Distribuição amostral da média e da proporção

CET173 – Probabilidade e Estatística
Engenharia Elétrica
Alunos:
- Sarah modesto
- Kevin Moura

2021.2

DISTRIBUIÇÃO AMOSTRAL

RELEMBRANDO

POPULAÇÃO

1. UM CONJUNTO DE ITENS OU EVENTOS SEMELHANTES QUE INTERESSA PARA ALGUMA QUESTÃO OU EXPERIMENTO
2. PODE SER REPRESENTADA POR UM MODELO COM PARÂMETROS

SE A PESQUISA É FEITA ATRAVÉS DE

AMOSTRAGEM PROBABILISTICA

GARANTINDO A

VARIBILIDADE INTERNA DA POPULAÇÃO ATRAVÉS DA QUANTIDADE SUFICIENTE DE DADOS

E A

REPRESENTATIVADE
EX: AO SELECIONAR PROPORCIONALMENTE PESSOAS POR ESTADO BRASILEIRO

É POSSIVEL REALIZAR INFERÊNCIA ESTATISTICA

AMOSTRA PROBABILÍSTICA

PODEMOS CALCULAR ESTATÍSTICAS TAIS QUAIS MEDIA AMOSTRAL, DESVIO AMOSTRAL E PROPORÇÃO AMOSTRAL

ESTATÍSTICAS (X, S, P)

VARIAVEIS ALEATORIAS -> MODELOS PROBABILISTICOS

O MODELO PROBABILÍSTICO DE UMA ESTATISTICA É CHAMADO DE DISTRIBUIÇÃO AMOSTRAL

PARÂMETRO θ

(DE UMA POPULAÇÃO)

AMOSTRAGEM PROBÁBILISTICA

T1
T2
.
.
.
TN

AMOSTRAS ESTATÍSTICAS

SEGUEM UMA DISTRIBUIÇÃO AMOSTRAL

MÉDIA

PROPORÇÃO

AO RETIRARMOS VÁRIAS, AMOSTRAS ALEATÓRIAS DE n ELEMENTOS DA POPULAÇÃO E CALCULARMOS AS PROPORÇÕES DE UM ATRIBUTO EM TODAS AS AMOSTRAS

$E(P) = \pi$

O VALOR ESPERADO DA PROPORÇÃO AMOSTRAL SERA IGUAL A PROPORÇÃO POPULACIONAL

$V(P) = \pi \cdot (1 - \pi) / n$

A VARIÂNCIA SERA IGUAL PROPORÇÃO POPULACIONAL VEZES O SEU COMPLEMENTAR DIVIDIDO PELO TAMANHO DA AMOSTRA

Exemplo: De acordo com pesquisa sobre a comunidade americana realizada pelo instituto de Censo dos Estados Unidos, 87% dos americanos com mais de 25 anos conquistaram um diploma do Ensino Médio. Suponha que vamos pegar uma amostra aleatória de 200 americanos nessa faixa etária e calcular a proporção da amostra com diploma do Ensino Médio.

Qual é a probabilidade da proporção de pessoas com diploma do Ensino Médio na amostra ser menor que 85%?

Parte 1: estabelecer a normalidade

Qual é o número esperado de pessoas na amostra com um diploma do Ensino Médio? $np = (200)(0{,}87) = 174$

Qual é o número esperado de pessoas na amostra sem diploma do Ensino Médio? $n(1-p) = (200)(1-0{,}87) = (200)(0{,}13) = 26$

A distribuição de amostragem de $\hat{p}$ é aproximadamente normal?

Número esperado de pessoas com diploma do Ensino Médio: $174 \geq 10$

Número esperado de pessoas sem diploma do Ensino Médio: $26 \geq 10$

Como as contagens esperadas de pessoas com e sem diploma do Ensino Médio na amostra são ambas iguais ou maiores que 10, a distribuição de amostragem de $\hat{p}$ é aproximadamente normal.

