Probabilidade e Estatística

Aula 7

Distribuição da Média Amostral

Leitura obrigatória:

Devore: Seções 5.3, 5.4 e 5.5

Inferência Estatística

- Na próxima aula vamos começar a parte de inferência estatística que tenta tirar conclusões sobre uma população desconhecida a partir de uma amostra.

- Em particular, queremos tirar conclusões sobre a média populacional, μ, partindo de informações de uma amostra.

  Ex: O Peso médio da população (μ) é maior do que 80 kg?

  Ex: A resistência média (μ) de vigas de um tipo de material é alta o suficiente para se adequar as normas?

  Ex: Um novo medicamento traz um benefício médio (μ) mais alto do que o benefício médio do medicamento antigo?

Objetivos

Precisamos saber como a média amostral ($\bar{X}$) se relaciona com a média populacional (μ)!

Antes de começarmos inferência estatística para a média, vamos obter a distribuição de probabilidade da média amostral ($\bar{X}$).

Para tanto, vamos aprender:

- A definição de amostra aleatória
- A distribuição da média amostral de uma amostra aleatória partindo de uma população normal.
- O famoso Teorema do Limite Central

Vocabulário Básico

POPULAÇÃO

Uma **população** consiste de *todos* os itens ou indivíduos sobre os quais desejamos tirar uma conclusão.

AMOSTRA

Uma **amostra** é uma *porção* da população selecionada para a análise.

PARÂMETRO

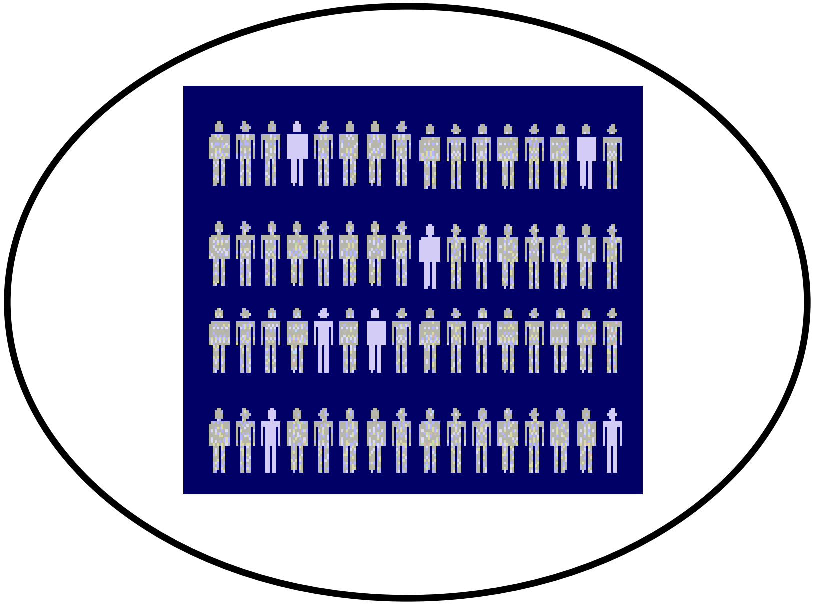
Um **parâmetro** é uma medida númerica que descreve a distribuição da *população*.

ESTATíSTICA

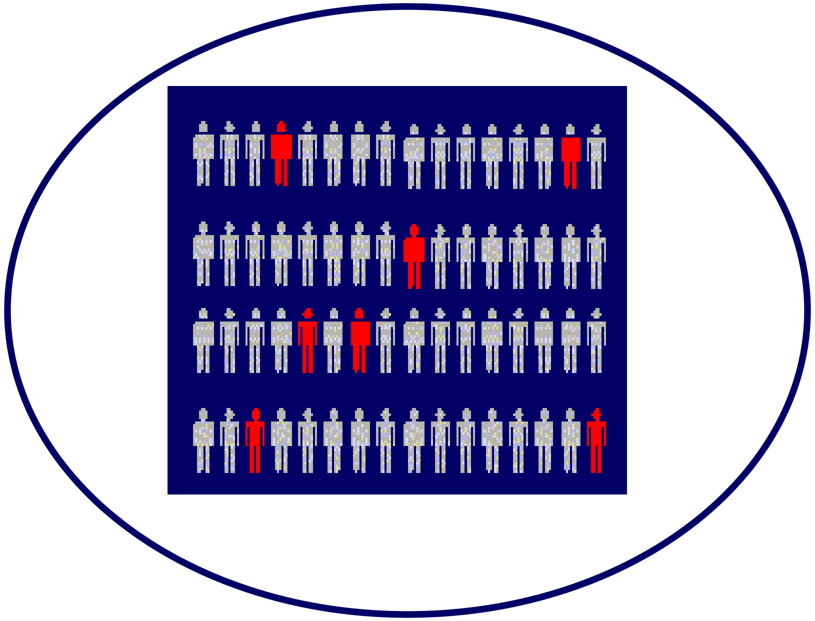
Uma **estatística** é uma medida númerica que descreve uma característica da *amostra*, ou seja, é <u>qualquer função da amostra</u>.

População vs. Amostra

População



Amostra



Medidas usadas para descrever populações são chamadas de **parâmetros**

Medidas computadas para dados amostrais são chamadas de **estatísticas**

Estatísticas: exemplo

- Suponha que você tem uma população de 4 alunos no curso.
- Tamanho da população $N = 4$
- Variável aleatória: $X =$ idade dos alunos
- Valores possíveis de X: $18, 20, 22, 24$ (anos)

- Quais são as amostras possíveis de tamanho 2 (de 2 alunos) com reposição?
- Quais são os valores possíveis para a média das amostras?

