hat{h}_{ii})}},$$

Em que h_{ii} é o i-ésimo elemento da diagonal da matriz H

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$$

Os resíduos de Pearson padronizados apresentam propriedade razoáveis de segunda ordem, mas podem ter distribuições muito divergentes da normal

Resíduo Componente da *Deviance*

São as raízes quadradas dos componentes da *deviance* com sinal igual a $y_i - \hat{\mu}_i$. Sabe-se que a *deviance* é dada por:

$$D_p = 2 \sum_{i=1}^n [y_i(\tilde{\theta}_i - \hat{\theta}_i)] + b(\hat{\theta}_i) - b(\tilde{\theta}_i)$$

Então temos:

$$r_i^D = \text{sinal } (y_i - \hat{\mu}_i) \cdot \sqrt{2y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)}$$

O resíduo r_i^D representa a distância da observação y_i ao valor ajustado $\hat{\mu}_i$, medida na escala da verossimilhança.

Resíduo Componente da Deviance

Quando r_i^D  temos que a i-ésima observação está mal ajustada pelo modelo

O resíduo componente da *deviance* possui uma versão padronizada que é dada por:

$$r_i^{D'} = \frac{r_i^D}{\sqrt{\phi (1 - \widehat{h}_{ii})}},$$

Em que h_{ii} é o i-ésimo elemento da diagonal da matriz H

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$$

Resíduo Quantílico

O resíduo quantílico aleatorizado baseia-se no método da transformação integral da probabilidade.

Seja y_i uma v.a. com fda $F(y_i; \mu_i; \phi)$. O método da transformação integral da probabilidade baseia-se no seguinte resultado:

$$u_i = F(y_i; \mu_i; \phi) \sim U(0,1).$$

Considerando que u_i tem distribuição uniforme entre 0 e 1, temos então:

$$\phi^{-1}(F(y_i; \mu_i; \phi)) \sim N(0,1).$$

Resíduo Quantílico

O **Resíduo Quantílico** aleatorizado pode ser definido por:

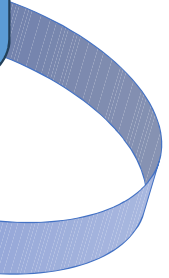
$$r_i^q = \phi^{-1} [F(y_i; \mu_i; \phi)]$$

De forma que se o modelo estiver corretamente especificados, estes resíduos seguiram uma distribuição normal padrão.


Se y_i fo uma v.a. discreta, então $F(y_i; \mu_i; \phi)$ é uma com fda com “saltos” em cada valor de y_i .

Análise de Resíduos

Os resíduos de Perason e componente da Deviance geralmente **não possuem boas aproximações com a distribuição normal**, ainda que o **modelo ajustado esteja correto**.



A avaliação da qualidade do ajuste baseado em gráficos de probabilidade (QQ-plot), podem não ser adequados.



Um tipo de resíduo que, por construção, tem **distribuição normal caso o modelo ajustado esteja correto** é o **resíduo quantílico aleatorizado**



PUC Minas
Virtual