UC Medição em ciências e representação gráfica

Noções de estatística aplicada à medição

Parte III – Amostragem

Prof. Simões

Objetivos dessa aula

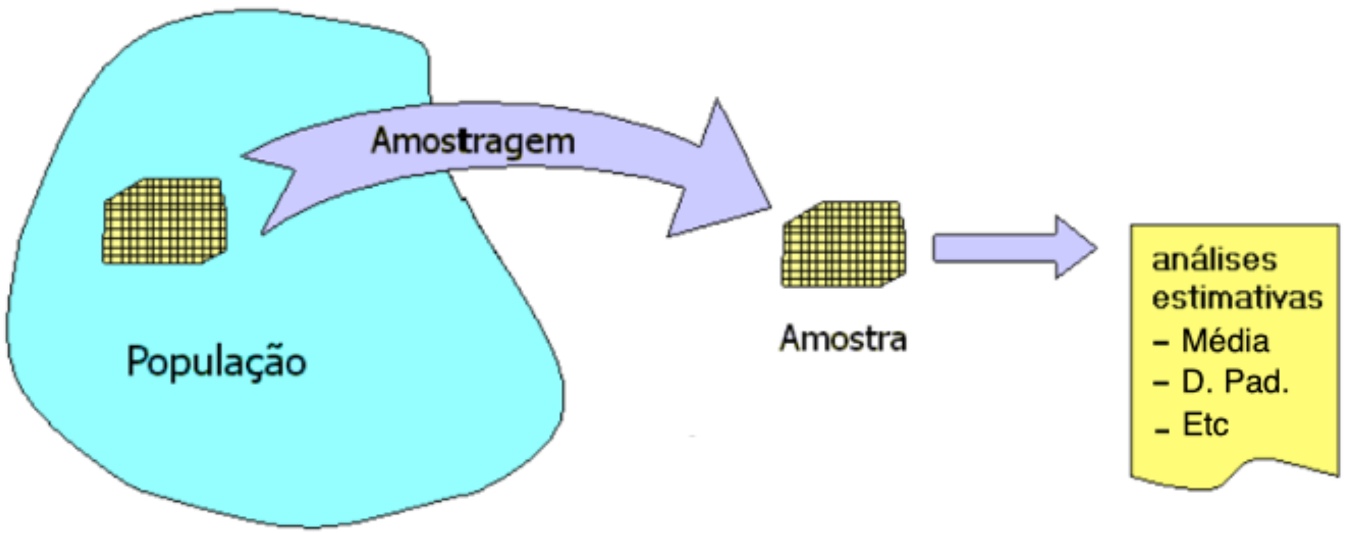
- Ao final dessa aula você deverá ser capaz de:
  - Explicar o que é amostragem e sua importância na metrologia
  - Compreender os parâmetros objetivos e subjetivos da amostragem
  - Identificar e diferenciar os casos de apliação de amostragem
  - Identificar a fórmula adequada para cada caso
  - Aplicar corretamente a fórmula adequada
  - Interpretar seu resultado

Problema típico

- Uma comunidade de 20000 casas precisa ser analisada quanto à conexão ou não à rede de esgotos. Estima-se que 30% das casas estejam conectadas. Quantas casas devem ser pesquisadas para uma confiança de 95% e um erro amostral de 2%?

Definição e importância

- Amostragem é a determinação da quantidade mínima de elementos de uma população a serem examinados, de modo que o resultado tenha significado estatístico

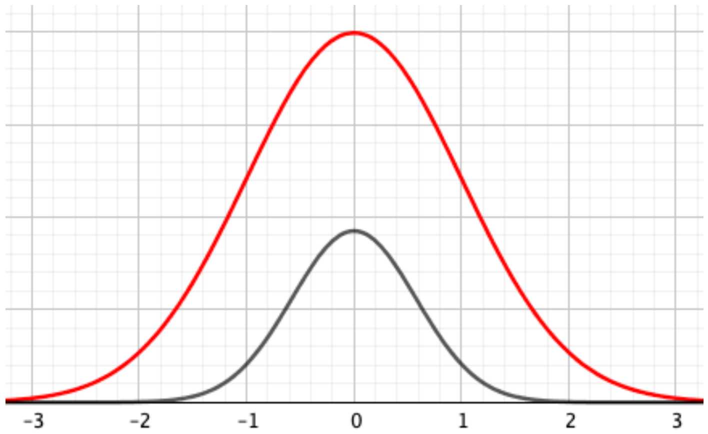


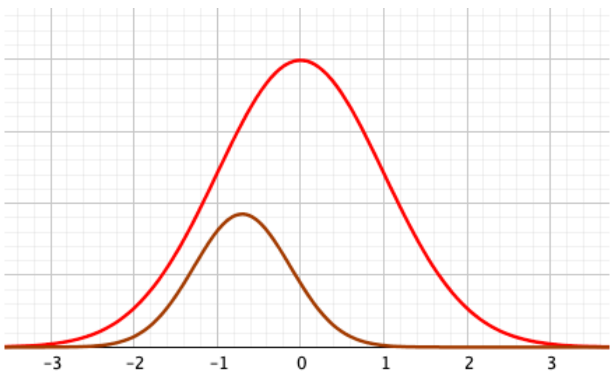
- Importância: uma amostra pequena produz resultados potencialmente incorretos, e uma amostra grande pode resultar em custos desnecessários.

