Universidade Federal do Rio Grande do Norte
Instituto Metrópole Digital
IMD0601 - Bioestatística

Estatística inferencial

Prof. Dr. Tetsu Sakamoto
Instituto Metrópole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br

Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601

> git pull
```

Aula passada

- **Simulação Monte Carlo**
  - **Função Distribuição Acumulada Inversa;**
  - **Amostragem por rejeição;**
- **MCMC**
  - **Algoritmo de Metropolis**

Estatística inferencial

- Fundamentos da Estatística Inferencial
  - Teoria da amostragem;
  - Distribuição amostral;
  - Estimativa de parâmetros;
  - Bootstraping;
- Teste de hipótese;

Teoria de amostragem

Relação entre a amostra e a população

- Envolve tópicos como a coleta, análise e interpretação dos dados coletados de uma amostra aleatória de uma população em estudo.
- Quando coletamos dados, estamos amostrando de uma população;

Teoria de amostragem

Amostra:

- Subconjunto de uma população;
- Boa amostra ➔ representativo de uma população em estudo;
    - Princípio da aleatoriedade;
    - Propriedade da independência entre os resultados;
    - Os resultados compartilham a mesma distribuição de probabilidade;

Teoria de amostragem

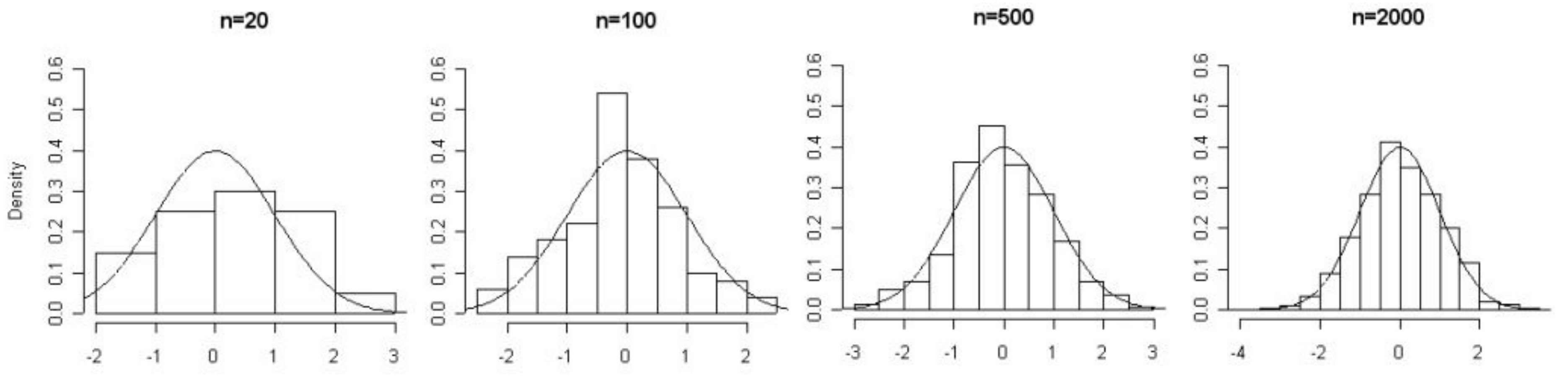
Tamanho da Amostra (n):

- Quanto maior a sua amostra, melhor é a sua amostragem;

Teoria de amostragem

Tamanho da Amostra (n):

- Quanto maior a sua amostra, melhor é a sua amostragem;



Efeito do tamanho da amostra na caracterização de uma população que segue uma distribuição normal.

Teoria de amostragem

Média amostral:

- Média da amostra;
- O mais comum das medidas de tendência central;
- Se pegarmos repetidos dados amostrados de uma mesma população e calcular para cada uma a média amostral ➔ distribuição da média amostral seguirá o **Teorema do Limite Central**;

Teoria de amostragem

Teorema do Limite Central:

- Lei das médias;
- A distribuição da média amostral de n dados aproxima-se de uma distribuição normal quando n é suficientemente alto;
- Se aplica a **qualquer** distribuição, mesmo que ela seja muito diferente de uma distribuição normal.

