Estatística e Probabilidade

Bacharelado em Sistemas de Informação

Aula 9: Amostragem e Estimativa

Prof. Dr. Samuel Sanches

AMOSTRAGEM ALEATÓRIA

★ **População**: todas observações concebivelmente possíveis (ou hipoteticamente possíveis).

★ **Amostra**: parte da população.

★ **Tipos de população**: finita e infinita

AMOSTRAGEM ALEATÓRIA POPULAÇÃO FINITA

★ **Exemplo:** Quantas amostras de tamanho **n** diferentes podem ser extraídas de uma população finita de tamanho **N**, se:

a) n = 2 e N = 12?

$$\binom{12}{2} = \frac{12 \cdot 11}{2} = 66 \text{ amostras diferentes}$$

b) n = 3 e N = 50?

$$\binom{50}{3} = \frac{50 \cdot 49 \cdot 48}{3!} = 19.600 \text{ amostras diferentes.}$$

AMOSTRAGEM ALEATÓRIA

- ★ Uma amostra de tamanho **n** extraída de uma população finita de tamanho **N** é aleatória se é escolhida de tal modo que cada uma das combinações **(N n)** amostras possíveis tem a mesma probabilidade **1/(N   n)** de serem escolhidas.

- ★ ou seja, podemos sempre escolher aleatoriamente alguém da nossa amostra, já que todas têm a mesma probabilidade de ocorrer.

- ★ Uma amostra de tamanho **n** de uma população infinita é aleatória se consiste em valores de variáveis aleatórias independentes que têm a mesma distribuição.

DISTRIBUIÇÕES AMOSTRAIS

- ★ De amostra para amostra, podemos ter variações na média, mediana e desvio-padrão.

- ★ São importantes para a inferência!

DISTRIBUIÇÕES AMOSTRAIS

★ **Exemplo**: Vamos construir a distribuição amostral de uma amostra aleatória de tamanho n = 2 extraída, sem reposição, de uma população finita de tamanho N = 5, cujos elementos são os números 3, 5, 7, 9 e 11.

$$\mu = \frac{3+5+7+9+11}{5} = 7$$

$$\sigma = \sqrt{\frac{(3-7)^2+(5-7)^2+(7-7)^2+(9-7)^2+(11-7)^2}{5}} = \sqrt{8}$$

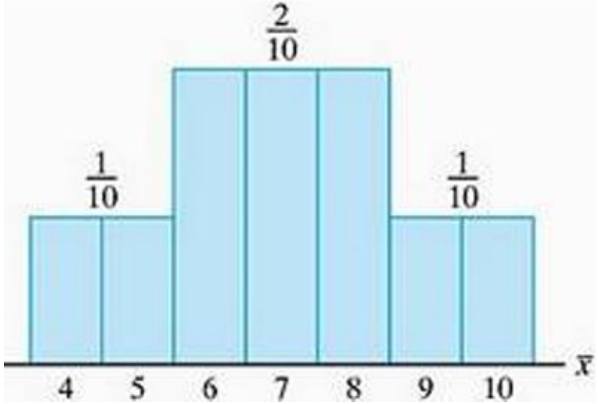
Tomando uma amostra aleatória de tamanho n = 2, temos (5   2) = 10 possibilidades: 3 e 5 (μ=4), 3 e 7 (μ=5), 3 e 9 (μ=6), 3 e 11 (μ=7), 5 e 7 (μ=6), 5 e 9 (μ=7), 5 e 11 (μ=8), 7 e 9 (μ=8), 7 e 11 (μ=9) e 9 e 11 (μ=10), sendo cada um com probabilidade 1/10.

DISTRIBUIÇÕES AMOSTRAIS

★ **Exemplo**: Vamos construir a distribuição amostral de uma amostra aleatória de tamanho n = 2 extraída, sem reposição, de uma população finita de tamanho N = 5, cujos elementos são os números 3, 5, 7, 9 e 11.

Montamos a distribuição amostral da média:

$\bar{x}$	*Probabilidade*
4	$\frac{1}{10}$
5	$\frac{1}{10}$
6	$\frac{2}{10}$
7	$\frac{2}{10}$
8	$\frac{2}{10}$
9	$\frac{1}{10}$
10	$\frac{1}{10}$



DISTRIBUIÇÕES AMOSTRAIS

★ **Exemplo**: Vamos construir a distribuição amostral de uma amostra aleatória de tamanho n = 2 extraída, sem reposição, de uma população finita de tamanho N = 5, cujos elementos são os números 3, 5, 7, 9 e 11.