Parte 2: calcular a média e o desvio-padrão da distribuição de amostragem

Qual é a média da distribuição de amostragem de $\hat{p}$?

Em média, a proporção amostral será igual à proporção populacional.

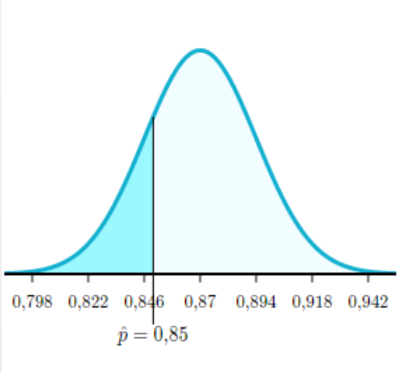
$$\mu_{\hat{p}} = p = 0{,}87$$

Qual é o desvio-padrão da distribuição de amostragem de $\hat{p}$?

$$\begin{aligned}
\sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}} \\
&= \sqrt{\frac{0{,}87(1-0{,}87)}{200}} \\
&= \sqrt{\frac{0{,}1131}{200}} \\
&= \sqrt{0{,}0005655} \\
&\approx 0{,}02378
\end{aligned}$$

A função "normalcdf" representa a função de densidade acumulada normal e encontra a área sob uma curva normal entre dois pontos dados.

normalcdf(0, 0.85, 0.87, 0.024) == 0.20
pnorm(0.85, 0.87, 0.024) == 0.20



Exemplo: Consideremos um processo de amostragem com n = 2 em uma urna que contém três tipos de fichas (2, 4 e 6) na mesma quantidade:

Combinações possíveis:

	2	4	6
2	4	6	8
4	6	8	10
6	8	10	12

$\sum y$

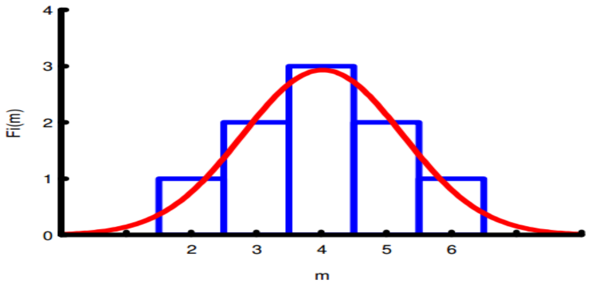
Médias possíveis:

	2	4	6
2	2	3	4
4	3	4	5
6	4	5	6

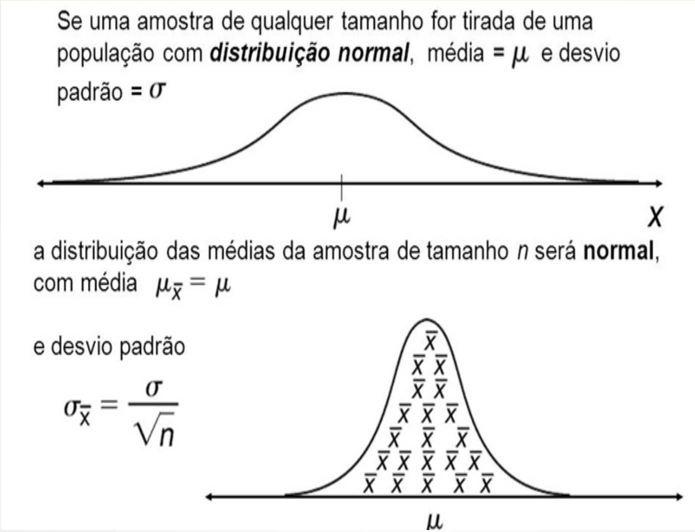
Frequência da média:

