

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Estat stica inferencial

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole 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

- **Teoria da amostragem**
 - **Amostra;**
 - **Tamanho da amostra;**
 - **Teorema Central do Limite;**
 - **Distribuição amostral;**
 - **Distribuição t de Student**
 - **Distribuição qui-quadrado**

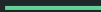
Estatística inferencial

- Estimativa de parâmetros;
 - Máxima Verossimilhança
 - Intervalo de confiança;
- Bootstrapping;

Estimativa de parâmetro

... para ajustar a um modelo

- Dado que você tenha dados → ajustar a um modelo;
- Modelo probabilístico padrão;
- Necessário estimar os melhores valores para os parâmetros desses modelos.

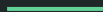


Estimativa de parâmetro

... para ajustar a um modelo

Duas abordagens:

- **Estimativa pontual** → Estimativa de um valor para um parâmetro populacional;
- **Estimativa intervalar** → intervalo de valores usados para estimar um parâmetro;



Estimativa de parâmetro

... para ajustar a um modelo

Exemplo:

- Amostra: Altura de pessoas (cm) → 165, 170, 170, 168, 178, 180, 168, 150, 172;
 - Estime o valor da média populacional → *estimativa pontual*;
 - Estime o intervalo de valores que a média populacional pode assumir → *estimativa intervalar*.
-

Estimativa pontual

Um valor para um parâmetro populacional

Muitos parâmetros são simples de serem calculados a partir dos dados:

- Média \rightarrow `mean(x)`;
- Variância \rightarrow `var(x)`;
- Desvio padrão \rightarrow `sd(x)`

Outros parâmetros são mais complicados de serem estimados:

- Alfa e beta de uma distribuição gama ou beta;
-

Estimativa pontual

Um valor para um parâmetro populacional

Para os parâmetros mais complexos
→ métodos mais sofisticados;

- Estimativa dos momentos;
- **Máxima Verossimilhança;**
- Mínimos quadrados;
- Métodos Bayesianos;

Máxima verossimilhança

$F(\theta|\text{dados}) \rightarrow$ Função de
probabilidade do parâmetro
dado os dados (verossimilhança)

Princípio \rightarrow um valor no intervalo de possíveis valores do parâmetro, a função de verossimilhança atinge o máximo;

**Estimativa da Máxima
Verossimilhança;**

Máxima verossimilhança

$F(\theta|\text{dados}) \rightarrow$ Função de probabilidade do parâmetro dado os dados (verossimilhança)

Analiticamente...

O valor máximo de uma função \rightarrow primeira derivada da função = zero;

Solução MV para uma binomial:

$$V(p) = p^{\sum x_i} (1 - p)^{n - \sum x_i}$$

$$v(p) = \log V(p)$$

$$v(p) = \sum x_i \log(p) + (n - \sum x_i) \log(1 - p)$$

$$v'(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p} = 0$$

$$\underline{p = \frac{\sum x_i}{n}}$$

Estimativa intervalar

Determina a acurácia e a
precisão das estimativas

Método alternativo para estimar um
parâmetro;

Provável intervalo de valores que o
parâmetro assume;

Normalmente realizamos as
estimativas pontuais e intervalares
em conjunto;

Intervalo de confiança;

Intervalo de confiança

O mais comum das estimativas
intervalar

IC = estimativa pontual \pm erro
padrão para a estimativa pontual *
teste estatístico

Intervalo simétrico

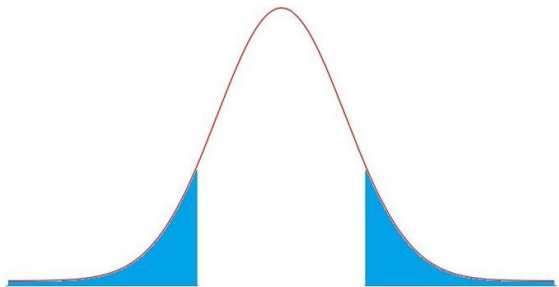
Meio do intervalo \rightarrow estimativa
pontual;

Limites do intervalo:

- Abaixo $\rightarrow EP - ErrP * \text{teste est.};$
 - Acima $\rightarrow EP + ErrP * \text{teste est.};$
-

