Estatística Avançada - Aula 05

Fundamentos da Inferência Estatística: Estimação

Kaique Matias de Andrade Roberto

Ciências Atuariais - Ciências Econômicas

HECSA - Escola de Negócios

FIAM-FAAM-FMU

Conteúdo

Conceitos que aprendemos em Aulas anteriores

Conceitos que aprendemos em Aulas anteriores

Na aula passada começamos a estudar Inferência Estatística. Em particular, falamos de:

- população e amostra;
- amostragem;
- distribuição amostral;
- tamanho da amostra.

Distribuição Amostral da Proporção

Distribuição Amostral da Proporção

Vamos considerar uma população em que a proporção de elementos portadores de certa característica p.

Distribuição Amostral da Proporção

Logo podemos definir uma variável aleatória X da seguinte maneira:

$$X = \begin{cases} 1 \text{ se o indivíduo for portador da característica} \\ 0 \text{ se o indivíduo não for portador da característica} \end{cases}$$

Distribuição Amostral da Proporção

Daí

$$E(X) = \mu = p \text{ e } \mathrm{Var}(X) = \sigma^2 = p(1 - p).$$

Distribuição Amostral da Proporção

Retirada uma AAS dessa população, e indicada por Y_n o total de indivíduos portadores da característica na amostra, já vimos que

$$Y_n \sim b(n, p).$$

Distribuição Amostral da Proporção

Vamos definir por $\hat{p}$ a proporção dos indivíduos portadores da característica na amostra, isto é,

$$\hat{p} = \frac{Y_n}{n}.$$

Distribuição Amostral da Proporção

Vimos que a distribuição binomial pode ser aproximada pela distribuição normal. O TLC nos dá uma outra justificativa para isso.

Distribuição Amostral da Proporção

Assim, $\overline{X}$ pode ser aproximada por uma normal, ou seja

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right).$$

Distribuição Amostral da Proporção

Se quisermos trabalhar com a Normal Padrão, temos

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N\left(0, 1\right).$$

Exemplo 2.1

Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e calculamos

$$\hat{p} = \text{ proporção de mulheres na amostra.}$$

Qual a probabilidade de que $\hat{p}$ difira de p em menos de $0,01$?

Propriedades dos Estimadores

Propriedades dos Estimadores

Vimos que a Inferência Estatística tem por objetivo fazer generalizações sobre uma população, com base nos dados de uma amostra.

Propriedades dos Estimadores

Salientamos que dois problemas básicos nesse processo são:

1. estimação de parâmetros; e
2. teste de hipóteses sobre parâmetros.

Propriedades dos Estimadores

Lembremos que parâmetros são funções de valores populacionais, enquanto estatísticas são funções de valores amostrais.

Propriedades dos Estimadores

Inicialmente vejamos a questão da estimação de um modo mais geral. Consideremos uma amostra $(X_1, X_2, ..., X_n)$ de uma variável aleatória que descreve uma característica de interesse de uma população.

Propriedades dos Estimadores

Seja θ um parâmetro que desejamos estimar, como por exemplo a média $\mu = E(X)$ ou a variância $\sigma^2 = \text{Var}(X)$.

Propriedades dos Estimadores

Definição 3.1

Um **estimador** T do parâmetro θ é qualquer função das observações da amostra, ou seja, $T = g(X_1, ..., X_n)$.

Notemos que, segundo essa definição, um estimador é o que chamamos antes de estatística, porém associando-o a um parâmetro populacional.

O problema da estimação é, então, determinar uma função $T = g(X_1, X_2, ..., X_n)$ que seja "próxima" de θ, segundo algum critério. O primeiro critério que iremos abordar é dado a seguir.

Definição 3.2

O estimador T é **não-viesado** para o parâmetro θ se $E(T) = \theta$. Caso T seja viesado, a diferença $V(T) = E(T) - \theta$ é chamada **viés** de T.

Propriedades dos Estimadores

Definição 3.3

Estimativa é o valor assumido pelo estimador em uma particular amostra.

Exemplo 3.4

No Exemplo 2.1 desta Aula, p é um parâmetro (neste caso, conhecido), $\hat{p}$ é um estimador enquanto que os valores de $\hat{p}$ são estimativas.

Exemplo 3.5

Vimos (Teorema do Limite Central) que, uma vez extraída da população X uma amostra $(X_1, ..., X_n)$, a média amostral $\overline{X}$ é um estimador não-viesado de $E(X) = \mu$.

Exemplo 3.6

Da mesma maneira, a proporção amostral $\hat{p}$ é um estimador não-viesado da proporção p de indivíduos de uma população que tem certa característica comum.

Exemplo 3.7

Considere uma população com N elementos e a variância populacional

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2.$$

Um possível estimador para σ^2, baseado numa AAS de tamanho n extraída dessa população, é

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2.$$

Mas **pode-se demonstrar que $\hat{\sigma}^2$ é um estimador viesado**.

Exemplo 3.8

Por outro lado, o estimador

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu)^2$$

é não-viesado para σ^2.

Propriedades dos Estimadores

Há outras propriedades dos estimadores (que não entraremos em grandes detalhes nessa disciplina). Por exemplo, podemos considerar a eficiência:

Definição 3.9

Se T_1 e T_2 são dois estimadores não-viesados de um mesmo parâmetro θ, diremos que T_1 é **mais eficiente** que T_2 se

$$\text{Var}(T_1) < \text{Var}(T_2).$$

Intervalo de Confiança

Intervalo de Confiança

Até agora, todos os estimadores apresentados foram pontuais, isto é, especificam um único valor para o estimador.

Intervalo de Confiança

Esse procedimento não permite julgar qual a possível magnitude do erro que estamos cometendo.