Considerações

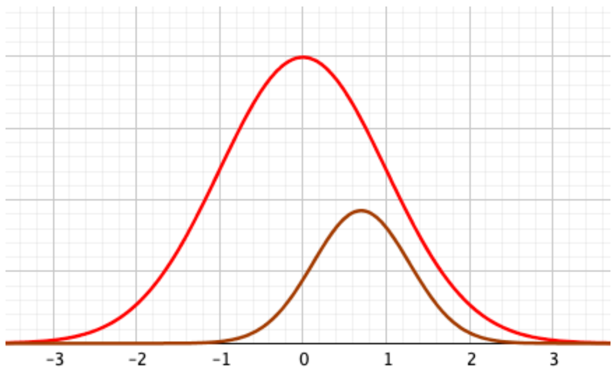
- Além dos aspectos matemáticos que consideraremos, é importante que a amostra não seja tendenciosa.
- Muitas vezes isso depende de fatores subjetivos: local da coleta, horário, etc.
- Idealmente, a amostra deve ser uma “população em menor escala”

Selecionada corretamente





Selecionadas incorretamente



Considerações

- Podemos dividir em dois grandes grupos os casos em que precisamos estimar o tamanho da amostra
  - Para determinar uma **média**: por exemplo, qual o diâmetro médio dos eixos produzidos por essa máquina?
  - Para determinar uma **proporção**: por exemplo, quantos por cento das casas nesse bairro recebe água tratada?
- Nos dois casos acima, temos que considerar que a população por ser **finita**, ou **infinita** (ou, pelo menos, estatísticamente muito grande)
- Além disso, podemos ou não conhecer previamente o **desvio padrão** populacional
- Em todos os casos, assumiremos que a distribuição é **normal**, a menos que o histórico indique que não. Como segurança, a população deve ser superior a **30 elementos**.

Casos a serem considerados

- Médias
  - Desvio padrão conhecido
    - População finita
    - População infinita

  - Desvio padrão desconhecido
    - População finita
    - População infinita

- Proporções
  - População finita
  - População infinita

Média, DP conhecido, população infinita

- A demonstração matemática das fórmulas abaixo não será vista nesse momento. Tomemos então como certo que:
- Para populações infinitas

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

Onde:
$n \Rightarrow$ quantidade mínima de elementos na amostra (arredondar sempre para mais)
$z \Rightarrow$ valor padronizado correspondente ao grau de confiança exigido
$\sigma \Rightarrow$ desvio padrão populacional
$E \Rightarrow$ erro da estimativa da média

Média, DP conhecido, população infinita

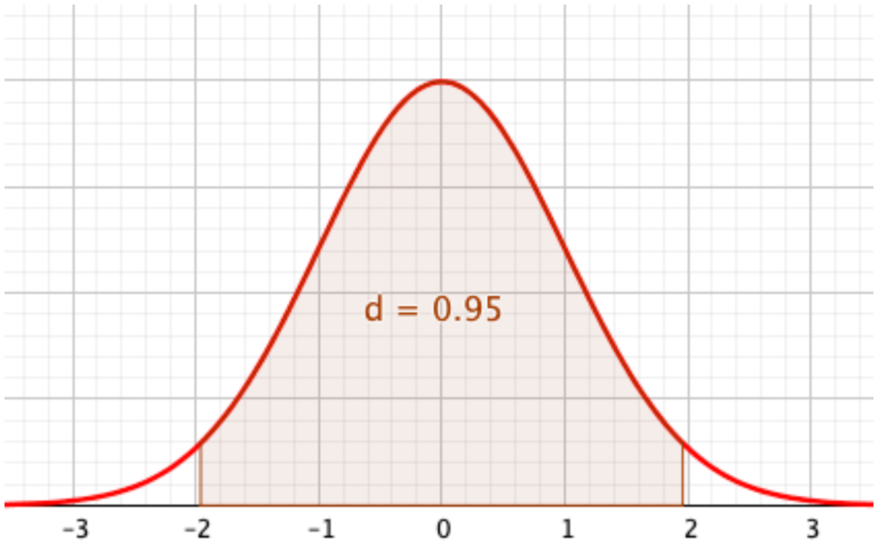
- Grau de confiança: representa quantos por cento da população espera-se que esteja dentro da margem de erro da estimativa da média tolerada
- Desvio padrão: quanto os valores da população se afastam em média da média populacional
- Erro da estimativa da média: quanto as médias dos grupos de elementos coletados (amostras) se afastam em média da média populacional real

