Intervalos de Confiança - Definições Iniciais

Prof. Eduardo Bezerra

CEFET/RJ

14 de Setembro de 2018

Roteiro

Roteiro

Exemplos de motivação

Exemplo 1

Foi feita uma pesquisa junto a uma amostra de tamanho $n = 500$ da comunidade de uma universidade acerca da criação de um bandejão. Cada entrevistado deveria dar uma de duas respostas possíveis:

- SIM (i.e., a favor da criação)
- NÃO (não favorável à criação)

Exemplos de motivação

Exemplo 1

Foi feita uma pesquisa junto a uma amostra de tamanho $n = 500$ da comunidade de uma universidade acerca da criação de um bandejão. Cada entrevistado deveria dar uma de duas respostas possíveis:

- SIM (i.e., a favor da criação)
- NÃO (não favorável à criação)

Deseja-se estimar a *proporção* de pessoas da comunidade como um todo (população) que são favoráveis à criação do bandejão.

Exemplos de motivação

Exemplo 1

Foi feita uma pesquisa junto a uma amostra de tamanho $n = 500$ da comunidade de uma universidade acerca da criação de um bandejão. Cada entrevistado deveria dar uma de duas respostas possíveis:

- SIM (i.e., a favor da criação)
- NÃO (não favorável à criação)

Deseja-se estimar a *proporção* de pessoas da comunidade como um todo (população) que são favoráveis à criação do bandejão.

Se 300 pessoas dizem sim, então uma estimativa natural para essa proporção é 60%. Entretanto, outra amostra poderia levar a uma estimativa diferente.

Exemplos de motivação

Exemplo 2

Em uma amostra grande de medições independentes do diâmetro de pistões, $\bar{x} = 14,0$ cm e $\sigma_{\bar{x}} = 0,1$ cm.

O valor 14,0 vem de uma distribuição normal, porque trata-se da média de um grande número de medições.

Exemplos de motivação

Exemplo 2

Em uma amostra grande de medições independentes do diâmetro de pistões, $\bar{x} = 14,0$ cm e $\sigma_{\bar{x}} = 0,1$ cm.

O valor 14,0 vem de uma distribuição normal, porque trata-se da média de um grande número de medições.

O diâmetro médio da população (μ) não é exatamente igual à média da amostra de 14,0 cm ($\bar{x}$).

Exemplos de motivação

Exemplo 2

Em uma amostra grande de medições independentes do diâmetro de pistões, $\bar{x} = 14,0$ cm e $\sigma_{\bar{x}} = 0,1$ cm.

O valor 14,0 vem de uma distribuição normal, porque trata-se da média de um grande número de medições.

O diâmetro médio da população (μ) não é exatamente igual à média da amostra de 14,0 cm ($\bar{x}$).

No entanto, pelo TLC, podemos utilizar o seu desvio padrão para determinar quão próximo esse valor está do valor μ. Por exemplo, é pouco provável que $|\bar{x} - \mu|$ seja maior do que três desvios padrão.

Roteiro

Estimador e Estimativa

Estimativa

Uma estimativa de um parâmetro desconhecido θ é um valor obtido a partir da amostra (por meio de uma estatística). Esse é um valor aproximado do parâmetro.

Estimador e Estimativa

Estimativa

Uma estimativa de um parâmetro desconhecido θ é um valor obtido a partir da amostra (por meio de uma estatística). Esse é um valor aproximado do parâmetro.

Estimador

Um estimador $\widehat{\theta}$ é uma estatística (i.e. função da amostra) que fornece estimativas para algum parâmetro da população.

Estimador e Estimativa - exemplo

Na pesquisa sobre o bandejão, considere que $X_1, X_2, \ldots, X_{500}$ seja a amostra colhida, em que

$$X_i = \begin{cases} 1 & \text{se a i-ésima pessoa na amostra responder SIM} \\ 0 & \text{se a i-ésima pessoa na amostra responder NÃO} \end{cases}$$

Se $Y_n = \sum_{i=1}^{n} X_i$, então Y_n tem distribuição binomial com parâmetros n e p, e o problema consiste em estimar p.

Estimador e Estimativa - exemplo

Na pesquisa sobre o bandejão, considere que $X_1, X_2, \ldots, X_{500}$ seja a amostra colhida, em que

$$X_i = \begin{cases} 1 & \text{se a i-ésima pessoa na amostra responder SIM} \\ 0 & \text{se a i-ésima pessoa na amostra responder NÃO} \end{cases}$$

Se $Y_n = \sum_{i=1}^{n} X_i$, então Y_n tem distribuição binomial com parâmetros n e p, e o problema consiste em estimar p.