Média e desvio-padrão da distribuição amostral das médias:

$$\mu_{\bar{x}} = 4 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 6 \cdot \frac{2}{10} + 7 \cdot \frac{2}{10} + 8 \cdot \frac{2}{10} + 9 \cdot \frac{1}{10} + 10 \cdot \frac{1}{10} = 7$$

$$\sigma_{\bar{x}}^2 = (4-7)^2 \cdot \frac{1}{10} + (5-7)^2 \cdot \frac{1}{10} + (6-7)^2 \cdot \frac{2}{10} + (7-7)^2 \cdot \frac{2}{10} + (8-7)^2 \cdot \frac{2}{10} + (9-7)^2 \cdot \frac{1}{10} + (10-7)^2 \cdot \frac{1}{10} = 3$$

$$\rightarrow \sigma_{\bar{x}} = \sqrt{3}$$

DISTRIBUIÇÕES AMOSTRAIS

- ★ Veja que:

  $\mu_{\bar{x}}$, a média da distribuição amostral de $\bar{x}$, é igual a μ, a média da população;

  $\sigma_{\bar{x}}$, o desvio-padrão da distribuição amostral de $\bar{x}$, é menor do que σ, o desvio-padrão populacional.

- ★ Para casos em que **n** e **N** são pequenos, conseguimos fazer sem problemas. Se n = 10 e N = 100, a combinação nos retorna mais de 17 trilhões de amostras! → Computador!!!!

- ★ Quando nossa população é uma **distribuição inteira** (**distribuição discreta uniforme**), ou seja, números inteiros, podemos fazer:

  $$\mu = \frac{N+1}{2} \qquad \sigma = \sqrt{\frac{N^2-1}{12}}$$

  para N = 1000, temos μ = 500,5 e σ = 288,67

ERRO-PADRÃO DA MÉDIA

★ Vimos que a média da distribuição amostral é igual a média da população que ela provém:

$$\mu_{\bar{x}} = \mu$$

★ E o desvio-padrão é menor, podemos escrever para uma população infinita (chama-se erro-padrão da média):

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

★ Para população finita:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

ERRO-PADRÃO DA MÉDIA

- ★ Erro-padrão da média é fundamental para sabermos:
  - Se for pequeno, temos grandes indícios de a média de uma amostra, estar próxima da média da população;
  - Se for grande, temos chance de a média de uma amostra ser consideravelmente diferente da média da população.
- ★ Cresce quando a variabilidade da população é grande e decresce quanto maior for o tamanho da amostra, ou seja, amostras pequenas têm maior chance de terem maior variabilidade.

ERRO-PADRÃO DA MÉDIA

★ **Exemplo:** Quando extraímos uma amostra aleatória de uma população infinita, o que acontece com o erro-padrão da média e, portanto, com o erro que poderíamos esperar quando utilizamos a média da amostra para estimar a média da população, se o tamanho da amostra é: **a)** aumentado de 50 para 200; **b)** diminuído de 360 para 40?

a) razão dos erros, quando **n** é quadruplicado, o erro é reduzido pela metade.

$$\frac{\frac{\sigma}{\sqrt{200}}}{\frac{\sigma}{\sqrt{50}}} = \frac{\sqrt{50}}{\sqrt{200}} = \sqrt{\frac{50}{200}} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

b) razão dos erros, quando **n** é dividido por 9, o erro é multiplicado por 3.

$$\frac{\frac{\sigma}{\sqrt{40}}}{\frac{\sigma}{\sqrt{360}}} = \frac{\sqrt{360}}{\sqrt{40}} = \sqrt{9} = 3$$

ERRO-PADRÃO DA MÉDIA

★ **Exemplo:** Encontre o valor do fator de correção (termo a mais multiplicado no erro-padrão para população finita) para N = 10.000 e n = 100.

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{10.000-100}{10.000-1}} = 0,995$$

note que o valor é tão próximo de 1 que o fator de correção pode ser omitido para fins práticos (amostra é pelo menos 5% da população, no exemplo, é 10%).

ERRO-PADRÃO DA MÉDIA

★ **Exemplo:** Nos slides 7 e 8, calculamos o erro-padrão da média para N = 5, n = 2 e σ = √8, onde obtivemos σ = √3, verifique utilizando a expressão para o erro-padrão da média para população finita, obtemos o mesmo valor.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{\sqrt{8}}{\sqrt{2}} \cdot \frac{\sqrt{5-2}}{5-1} = \frac{\sqrt{8}}{\sqrt{2}} \cdot \frac{\sqrt{3}}{4} = \sqrt{\frac{8}{2} \cdot \frac{3}{4}} = \sqrt{3}$$

ERRO-PADRÃO DA MÉDIA

★ **Exemplo:** Uma simulação com N = 1.000, n = 15 e σ = 288,67, encontre o valor para o desvio-padrão das médias amostrais.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{288{,}67}{\sqrt{15}} \cdot \sqrt{\frac{1.000-15}{1.000-1}} = 74{,}01$$

TEOREMA DO LIMITE CENTRAL

- Se a média de uma amostra aleatória de tamanho **n** de uma população infinita com a média **μ** e desvio-padrão **σ** e se **n** é grande:

