

# Introdução à Bioestatística

Marcelo Goulart Correia

Instituto Nacional de Cardiologia

March 7, 2016

- 1 Introdução à Estatística Inferencial
- 2 Amostragem
- 3 Hipótese
- 4 Estimativa de confiança

- Fazer inferências acerca da população a partir de uma amostra dela extraída
  - Dimensionar a amostragem de maneira correta
  - Verificar o comportamento dos dados
  - Escolher a técnica correta
  - Inferir informações sobre os resultados

- Dimensionar a amostragem de maneira correta
  - Estudar relações existentes entre a população e a amostra
    - Custo reduzido
    - Maior rapidez
    - Diversas possibilidades

- Dimensionar a amostragem de maneira correta
  - Tipos de amostras
    - Aleatória simples (Com ou sem reposição)
    - Aleatória estratificada
    - Sistemática
    - Conglomerados

- Dimensionar a amostragem de maneira correta
  - Aleatória simples (Com ou sem reposição)
    - População  $E$  com  $N$  elementos - Sem reposição:
      - Probabilidade de  $e$  ( $e \in E$ ) ser escolhido em primeiro lugar  $\rightarrow 1/N$
      - Probabilidade de  $e$  ( $e \in E$ ) ser escolhido em segundo lugar  $\rightarrow 1/N-1$
      - ...
      - Probabilidade de  $e$  ( $e \in E$ ) ser escolhido em  $i$ -ésimo lugar  $\rightarrow 1/(N-i+1)$
    - População  $E$  com  $N$  elementos - Com reposição:
      - Probabilidade de  $e$  ( $e \in E$ ) ser escolhido em qualquer lugar  $\rightarrow 1/N$

- Dimensionar a amostragem de maneira correta
  - Aleatória estratificada
    - Divide-se a população de  $N$  indivíduos em  $k$  estratos
    - A partir de critérios importantes do estudo
    - Alocação proporcional e Alocação ótima

- Dimensionar a amostragem de maneira correta
  - Aleatória estratificada - Alocação proporcional
    - É quando o tamanho da amostra de cada estrato é proporcional ao tamanho do estrato

$$n_i = n * \frac{N_i}{N} \quad (1)$$

- Aleatória estratificada - Alocação ótima
  - Define-se os tamanhos dos estratos seguindo alguns critérios:
    - Minimização de variância
    - Minimização do custo da amostragem



- Dimensionar a amostragem de maneira correta
  - Aleatória sistemática
    - Utilizado quando os elementos da população estão organizados em fichas ou listagens
    - Determina-se um intervalo de retirada ( $k = N/n$ )
    - Seleciona-se um número aleatoriamente de 1 até N para iniciar a amostragem
    - Designa-se para a amostragem o elemento que estiver contido em cada ponto do intervalo de escolha

- Dimensionar a amostragem de maneira correta
  - Aleatória por conglomerados
    - Sortear aleatoriamente regiões e subregiões
    - Coletar todos os elementos desse subconjunto
    - É uma forma de simplificar e baratear a amostragem aleatória simples

- O cálculo do tamanho amostral basea-se em alguns critérios:
  - População
  - Objetivo da pesquisa
  - Tipo de informação da pesquisa

- Tipos de cálculo amostral
  - Proporção de eventos (População x Amostra)
  - Proporção de eventos (Grupo I x Grupo II)
  - Proporção de eventos (Ensaio de não-inferioridade)
  - Comparação de médias
  - Comparação de médias pareadas
  - Comparação de duas curvas de sobrevivência
  - ...