m	$F_i(m)$
2	1
3	2
4	3
5	2
6	1

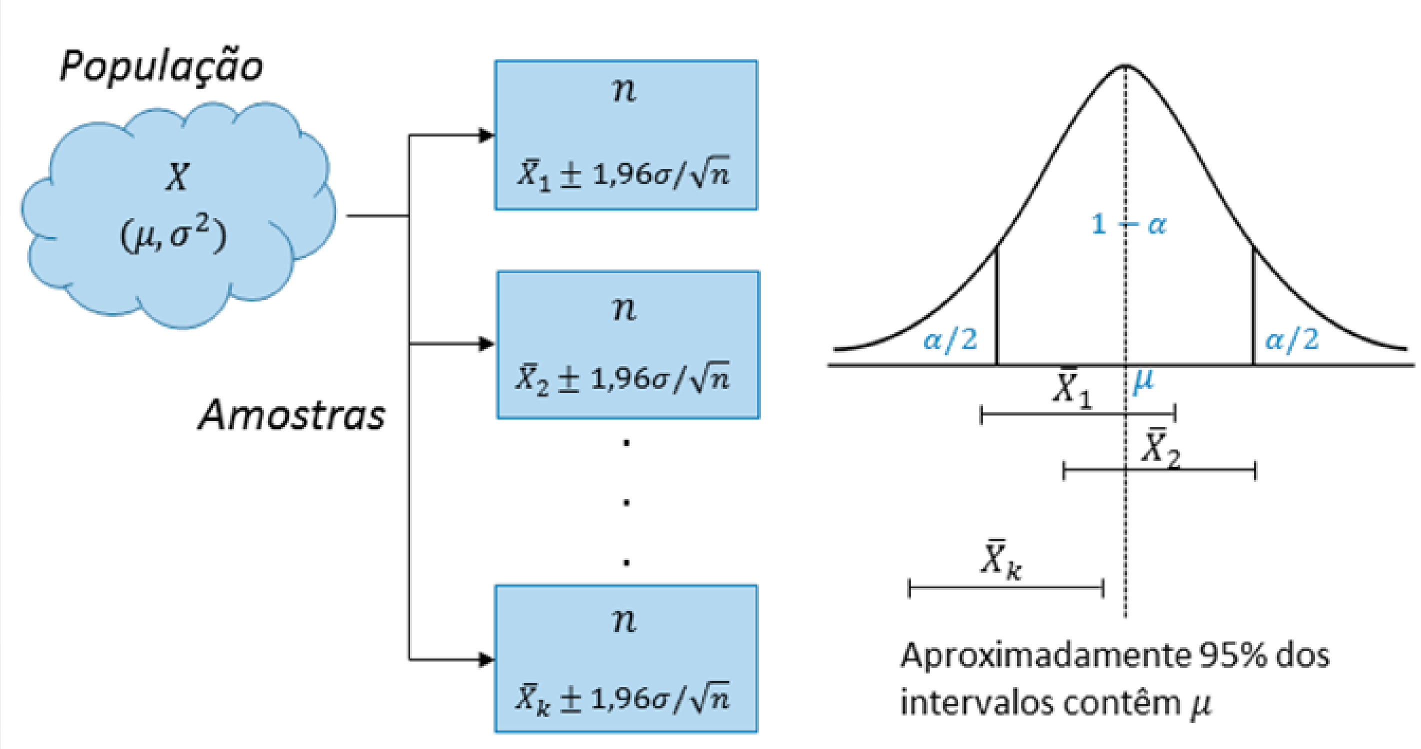
Histograma:

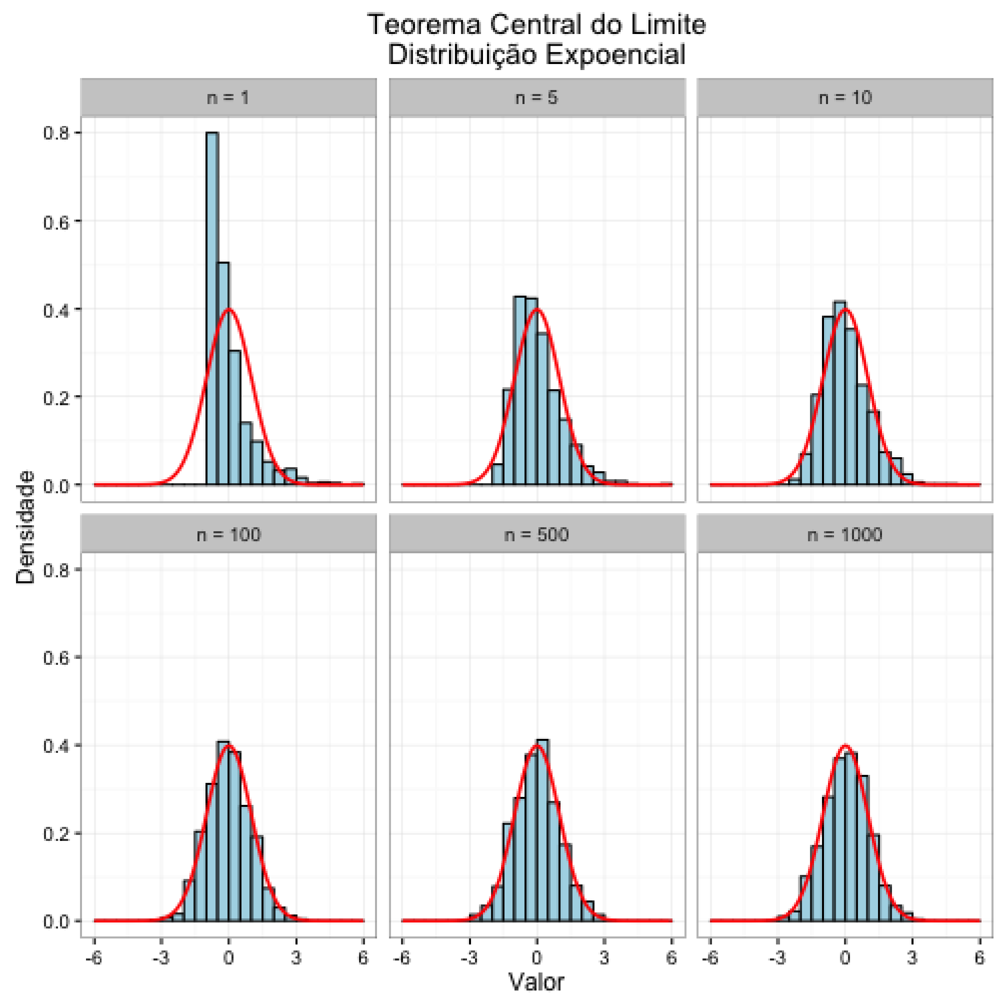


DAM - 2021 CET173



Definição: Considere uma amostra aleatória simples de tamanho n extraída de uma população qualquer com média μ e desvio padrão. Quando n é grande, a distribuição amostral da média $\bar{X}$ se aproxima da distribuição normal com média μ e desvio padrão $\sigma/\sqrt{n}$.


População
X
(μ, σ^2)
Amostras
n
$\bar{X}_1 \pm 1{,}96\sigma/\sqrt{n}$
n
$\bar{X}_2 \pm 1{,}96\sigma/\sqrt{n}$
n
$\bar{X}_k \pm 1{,}96\sigma/\sqrt{n}$
$1-\alpha$
$\alpha/2$
$\alpha/2$
μ
$\bar{X}_1$
$\bar{X}_2$
$\bar{X}_k$
Aproximadamente 95% dos intervalos contêm μ



DAM - 2021 CET173

$\bar{x}$

$$\bar{X} \sim N\left(\mu\,,\ \frac{\sigma^2}{n}\right)$$

O valor esperado das médias amostrais é igual a média da população

$$E(\bar{x}) = \boldsymbol{\mu}$$

Variância das médias amostrais:

$$\mathbf{V}(\bar{x}) = \frac{\boldsymbol{\sigma}^{\mathbf{2}}}{\boldsymbol{n}}$$

Desvio padrão das medias amostrais (ou erro padrão):

$$\boldsymbol{\sigma}(\bar{x}) = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$$