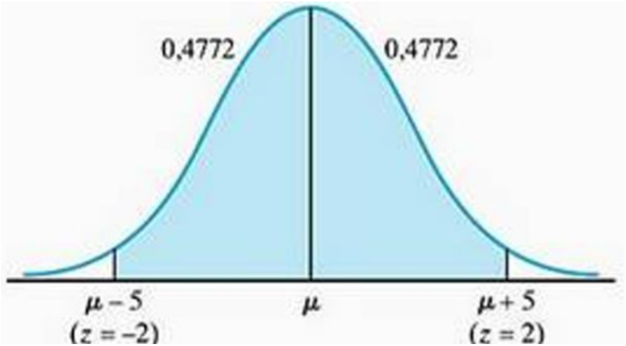
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- Podemos dizer que temos aproximadamente a distribuição normal padrão.

- Pode-se utilizar para populações finitas, desde que **n** seja grande, esse grande não podemos dizer, mas em geral n = 30 já é suficiente.

TEOREMA DO LIMITE CENTRAL

★ **Exemplo:** Usando o teorema do limite central, qual é a probabilidade de o erro ser inferior a 5 quando usamos a média de uma amostra aleatória de tamanho n = 64 para estimar a média de uma população infinita com σ = 20?



inferior a 5, quer dizer entre menos 5 e mais 5:

$$z = \frac{-5}{20/\sqrt{64}} = -2 \quad \text{e} \quad z = \frac{5}{20/\sqrt{64}} = 2$$

pela tabela para z = 2,00 temos 0,4772, sendo assim a probabilidade será: 0,4772 + 0,4772 = 0,9544

Tabela: https://professorguru.com.br/wa_files/tabelas-normal-padrao-de-0-a-z.pdf

ESTIMATIVA DE MÉDIAS

★ **Inferência** é classificado como **estimativa** para determinar (com limites razoáveis) os valores dos parâmetros populacionais.

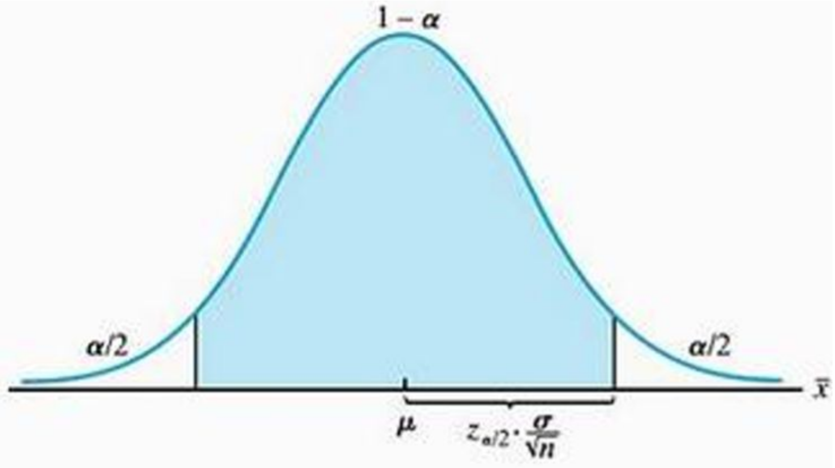
★ **Exemplo**: Dados de 36 pessoas, de tempo (minutos) de montagem de um determinado móvel.

17	13	18	19	17	21	29	22	16	28
21	15	26	23	24	20	8	17	17	21
32	18	25	22	16	10	20	22	19	14
30	22	12	24	28	11				

média da amostra é 19,9, como não temos mais informações podemos assumir que é a média da população.

Essa estimativa é denominada de **estimativa pontual**, pois consiste em um único número, mais comum, porém com **margem para dúvidas**, quanta informação se baseia a estimativa, qual o tamanho da distribuição amostral da média

ERRO MÁXIMO DE ESTIMATIVA



- ★ Sendo **zα** e a área à direita sob curva normal é α, portanto, a área sob a curva normal entre **-zα/2** e **zα/2** é **1 - α**.

- ★ Assim, quando utilizamos a média amostral como uma estimativa da média populacional, a probabilidade é **1 - α** de que essa estimativa vá diferir para um ou para o outro lado por, no máximo:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- ★ Muito usados para **1 - α** são 0,95 e 0,99 e os correspondentes de **α/2** são 0,025 (z = 1,96) e 0,005 (z = 2,575).

ERRO MÁXIMO DE ESTIMATIVA

★ **Exemplo**: Utilizando a média de uma amostra aleatória de tamanho **n = 150** para estimar a aptidão mecânica média. Se, com base na experiência pode-se supor que **σ = 6,2** para tais dados, o que pode-se afirmar, com uma probabilidade de **0,99**, sobre o erro máximo da estimativa?

então $1 - \alpha = 0{,}99 \rightarrow \alpha = 0{,}01 \rightarrow \alpha/2 = 0{,}005$:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = 2{,}575 \cdot \frac{6{,}2}{\sqrt{150}} \approx 1{,}30$$

podemos afirmar com 0,99 de probabilidade que o erro será de no máximo 1,30.

