REVISÃO DE INFERÊNCIA ESTATÍSTICA BÁSICA

ESTIMAÇÃO
INTERVALO DE CONFIANÇA
TESTE DE HIPÓTESES

AIRLANE P. ALENCAR - IME-USP

PARÂMETRO E ESTIMADORES

Amostragem

População
Parâmetros
μ
σ
p

Amostra
Estimadores
$\bar{X}$
S^2
$\hat{P}$

Inferência

MÉTODOS DE ESTIMAÇÃO

- **Amostra aleatória: $X_1,\ldots X_n$**
- **Método dos momentos:** Utiliza momentos amostrais para estimar momentos populacionais (Bussab e Morettin).
- O primeiro momento populacional é E(X)=μ e o primeiro momento amostral é $\bar{X}$. O método dos momentos estima o 1o momento populacional usando o 1o momento amostral: $\hat{\mu}=\bar{X}$
- $$E\left(\hat{X}^{2}\right)=\frac{\sum_{i=1}^{n} X^{2}}{n}$$

$$\hat{\sigma}^{2}=\widehat{E\left(X^{2}\right)}-\hat{\mu}^{2}=\frac{\sum_{i=1}^{n} X^{2}}{n}-\bar{X}^{2}=\frac{\sum_{i=1}^{n}\left(X-\bar{X}\right)^{2}}{n}$$

MÉTODOS DE ESTIMAÇÃO

- **Amostra aleatória: $X_1,\ldots X_n$**
- **Método de mínimos quadrados:** Minimiza S= soma dos quadrados da diferença entre as observações e seu parâmetro de locação:

$$S=\sum_{i=1}^{n}(X_i-\mu)^2 \qquad \frac{dS}{d\mu}=\sum_{i=1}^{n}-2(X_i-\mu)$$

A derivada é zero se $\sum_{i=1}^{n} X_i=n\hat{\mu}\Rightarrow\hat{\mu}=\bar{X}$

Como $\frac{d^2S}{d\mu^2}=2n>0$, então $\hat{\mu}=\bar{X}$ é ponto de mínimo.

MÉTODOS DE ESTIMAÇÃO

Método de Máxima Verossimilhança: Assume que cada X segue distribuição que depende dos parâmetros e escreve a verossimilhança que corresponde à densidade conjunta de $X_1,...,X_n$.

Se cada X ~ N(μ, σ²), a densidade de X é

$$f(x1,\ldots,xn;\ \mu,\ \sigma^2) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(\frac{-\sum\left(x_i-\mu\right)^2}{2\sigma^2}\right)$$

A função de verossimilhança L(μ, σ²) é igual a essa densidade vista como função de μ, σ².

$$L(\mu,\ \sigma^2) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(\frac{-\sum\left(x_i-\mu\right)^2}{2\sigma^2}\right)$$

MÉTODOS DE ESTIMAÇÃO

L(μ, σ²) = $$\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(\frac{-\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{2\sigma^2}\right)$$

Melhor maximizar log de L:

log L(μ, σ²) = $$-\frac{n}{2}\ln\left(2\pi\right)-\frac{n}{2}\ln\left(\sigma^2\right)-\left(\frac{\sum\left(x_i-\mu\right)^2}{2\sigma^2}\right)$$

Encontre μ que maximize log L.

PESO MÉDIO DE CRIANÇAS DE 1 ANO

Probabilidade: Parâmetros conhecidos

O peso médio populacional de crianças de 1 ano é de 10 quilos com desvio padrão populacional 1 quilo. Também podemos assumir que a distribuição dos pesos é gaussiana.

a) Para uma criança com essa idade, qual a probabilidade de que ela tenha menos que 8 quilos? E mais que 12 quilos?

b) Encontre um valor de peso x tal que a probabilidade do peso ser menor que x seja de 99%. Esse valor x é o quantil 99%.

ESTIMADOR

Na prática, os **parâmetros mas não são conhecidos**.

Os parâmetros são fixos, como o peso médio populacional (inferência clássica).

Suponha que não conhecemos o peso médio mas conhecemos o desvio padrão.

Temos que **estimar o peso médio**.

Estimador $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ é **variável aleatória**.

Qual sua esperança? E variância? E erro padrão?

É não viesado para estimar μ? É consistente?

Rever Esperança e Variância cap. 6 em Bussab e Morettin

PESO MÉDIO

A partir de uma amostra aleatória de 25 crianças de 1 ano, obteve-se peso médio amostral de 11 quilos e o desvio padrão populacional igual a 1 quilo (isso não ocorre na prática).

c) Qual o intervalo de confiança para o peso médio com coeficiente de confiança de 95%?

d) Teste se o peso médio é maior que 10 quilos ou não com nível de significância de 5%.

