/ /

HW 11

① Uma transformação de estabilização de variância ANOVA estabiliza as variâncias da seguinte maneira aproximada. Deixe Y ter média θ e variância $v(\theta)$.

(a) Use argumentos como da Sessão 10.1.3 [CB] para mostrar que para mostrar que um termo da aproximação de Taylor da variância de $g(Y)$ é dada por $Var(g(Y)) = [\frac{\partial}{\partial\theta} g(\theta)]^2 v(\theta)$.

(b) Mostre que a aproximação da variância de $g^*(Y)$ é independente de θ, onde $g^*(Y) = \int [1/\sqrt{v(y)}]\, dy$.

ⓐ A aproximação de serie de Taylor de primeira ordem é

$$Var[g(Y)] \approx [g'(\theta)]^2 \cdot Var\, Y = [g'(\theta)]^2 \cdot v(\theta).$$

ⓑ Se escolhemos $g(Y) = g^*(Y) = \int_a^{Y} \frac{1}{\sqrt{v(x)}} dx$, então

$$\frac{\partial}{\partial\theta} g^*(\theta) = \frac{\partial}{\partial\theta} \int_a^{\theta} \frac{1}{\sqrt{v(x)}} dx = \frac{1}{\sqrt{v(\theta)}},$$

pelo TFC (teo fundamental do calculo). Então, para qualquer θ,

$$Var[g^*(Y)] \approx \left[\frac{1}{\sqrt{v(\theta)}}\right]^2 \cdot v(\theta) = 1.$$

/ /

② No sentido do exercício anterior verifique se as transformações são estabilização-de-variância

(a) $Y \sim$ Poisson(λ) $g^*(y) = \sqrt{y}$

(b) $Y \sim$ binomial(n,p), $g^*(y) = \sin^{-1}(\sqrt{y/n})$

(c) Y tem variância $V(\theta) = K\theta^2$ para alguma constante K, $g^*(y) = \log(y)$.

(a) Temos que,

$$V(\lambda) = \lambda, \quad g^*(y) = \sqrt{y}, \quad \frac{\partial g^*(\lambda)}{\partial \lambda} = \frac{1}{2\sqrt{\lambda}}$$

Daí,

$$\text{Var}\, g^*(Y) \approx \left(\frac{\partial}{\partial \lambda} g^*(\lambda)\right)^2 \cdot V(\lambda) = \frac{1}{4}$$ independente de λ

(b) Para usar a aproximação em série de Taylor precisamos expressar todo termo em termos de $\theta = EY = np$. Então $V(\theta) = \theta(1-\theta/n)$ e

$$\left(\frac{\partial}{\partial \theta} g^*(\theta)\right)^2 = \left(\frac{1}{\sqrt{1-\frac{\theta}{n}}} \cdot \frac{1}{2\sqrt{\theta/n}} \cdot \frac{1}{n}\right)^2 = \frac{1}{4n\theta(1-\theta/n)}.$$

Além disso,

$$\text{Var}[g^*(Y)] \approx \left(\frac{\partial}{\partial \theta} g^*(\theta)\right)^2 V(\theta) = \frac{1}{4n}.$$ independente de θ e p.

(c) $V(\theta) = K\theta^2$, $\frac{\partial g^*(\theta)}{\partial \theta} = \frac{1}{\theta}$ e $\text{Var}\, g^*(Y) \approx \left(\frac{1}{\theta}\right)^2 \cdot K\theta^2 = K$ que não depende de θ.

(3) A família de transformações Box-Cox é definida por

$$g^*_\lambda(Y) = \begin{cases} (Y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \log Y, & \lambda = 0 \end{cases}$$

onde λ é um parâmetro livre.

(a) Mostre que, para cada Y, $g^*_\lambda(Y)$ é contínua em λ. Em particular, mostre que

$$\lim_{\lambda \to 0} (Y^\lambda - 1)/\lambda = \log Y$$

(b) Encontre a função $v(\theta)$, a variância aproximada de Y, que estabiliza $g^*_\lambda(Y)$.

