

Plano Aula 17 e 18

(cont.) Inferência Estatística

Exemplo 1: Média amostral, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, em que X_1, \dots, X_n uma amostra aleatória de $X_i \sim \text{Normal}(\mu, \sigma^2)$ e σ^2 conhecido:

- Qual a distribuição amostral de \bar{X} ?
- \bar{X} é um bom estimador para a média populacional μ ?
- Como usar $\text{Var}(\bar{X})$ para fornecer um grau de certeza sobre usarmos \bar{X} para representar/estimar μ ?

Estimação Pontual (Bussab e Morettin - Capítulo 11)

- Estatísticas: Estimador *versus* Estimativa.

Definição (Estimador): Um estimador T do parâmetro θ é qualquer função das observações da amostra, $T = g(X_1, \dots, X_n)$.

Definição (Estimativa): Uma estimativa é um particular valor do estimador. Para uma amostra observada x_1, \dots, x_n uma estimativa t do parâmetro θ é dada por $t = g(x_1, \dots, x_n)$.

- Propriedades dos estimadores (Bussab e Morettin - Seção 11.2)
 - **Vies** e o Erro Quadrático Médio (EQM); **Consistência** e **Eficiência**.

(cont.) **Exemplo 1:** E para a média amostral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ se σ^2 desconhecido?

Exemplo 2: Para a variância amostral $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - E(X))^2$? E para $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - E(X))^2$?

Exemplo 3: E para a proporção amostral $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$?

Introdução à estimação intervalar

Estimação pontual \times estimação intervalar

Intervalos de Confiança (IC) (Bussab e Morettin - Seção 11.6)

Definição (Intervalo de confiança (IC)): Seja T um estimador para o parâmetro θ , o IC ao nível $(1 - \alpha) \times 100\%$ para θ será denotado pelo intervalo

$$IC(\theta; 1 - \alpha) = (t_1(T), t_2(T)),$$

para dois valores $t_1(T)$ e $t_2(T)$ tais que $P[t_1(T) < \theta < t_2(T)] = 1 - \alpha$. (Se conhecida a distribuição amostral de T , será sempre possível achar $t_1(T)$ e $t_2(T)$).

- Esse é um tipo de estimação intervalar (o mais popular em inferência paramétrica clássica)
 - Veremos todas as situações de intervalos nos **slides dessa semana**.

Erro padrão de um Estimador (Bussab e Morettin - Seção 11.7)

Definição (**Erro padrão**): denominamos *erro padrão* do estimador T (para o parâmetro θ) a quantidade $EP(T) = \sqrt{Var(T)}$.

Definição (**Erro padrão estimado**): $ep(T) = \widehat{EP}(T) = \sqrt{\widehat{Var}(T)}$.

- ...cont. **Exemplo 1**: Média amostral \bar{X} . Calcular $EP(\bar{X})$. E $ep(\bar{X})$?
- ...cont. **Exemplo 3**: Proporção amostral \hat{p} . $EP(\hat{p})$ e $ep(\hat{p})$?

IC para uma média populacional μ (supondo σ^2 conhecido ou $n > 30$)

Iniciaremos com o IC para uma média populacional μ ; * Resultado importante na construção de IC para uma média populacional: + No **Exemplo 1**, supondo σ^2 conhecido (ou $n > 30$), então

$$\bar{X} \sim Normal(\mu, \sigma^2/n)$$

se $X \sim Normal(\mu, \sigma^2)$. Também

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim Normal(0, 1).$$

IC para uma média populacional μ (supondo σ^2 desconhecido e $n \leq 30$)

Estimação de σ^2

- Se desconhecemos a variância populacional, podemos estimá-la usando o estimador $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ (porquê?)
- Nesse caso S^2 é uma variável aleatória (v.a.). (Sabemos qual a distribuição amostral de S^2 ?)
- Qual a distribuição amostral da transformação $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$?

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim ?$$

Distribuição (de probabilidade) t de Student (Bussab e Morettin - Seção 7.7.3)

Teorema (**Distribuição t -Student, nossa versão**): Seja X_1, \dots, X_n uma amostra aleatória da v.a. $X \sim Normal(\mu, \sigma^2)$, então (dadas algumas outras suposições para S que omitimos aqui)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

em que $t_{(n-1)}$ denota a distribuição de probabilidade t -Student com $n - 1$ graus de liberdade (g.l.).