ERRO MÁXIMO DE ESTIMATIVA

- ★ Como não podemos de fato afirmar que a média amostral será no máximo diferente da média populacional pelo nosso erro, já que ele tem a ver com o método utilizado, costuma-se dizer **confiança** em vez de probabilidade.

- ★ Em geral, fazemos afirmações probabilísticas sobre valores futuros de variáveis aleatórias (como, o erro potencial de uma estimativa) e afirmações de confiança, assim que os dados tenham sido obtidos.

- ★ Muitas vezes não temos o desvio-padrão da população (é raro), pode-se usar desvio-padrão amostral (s) desde que a amostra seja grande, $n \geq 30$.

ERRO MÁXIMO DE ESTIMATIVA

★ **Exemplo**: Utilizando a média de uma amostra aleatória de tamanho **n = 36**. Se **s = 5,73** para tais dados, o que pode-se afirmar, com uma probabilidade de **95%**, sobre o erro máximo da estimativa?

então 1 - α = 0,95 → α = 0,05 → α/2 = 0,025:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = 1{,}96 \cdot \frac{5{,}73}{\sqrt{36}} \approx 1{,}87$$

podemos afirmar com 0,95 de probabilidade que o erro será de no máximo 1,87, ou não, e não sabemos qual é o caso, mas para uma aposta equilibrada de que esse é o máximo erro, temos 95 contra 5 (19 contra 1).

ERRO MÁXIMO DE ESTIMATIVA

★ Em muitos casos, o interesse é saber o tamanho da amostra necessária para a precisão desejada, assim, resolvendo para **n**:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$n = \left[\frac{z_{\alpha/2} \cdot \sigma}{E}\right]^2$$

ERRO MÁXIMO DE ESTIMATIVA

★ **Exemplo**: Utilizando a média de uma amostra aleatória e quer estimar o tempo médio que os alunos gastam para ir de uma sala de aula a outra, queremos ser capazes de afirmar com **0,95** de probabilidade de que o erro será de no máximo **0,30** minuto. Sabe-se que **σ = 1,50** minuto, qual é o tamanho da amostra necessária?

E = 0,30, σ = 1,50, e $1 - \alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow \alpha/2 = 0{,}025$, z = 1,96

$$n = \left(\frac{1{,}96 \cdot 1{,}50}{0{,}30}\right)^2 \approx 96{,}04$$

assim a amostra teria tamanho 97 para a estimativa.

ERRO MÁXIMO DE ESTIMATIVA

★ Normalmente, começamos com uma amostra pequena e com o seu desvio-padrão verificamos se precisamos de mais dados.

★ Para amostras aleatórias grandes de populações infinitas:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

★ como a probabilidade de **1 -** α, temos:

$$-z_{\alpha/2} < z < z_{\alpha/2} \longrightarrow -z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \longrightarrow \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

★ (desigualdades, trocam sinal quando multiplicadas por -1; 4 é maior que 3, mas -4 é menor que -3).

ERRO MÁXIMO DE ESTIMATIVA

- ★ **Intervalo de confiança** para média populacional:

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- ★ As <u>extremidades</u> são os **limites de confiança**.

- ★ <u>1 - α</u> é o **grau de confiança**, que são muito comuns de 95% (z = 1,96) e 99% (z = 2,575)

- ★ Estimativa dada sob a forma de intervalo de confiança são chamadas **estimativas intervalares**

ERRO MÁXIMO DE ESTIMATIVA

★ **Exemplo**: Sendo, **n = 150** e **σ = 6,2** e a média amostral **69,5**, calcule um intervalo de confiança de **95%** para estes dados.

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \longrightarrow 69{,}5 - 1{,}96 \cdot \frac{6{,}2}{\sqrt{150}} < \mu < 69{,}5 + 1{,}96 \cdot \frac{6{,}2}{\sqrt{150}}$$

$$68{,}5 < \mu < 70{,}5$$

★ Podemos dizer com 95% de confiança que a média está entre estes valores. Para 99%: 68,2 < μ < 70,8.

★ Quando aumentamos o grau de confiança, o intervalo de confiança se torna mais amplo, dando-nos, assim, menos informação sobre a grandeza que estamos procurando estimar.

ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

- **Estatística t**: desconhecendo desvio da população, utilizamos da amostra

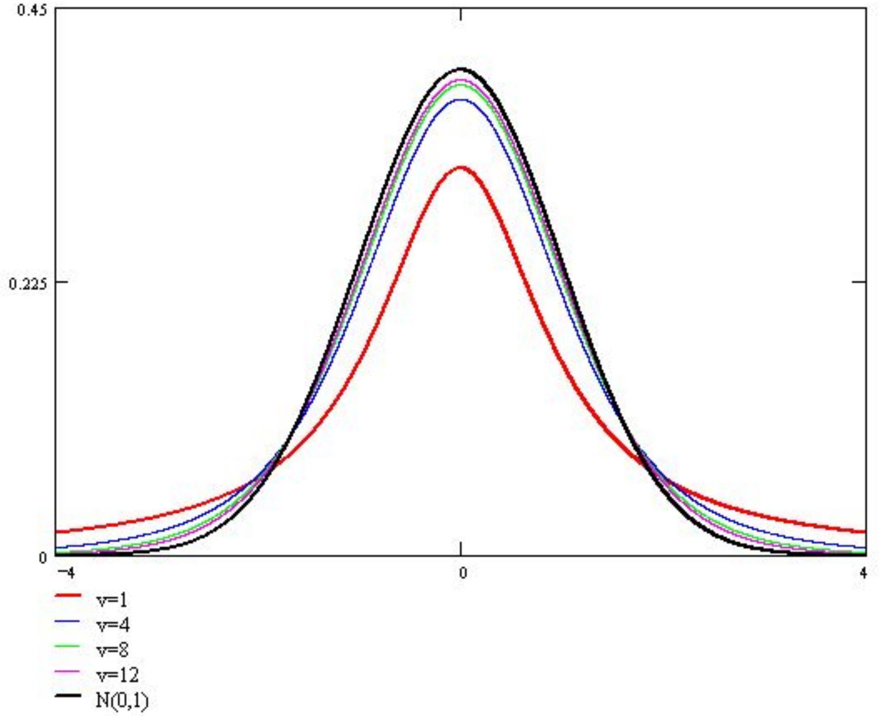
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- **Distribuição t**, ou **distribuição t de Student** ou **distribuição de Student**, originária de W. S. Gosset, que assinava os trabalhos com o nome Student.

- Semelhante à distribuição normal, a t depende do número de **graus de liberdade**, **n - 1**, tamanho da amostra subtraído 1.

- Nova tabela: https://sites.icmc.usp.br/francisco/SME0123/listas/Tabela_Dist_t.pdf

ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)



ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

★ Da mesma maneira, temos o **intervalo de confiança** quando σ é desconhecido:

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

★ Vale lembrar que, para esse intervalo, assume-se que a amostra tenha sido extraída de uma população normal, ou que pelo menos tenha o formato de uma.

ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

★ **Exemplo**: A média de batimentos cardíacos de **12** pessoas aumentou em **27,33** batimentos por minuto, com desvio-padrão de **4,28**. Construa um intervalo de **99%** de confiança para o verdadeiro aumento médio dessa taxa.

n = 12, média amostral = 27,33, s = 4,28 e 1 - α = 0,99 $\rightarrow$ $\alpha/2$ = 0,005, pela tabela t0,005 = 3,106 (12 - 1 = 11 graus de liberdade)

$$27{,}33 - 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}} < \mu < 27{,}33 + 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}} \longrightarrow 23{,}49 < \mu < 31{,}17$$

https://sites.icmc.usp.br/francisco/SME0123/listas/Tabela_Dist_t.pdf

ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

★ **Exemplo**: Em 8 locais a pintura na rua estragou depois de passarem 14,26; 16,78; 13,65; 11,53; 12,64; 13,37; 15,60 e 14,96 milhões de veículos, um gráfico mostra que se assemelha com a normal, sendo assim temos a média amostral 14,10 e desvio s = 1,67. Calcule com 95% de confiança para a média da população considerada.

n = 8, média amostral = 14,10, s = 1,67 e 1 - α = 0,95 → α/2 = 0,025, pela tabela t0,025 = 2,365 (8 - 1 = 7 graus de liberdade)

$$14{,}10 - 2{,}365 \cdot \frac{1{,}67}{\sqrt{8}} < \mu < 14{,}10 + 2{,}365 \cdot \frac{1{,}67}{\sqrt{8}} \longrightarrow 12{,}70 < \mu < 15{,}50$$

https://sites.icmc.usp.br/francisco/SME0123/listas/Tabela_Dist_t.pdf

ESTIMATIVA DE MÉDIAS (σ DESCONHECIDO)

★ **Exemplo**: Supondo que a média amostral seja 27,33, está sendo usada como uma estimativa do verdadeiro valor, e **s = 4,28** e **n = 12**. O que pode ser dito sobre o erro máximo desse valor com 99% de confiança?