Intervalo de Confiança

Daí, surge a ideia de construir os intervalos de confiança, que são baseados na distribuição amostral do estimador pontual.

Exemplo 4.1

Suponha que queiramos estimar a média μ de uma população qualquer, e para tanto usamos a média $\overline{X}$ de uma amostra de tamanho n.

Intervalo de Confiança

Do Teorema do Limite Central,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Daqui podemos determinar a probabilidade de cometermos erros de determinadas magnitudes.

Intervalo de Confiança

Convém lembrar que μ não é uma variável aleatória e sim um parâmetro, e a expressão que obtemos deve ser interpretada da seguinte maneira:

Intervalo de Confiança

Se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma

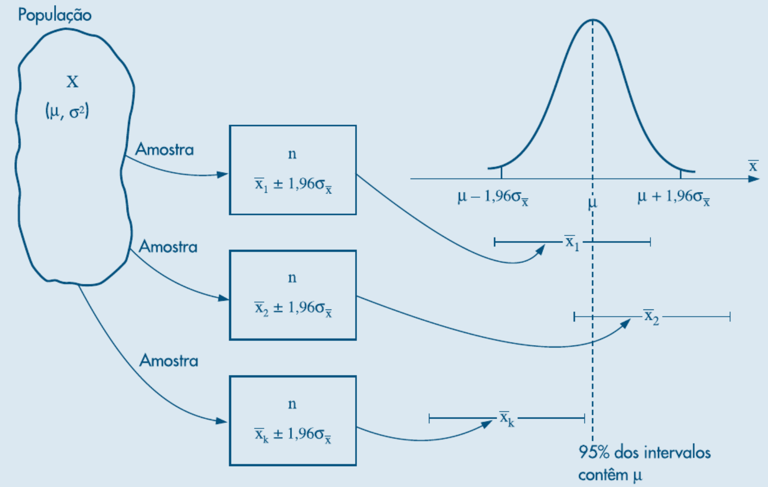
$$]\overline{X} - z_\gamma \sigma_{\overline{X}}, \overline{X} + z_\gamma \sigma_{\overline{X}}[\text{ com } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

todos baseados em amostras de tamanho n, 95% deles conteriam o parâmetro μ.

Intervalo de Confiança

Dizemos que γ é o coeficiente de confiança. Na figura estão esquematizados o funcionamento e o significado de um intervalo de confiança (IC) para μ com $\gamma = 0.95$ e σ^2 conhecido.

Intervalo de Confiança



Escolhida uma amostra e encontrada sua média $\overline{X}_0$, e admitindo-se σ_X conhecido, podemos construir o intervalo

$$]\overline{X} - z_\gamma \sigma_{\overline{X}}, \overline{X} + z_\gamma \sigma_{\overline{X}}[.$$

Esse intervalo **pode ou não** conter o parâmetro μ, mas pelo exposto acima **temos** 95% **de confiança** de que contenha.

Exemplo 4.2

Uma máquina enche pacotes de café com uma variância igual a $100g$. Ela estava regulada para encher os pacotes com $500g$, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou uma média igual a $485g$. Vamos construir um intervalo de confiança com 95% de confiança para μ.

Exemplo 4.3

Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas sobre determinado produto, e 60% delas preferiram a marca A. Vamos construir um intervalo de confiança com 95% de confiança para p com $\gamma = 0.95$.

Exemplo 4.4

Suponha que em $n = 400$ provas obtemos $k = 80$ sucessos. Vamos obter um intervalo de confiança para p com $\gamma = 0.90$.

Intervalo de Confiança

Se a variância populacional σ^2 não for conhecida, podemos substituir σ_X por $S/\sqrt{n}$. Para n grande, da ordem de 100, o intervalo de confiança com essa modificação, pode ainda ser usado.

Exemplo 4.5

Voltando ao exemplo da máquina que enche pacotes de café, agora suponha que ela estava regulada para encher os pacotes com $500g$, em média (note que agora não temos informações sobre a variância). Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 150 pacotes apresentou uma média igual a $485g$ com variância amostral $s^2 = 72$. Vamos construir um intervalo de confiança com 95% de confiança para μ.

Intervalo de Confiança

Para n não muito grande, a distribuição normal não pode mais ser usada e terá de ser substituída pela distribuição t de Student, que estudaremos após lidarmos com testes de hipóteses com variância conhecida.

Comentários Finais

Comentários Finais

Em resumo, na aula de hoje nós:

- distribuição amostral da proporção;
- estimadores e estimativas;
- propriedades dos estimadores;
- intervalo de confiança.

Comentários Finais

Nas próximas aulas nós vamos focar em Testes de Hipóteses. Mais especificamente vamos:

- definir o que é um Teste de Hipóteses;
- falaremos do Teste para Média com Variância conhecida;
- falaremos do Teste para proporção.

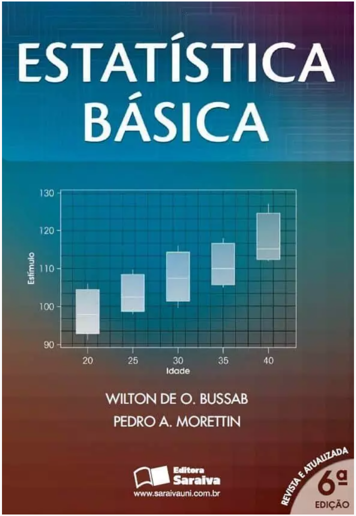
Comentários Finais

ATIVIDADE PARA ENTREGAR (E COMPOR A NOTA N1)

Em grupos de até 4 integrantes resolva os Exercícios 5.1-5.6.

Referências

Referências



Bons Estudos!