Introdução à Inferência Estatística – Aula 01

Statistics for Business and Economics 7 edição, by Paul Newbold , William Carlson , Betty Thorne (cap. Sampling and Sampling Distributions)

Cap 10 Bussab e Morettin

Statistics for Economics, Accounting and Business Studies, capítulo 2, Barrow

Marislei Nishijima

Instrumentos de Estatística

- **Estatísticas descritivas**
- Coletando, Apresentando e Descrevendo Dados
- **Estatística inferencial**
- Tirar conclusões e / ou tomar decisões sobre uma população com base apenas em dados amostrais

População e Amostras

População é o conjunto de todos os itens ou indivíduos de interesse

Exemplos:

Todos os prováveis eleitores na próxima eleição

Todos os faturamentos de empresas em novembro

CENSO populacional

Uma amostra é um subconjunto da população

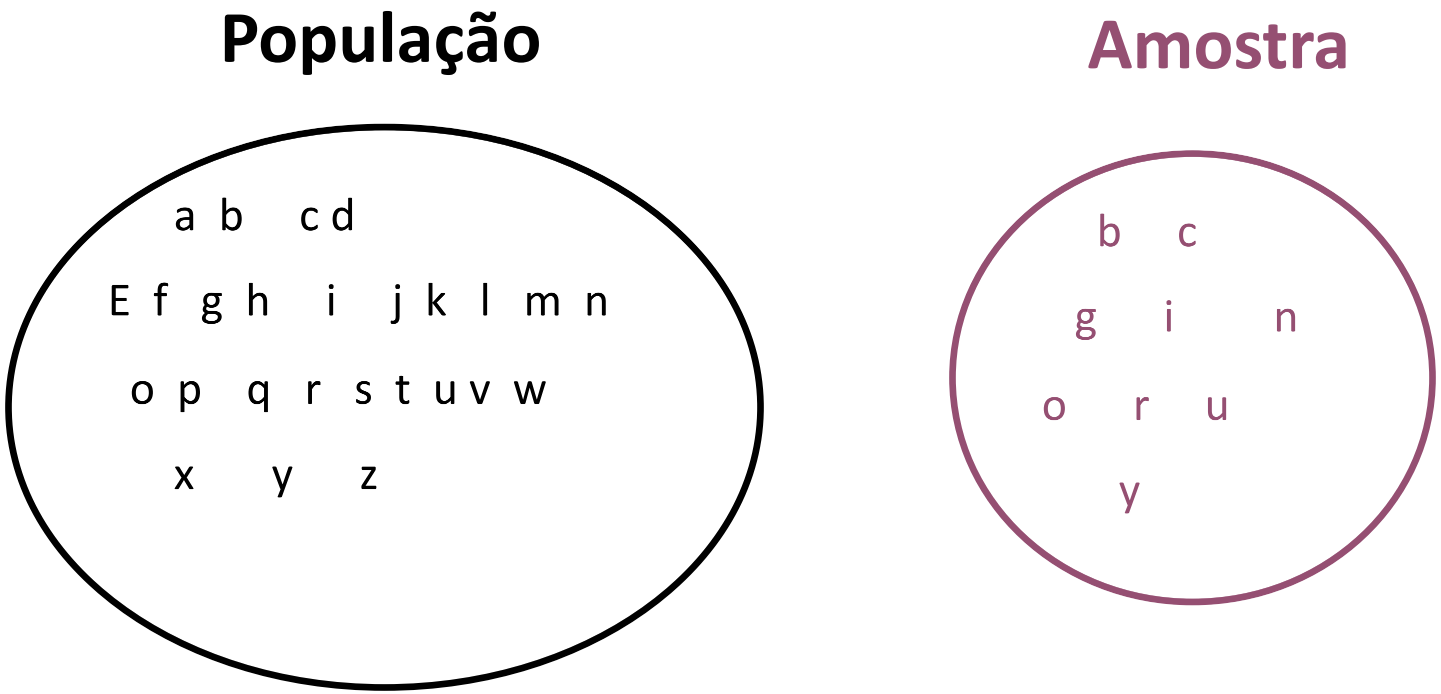
Exemplos:

