Estatísticas Ancilares

Casella e Berger

(1)

Até o momento trabalhamos com as estatísticas suficientes. Estas, de certo modo, contêm todas as informações sobre o que estão disponíveis na amostra. ↳ (pode ser vetor, denotado por $\underset{\sim}{\theta}$)

Agora, apresentaremos uma espécie diferente de estatística, que tem um propósito complementar.

Definição: (Estatística Ancilar)

Uma estatística $A(\underset{\sim}{X})$ cuja distribuição não depende do parâmetro θ é chamada estatística ancilar. ↳ (pode ser vetor, $\underset{\sim}{\theta}$)

Comentários: Sozinha, uma estatística ancilar não contém informações sobre θ ou $\underset{\sim}{\theta}$.

Paradoxalmente, uma estatística ancilar, quando utilizada juntamente com outras estatísticas, algumas vezes contém informações valiosas para inferências sobre θ ($\underset{\sim}{\theta}$).

Vejamos uns exemplos de estatísticas ancilares!!

Exemplo 0 : Seja $X_1, \dots, X_n$ a.a. $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$. ① ②

A estatística $T_1 = X_1 - X_2 \sim N(0, 2)$, portanto

T_1 é uma estatística ancilar.

Outra estatística ancilar é $T_2 = X_1 + \dots + X_{n-1} - (n-1)X_n \sim N(0, n(n-1))$, pois $\sum_{i=1}^{n-1} X_i \sim N((n-1)\theta, (n-1))$

e $(n-1) X_n \sim N((n-1)\theta, (n-1)^2)$.

Além disso, a variância amostral $S^2 \sim (n-1)^{-1} \chi^2_{n-1}$

ou $(n-1) S^2 \sim \chi^2_{n-1}$ é também uma

estatística ancilar.

Lembrete !!

Note que $T_4 = \dfrac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$. Neste

caso, não podemos discutir sobre ancilaridade,

pois T_4 não é uma estatística, depende de

μ.

Exemplo 1: Estatística ancilar da família de locação

Seja $X_1, \dots, X_n$ aa da família de locação com fda $F(x-\theta)$, $-\infty < \theta < \infty$ e conhecido como parâmetro de locação.

Mostraremos que $R = X_{(n)} - X_{(1)}$ é uma estatística ancilar, onde R é a amplitude.

Lembrete: Teorema

Seja $f(\cdot)$ qualquer fdp e seja η um número real qualquer, e σ um número real qualquer positivo. Então X é uma variável aleatória com fdp $\left(\frac{1}{\sigma}\right) f\left(\frac{x-\eta}{\sigma}\right)$ se e somente se, existir uma variável aleatória Z com fdp $f(z)$ e

$X = \sigma Z + \eta$.

→ Se definir $\sigma = 1$ no teorema gera um resultado (somente) para as famílias de locação, e se definir $\eta = 0$ gera um resultado (somente) para as famílias de escala.

Assim, pelo teorema, sabemos que $X_i = Z_i + \theta$,

onde $Z_i \sim F(\cdot)$ que não depende de θ, $i = 1, \dots, n$.

Portanto,

$$F_R(r|\theta) = P(R \leq r) = P(X_{(n)} - X_{(1)} \leq r)$$

$$= P(\max\{X_1, \dots, X_n\} - \min\{X_1, \dots, X_n\} \leq r) =$$

$$P(\max\{Z_1+\theta, \dots, Z_n+\theta\} - \min\{Z_1+\theta, \dots, Z_n+\theta\} \leq r)$$

$$= P(\max\{Z_1, \dots, Z_n\} - \min\{Z_1, \dots, Z_n\} \leq r)$$

$$= P(Z_{(n)} - Z_{(1)} \leq r)$$ que não depende de θ e concluímos que R é uma estatística ancilar.

Exemplo α: Seja $X_1, \dots, X_n$ aa $X \sim U(\theta - \frac{1}{\alpha}, \theta + \frac{1}{\alpha})$, $-\infty < \theta < \infty$. Mostre que $R = X_{(n)} - X_{(1)}$ é uma estatística ancilar.

Sabemos que a família de distribuições $U(\theta - \frac{1}{\alpha}, \theta + \frac{1}{\alpha})$ pertence à família de locação (exercício (pedido/dado) anteriormente)!!

Então, pelo exemplo 1, temos que r é uma (4)

estatística ancilar.

Exemplo 3: Estatística ancilar da família de escala

Seja $X_1, \dots, X_n$ aa da família de escala com fda $F\left(\frac{x}{\sigma}\right)$, $\sigma > 0$ e conhecido como parâmetro de escala. Mostraremos que qualquer estatística que dependa da amostra somente por meio dos $n-1$ valores $\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}$ é uma estatística ancilar. Por exemplo,

$$\frac{X_1 + \dots + X_n}{X_n} = \frac{X_1}{X_n} + \dots + \frac{X_{n-1}}{X_n} + 1$$

é uma estatística ancilar.

Assim, pelo teorema, sabemos que $X_i = \sigma Z_i$, onde $Z_i \sim F(\cdot)$ que não depende de σ, $i = 1, \dots, n$.

Portanto, a fda conjunta de $\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}$ é dada por

$$P\left(\frac{X_1}{X_n} \le y_1, \dots, \frac{X_{n-1}}{X_n} \le y_{n-1}\right) = P\left(\frac{\cancel{\sigma} Z_1}{\cancel{\sigma} Z_n} \le y_1, \dots, \frac{\cancel{\sigma} Z_{n-1}}{\cancel{\sigma} Z_n} \le y_{n-1}\right)$$

$= P\left(\frac{Z_1}{Z_n} \leq y_1, \ldots, \frac{Z_{n-1}}{Z_n} \leq y_{n-1} \right)$ que não depende

de σ. Assim, a distribuição de $\frac{X_1}{X_n}, \ldots, \frac{X_{n-1}}{X_n}$ é independente de σ, assim como a distribuição de qualquer função dessas quantidades.

Exemplo 4: Sejam X_1 e X_2 observações i.i.d $N(0, \sigma^2)$. A partir do resultado acima (exemplo 3), temos que $\frac{X_1}{X_2}$ é uma estatística ancilar, ou seja, $\frac{X_1}{X_2}$ tem uma distribuição que é a mesma para cada valor de σ.

Note que se $\sigma = 1$, $\frac{X_1}{X_2}$ é a razão de duas variáveis aleatórias independentes e normais padrão, e podemos verificar que $\frac{X_1}{X_2} \sim \text{Cauchy}(0,1)$ (apêndice Casella e Berger - distribuições contínuas). Assim, para qualquer $\sigma > 0$, a distribuição de $\frac{X_1}{X_2}$ é esta mesma distribuição de Cauchy.