1 - α = 0,99 $\rightarrow$ α/2 = 0,005, pela tabela t0,005 = 3,106 (12 - 1 = 11 graus de liberdade)

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3{,}106 \cdot \frac{4{,}28}{\sqrt{12}} \approx 3{,}84$$

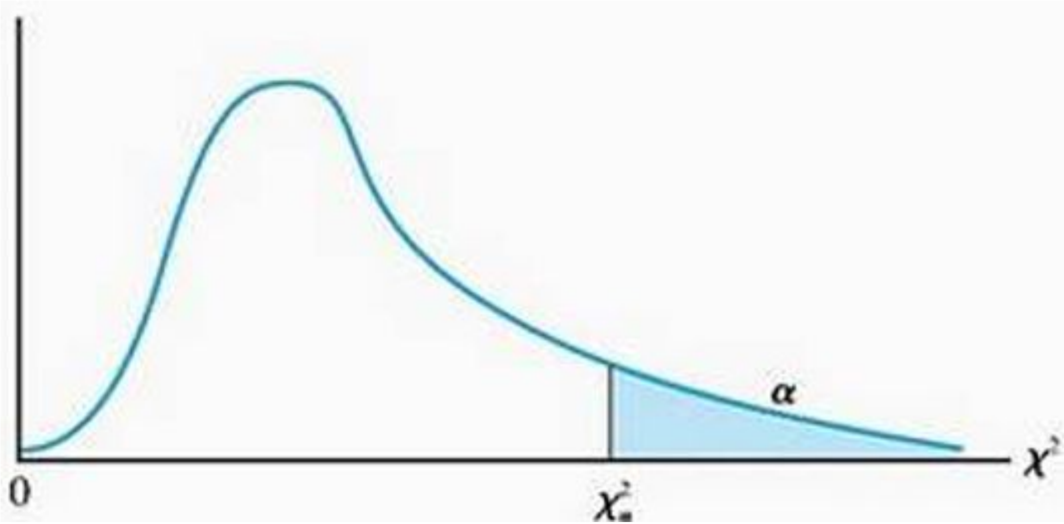
https://sites.icmc.usp.br/francisco/SME0123/listas/Tabela_Dist_t.pdf

ESTIMATIVA DE DESVIOS-PADRÃO

- ★ Até então vimos métodos para a média da população.
- ★ Intervalo de confiança para σ baseado em s, lembrando que a amostra tenha aproximadamente a forma da distribuição normal.
- ★ **Estatística qui-quadrado**: $$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$
- ★ Que é a variável aleatória contínua para a distribuição qui-quadrado, n - 1 continua sendo os graus de liberdade.

ESTIMATIVA DE DESVIOS-PADRÃO

- ★ Domínio (valores) para números reais e não-negativos.



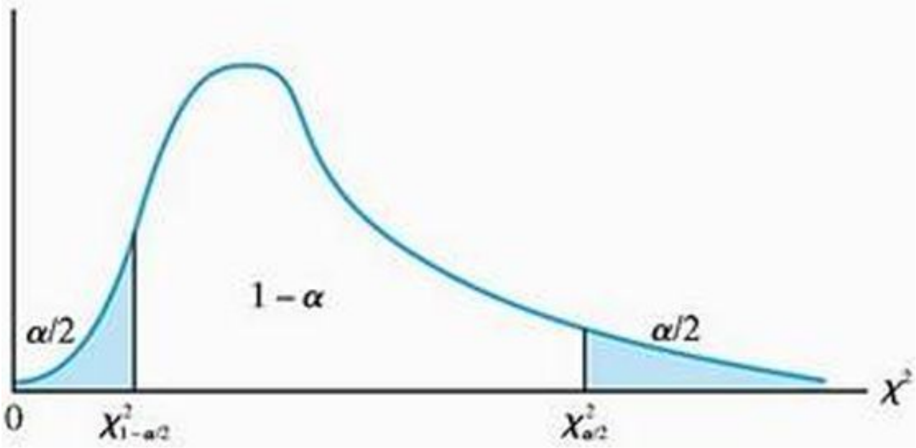
ESTIMATIVA DE DESVIOS-PADRÃO

- Analogamente aos outros casos, temos o intervalo de confiança para σ^2:

$$\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}$$

$$\chi^2_{1-\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\alpha/2}$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$



- Tabela: https://www.ime.unicamp.br/~cnaber/Tabela%20da%20Qui-quadrado.pdf

ESTIMATIVA DE DESVIOS-PADRÃO

★ **Exemplo**: Para n = 12 e s = 4,28, construa um intervalo de 99% de confiança para σ.

Não sabemos se os dados vem de uma distr. normal, resultado fica a mercê dessa informação, com **n = 12**, **s = 4,28**, 1 - α = 0,99 → $\boldsymbol{\alpha}$**/2 = 0,005** e **1 -** $\boldsymbol{\alpha}$**/2 = 0,995**, com 12 - 1 = 11 graus de liberdade, pela tabela $\chi^2$0,995 = 2,603 e $\chi^2$0,005 = 26,757:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \longrightarrow \frac{11(4{,}28)^2}{26{,}757} < \sigma^2 < \frac{11(4{,}28)^2}{2{,}603}$$

$$7{,}53 < \sigma^2 < 77{,}41$$

Então: 2,74 < σ < 8,80 com 99% de confiança!