1000 eleitores selecionados aleatoriamente para entrevista

Faturamentos aleatórios selecionados para auditoria

PNAD

População vs Amostra

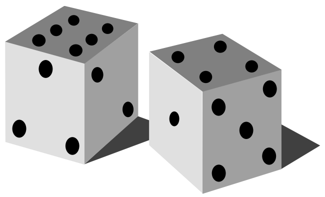


Por que fazer amostras?

- Menos demorado que um CENSO
- Mais barato de administrar do que um CENSO
- É possível obter resultados estatísticos de precisão suficientemente alta com base em amostras.

Amostra Aleatória Simples

- Todo objeto na população tem a mesma chance de ser selecionado
- Objetos são selecionados independentemente
- As amostras podem ser obtidas em uma tabela de números aleatórios ou em geradores de números aleatórios por computador



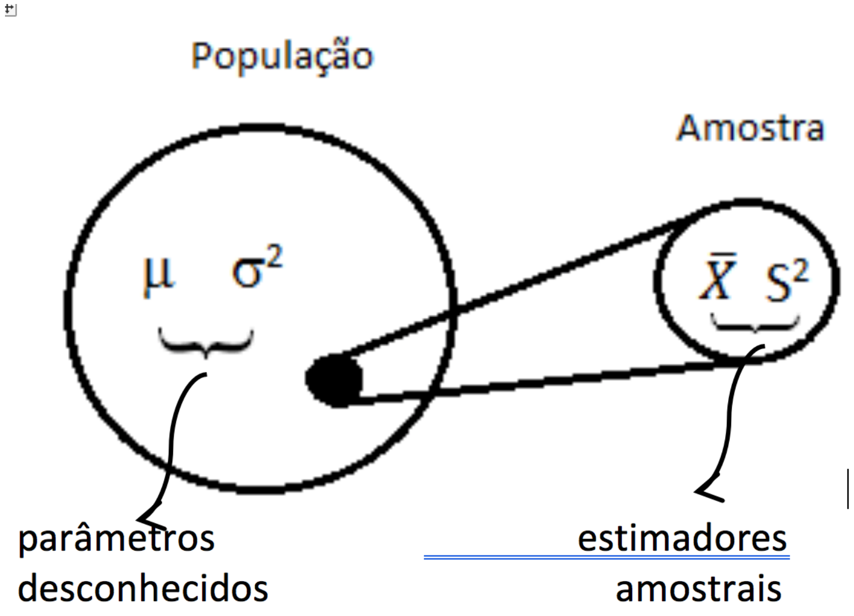
- Uma amostra aleatória simples é o ideal, quando se compara com outros métodos de amostragem