Estatísticas: exemplo

Considere todas as amostras possíveis de tamanho $n = 2$

1ª Obs.	2ª Observação			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 amostras diferentes são possíveis para a amostragem com reposição

Todas tem a mesma chance de serem sorteadas.

Estatísticas

- A partir de uma população selecionamos um subconjunto de observações $\{X_1, X_2, \ldots X_n\}$: uma amostra de tamanho n.

- Como a amostra ainda não foi "retirada" da população, os valores $\{X_1, X_2, \ldots X_n\}$ são variáveis aleatórias.

- Qualquer função calculada a partir da amostra $\{X_1, X_2, \ldots X_n\}$ é uma estatística!
  - Ex: média amostral
  - Ex: desvio-padrão amostral
  - Ex: mediana
  - Ex: raíz quadrada do maior valor

Definição!

Estatísticas

- Existe incerteza no valor da estatística antes de obter os dados, ou seja, a estatística é uma variável aleatória.
  - Letras maísculas:variável aleatória estatística (antes)
  - Letras minúsculas: valor que a estatística assume (depois).

- A estatística que estamos interessados nesta aula é a média amostral, $\bar{X}_n$, definida como:

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Estatísticas

Exemplo: idade de alunos que vimos anteriormente. A média amostral com uma amostra de tamanho 2 é:

- Antes de selecionar alunos:

$$\bar{X}_2 = \frac{X_1 + X_2}{2}$$

- Depois de selecionar os alunos: para cada uma das amostras possíveis, podemos calcular a idade média (idade média amostral):
  - Se $x_1 = 18\ e\ x_2 = 20 \Rightarrow \bar{x}_2 = 19$.
  - Se $x_1 = 18\ e\ x_2 = 18 \Rightarrow \bar{x}_2 = 18$.
  - …

Média amostral

- Se a média amostral é uma variável aleatória, qual é a sua **distribuição de probabilidade**? (fdp, FDA, fmp?)

- A **distribuição de probabilidade** da estatística depende do método de amostragem e da distribuição da população.

- No exemplo da idade, assumimos que a amostragem era por sorteio com reposição e que a população possui apenas 4 valores possíveis (18, 20, 22 e 24). Como selecionamos por sorteio, a chance de cada um desses valores é igual!

Distribuição da Média Amostral

Qual é a distribição de probabilidade da idade média de um conjunto de 2 alunos, $\bar{X}_2$?

seja $(x_1, x_2) = (idade\ do\ 1º\ sorteado, idade\ do\ 2º\ sorteado)$

$\bar{x}_2$	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Resultados	(18,18)	(20,18) ou (18,20)	(20,20) ou (22,18), ou (28,22)	(22,20) ou (20,22) ou (24,18) ou (18,24)	(22,22) ou (24,20) ou (20,24)	(22,24) ou (24,22)	(24,24)
$p(\bar{x}_2)$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

- OBS: A distribuição de $\bar{X}_3$ seria diferente... A de $\bar{X}_4$ também...

Distribuição da Média Amostral

Exercício: Um fabricante de celulares vende 3 tipos de modelos de celulares diferentes. O preço do modelo 1 é R$ 80, o do modelo 2 é R$ 100 e o do modelo 3 é R$120 reais. 20%, 30% e 50% dos consumidores escolhem os modelos 1, 2 e 3 respectivamente.

Seja $X = preço\ de\ compra\ de\ um\ celular$, então:

x	80	100	120
$p(x)$	.2	.3	.5

$\mu = 106,\ \sigma^2 = 244$

Suponha que, em certo dia, apenas 2 celulares são vendidos. Assuma que a escolha dos modelos entre clientes seja independente. Qual é a distribuição do preço médio de venda dos 2 celulares?

–

Distribuição da Média Amostral

Solução: A distribuição de probabilidade na população (ou quando apenas 1 consumidor é selecionado) é bem simples. Podemos obter a distribuição da média listando todas as combinações possíveis entre X_1 (preço de venda do celular 1) e X_2 (preço de venda do celular 2). Usamos a independência para obter a probabilidade das combinações possíveis.

x	80	100	120
$p(x)$	.2	.3	.5

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$
80	80	.04	80
80	100	.06	90
80	120	.10	100
100	80	.06	90
100	100	.09	100
100	120	.15	110
120	80	.10	100
120	100	.15	110
120	120	.25	120

Distribuição da Média Amostral

Solução: Partindo da lista de resultados possíveis e suas probabilidades, escrevemos a função massa de probabilidade da média.

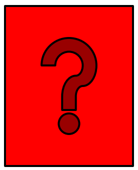
x_1	x_2	$p(x_1, x_2)$	$\bar{x}$
80	80	.04	80
80	100	.06	90
80	120	.10	100
100	80	.06	90
100	100	.09	100
100	120	.15	110
120	80	.10	100
120	100	.15	110
120	120	.25	120

$\bar{x}$	80	90	100	110	120
$p_{\bar{X}}(\bar{x})$	.04	.12	.29	.30	.25

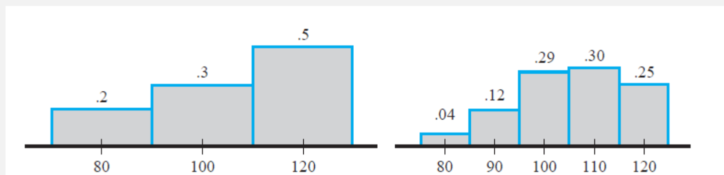
com $\mu_{\bar{X}_2} = \mu = 106$ e $\sigma_{\bar{X}_2} = 122 = \frac{244}{2} = \frac{\sigma}{2}$