Exemplo

- O engenheiro de produção responsável pela fabricação de um componente precisa determinar o **tempo médio** para uma operação de usinagem. O dado histórico do desvio padrão do equipamento é de 40 s. O erro admissível nessa análise é de 14 s com um grau de confiança de 95%. Calcule qual a amostra mínima (ou seja, quantas operações ele deve cronometrar para ter uma média significativa).

Exemplo

- O grau de confiança representa quantos por cento da população ele espera que esteja dentro da margem de erro tolerada.



$$95\ \% \Rightarrow \ \alpha = 0{,}950 \Rightarrow \frac{\alpha}{2} = 0{,}475$$

$$Tabela \Rightarrow z_{\frac{\alpha}{2}=0{,}4750} = 1{,}96$$

Exemplo

- Aplicando os valores na fórmula, teremos:

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

$$n = \left(\frac{1{,}96 \cdot 40}{14}\right)^2 \Rightarrow n = 32\, placas$$

- Interpretação: ao calcular a média em uma amostra de 32 operações, o engenheiro pode afirmar que, estatisticamente, 95% das operações terão a média dentro da margem de erro estabelecida.

Exercício proposto

- Repita o cálculo anterior, com uma confiança de 90% e de 99%, e analise os resultados

Respostas: 23 e 54 placas

Exercício proposto

- Repita o cálculo anterior, com uma confiança de 90% e de 99%, e analise os resultados

$$90\,\% \Rightarrow \ \alpha = 0{,}900 \Rightarrow \frac{\alpha}{2} = 0{,}4500$$

$$Tabela \Rightarrow z_{\frac{\alpha}{2}=0{,}4500} = 1{,}65$$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 \Rightarrow n = \left(\frac{1{,}65 \cdot 40}{14}\right)^2$$

$$n = 23\ placas$$

$$99\,\% \Rightarrow \ \alpha = 0{,}990 \Rightarrow \frac{\alpha}{2} = 0{,}4950$$

$$Tabela \Rightarrow z_{\frac{\alpha}{2}=0{,}4950} = 2{,}58$$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 \Rightarrow n = \left(\frac{2{,}58 \cdot 40}{14}\right)^2$$

$$n = 54\ placas$$

Conclusão: o tamanho da amostra é maior, caso queiramos uma maior confiança.

Exercício proposto

- Um fabricante de esterilizadores afirma que seu equipamento apresenta um desvio padrão de 30°C na temperatura de operação. Determinar quantas medições são necessárias para uma estimativa estatisticamente válida da média da temperatura desse equipamento, com um nível de confiança de 90% e erro máximo de 6°C.

Resposta: 69 medições

Média, DP conhecido, população finita

- Para populações finitas, a fórmula fica corrigida para

$$n = \frac{N \cdot z^2 \cdot \sigma^2}{(N-1) \cdot E^2 + z^2 \cdot \sigma^2}$$

Onde:

$N \Rightarrow \text{tamanho da população}$

Exercício proposto

- Um cliente encomendou de certa indústria um lote de 5000 unidades de certo tipo de peça. O engenheiro responsável pela qualidade do produto pretende examinar uma amostra de peças desse lote para fazer uma estimativa de seu comprimento médio antes de entregá-lo ao cliente. Sabe-se que o desvio padrão do processo de fabricação desse componente é de 3,0 mm. Com uma confiança de 95% e um erro de estimativa de 1,0 mm, determinar o tamanho mínimo que deve ter essa amostra.

Resposta: 35 peças

Exercício proposto

Média, DP desconhecido

- Caso o desvio padrão populacional seja desconhecido, pode ser adotada uma das duas regras práticas abaixo

1. Estudo piloto: realizamos uma coleta de pelo menos 31 elementos para o cálculos da média e desvio padrão, e adotamos esse valor nas mesmas fórmulas anteriores
2. Adotamos como desvio padrão populacional estimado o seguinte valor

$$\sigma = \frac{\text{Amplitude}}{4} = \frac{\text{Maior valor} - \text{Menor valor}}{4}$$

Exercício proposto

- Um termômetro registra a temperatura mínima e máxima de um equipamento de refrigeração, que foram, respectivamente, 12°C e 18°C. Quantas medições devem ser feitas para se obter uma média que represente 99% de confiança, com um erro máximo de 1°C?