Teoria de amostragem

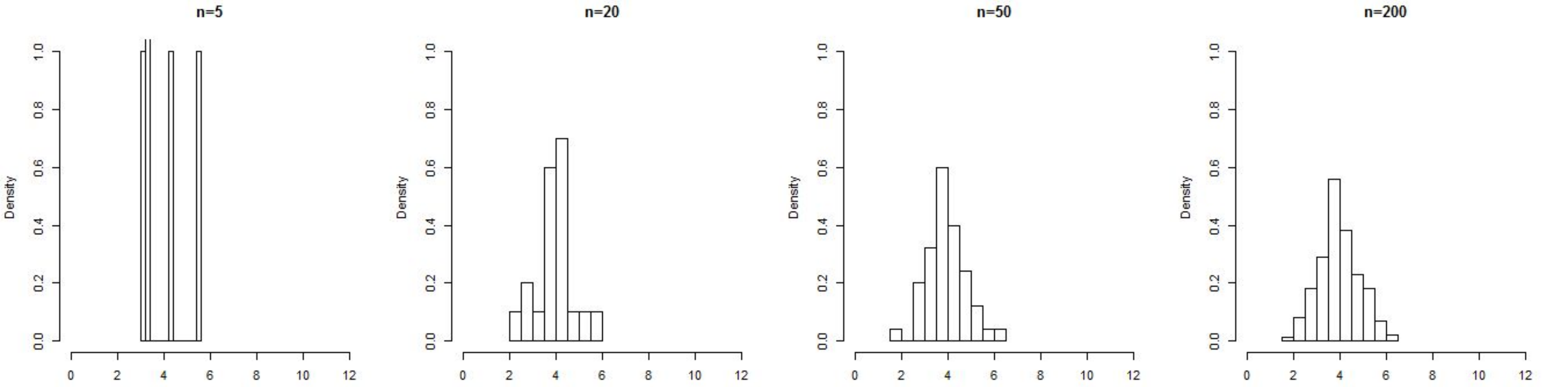
Teorema do Limite Central:

- Variando o número de repetições dos dados (tamanho dos dados = 20):

Teoria de amostragem

Teorema do Limite Central:

- Variando o número de repetições dos dados (tamanho dos dados = 20):



Teoria de amostragem

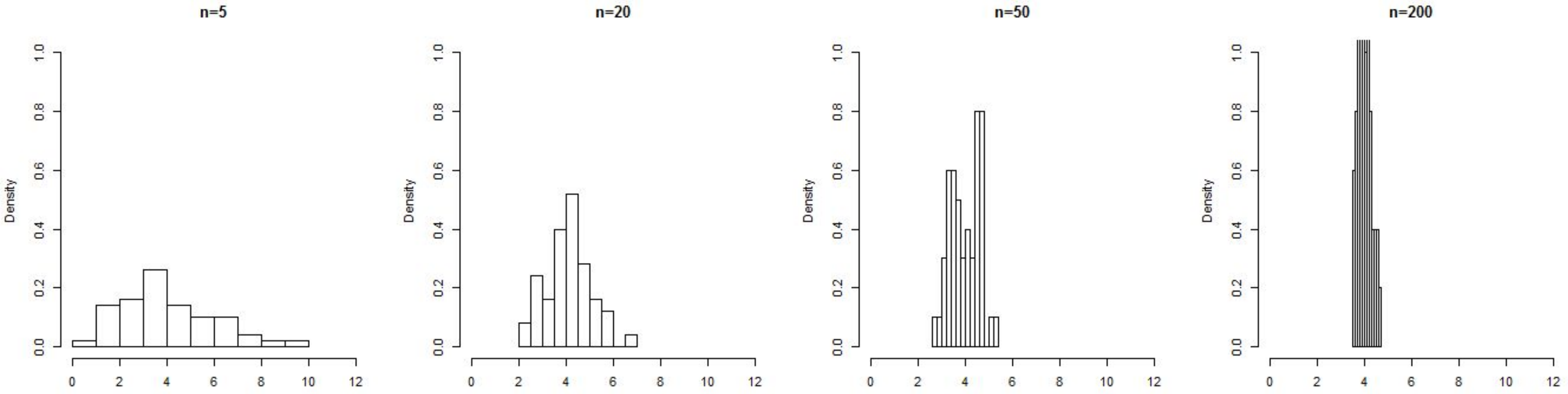
Teorema do Limite Central:

- Variando o tamanho da amostra dos dados (# repetições = 50):

Teoria de amostragem

Teorema do Limite Central:

- Variando o tamanho da amostra dos dados (# repetições = 50):



Teoria de amostragem

Distribuição da média amostral:

- Dispersão da distribuição diminui à medida que o tamanho amostral (n) aumenta;
- **Erro padrão das médias** (Desvio padrão aplicada amostragens repetidas da média amostral) é inversamente proporcional a raiz quadrada do tamanho amostral.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Aumentando o tamanho amostral, aumenta a precisão ➔ implicações nos testes.