–

Distribuição da Média Amostral

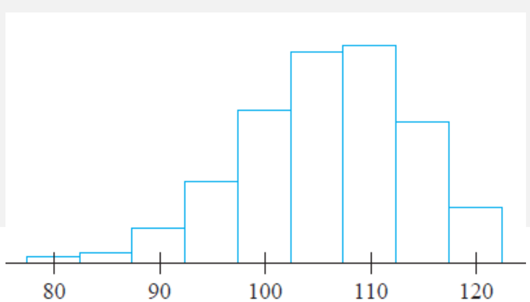


Solução: Graficamente, temos:



E se quisséssemos a distribuição do preço médio de venda de 4 celulares? O mesmo procedimento levaria a:

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\bar{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625



com $\mu_{\bar{X}_4} = \mu = 106$ e $\sigma_{\bar{X}_4} = 61 = \frac{\sigma}{4}$

Amostra Aleatória

- Um método amostral que facilita a obtenção da distribuição de estatísticas é a seleção de uma **amostra aleatória**.

- Uma amostra é **amostra aleatória** se for iid (independente e identicamente distribuída):
  - **Independente:** a coleta de uma observação independe da outra.
  - **Identicamente distribuída***:* cada valor tem a mesma distribuição de probabilidade!

Definição!

Como obter amostra aleatória?

- **Amostra por sorteio com reposição**
- **Amostra por sorteio sem reposição de população infinita**
- **Amostra por sorteio sem reposição de população suficientemente grande (no máximo 5% da população)**

Média Amostral: distribuição

Propriedade: Seja Y uma v.a. definida como a combinação linear de 2 outras variáveis aleatórias, tal que:

$$Y = aX_1 + bX_2$$

em que a e b são constantes.

Então:

- O valor esperado de Y, ou a média de Y, é:

$$E(Y) = E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$$

- Se X_1 e X_2 são independentes:

$$V(Y) = a^2V(X_1) + b^2V(X_2)$$

Média Amostral: distribuição

Exercício: Seja $\{X_1, X_2, \dots, X_n\}$ uma amostra aleatória de uma v.a. X com média μ e variância σ^2.

A média da amostra aleatória é:

$$\bar{X}_n = \frac{(X_1 + X_2 + \cdots + X_n)}{n}$$

a) Calcule o valor esperado da média amostral, $E(\bar{X}_n)$.

b) Calcule a variância da média amostral, $V(\bar{X}_n)$.

Média Amostral: distribuição

Solução:

a) O valor esperado da média amostral é:

$$\boldsymbol{E(\bar{X}_n)} = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{\mathrm{n}}\{E(X_1) + E(X_2) + \cdots + E(X_n)\}$$

$$= \frac{1}{n}\{\mu + \mu + \cdots + \mu\} = \frac{1}{n}\{n\mu\} = \boldsymbol{\mu}$$

$\therefore \boldsymbol{E(\bar{X}_n) = \mu}$

Ou seja, a média (ou o valor esperado) da média amostral ($\bar{X}$) é igual a média populacional (μ)

Média Amostral: distribuição

Solução:

b) A variância da média amostral é:

$$V(\bar{X}_n) = V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{1}{\mathrm{n}^2} V(X_1 + \cdots + \mathrm{X_n})$$

Como a amostra é aleatória, as variáveis aleatórias $\{X_1, \ldots, X_n\}$ são independentes, de forma que:

$$V(\bar{X}_n) = \frac{1}{\mathrm{n}^2} V(X_1 + \cdots + \mathrm{X_n}) = \frac{1}{n^2}\{V(X_1) + \cdots + V(X_n)\}$$

$$= \frac{1}{n^2}\{\sigma^2 + \cdots + \sigma^2\} = \frac{1}{n^2}\{n * \sigma^2\} = \frac{\sigma^2}{n}$$

E o desvio-padrão é: $\boldsymbol{\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}}$

∴o desvio-padrão da distribuição da média amostral é $\boldsymbol{\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}}$

Média Amostral: distribuição

Exercício: Seja $\{X_1, X_2, \dots, X_n\}$ uma amostra aleatória de uma v.a. X com média μ e variância σ^2.

O valor total da amostra é:

$$T_0 = X_1 + X_2 + \cdots + X_n$$

a) Calcule o valor esperado da total da amostra, $E(T_0)$.

b) Calcule a variância do total da amostra, $V(T_0)$.

Resposta: $n\mu$ e $n\sigma^2$.

Média Amostral: distribuição

Exercício: Em um teste de fadiga à tração de certo material, o número esperado de ciclos para a o início de uma trinca é $\mu = 28\,000$ e o desvio padrão do nº de ciclos é $\sigma = 5000$.

Assuma que $X_1, X_2, \ldots, X_{25}$ são itens de uma amostra aleatória, em que cada X_i é o nº de ciclos para um corpo de prova diferente.

Qual é o valor esperado do nº de ciclos médio da amostra? E o desvio padrão do nº de ciclos médio?

Solução:

Como o tamanho da amostra é $n = 25$ e a amostra é aleatória:

$E(\bar{X}_{25}) = E(X) = \mu = 28000$ e $\sigma_{\bar{X}_{25}} = \frac{\sigma}{\sqrt{n}} = \frac{5000}{5} = 1000$

Média Amostral: distribuição

Propriedade: A combinação linear de variáveis aleatórias **normais independentes** tem **distribuição normal**!

