Paradigma frequentista de Inferência Estatística

Ideias sobre distribuição amostral

Métodos Estatísticos em Pesquisa Científica (MEPC)

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística
Laboratório de Estatística e Geoinformação

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Introdução

Já discutimos os conceitos de **população**, **amostra** e **inferência**.

- **População**: conjunto de todos os elementos que compartilham alguma característica comum que temos interesse em estudar.
- **Amostra**: subconjunto da população.
- **Inferência**: ramo da Estatística que tem como objetivo estudar a **população** por meio de evidências fornecidas por uma **amostra**.

Introdução

- Muitas vezes estamos interessados em quantidades populacionais, contudo trabalhar com a população pode ser custoso ou até mesmo impossível.

- A solução é trabalhar com um subconjunto da população, isto é, uma amostra.

- O objetivo das técnicas de **amostragem** é gerar um subconjunto que seja **representativo** em relação a população para estimar quantidades de interesse (uma média, uma variância, uma proporção, etc).

Introdução

- Contudo é intuitivo notar que, caso se repita o processo de amostragem, uma amostra diferente da inicial será obtida.
- Consequentemente, as medidas de interesse calculadas (média, variância, etc.) em diferentes amostras não serão iguais.
- Isto quer dizer que mesmo o procedimento de amostragem estando correto sempre haverá aleatoriedade envolvida e os valores calculados com base na amostra são **candidatos** à quantidade na população.

Introdução

- Devido à natureza aleatória, todas as quantidades associadas à amostra devem receber tratamento probabilístico.

- Levando isso em conta, são objetivos da inferência estatística:
  1. Estimar quantidades com base apenas na amostra (valor pontual).
  2. Avaliar o quão preciso ou creditável é o valor estimado (intervalo de confiança).
  3. Decidir sobre possíveis valores da quantidade baseado apenas na amostra (teste de hipótese).

Distribuição amostral

Distribuição amostral

- Estimativas são **variáveis aleatórias** (sabemos o que pode acontecer, mas não o que vai acontecer).
- Variáveis aleatórias têm distribuição de probabilidade.
- A **distribuição de probabilidades de estimativas** é chamada de **distribuição amostral**.
- Para estudar um **parâmetro**, usamos a distribuição amostral.
- No **paradigma frequentista** pensamos no que aconteceria se diversas amostras fossem tomadas e em cada amostra a quantidade de interesse fosse obtida.

Distribuição amostral

- Imagine que:
  - Coletamos diversas amostras.
  - Em cada amostra calculamos o estimador de interesse (uma média, por exemplo).
  - Se obtivermos a distribuição empírica deste estimador, temos tudo que precisamos para fazer inferência.
- A distribuição amostral pode ser usada para avaliar o que aconteceria se o estudo fosse replicado um grande número de vezes.
- A distribuição amostral é o objeto de inferência (frequentista).
  - A **estimativa pontual** é um resumo da distribuição amostral.
  - Intervalos entre **quantis** representam a incerteza sobre o valor estimado.

Teorema Central do Limite

Teorema Central do Limite

- A média é uma das quantidades de maior interesse em contextos práticos.
- A distribuição amostral da média é conhecida graças ao Teorema Central do Limite (TCL).
- Segundo o teorema, quanto maior o tamanho da amostra, a distribuição da média amostral se comporta segundo um modelo Normal.

Teorema Central do Limite

- Suponha uma amostra aleatória de tamanho n retirada de uma população com média μ e variância σ^2 finita.

- Note que o modelo da variável aleatória não é especificado.

- A amostra $(Y_1, Y_2, \cdots, Y_n)$ consiste de n variáveis aleatórias independentes e identicamente distribuidas.

Teorema Central do Limite

Segundo o teorema:

$$\left(\frac{\bar{Y}-\mu}{\sigma\sqrt{n}}\right) \xrightarrow{D} Z \sim N(0{,}1), \text{ para } n \to \infty$$

De forma alternativa:

$$\bar{Y} \sim N(\mu; \sigma^2/n)$$

O resultado pode ser generalizado para proporções:

$$\hat{p} \sim N\left(p; \frac{p(1-p)}{n}\right)$$

Ilustração computacional

Ilustração computacional

- Um questionário foi aplicado a uma turma com diversas questões sobre características dos alunos.