Teoria de amostragem

Distribuição da média amostral:

- Dispersão de uma amostra é normalmente medido pelo desvio padrão;
- Desvio padrão ➔ raiz quadrada da média do quadrado dos desvios.
- Correção de Bessel;

$$\boldsymbol{s} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

Distribuição amostral

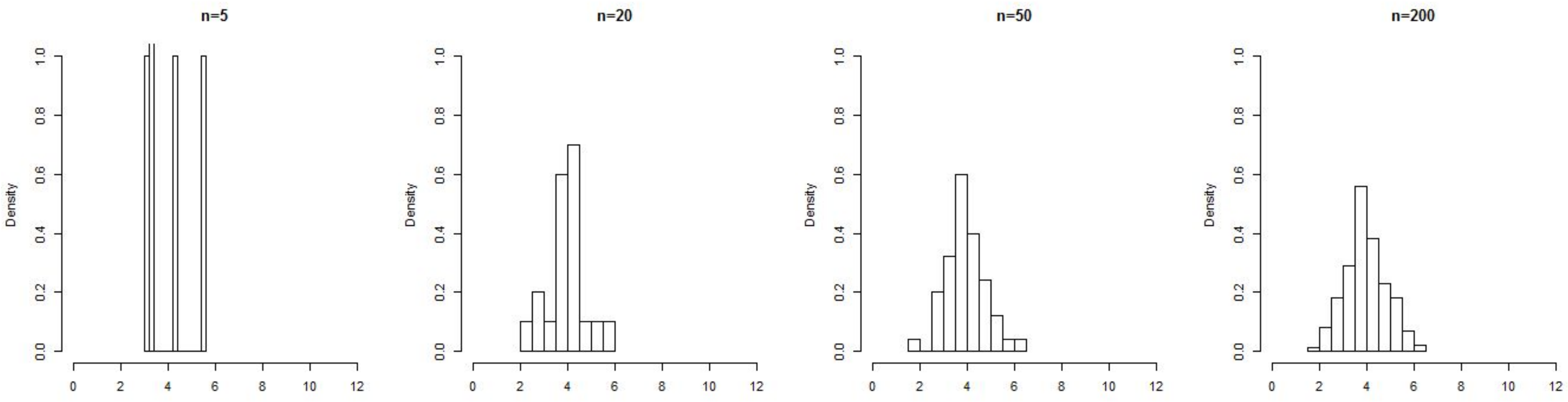
Distribuição das medidas estatísticas

Distribuição das medidas estatísticas → média, mediana, moda, ...