Sejam $X_1 \sim N(\mu_1, \sigma_1)$ e $X_2 \sim N(\mu_2, \sigma_2)$, então:

$$Y = aX_1 + bX_2 \sim N(\mu_Y, \sigma_Y)$$

com:

$$\mu_Y = a\mu_1 + b\mu_2$$

$$\sigma_Y = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$$

Média Amostral: distribuição

Propriedade: Seja $\bar{X}_n$ a média amostral de uma amostra aleatória $\{X_1, \dots, X_n\}$ com tamanho n selcionada de uma v.a. X com média μ e variância σ² :

Então:

$$\boldsymbol{E(\bar{X}_n) = \mu} \textbf{ e } \boldsymbol{V(\bar{X}_n) = \frac{\sigma^2}{n}}$$

E se $\{X_1, \dots, X_n\}$ é amostra aleatória de v.a. X com distribuição normal: $X \sim N(\mu, \sigma)$:

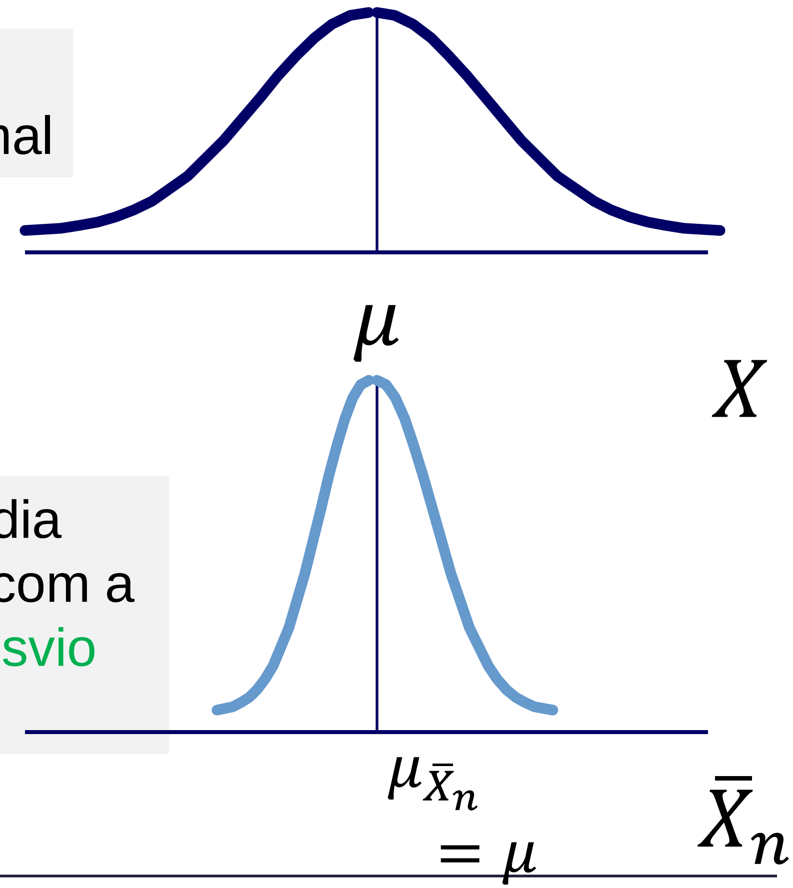
$$\boldsymbol{\Rightarrow \bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Distribuição da Média Amostral