★ Tabela: https://www.ime.unicamp.br/~cnaber/Tabela%20da%20Qui-quadrado.pdf

ESTIMATIVA DE DESVIOS-PADRÃO

★ **Exemplo**: Um estudo sobre lubrificantes industriais, se quer estudar quantos ciclos (abrir e fechar) uma porta, começará a chiar, com n = 15, temos os dados:

4295	4390	4338	4426	4698
4405	4694	4468	4863	4230
4664	4494	4535	4479	4600

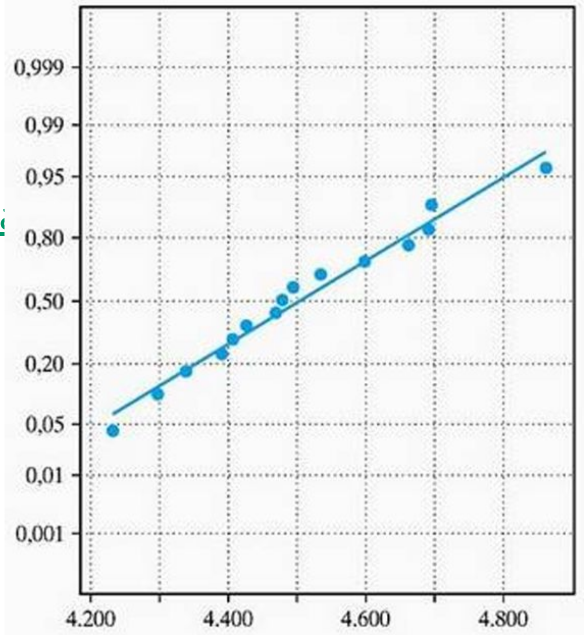
a) Verifique se é razoável tratar esses dados como uma amostra de uma população normal.
b) Se for razoável, construa um intervalo de 95% de confiança para σ.

ESTIMATIVA DE DESVIOS-PADRÃO

4295	4390	4338	4426	4698
4405	4694	4468	4863	4230
4664	4494	4535	4479	4600

★ **Exemplo**: **a)** Verifique se é razoável tratar esses dados como uma amostra de uma população normal.

Com os dados, fazemos um gráfico de probabilidades
Tendo uma reta, podemos afirmar que os dados seguem uma distribuição normal
(https://mathcracker.com/pt/fabricante-grafico-probabilida



ESTIMATIVA DE DESVIOS-PADRÃO

★ **Exemplo**: **b)** Se for razoável, construa um intervalo de 95% de confiança para σ.

Pelos dados calculamos s = 172,3, n = 15, 1 - α = 0,95 $\rightarrow$ $\alpha/2$ = 0,025 e 1 - $\alpha/2$ = 0,975, com 15 - 1 = 14 graus de liberdade, pela tabela $\chi^2 0{,}975$ = 5,629 e $\chi^2 0{,}025$ = 26,119:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \longrightarrow \frac{14(172{,}3)^2}{26{,}119} < \sigma^2 < \frac{14(172{,}3)^2}{5{,}629}$$

$$15.913 < \sigma^2 < 73.836$$

Então: 126,1 < σ < 271,7 com 95% de confiança!

★ Tabela: https://www.ime.unicamp.br/~cnaber/Tabela%20da%20Qui-quadrado.pdf

ESTIMATIVA DE DESVIOS-PADRÃO

★ Quando a amostra é grande, n ≥ 30, podemos utilizar para o **intervalo de confiança de grandes amostras para σ**:

$$\frac{s}{1+\frac{z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{s}{1-\frac{z_{\alpha/2}}{\sqrt{2n}}}$$

★ Para comparação (antes):

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

ESTIMATIVA DE DESVIOS-PADRÃO

★ **Exemplo**: Usando aquele exemplo que montamos as classes (do tempo de espera do geiser), tínhamos s = 14,35, n = 110, construa um intervalo de 95% de confiança para σ da população amostrada.

n é grande, então, $1 - \alpha = 0{,}95 \rightarrow \alpha/2 = 0{,}025$, $z_{0,025} = 1{,}96$ (slide 19 dessa aula):

$$\frac{s}{1 + \frac{z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{z_{\alpha/2}}{\sqrt{2n}}} \longrightarrow \frac{14{,}35}{1 + \frac{1{,}96}{\sqrt{220}}} < \sigma < \frac{14{,}35}{1 - \frac{1{,}96}{\sqrt{220}}}$$

Então: 12,68 < σ < 16,53 com 95% de confiança (aposta 95 para 5 ou 19 para 1)!