Distribuição cuja forma modifica com o tamanho amostral (n);

Grau de liberdade → é um dos parâmetros que define o formato da distribuição. Ele é, em parte, baseado no tamanho amostral,

Distribuição amostral das médias amostrais



Distribuição amostral das médias amostrais tendem a seguir uma distribuição normal;

Normalizar as médias amostrais com Zscore; $Z = \frac{X-\mu}{\sigma}$ $t = \frac{X-\mu}{s}$

Distribuição amostral das médias amostrais

Distribuição t de Student

- Descreve a Distribuição amostral das médias amostrais quando a variância populacional é desconhecida;
- Base do teste t, que compara as médias de diferentes amostras.
- Modificação da distribuição normal padrão que considera a variabilidade do valor do desvio padrão, já que o verdadeiro sigma não é conhecido;

$$t = \frac{X - \mu}{s}$$

$$Z = \frac{X - \mu}{\sigma}$$

Distribuição amostral

Distribuição t de Student

- Se os valores $x_1, x_2, \ldots, x_n$ seguem uma distribuição normal, então a distribuição do t-score dos valores correspondentes é uma distribuição t com (n - 1) graus de liberdade;

$$t = \frac{X - \mu}{s}$$

- Também é aplicável na distribuição amostral da média amostral;

Média amosral

$$t = \frac{\bar{X} - \mu}{s(\bar{X})}$$

Erro padrão
((s)/(n)½)

Distribuição amostral

Grau de liberdade

Conceito aplicado quando se quer estimar um parâmetro;

Número de observações independentes utilizadas para estimar um parâmetro;

Estimar a variância da altura de uma população de plantas, considerando que a média é conhecida (média = 6);

Uma amostra: $h = \{8\} \rightarrow V = (8-6)^2 = 4$; GL = 1;

Duas amostras: $h = \{8, 5\} \rightarrow V = ((8-6)^2+(5-6)^2)/2 = 2{,}5$; GL = 2

Distribuição amostral

Grau de liberdade