População com Distribuição normal

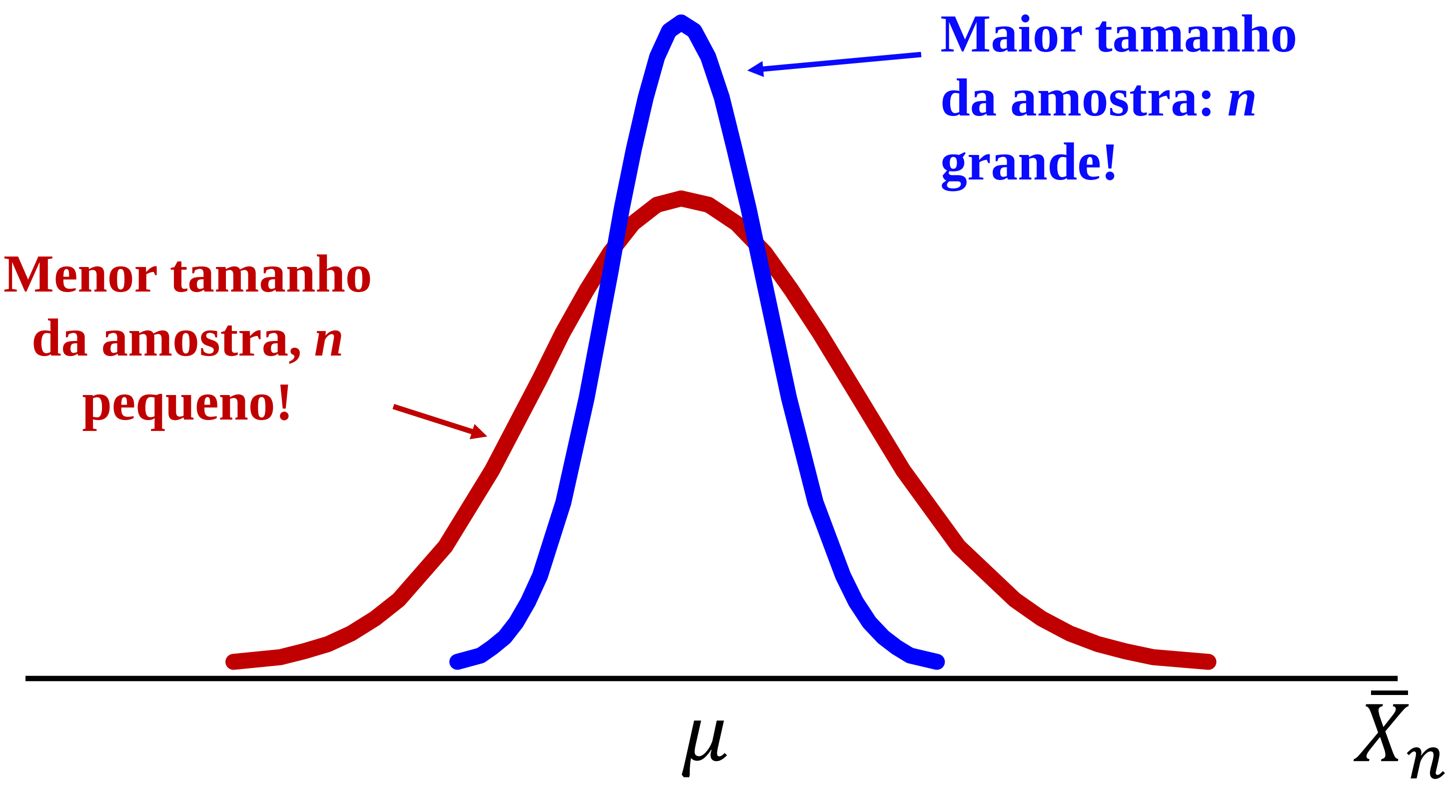
$$\mu_{\bar{X}_n} = \mu_X = \mu$$

Distribuição da média amostral é normal com a mesma média e desvio padrão menor!



Distribuição da Média Amostral

A medida que n cresce, $\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}$ diminui!



Distribuição da Média Amostral

Exercício: O diâmetro interno de um pistão selecionado ao acaso é uma **variável aleatória normal** com valor médio $\mu = 12$cm e desvio-padrão $\sigma = 0.04cm$.

a) Seja $\bar{X}_{16}$ o diâmetro médio para uma amostra aleatória de tamanho *n=16* pistões. Qual é a distribuição de $\bar{X}_{16}$? ? Faça o gráfico da função densidade de $\bar{X}_{16}$ e indique onde está centrada a distribuição da média amostral $\bar{X}_{16}$, e valor do desvio-padrão da média amostral $\bar{X}_{16}$.

b) Repita a letra a para um amostra com $n = 64$ pistões, isto é, obtenha a distribuição de $\bar{X}_{64}$

c) Para qual dos dois tamanhos de amostra, $n = 16$ ou $n = 64$, a probabilidade da média estar a menos do que 0.01 cm de distância de 12 cm é menor?

Distribuição da Média Amostral

Exercício: Solução:

Seja X o diâmetro interno de um pistão selecionado ao acaso. Segundo o encunciado: $X \sim N(\mu = 12, \sigma = 0.04)$.

a) Como a população tem distribuição normal, a média amostral de uma amostra com tamanho n tem distribuição: $\bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Assim, para uma amostra 16 pistões: $\bar{X}_{16} \sim N\left(12, \frac{0.04}{\sqrt{16}}\right) = N(12{,}0.01)$.

b) De forma similar, como $n = 64$: $\bar{X}_{64} \sim N\left(12, \frac{0.04}{\sqrt{64}}\right) = N(12{,}0.005)$.

c) Queremos calcular $P($média amostral a menos de 0.01 de distância da média populacional$)$ = $P(|\bar{X}_n - \mu| < 0.01) = P(|\bar{X}_n - 12| < 0.01) =$ $P(-0.01 < \bar{X}_n - 12 < 0.01) = P(11.99 < \bar{X}_n < 12.01)$ para $n = 16$ e $n = 64$.

–

Distribuição da Média Amostral

Exercício: **Solução:**

c) $P(11.99 < \bar{X}_n < 12.01) = ??$ para $n = 16$ e $n = 64$.

Como definimos nas letras a e b, $\bar{X}_n$ tem distribuição normal. Para calcular esta probabilidade devemos calcular o escore Z associado a 11.99 e 12.01 para então olhar a probabilidade acumulada na tabela da normal padrão.

Daí, para $n = 16$:

$$P(11.99 < \bar{X}_{16} < 12.01)$$
$$= P\left(Z < \frac{12.01 - 12}{0.01}\right) - P\left(Z < \frac{11.99 - 12}{0.01}\right)$$
$$= P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826$$

E, para $n = 64$:

– ...

Distribuição da Média Amostral

Exercício: Solução:

c) $P(11.99 < \bar{X}_n < 12.01) = ??$ para $n = 16$ e $n = 64$.

E, para $n = 64$:

$$P(11.99 < \bar{X}_{64} < 12.01)$$
$$= P\left(Z < \frac{12.01 - 12}{0.005}\right) - P\left(Z < \frac{11.99 - 12}{0.005}\right)$$
$$= P(Z < 2) - P(Z < -2) = 0.9772 - 0.0228 = 0.9554$$

Sugestão: Façam o gráfico das distribuições de $\bar{X}_{16}$ e $\bar{X}_{64}$. Observem que como não a distribuição da média amostral com 64 pistões tem um desvio-padrão bem menor (a metade!) do que o desvio-padrão da média com 16 pistões, não era necessário fazer os cálculos para responder à questão!