Um possível estimador de p é

$$\widehat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\text{quantidade de SIM}}{\text{quantidade de pessoas da amostra}}$$

A estimativa para $\widehat{p}$ na pesquisa sobre o bandejão foi de 60%.

Estimador não tendencioso

Estimador não tendencioso

Se θ é um parâmetro populacional, e $\widehat{\theta}$ é um de seus estimadores, dizemos que $\widehat{\theta}$ é um estimador não tendencioso (*unbiased estimator*) de θ se, para todo θ,

$$\mathrm{E}(\widehat{\theta}) = \theta.$$

Estimador não tendencioso

Estimador não tendencioso

Se θ é um parâmetro populacional, e $\widehat{\theta}$ é um de seus estimadores, dizemos que $\widehat{\theta}$ é um estimador não tendencioso (*unbiased estimator*) de θ se, para todo θ,

$$\mathrm{E}(\widehat{\theta}) = \theta.$$

Sinônimos:

- estimador não-viesado,
- estimador não viciado,
- estimador não-enviesado.

Estimador não tendencioso - exemplo 01

Já vimos que, dada uma amostra $(X_1, X_2, \ldots, X_n)$ de uma população com média μ e variância σ^2, então $\overline{X}$ tem distribuição normal, i.e.,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Então, $\overline{X}$ é um **estimador não-tendencioso** de μ (i.e., a média amostral é um estimador não tendencioso da média populacional).

Estimador não tendencioso - exemplo 02

Já vimos que, dada uma amostra $(X_1, X_2, \ldots, X_n)$ de uma população, a proporção amostral é uma v.a. que segue uma distribuição normal, i.e.,

$$\widehat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Então, $\widehat{p}$ é um **estimador não-tendencioso** de p, a proporção de elementos da população que têm uma certa característica em comum.

Estimador tendencioso - exemplo 03

É possível mostrar que o valor esperado da variância amostral é dado pela expressão

$$\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2$$

Estimador tendencioso - exemplo 03

É possível mostrar que o valor esperado da variância amostral é dado pela expressão

$$\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2$$

Portanto, S^2 é um **estimador tendencioso** de σ^2, posto que $\mathrm{E}(S^2) \neq \sigma^2$.

Intervalos de confiança

Na pesquisa sobre o bandejão, 60% é uma estimativa para o parâmetro *proporção da população favorável ao bandejão*.

Em geral,

- essa estimativa corresponde a um **único valor** para estimar um parâmetro.
- há uma diferença (para mais ou para menos) entre essa estimativa e o valor do parâmetro.

Conhecer apenas a estimativa não permite julgar qual a **magnitude do erro** cometido.

Intervalos de confiança

Na pesquisa sobre o bandejão, 60% é uma estimativa para o parâmetro *proporção da população favorável ao bandejão*.

Em geral,

- essa estimativa corresponde a um **único valor** para estimar um parâmetro.
- há uma diferença (para mais ou para menos) entre essa estimativa e o valor do parâmetro.

Conhecer apenas a estimativa não permite julgar qual a **magnitude do erro** cometido.

Assim, é necessário construir um intervalo de confiança em torno da estimativa.

Roteiro

Intervalos de confiança

Uma intervalo de confiança (ou estimativa intervalar) corresponde a um par de números. Por exemplo, no caso em que o estimador é $\hat{p}$, temos:

$$[\hat{p} - \epsilon, \hat{p} + \epsilon]$$

O valor ϵ é o erro amostral ou margem de erro.

Intervalos de confiança

Uma intervalo de confiança (ou estimativa intervalar) corresponde a um par de números. Por exemplo, no caso em que o estimador é $\hat{p}$, temos:

$$[\hat{p} - \epsilon, \hat{p} + \epsilon]$$

O valor ϵ é o erro amostral ou margem de erro.

Perguntas:

- Como encontrar ϵ nesse caso da proporção?

Intervalos de confiança

Uma intervalo de confiança (ou estimativa intervalar) corresponde a um par de números. Por exemplo, no caso em que o estimador é $\hat{p}$, temos:

$$[\hat{p} - \epsilon, \hat{p} + \epsilon]$$

O valor ϵ é o erro amostral ou margem de erro.

Perguntas:

- Como encontrar ϵ nesse caso da proporção?
- Como encontrar ϵ para qualquer estatística (não apenas para a proporção)?