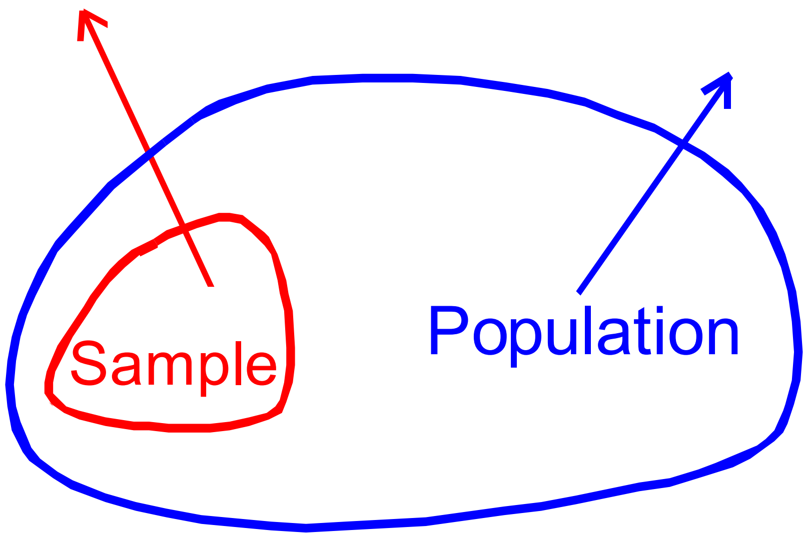
Inferência Estatística

- Fazer afirmações sobre uma população com base em resultados obtidos em amostra

Estatísticas da Amostra → Parâmetros Populacionais

(conhecidos) **Inferência** (desconhecidos)





Inferência Statística

Tirar conclusões e / ou tomar decisões sobre uma população com base nos resultados da amostra.

- **Estimação**
  - por exemplo, estimar o peso médio da população usando o peso médio da amostra
- **Teste de Hipótese**
  - por exemplo, usar evidências da amostra para testar a alegação de que o peso médio da população é de 70kg

Distribuição da População (Lembrando)

Variável aletória que descreve uma população é uma variável *X* (a população alvo) que segue uma distribuição *f(x)*. Isto é, os valores que X pode assumir estão associados com determinadas probabilidades.

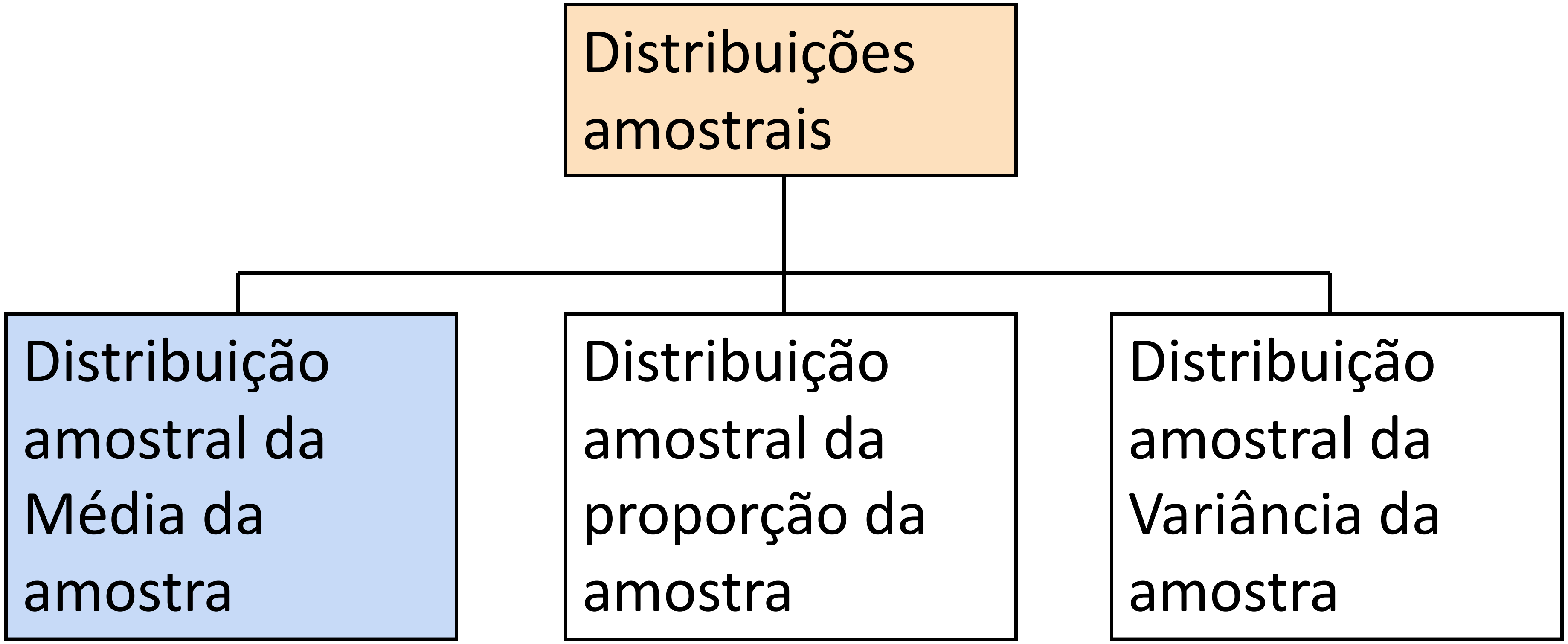
Ex: X é o número de caras obtidas depois de lançar uma moeda 50 vezes. Com moeda não viciada, o número de caras segue uma distribuição binomial, $X \sim b(25,p)$, em que $p=0.5$.