- A distribuição t de Student também possui **valores tabelados**, como a distribuição **normal padrão**. Qual a relação entre essas distribuições?
- Como usar a distribuição t de Student para construir um IC para μ ? Quais as suposições necessárias? Como interpretar os resultados?

Intervalo de confiança para a Variância

- Suponha que agora queremos estimar uma variância populacional σ^2 .
- Exemplo: Estimar a variabilidade dos retornos de certa aplicação financeira.
 - Qual o estimador pontual “natural” para o problema? E como calcular um IC para σ^2 ?

(...continuação) **Estimação de σ^2**

- Se desconhecemos a variância populacional, podemos estimá-la usando o estimador $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ (porquê?)
- Nesse caso S^2 é uma variável aleatória (v.a.). (Sabemos qual a distribuição amostral de S^2 ?)

Distribuição (de probabilidade) Qui – Quadrado (Bussab e Morettin - pág. 358)

Teorema (Distribuição Qui-Quadrado, nossa versão): Seja X_1, \dots, X_n uma amostra aleatória da v.a. $X \sim Normal(\mu, \sigma^2)$ e $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$, então podemos escrever uma quantidade Q tal que (dadas algumas outras suposições que omitimos aqui)

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

em que $\chi^2_{(n-1)}$ denota a distribuição de probabilidade Qui-Quadrado com $n-1$ graus de liberdade (g.l.).

- A distribuição χ^2 **valores tabelados**, assim como a distribuição **normal padrão** e a t . A diferença é que Q só assume valores positivos.
- Como usar a distribuição de Q para construir um IC para σ^2 ? **Quais as suposições necessárias? Como interpretar os resultados?**

Intervalo para uma proporção (populacional)

- Suponha que agora queremos estimar uma proporção populacional π .
- **Exemplo:** Estimar a proporção de pessoas infectadas por um certo vírus numa população.
 - Qual o estimador pontual “natural” para o problema? E como calcular um IC para π ?
- Quais as **suposições necessárias**? Como **interpretar os resultados**?

Usando o teorema central do limite

- $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim Normal(0, 1)$ se $X \sim Normal(\mu, \sigma^2)$, para σ^2 conhecido, ou
- $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim Normal(0, 1)$ se o tamanho amostral for grande, $n \gg 30$.

No caso da proporção amostral X não será normal Para uma amostra aleatória X_1, \dots, X_n da v.a. $X \sim Bernoulli(\pi)$ temos que $\sum_{i=1}^n X_i \sim Binomial(n, \pi)$. Das propriedades da distribuição binomial sabemos que $E(\sum_{i=1}^n X_i) = np$ e $V(\sum_{i=1}^n X_i) = np(1-p)$.

Assim, para um tamanho de amostra suficientemente grande ($n \gg 30$)

$$Z = \frac{(\sum_{i=1}^n X_i) - np}{\sqrt{np(1-p)}} \sim Normal(0, 1)$$

ou ainda usando $p = \sum_{i=1}^n X_i/n$

$$Z = \frac{(\sum_{i=1}^n X_i/n) - p}{\sqrt{\frac{p(1-p)}{n}}} \sim Normal(0, 1)$$

Dimensionamento do tamanho de amostra n

Chamamos de erro de estimação a metade da amplitude do intervalo,

- no caso de IC para μ com σ^2 conhecido, $E = z_{\alpha/2} \times \sigma/\sqrt{n}$,
- no caso de IC para μ com σ^2 desconhecido e n pequeno, $E = t_{(n-1);\alpha/2} \times s/\sqrt{n}$,
- e no caso de IC para π , $E = z_{\alpha/2} \times \sqrt{p(1-p)/n}$.

Como calcular o **tamanho mínimo de uma amostra** para uma **confiança** $1 - \alpha$ especificada e um **erro máximo** E também fixado?

Ler slides e ver vídeos da semana 9.

Fazer lista de exercícios 2-4.

Fazer o Quiz da semana 9 - VALE NOTA!!!