- Uma das questões perguntava a altura dos alunos. Consideraremos que a turma é uma população e temos interesse em fazer inferência sobre a altura média desta população.

- Para isso:
  1. Tomamos diversas amostras.
  2. Para cada amostra calculamos a média.
  3. Considerando o vetor de médias, construimos a distribuição amostral.
  4. Com base na distribuição amostral empírica, fazemos inferência (estimativa pontual e intervalar).

- Neste caso, sabemos a verdadeira altura média. Logo, podemos verificar se nossa inferência foi bem sucedida.

Ilustração computacional

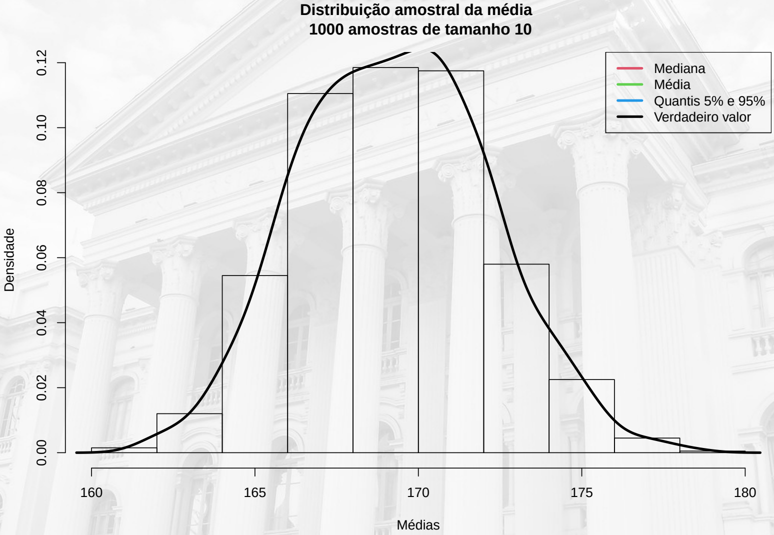


Ilustração computacional