- Parâmetros da amostragem
  - Alfa ( $1 - \alpha$ )  $\rightarrow$  Nível de confiança
    - Probabilidade de apontar um verdadeiro positivo
  - Beta ( $1 - \beta$ )  $\rightarrow$  Poder estatístico
    - Probabilidade de apontar um verdadeiro negativo

### Possibilidades envolvidas em um teste de hipóteses

Realidade		Decisão	Aceitar $H_0$	Rejeitar $H_0$
$H_0$ é verdadeira			<b>Decisão correta</b> $1 - \alpha = P(\text{Aceitar } H_0 / H_0 \text{ é V}) = P(H_0 / H_0)$	<b>Erro do Tipo I</b> $\alpha = P(\text{Erro do tipo I}) = P(\text{Rejeitar } H_0 / H_0 \text{ é V}) = \text{Nível de significância do teste} = P(H_1 / H_0)$
$H_0$ é falsa			<b>Erro do Tipo II</b> $\beta = P(\text{Erro do tipo II}) = P(\text{Aceitar } H_0 / H_0 \text{ é falsa}) = P(\text{Aceitar } H_0 / H_1 \text{ é V}) = P(H_0 / H_1)$	<b>Decisão correta</b> $1 - \beta = P(\text{Rejeitar } H_0 / H_0 \text{ é falsa}) = P(H_1 / H_1) = \text{Poder do teste.}$

- Hipótese nula ( $H_0$ )?!
  - Resposta presumida e provisória que poderá ser rejeitada ou não rejeitada
  - Grosseiramente, afirmar que (há igualdade de valores, valores maiores, menores,...) entre os grupos estudados

- Valor de P!
  - É a probabilidade de uma determinada estatística teste seja igual ou maior que o valor observado
  - O que isso significa?!
  - A probabilidade de se obter um fenômeno (diferença, inferioridade, superioridade,...) **igual ou maior** a encontrada num estudo, dado que ambos os grupos analisados sejam de mesma eficácia é de X%

- Valor de P - Cuidados!
  - Significância Estatística x Significância Clínica
  - Amostras muito grandes → Qualquer diferença torna-se significativa (Estatística)
  - Amostras muito diferentes → Valor de p menor
  - Valor de P pequeno não garante que a pesquisa esteja correta!



- Afirmações incorretas sobre o valor de P

- Se  $P = 0,05$ , então a hipótese nula só tem 5% de chance de ser verdadeira
- Uma diferença não significativa ( $p \geq 0,05$ ) significa que **NÃO** existe diferença entre os grupos
- Um achado estatisticamente significativo é clinicamente importante
- Estudos com mesmo valor de P fornecem a mesma evidência contra a hipótese nula
- Valor de P é a probabilidade do resultado ser incorreto por erro de amostragem
- $P = 0,05$  e  $P \leq 0,05$  são as mesmas coisas
- É apropriado escrever o valor de P como inequidade ( $P = 0,015 \rightarrow P \leq 0,02$ )
- $P = 0,05$  significa que se rejeitarmos a hipótese nula, a probabilidade de erro tipo I será igual a 5%
- Se o limite de significância é  $P = 0,05$ , então a chance de encontrarmos um erro tipo I será de 5%
- Conclusões de cunho científico ou políticas de tratamento devem ser baseadas apenas no valor de P
- Valor de P é a probabilidade da hipótese nula ser **VERDADEIRA**

- Atribui um conjunto de valores como estimativa do parâmetro na forma de intervalo
- Quando um intervalo tem probabilidade de  $100*\alpha\%$  do parâmetro não estar incluído nesse intervalo, dizemos então que o intervalo de confiança tem significância é de  $\alpha\%$
- O intervalo de confiança pode ser calculado para:
  - Média
  - Variância
  - Proporção
- Grau de liberdade = Dimensão da amostra - Número de parâmetros estatísticos avaliados

- Os mais utilizados:
  - Média (variância conhecida)

$$\mu = \bar{X} \pm z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} \quad (2)$$

- Diferença entre médias (caso geral)

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2; 1-\alpha/2} * \hat{S} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (3)$$

$$\hat{S}^2 = \frac{(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2}{n_1+n_2-2} // \hat{S}_n = S_n \sqrt{\frac{n}{n-1}}$$

- Proporção

$$p = \hat{p} \pm z_{1-\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (4)$$

- Exercício