Resposta: 15 medições

Exercício proposto

- Um termômetro registra a temperatura mínima e máxima de um equipamento de refrigeração, que foram, respectivamente, 12°C e 18°C. Quantas medições devem ser feitas para se obter uma média que represente 99% de confiança, com um erro máximo de 1°C, considerando que o lote de equipamentos a serem avaliados é de 100 unidades.

Resposta: 15 medições

Exercício proposto

- As frutas produzidas em um pomar têm massa entre 150 g e 195 g. É necessário determinar uma amostra para se obter uma estimativa da média populacional com um grau de 96 % de confiança e com um erro máximo de 5 gramas na estimativa da média. Quantas frutas devem ser pesadas?

Resposta: 23 frutas

Exercício proposto

- As frutas produzidas em um pomar têm massa entre 150 g e 195 g. É necessário determinar uma amostra para se obter uma estimativa da média populacional com um grau de 96 % de confiança e com um erro máximo de 5 gramas na estimativa da média. Quantas frutas devem ser pesadas?

$$96\,\% \Rightarrow \alpha = 0{,}96 \Rightarrow \frac{\alpha}{2} = \frac{0{,}96}{2} = 0{,}48 \qquad Tabela \Rightarrow z_{\frac{\alpha}{2}=0{,}4500} = 2{,}08$$

$$\sigma = \frac{\text{Amplitude}}{4} = \frac{195 - 150}{4} = 11{,}3\ g$$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 \Rightarrow n = \left(\frac{2{,}08 \cdot 11{,}3}{5}\right)^2$$

$$n = 23\ frutas$$

Amostra para determinação da proporção

- Em muitos casos, não desejamos o valor de uma variável propriamente, mas sua proporção em relação a uma população.
- Por exemplo: uma máquina fabrica resistores e é necessário estimar a proporção de resistores fora da especificação produzidos, com um erro máximo de 3,2% e uma confiança de 98%. Em um estudo anterior constatou-se que 4,8% dos resistores produzidos não atendem às especificações.

Proporções

- Nesses casos, usamos as seguintes fórmulas
- População infinita

$$n = \frac{z^2 \cdot p \cdot q}{E^2} \qquad n = \frac{z^2 \cdot 0{,}25}{E^2}$$

Usamos essa fórmula quando a proporção populacional é desconhecida

- População finita

$$n = \frac{N \cdot z^2 \cdot p \cdot q}{(N-1) \cdot E^2 + z^2 \cdot p \cdot q}$$

Onde:

$p \Rightarrow$ proporção populacional

$q \Rightarrow$ complemento de p, isto é, $q = 1 - p$

Exercício proposto

- Uma máquina fabrica resistores e é necessário estimar a proporção de resistores fora da especificação produzidos, com um erro máximo de 3,2% e uma confiança de 98%. Em um estudo anterior constatou-se que 4,8% dos resistores produzidos não atendem às especificações.

Resposta: 243 resistores

Exercícios propostos

1. Verificar a amostra necessária no cenário anterior, supondo que não há informação histórica sobre a proporção de peças rejeitadas. Resposta: 1326
2. Verificar a amostra necessária para uma proporção de rejeição de 4,8%, com uma confiança de 90%. Resposta: 122
3. Verificar a amostra necessária para uma confiança de 90%, quando não se conhece previamente a proporção de rejeição. Resposta: 665
4. Uma comunidade de 20000 casas precisa ser analisada quanto à conexão ou não à rede de esgotos. Estima-se que 30% das casas estejam conectadas. Quantas casas devem ser pesquisadas para uma confiança de 95% e um erro amostral de 2%? Resposta: 1833

Exercícios propostos

5. Um engenheiro eletricista de uma fábrica de aparelhos eletrônicos deseja estimar a proporção populacional dos aparelhos de som que apresentam algum tipo de defeito durante o período de garantia. Utilizando uma confiança de 95% e um erro de estimativa de 4%, determinar a quantidade mínima de aparelhos desse tipo que devem ser testados, se:
   a) Em um estudo anterior, constatou-se que 25% dos aparelhos apresentam defeitos durante o período de garantia
   b) Não se tem qualquer informação que possa sugerir o valor de p.

Respostas: 451; 601

Conclusão

- Ao final dessa aula você deve ser capaz de:
  - Explicar o que é amostragem e sua importância na metrologia
  - Compreender os parâmetros objetivos e subjetivos da amostragem
  - Identificar e diferenciar os casos de apliação de amostragem
  - Identificar a fórmula adequada para cada caso
  - Aplicar corretamente a fórmula adequada
  - Interpretar seu resultado