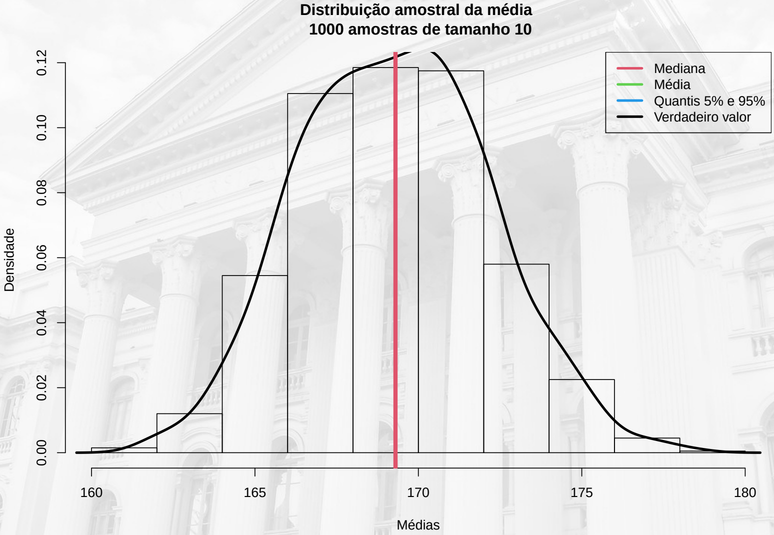


Ilustração computacional

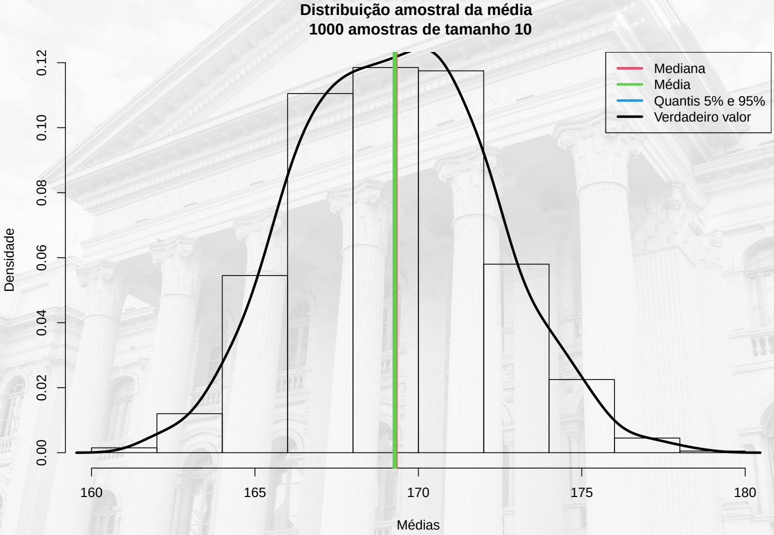


Ilustração computacional

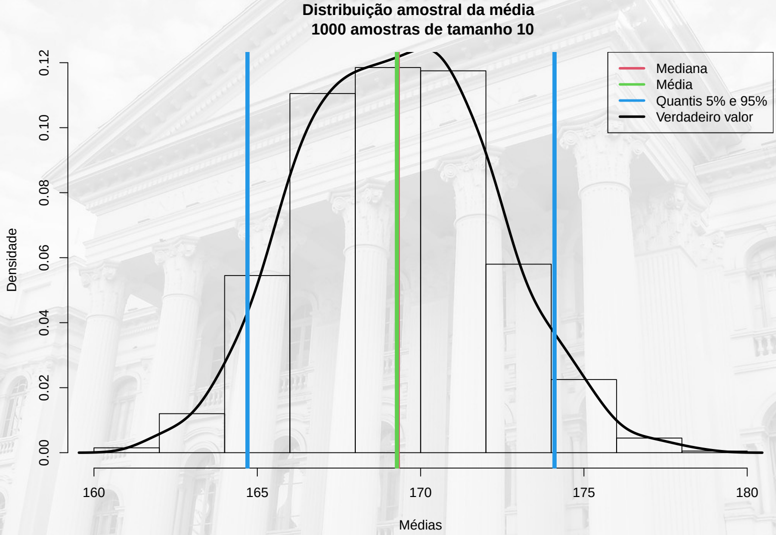


Ilustração computacional

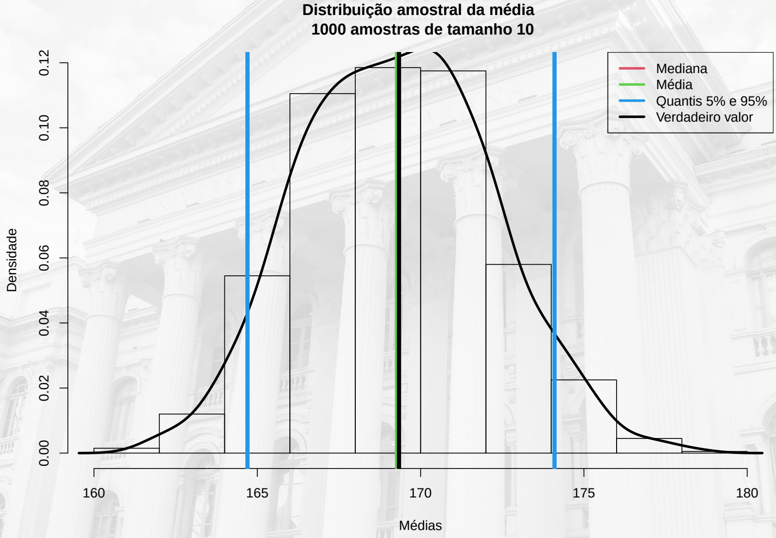


Ilustração computacional

- Os resultados mostram que não precisamos olhar a população para ter uma estimativa satisfatóriamente próxima do verdadeiro valor do parâmetro de interesse.
- Contudo esta estratégia é inviável na prática, pois necessita de várias amostras.
- Verificamos que a distribuição amostral é simétrica.
- O teorema central do limite garante que esta distribuição é Normal.