- Hipóteses nula e alternativa,
- Estatística para o teste,
- Região crítica
- Valor p

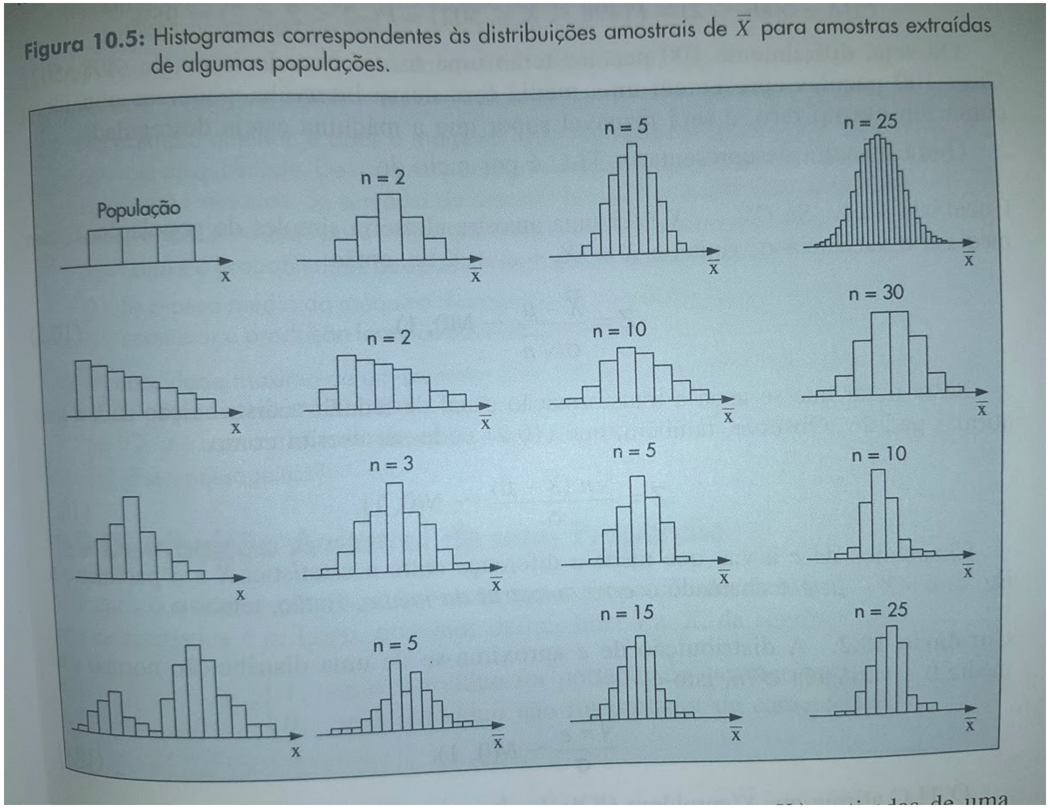
DESVIO PADRÃO DESCONHECIDO

Quando vale ?

Se cada X ~ Normal, ok.

E se n for bem grande? Teorema Central do Limite

(Ler cap. 10 BM)



VARIÂNCIA DESCONHECIDA

Estimar a variância σ^2 usando a variância amostral

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Calcule a esperança.

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n}(X_i - \mu)^2 + (\mu - \bar{X})^2 + 2(X_i - \mu)(\mu - \bar{X})}{n-1} = \dots$$

VARIÂNCIA DESCONHECIDA

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

E se substituo σ^2 pelo estimador não viesado S^2, o que ocorre?

Gosset = Student(1908). Pesquisem.

- $\frac{(n-1)S^2}{\sigma^2}$ tem dist. qui quadrado com n-1 graus de liberdade: se X normal
- S^2 e $\bar{X}$ independentes

Teorema: $$\frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$ Cauda + pesada!

PESO MÉDIO

Agora temos amostra aleatória de 25 crianças de 1 ano, obteve-se peso médio de 11 quilos e **desvio padrão amostral 1kg**.

c') Qual o intervalo de confiança para o peso médio com coeficiente de confiança de 95%?

d') Teste se o peso médio é maior que 10 quilos ou não com nível de significância de 5%.

- Hipóteses nula e alternativa,
- Estatística para o teste,
- Região crítica
- Valor p

PROPORÇÃO

Pré requisitos: Qual a distribuição da proporção amostral?

Quais as suposições necessárias para obtê-la?

Um instituto de pesquisa de opinião entrevistou amostra aleatória de 400 fãs de basquete e descobriu que 68% são a favor da aplicação de testes para detectar uso de drogas (doping) nos jogadores.

a) Apresente o intervalo de confiança para a proporção de fãs a favor dos testes com coeficiente de confiança de 95%. Qual a interpretação do coeficiente de confiança 95%?

b) Considerando o nível de confiança de 95%, qual deveria ser o tamanho da amostra para que o erro amostral fosse de 2 pontos percentuais para baixo e para cima? Calcule esse tamanho de amostra sem usar a informação da pesquisa que foi realizada.

REFERÊNCIA

Bussab e Morettin. Estatística básica. Saraiva.