Intervalo de confiança

O mais comum das estimativas intervalar



Tamanho do IC

- erro padrão
- teste estatístico

Erro padrão:

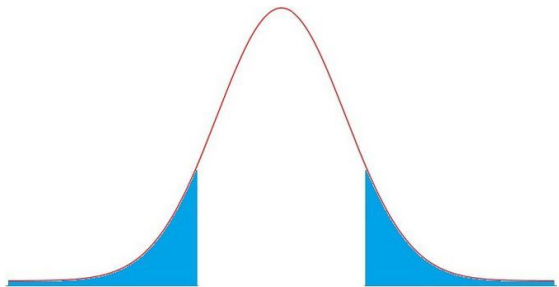
- Quanto menor \rightarrow menor o intervalo, mais preciso;

Teste estatístico:

- $\alpha \rightarrow$ Grau de erro, 5%;
 - Nível do IC = $1-\alpha$
-

Intervalo de confiança

O mais comum das estimativas
intervalar

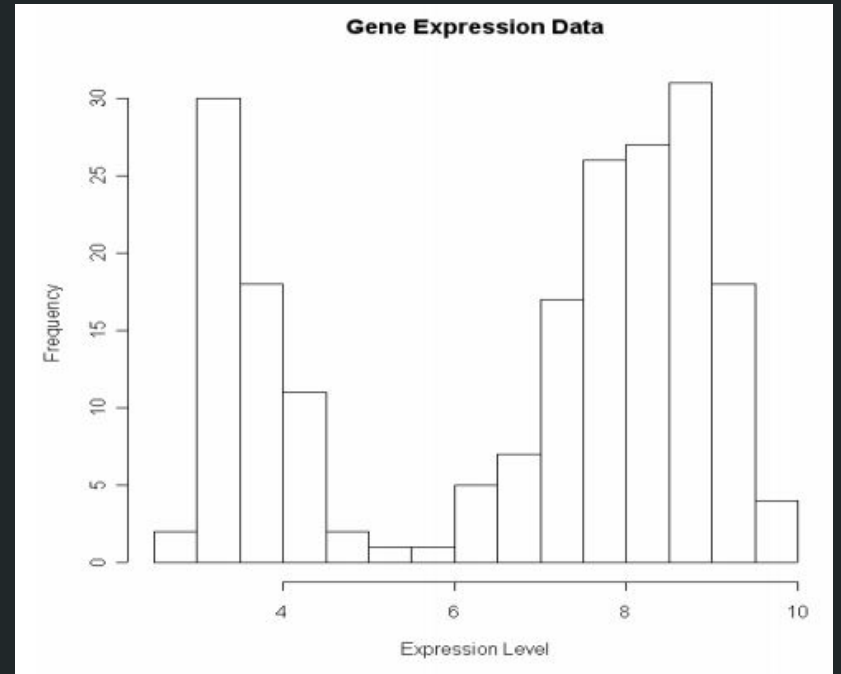


Interpretação: $IC_{95\%} = [-0.301, 0.765]$

- Se repetirmos o experimento várias vezes e calculássemos o IC, espera-se que 95% dos ICs conteria o valor verdadeiro;
- Confiança de 95% de que o verdadeiro valor esteja neste intervalo;

Bootstrapping

Lidando com dados irregulares



Bootstrapping

Lidando com dados irregulares

Desenvolvido por Brad Efron;

Encontrar desvio padrão ou o intervalo de confiança de uma métrica (média, mediana, ...) em dados irregulares;

Utilizado nas áreas de filogenia molecular e análise de expressão gênica;

Paramétrico ou **não paramétrico**;

Bootstrapping

Como funciona?

Amostrar repetidamente dos dados originais;

- A amostragem é feita com reposição;
- O tamanho da amostra é a mesma do original;
- Para cada amostragem do bootstrap, a métrica é calculada;
- Normalmente é obtido 1000 amostras de bootstrap.
- Mede-se a dispersão;

Revisão

Estimativa de parâmetro:

- **Estimativa pontual** - Máxima verossimilhança
- **Estimativa intervalar** - intervalo de confiança

Bootstrapping;