Ilustração TLC

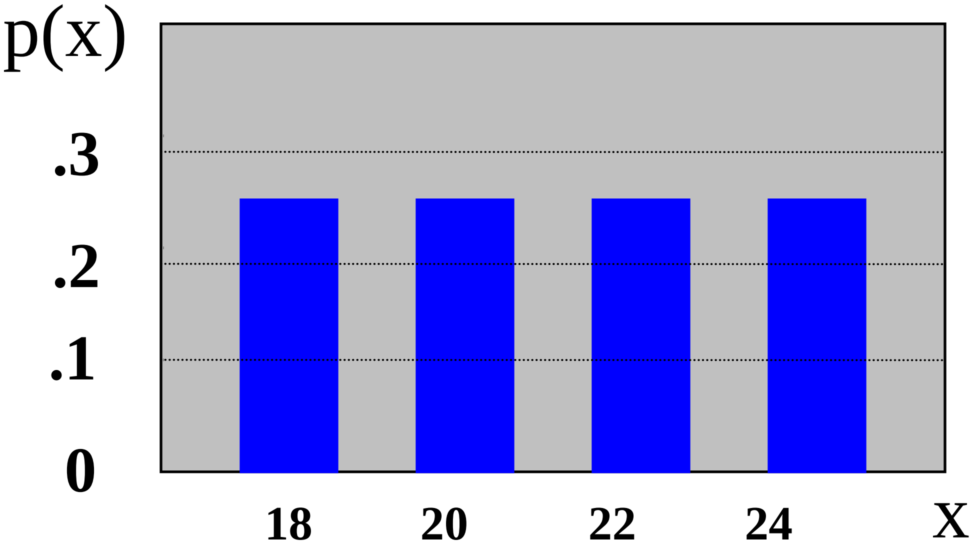
- E se a variável aletória não for normal?? Por exemplo, se ela possuir uma distribuição discreta?

- Suponha que você tem uma população de 4 alunos no curso.
  - Tamanho da população N=4
  - Variavel aleatória, X: idade de um aluno selecionado aleatoriamente
  - Valores possíveis de X: 18, 20, 22, 24 (anos)

Ilustração TLC

Distribuição da **População** da idade de 1 aluno selecionado aleatoriamente:

$$\mu = \sum_i x_i p(x_i) = (18 + 20 + 22 + 24)/4 = 21$$



$$\sigma = \sqrt{\sum_i (x_i - \mu)^2 p(x_i)} = 2.236$$

Ilustração TLC

Considere todas as amostras possíveis de tamanho n=2

1ª Obs.	2ª Observação			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	29,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 amostra possíveis (amostragem com reposição)

16 Médias Amostrais: idade média para cada amostra de 2 alunos

1ª Obs.	2ª Observação			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Ilustração TLC

Distribuição de média amostral de tamanho 2:

1ª Obs.	2ª Observação			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

16 amostras diferentes são possíveis

$\bar{x}_2$	$p(\bar{x}_2)$
18	$1/16$
19	$2/16$
20	$3/16$
21	$4/16$
22	$3/16$
23	$2/16$
24	$1/16$

Ilustração TLC

Distribuição de média amostral de tamanho 2:

16 Médias Amostrais

1ª Obs.	2ª Observação			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

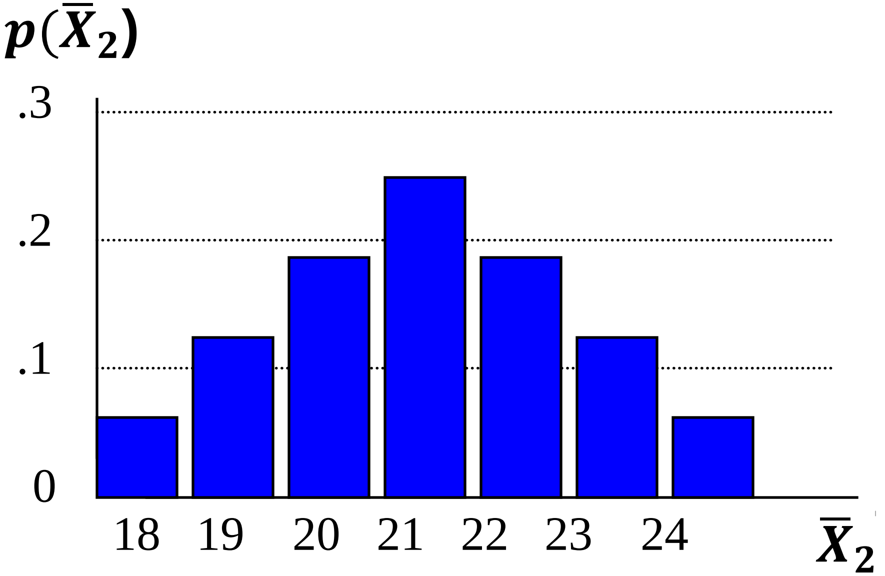


Ilustração TLC

Medidas para a distribuição de $\bar{X}_2$:

$$\mu_{\bar{X}_2} = \sum \bar{x}_2 p(\bar{x}_2) =$$

$$= \frac{(18 + 2 * 19 + 3 * 20 + 4 * 21 + 3 * 22 + 2 * 23 + 24)}{16} = 21$$

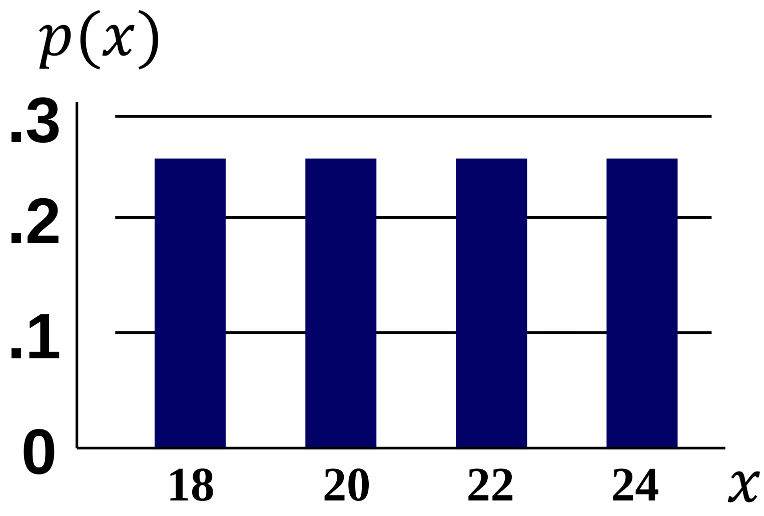
$$\sigma_{\bar{X}_2} = \sqrt{\sum \left(\bar{x}_2 - \mu_{\bar{X}_2}\right)^2 p(\bar{x}_2)} =$$

$$= \sqrt{\frac{(18-21)^2 + 2(19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Ilustração TLC

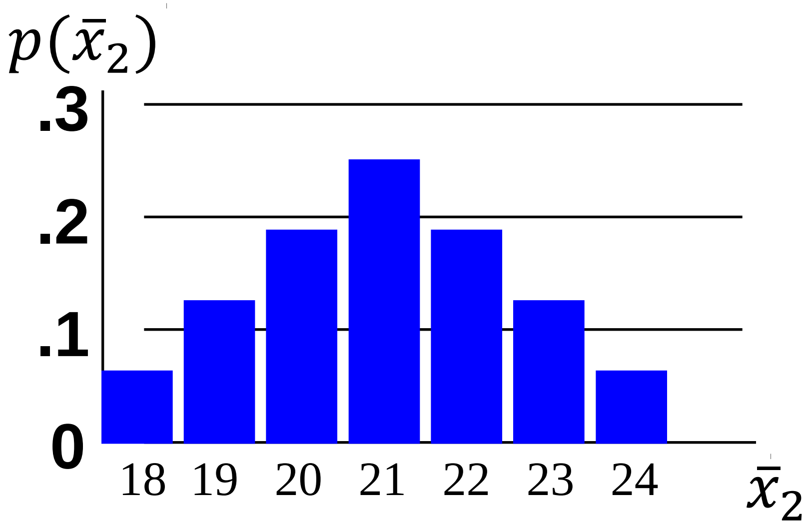
Distribuição da População com N = 4

$\boxed{\mu = 21 \text{ e } \sigma = 2.236}$



Distribuição média amostral com amostras de tamanho n = 2

$\boxed{\mu_{\bar{X}_2} = 21 \text{ e } \sigma_{\bar{X}_2} = 1.58}$



Teorema do Limite Central

Teorema do Limite Central

Seja $\{X_1, X_2, \dots, X_n\}$ uma amostra aleatória (iid) de uma v.a. X que tem qualquer distribuição com média, μ, e variância, σ^2, finita ($0 < \sigma^2 < \infty$):

Se $n \to \infty$, então:

$$\boldsymbol{\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

$$\boldsymbol{T_0 = X_1 + X_2 + \cdots + X_n \sim N\left(n\mu, \sqrt{n}\sigma\right)}$$

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

Teorema do Limite Central

- O Teorema do Limite Central garante que se cada amostra for grande o suficiente (n indo para infinito), a distribuição da média amostral é aproximadamente normal.

- **E isto é verdade independentemente do formato da distribuição de X!**

- Uma razão para distribuições com formato de sino (normais) aparecerem tantas vezes na natureza…

Teorema do Limite Central

Regras de bolso:

- Para a maior parte das distribuições, $n > 30$ implica em uma distribuição da média amostral quase normal.

- Para distribuições praticamente simétricas, $n > 15$ implica em uma distribuição da média amostral quase normal.

Teorema visto anteriormente:

- Para populações com distribuição normal, a distribuição da média amostral sempre é normal para qualquer $n \geq 1$!

Teorema do Limite Central

Exercício: Sejam $X_1, X_2, \dots, X_{100}$ os pesos líquidos reais de 100 sacos de fertilizantes de 50 lb selecionados aleatoriamente.

a) Se o peso esperado de cada saco for 50 lb e a variância 1 lb^2, calcule a probabilidade de a média amostral estar entre 49.75 lb e 50.25 (aproximadamente), usando o teorema do limite central.

b) Se o peso esperado for de 49.8 lb e não 50 lb, de modo que, na média, os sacos não estejam muito cheios, calcule a mesma probabilidade do item anterior. Assuma mesma variância (1 lb^2) por saco.

Teorema do Limite Central

Solução: Seja X = o peso líquido real de um saco de fertilizante e $\{X_1, \ldots, X_{100}\}$ é amostra aleatória de tamanho 100 de X.

a) Pelo enunciado, $E(X) = \mu = 50$ e $V(X) = \sigma^2 = 1$. Queremos calcular a probabilidade: $P(49.75 < \bar{X}_{100} < 50.25)$. Como temos uma amostra aleatória grande (100 sacos de fertilizante) de uma população com variância finita, pelo teorema do limite central, podemos aproximar a distribuição de $\bar{X}_{100}$ por uma distribuição normal com média e desvio-padrão: $\mu_{\bar{X}_{100}} = \mu = 50$ e $\sigma_{\bar{X}_{100}} = \frac{\sigma}{\sqrt{100}} = \frac{1}{10} = 0.1$.