Exemplo: médias em controle de qualidade

Uma montadora de veículos realiza testes de controle de qualidade na espessura da pintura em diferentes pontos das partes de seus carros, uma vez que ocorrem algumas variabilidades no processo de pintura. Uma determinada parte do carro deve ter uma espessura ideal de $2\ \mathrm{mm}$. A distribuição de espessuras nessa parte é assimétrica à direita com uma média de $2\ \mathrm{mm}$ e um desvio-padrão de $0{,}5\ \mathrm{mm}$.

n = 100

Assumindo que a média declarada e o desvio-padrão das espessuras estejam corretos, qual é a probabilidade de a média amostral das espessuras de 100 pontos variar no máximo $0{,}1\ \mathrm{mm}$ do valor ideal?

Parte 1: estabelecer a normalidade

Qual é a forma da distribuição de amostragem da média amostral da espessura?

Escolha 1 resposta:

- ◯ Assimétrica à direita
- ◯ Aproximadamente normal
- ◯ Não contém informações suficientes

Como $n = 100 \geq 30$, o teorema central do limite se aplica. Ainda que a população de espessuras seja assimétrica à direita, a média amostral estará normalmente distribuída, uma vez que o tamanho da amostra é grande.

Parte 2: calcular a média e o desvio-padrão da distribuição de amostragem

Qual é a média da distribuição de amostragem de $\bar{x}$?

Em geral, as médias amostrais serão iguais à média populacional.

$$\mu_{\bar{x}} = \mu = 2 \text{ mm}$$

Qual é o desvio-padrão da distribuição de amostragem de $\bar{x}$?

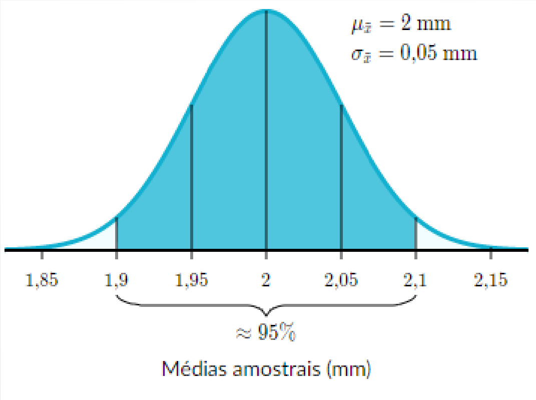
É razoável supor que há mais de 1.000 pontos na parte do carro, então podemos usar tranquilamente esta fórmula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}5}{\sqrt{100}} = \frac{0{,}5}{10} = 0{,}05 \text{ mm}$$

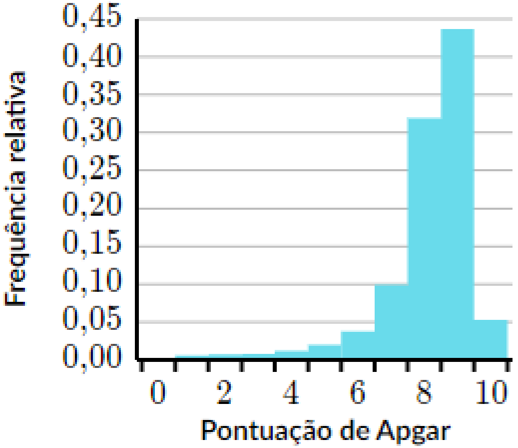
Parte 3: usar cálculos normais para encontrar a probabilidade em questão

Em qualquer distribuição normal, sabemos que aproximadamente 68% dos dados estão dentro de um desvio-padrão da média, 95% dos dados estão dentro de dois desvios-padrão da média e $99{,}7\%$ dos dados estão dentro de três desvios-padrão da média.

Já estabelecemos que a média amostral das espessuras $\bar{x}$ é normalmente distribuída com $\mu_{\bar{x}} = 2\ \mathrm{mm}$ e $\sigma_{\bar{x}} = 0{,}05\ \mathrm{mm}$. "A variação de no máximo $0{,}1\ \mathrm{mm}$ do valor ideal" é exatamente dois desvios-padrão acima e abaixo da média.