(a) $g^*_\lambda(Y)$ é claramente contínua com possível exceção para $\lambda = 0$. Para este caso usando L'Hospital temos

$$\lim_{Y \to 0} \frac{Y^\lambda - 1}{\lambda} = \lim_{Y \to 0} \frac{(\log Y) Y^\lambda}{1} = \log Y$$

(b) Pelo problema (1), queremos encontrar $v(\lambda)$ que satisfaz

$$\frac{Y^\lambda - 1}{\lambda} = \int_a^Y \frac{1}{\sqrt{v(x)}} dx,$$

derivando

$$\frac{\partial}{\partial Y}\left(\frac{Y^\lambda - 1}{\lambda}\right) = \frac{\partial}{\partial Y} \int_a^Y \frac{1}{\sqrt{v(x)}} dx = \frac{1}{\sqrt{v(Y)}}$$

Por isso, $V(Y) = Y^{-2(\lambda-1)}$. Pelo problema ①

$$Var\left(\frac{Y^\lambda - 1}{\lambda}\right) \approx \left(\frac{\partial}{\partial Y} \frac{\theta^\lambda - 1}{\lambda}\right)^2 V(\theta) = \theta^{2(\lambda-1)} \theta^{-2(\lambda-1)} = 1$$

(4) Suponha que v.a. Y_{ij} são observados de acordo com a ANOVA de um lado. Mostre que, ~~without~~ sem nenhuma restrição nos parametros, este modelo não é identificavel por exibir duas coleções distintas de parametros que levam exatamente à mesma distribuição dos Y_{ij}'s.

Para o modelo

$$Y_{ij} = \mu + \tau_{i} + \varepsilon_{ij}, \quad i=1,\dots,k, \; j=1,\dots,n_i,$$

tome $k=2$. A configuração de parametro

$$(\mu, \tau_1, \tau_2) = (10, 5, 2)$$
$$(\mu, \tau_1, \tau_2) = (7, 8, 5)$$

tem os mesmos valores para $\mu+\tau_1$ e $\mu+\tau_2$, então tem as mesmas distribuições para Y_1 e Y_2.

/ /

(5) Mostre que sobre a hipótese de ANOVA um lado, para algum conjunto de constantes $a = (a_1, \ldots, a_n)$, a quantidade $\sum a_i \bar{Y}_i$ é normalmente distribuída com média $\sum a_i \theta_i$ e variância $\sigma^2 \sum a_i^2 / n_i$

Sobre a hipótese de ANOVA um lado, $Y_{ij} \sim N(\theta_i, \sigma^2)$. Além disso,

$$\bar{Y}_i \sim N(\theta_i, \sigma^2/n_i)$$

$$a_i \bar{Y}_i \sim N(a_i\theta_i, a_i^2\sigma^2/n_i)$$

$$\cancel{\sum a_i \bar{Y}_i \sim N(a_i\theta_i, a}$$

$$\sum a_i \bar{Y}_i \sim N\left(\sum a_i\theta_i, \sigma^2 \sum a_i^2/n_i\right)$$

/ /

⑥ Mostre como construir um teste-t para (considerando a seção 11.2.7 do [CB])

(a) $H_0: \sum a_i \theta_i = \delta$ vs $H_1: \sum a_i \theta_i \neq \delta$.

(b) $H_0: \sum a_i \theta_i \leq \delta$ vs $H_1: \sum a_i \theta_i > \delta$, onde δ é uma constante

ⓐ Pelo exercício anterior

$$T = \sum a_i \bar{Y}_i \sim N\left(\sum a_i \theta_i, \sigma^2 \sum a_i^2\right)$$

e sobre H_0, $ET = \delta$, sobre H_0 (usando as equações 11.2.6 e 11.2.7 do [CB])

$$\frac{\sum a_i \bar{Y}_i - \delta}{\sqrt{S_p^2 \sum a_i^2}} \sim t_{N-k},$$

onde $N = \sum n_i$. Além disso, o teste é rejeitado se

$$\frac{\left|\sum a_i \bar{Y}_i - \delta\right|}{\sqrt{S_p^2 \sum a_i^2 / n_i}} > t_{N-k, \frac{\alpha}{2}}.$$

ⓑ Similarmente para $H_0: \sum a_i \theta_i \leq \delta$ vs $H_1: \sum a_i \theta_i > \delta$, rejeitamos H_0 se

$$\frac{\sum a_i \bar{Y}_i - \delta}{\sqrt{S_p^2 \sum a_i^2 / n_i}} > t_{N-k, \alpha}.$$