Conceito aplicado quando se quer estimar um parâmetro;

Número de observações independentes utilizadas para estimar um parâmetro;

Estimar a variância da altura de uma população de plantas, considerando que a média não é conhecida;

Duas amostras: $h = \{8, 5\} \rightarrow M = (8+5)/2 = 6{,}5$

$$V = ((8-6{,}5)^2+(5-6{,}5)^2)/2 = 2{,}25;\ GL = 1$$

Observações não independentes

Distribuição amostral

Grau de liberdade

Conceito aplicado quando se quer estimar um parâmetro;

Número de observações independentes utilizadas para estimar um parâmetro;

Grau de liberdade corresponde normalmente ao número de observações menos o número de parâmetros estimados para estimar o parâmetro em questão.

No exemplo anterior, tivemos que estimar a média para depois estimar a variância, por isso, GL = 2 - 1 = 1;

Distribuição amostral

Distribuição t de Student

- Simétrica como a normal;
- Média = 0;
- Desvio padrão > 1;

Média amosral

$$t = \frac{\bar{X} - \mu}{s(\bar{X})}$$

Erro padrão
((s)/(n)½)

Distribuição amostral

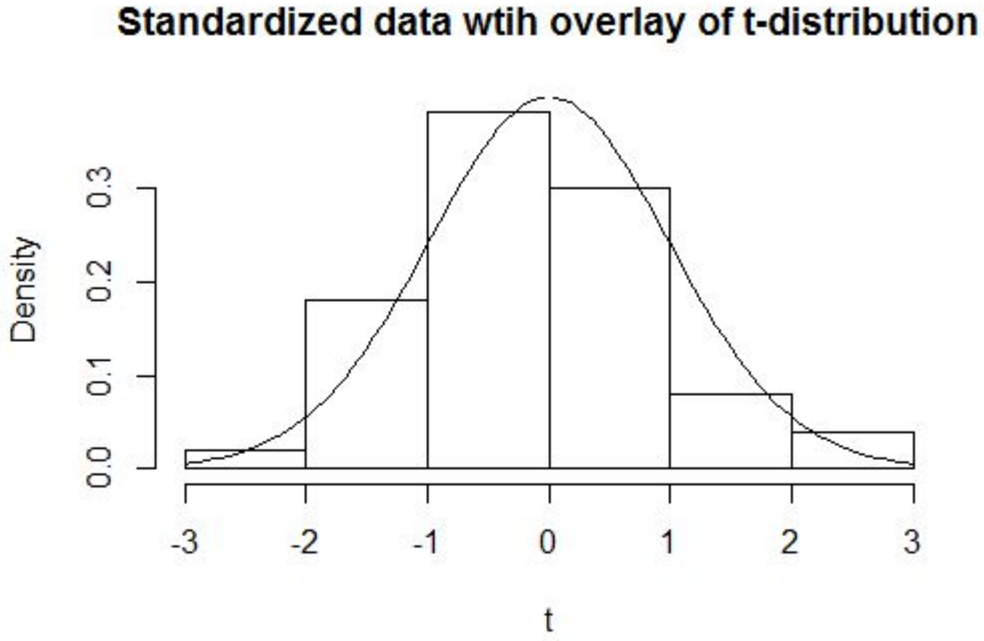
Distribuição t de Student

- Simétrica como a normal;
- Média = 0;
- Desvio padrão > 1;

Média amosral

$$t = \frac{\bar{X} - \mu}{s(\bar{X})}$$

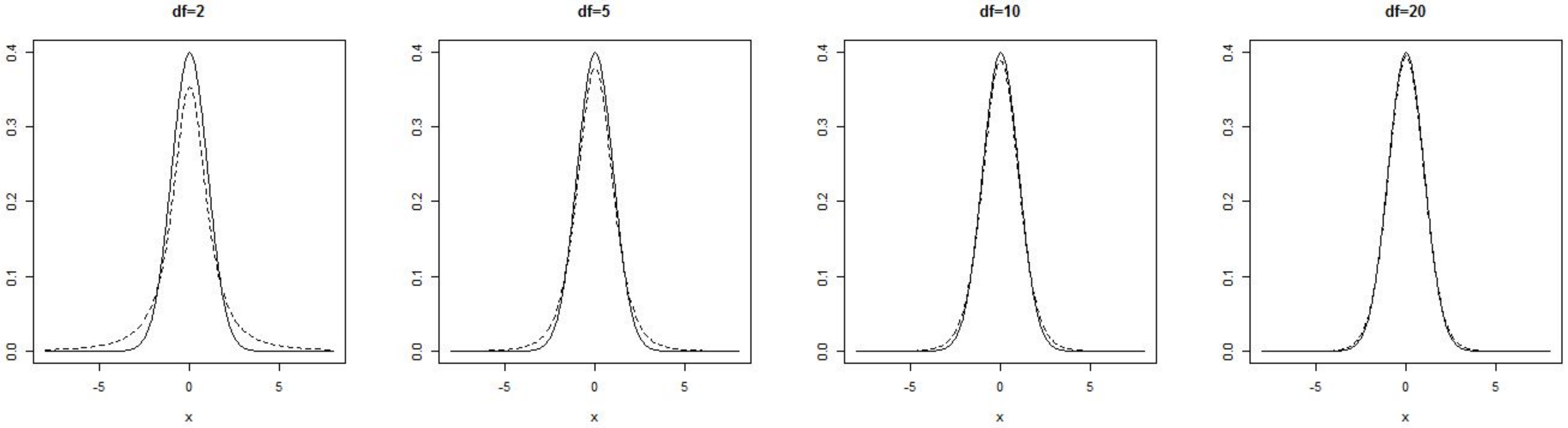
Erro padrão
((s)/(n)½)



Distribuição amostral

Distribuição t de Student

- Efeito do grau de liberdade ➜ Linha contínua: Distribuição normal; pontilhada: distribuição-t;



Distribuição amostral das variâncias amostrais

- Não existe um padrão como o que ocorre nas distribuições das médias amostrais;
- Mas...
  - Soma do quadrado dos valores que seguem uma distribuição N(0, 1) → **Distribuição qui-quadrado**

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$