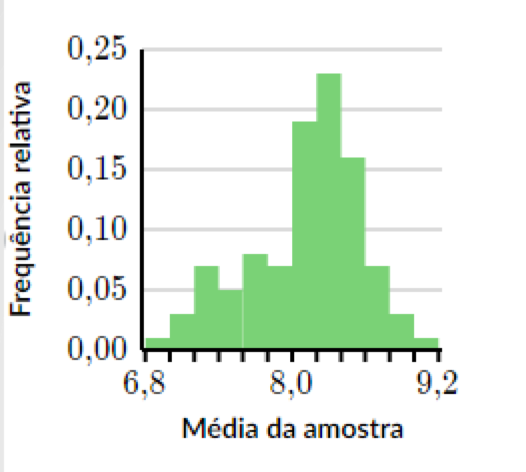


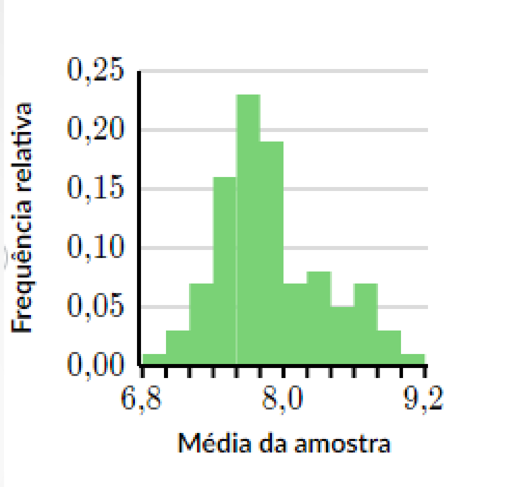
A Dra. Virginia Apgar desenvolveu um teste de saúde infantil que hoje é conhecido como "teste de Apgar". Ele avalia a saúde de uma criança em uma escala de 0 a 2 em cinco indicadores diferentes e a maior pontuação total possível no teste é 10. Aqui está a distribuição de pontuações de Apgar com base em mais de 50.000 registros de nascimento:

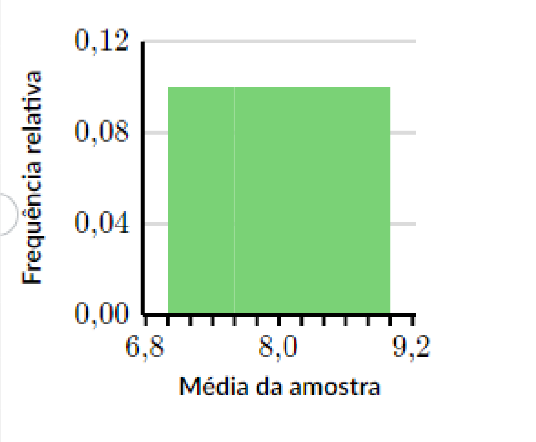


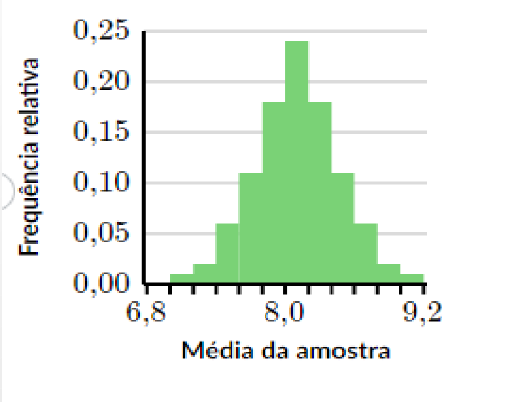
suponha que seleccionamos amostras de n = 30.

Qual gráfico mostra a aproximação mais razoável da distribuição amostral de $\bar{x}$?









Simulador: http://onlinestatbook.com/stat_sim/sampling_dist/index.html