Assim:

$$P(49.75 < \bar{X}_{100} < 50.25) = P(\bar{X}_{100} < 50.25) - \mathrm{P}(\bar{X}_{100} < 49.75)$$

$$= \mathrm{P}\left(\mathrm{Z} < \frac{50.25 - 50}{0.1}\right) - \mathrm{P}\left(\mathrm{Z} < \frac{49.75 - 50}{0.1}\right)$$

$$= P(Z < 2.5) - P(Z < -2.5) = 0.9938 - 0.0062 = 0.9876$$

Teorema do Limite Central

Solução: Seja X = o peso líquido real de um saco de fertilizante e $\{X_1, \dots, X_{100}\}$ é amostra aleatória de tamanho 100 de X.

a)

b) Agora $E(X) = \mu = 49.8$ e a variância não se altera. $P(49.75 < \bar{X}_{100} < 50.25)$. Seguindo o mesmo raciocínio do item anterior, pelo TLC: $\bar{X}_{100} \sim N(49.8, 0.1)$.

Assim:

$$P(49.75 < \bar{X}_{100} < 50.25) = P(\bar{X}_{100} < 50.25) - \mathrm{P}(\bar{X}_{100} < 49.75)$$

$$= \mathrm{P}\left(\mathrm{Z} < \frac{50.25 - 49.8}{0.1}\right) - \mathrm{P}\left(\mathrm{Z} < \frac{49.75 - 49.8}{0.1}\right)$$

$$= P(Z < 4.5) - P(Z < -0.5) = 1 - 0.3085 = 0.6915.$$

Observem como as probabilidades mudaram devido à mudança na média populacional igual a 0.2. Parece pouco, mas como o tamanho da amostra é muito grande, 0.2 corresponde a 2 vezes o desvio-padrão da média amostral!

Teorema do Limite Central

Aplicação do teorema do limite central.

Um item de um lote é selecionado aleatoriamente. Defina Sucesso = 'o item não tem defeito'. A variável aleatória X tem distribuição Bernoulli, isto é:

x	0	1
$p(x)$	$(1-p)$	p

Selecionamos uma amostra aleatória de itens do lote, $\{X_1, X_2, \dots, X_n\}$.

Qual a distribuição da proporção amostral de itens não defeituosos na amostra de tamanho n? Suponha que n é suficientemente grande.

Teorema do Limite Central

Aplicação do teorema do limite central

X_i é igual a 1 se o i-ésimo item não é defeituoso.

Assim, **a proporção amostral de itens defeituosos, $\hat{p}$**, é:

$$\hat{p} = \bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

Pelo teorema do limite central, se n é suficientemente grande:

$$\hat{p} = \bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$$

Precisamos calcular μ_X e σ_X!

Teorema do Limite Central

Aplicação do teorema do limite central

Como X é uma variável de Bernoulli, lembrando de distribuições discretas:

$$\mu_X = E(X) = p$$

e

$$\sigma_X = \sqrt{V(X)} = \sqrt{p(1-p)}$$

Assim, se n é suficientemente grande:

$$\boldsymbol{\hat{p} = \bar{X}_n \sim N\left(p, \frac{\sqrt{p(1-p}}{\sqrt{n}}\right)}$$

Teorema do Limite Central

Propriedade: Se $\{X_1, X_2, \dots, X_n\}$ é uma amostra aleatória de uma variável aleatória de Bernoulli com n grande*, a distribuição da proporção amostral sucessos, $\hat{p}$, é:

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

em que p é a probabilidade de sucesso!

*nesse caso, a aproximação é boa se:

$$n * p > 5 \ e \ n * (1 - p) > 5.$$

Teorema do Limite Central

Exercício: O primeiro trabalho de um curso de informática envolve a execução de um programa curto.

Se a experiência anterior indica que 40% de todos os alunos *não* cometerão erros de programação, calcule a probabilidade (aproximada) de que, em uma classe de 50 alunos, pelo menos 25 cometerão erros.

Teorema do Limite Central

Solução:

Seja Sucesso = 'aluno não comete erro de programação'.

$X_i = 1$ se o aluno i não comete erro, com $i = 1, \dots, 50$ alunos, tal que, X_i é v.a. Bernoulli com $p = 0.4$.

Pelo Teorema do Limite Central:

$$\hat{p} \sim N\left(0.4, \sqrt{\frac{0.4 * 0.6}{50}}\right) = N(0.4, 0.07)$$

Proporção de alunos que não cometem erro

Teorema do Limite Central

Solução:

$$\hat{p} \sim N\left(0.4, \sqrt{\frac{0.4 * 0.6}{50}}\right) = N(0.4, 0.07)$$

Se pelo menos 25 alunos cometem erro então:

nº de alunos que cometem erro: 25, 26, 27, ..., 49, 50

nº de alunos que não cometem erro: 25, 24, 23, ..., 1, 0

proporção de alunos que não cometem erro: $\hat{p} \leq \frac{25}{50} = 0.5$

$$P(pelo\ menos\ 25\ alunos\ cometem\ erro) = P(\hat{p} \leq 0.5) =$$

$$= P\left(Z \leq \frac{0.5 - 0.4}{0.07}\right) = P(Z \leq 1.45) = 0.9265$$

Resumo

Nesta aula, aprendemos

- Estatísticas
- Distribuição de uma estatística especial: a média amostral
- Teorema do Limite Central