Ilustração computacional

- Na prática usamos uma distribuição normal centrada na estimativa de uma única amostra.
- Com base nesta distribuição amostral estimada, fazemos inferência.
- Os quantis desta distribuição garantem a confiança. Neste caso tomaremos os quantis 5% e 95% da distribuição estimada.
- Se replicarmos o procedimento 100 vezes, esperamos que em 10 vezes o intervalo dado pelos quantis não contenham o valor do parâmetro.

Ilustração computacional

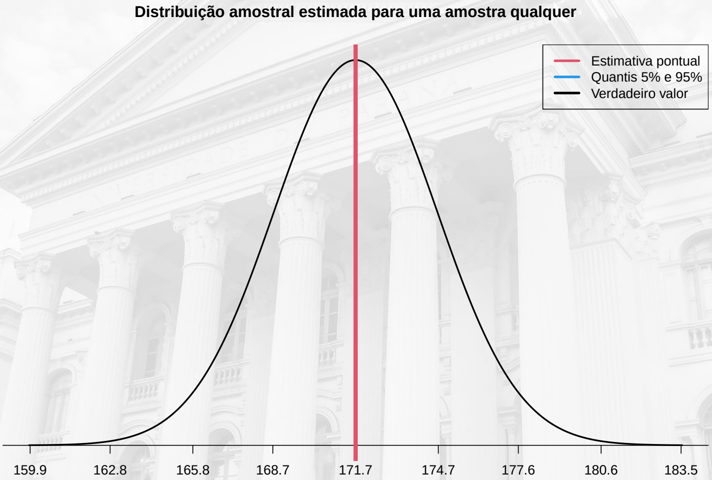


Ilustração computacional

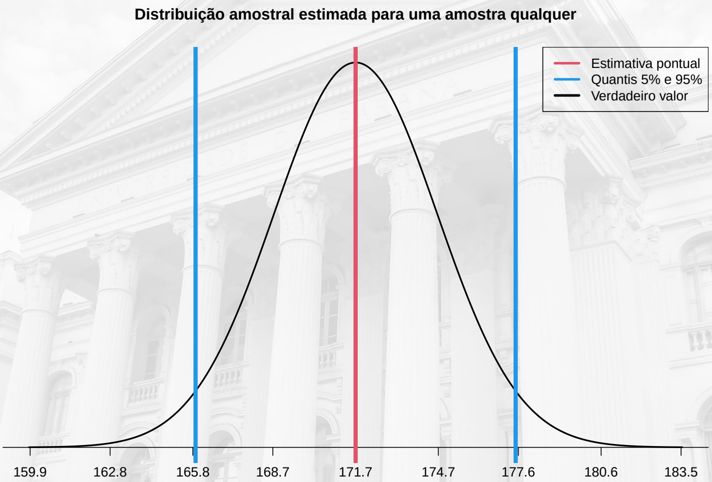
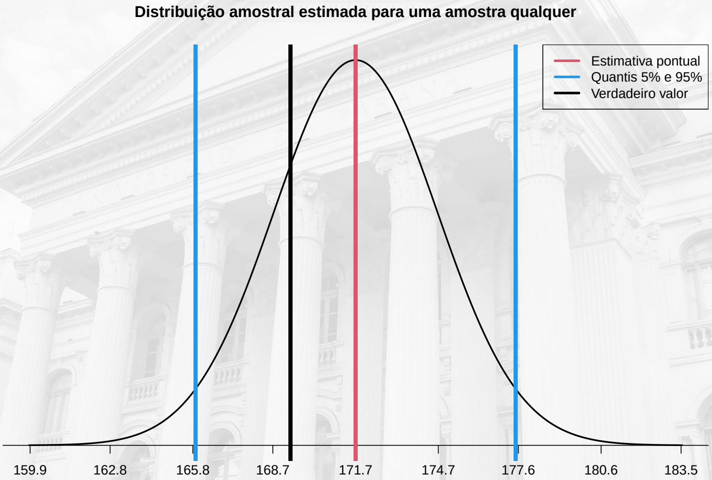


Ilustração computacional



O que foi visto:

- Conceitos importantes para inferência estatística.
- Ideia de distribuição amostral.
- Distribuição amostral da média.
- Ilustração computacional.