ESTIMATIVA DE PROPORÇÕES

★ Até agora, os dados eram medidos (tempo do geiser), agora queremos **dados de contagem**, que contamos e não medimos, como: quantos pneus duram mais do que 60.000 km, etc.

★ Vamos olhar para o parâmetro binomial p, a probabilidade de sucesso numa tentativa isolada.

★ **Proporção amostral**: x/n, x vezes que o evento ocorreu em n tentativas (total), como: 54 dentre 120, 54/120 = 0,45 → 45%.

★ Validade: $n*p > 5$ e $n*(1-p) > 5$
$n = 50$: $0{,}10 < p < 0{,}90$; $n = 100$: $0{,}05 < p < 0{,}95$; $n = 200$: $0{,}025 < p < 0{,}975$

ESTIMATIVA DE PROPORÇÕES

★ **Intervalo de confiança de grandes amostras para p**:

$$\hat{p} = \frac{x}{n} \qquad \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

★ **Erro máximo de estimativa**:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

★ **Tamanho da amostra**:

$$n = p(1-p)\left[\frac{z_{\alpha/2}}{E}\right]^2$$

sem conhecimento de p, utilizamos p(1-p) = 1/4

ESTIMATIVA DE PROPORÇÕES

★ **Exemplo**: Numa amostra aleatória, 136 dentre 400 pessoas que tomaram uma vacina contra gripe sentiram algum efeito colateral. Construa um intervalo de 95% de confiança para a verdadeira proporção das pessoas que experimentam efeito colateral com a referida vacina.

n = 400, p^=136/400 = 0,34, 1 - α = 0,95 $\rightarrow$ α/2 = 0,025, z0,025 = 1,96 (slide 19 dessa aula):

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0{,}34 - 1{,}96\sqrt{\frac{(0{,}34)(0{,}66)}{400}} < p < 0{,}34 + 1{,}96\sqrt{\frac{(0{,}34)(0{,}66)}{400}}$$

$$0{,}294 < p < 0{,}386$$

ESTIMATIVA DE PROPORÇÕES

★ **Exemplo**: Numa amostra aleatória de 250 pessoas entrevistadas ao saírem de uma votação, 145 afirmaram ter votado para a reeleição. Ao nível de 99% de confiança, o que podemos dizer sobre o erro máximo, se tomarmos p^=145/250 = 0,58 como uma estimativa da verdadeira proporção de votos para reeleição?

n = 250, p^= 0,34, 1 - α = 0,99 $\rightarrow$ α/2 = 0,005, z0,005 = 2,575 (slide 19 dessa aula):

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow E = 2{,}575\sqrt{\frac{(0{,}58)(0{,}42)}{250}} \approx 0{,}080$$

p = 0,34 temos um erro de + ou - 0,080 com 99% de confiança.

ESTIMATIVA DE PROPORÇÕES

★ **Exemplo**: Suponha que se quer estimar qual é a proporção de caminhões que trafegam entre duas cidades com carga excessiva, afirmar com 0,95 de probabilidade que o erro não vá exceder 0,04. Qual deve ser o tamanho da amostra para **a)** sabe que a verdadeira proporção está no intervalo de 0,10 a 0,25; **b)** não tem a menor ideia sobre qual poderia ser a verdadeira proporção?

a) E = 0,04 e p = 0,25, 1 - α = 0,95 $\rightarrow$ α/2 = 0,05, z0,025 = 1,96 (slide 19 dessa aula):

$$n = p(1-p)\left[\frac{z_{\alpha/2}}{E}\right]^2 \longrightarrow n = (0{,}25)(0{,}75)\left(\frac{1{,}96}{0{,}04}\right)^2 \approx 450{,}19$$

ser de 451.

ESTIMATIVA DE PROPORÇÕES

★ **Exemplo**: Suponha que se quer estimar qual é a proporção de caminhões que trafegam entre duas cidades com carga excessiva, afirmar com 0,95 de probabilidade que o erro não vá exceder 0,04. Qual deve ser o tamanho da amostra para **a)** sabe que a verdadeira proporção está no intervalo de 0,10 a 0,25; **b)** não tem a menor ideia sobre qual poderia ser a verdadeira proporção?

b) E = 0,04 e p(1-p) = 1/4, 1 - α = 0,95 $\rightarrow$ $\alpha/2$ = 0,05, z0,025 = 1,96 (slide 19 dessa aula):

$$n = p(1-p)\left[\frac{z_{\alpha/2}}{E}\right]^2 \longrightarrow n = \frac{1}{4}\left(\frac{1{,}96}{0{,}04}\right)^2 = 600{,}25$$

ser de 601 (compare com resultado de **a**).

EXERCÍCIOS

- **Lista 3 de Exercícios → Parte 3**

  https://drive.google.com/file/d/1n2YRbGJp-om4VG5FwZhNdB0ZB9M_uz-z/view?usp=share_link

★ Muito obrigado pela atenção!