Ex: Suponha que uma máquina automática de encher pacotes de café esteja regulada para essa finalidade de acordo com uma distribuição normal com media de 500 gramas e desvio padrão de 100 gramas, ou $X \sim N(500,\ 100^2)$.

Distribuições Amostrais

- Uma distribuição amostral é uma distribuição de todos os valores possíveis de uma estatística para uma determinada amostra de tamanho selecionada de uma população

Distribuições de amostras de médias de amostras



Desenvolvendo uma distribuição de amostragem

- **Assuma que existа uma população …**
- Tamanho da **população** N=4
- Variável aleatória, X,
  é a idade dos indivíduos
- Valores de X:
  18, 20, 22, 24 (anos)

Desenvolvendo uma distribuição de amostragem

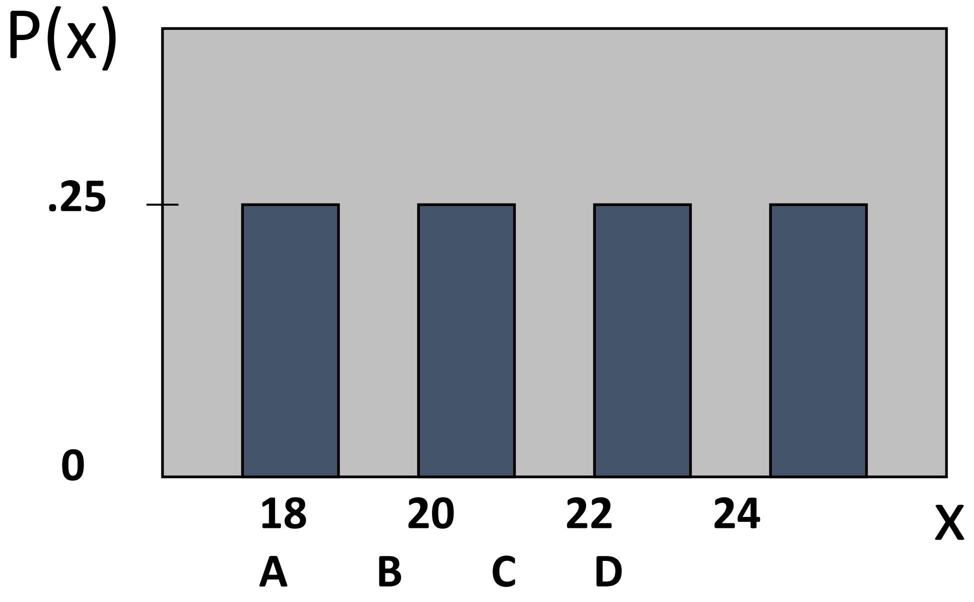
(cont.)

Resumo das medidas para a distribuição da população:

$$\mu = \frac{\sum X_i}{N}$$

$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Distribuição Uniforme

Desenvolvendo uma distribuição de amostragem

(cont.)

Considere todas as possíveis amostra de tamanho n = 2

1st Obs	2nd Observation			
	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 possíveis amostras (com reposição)

16 Médias de Amostras

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Desenvolvendo uma distribuição de amostragem

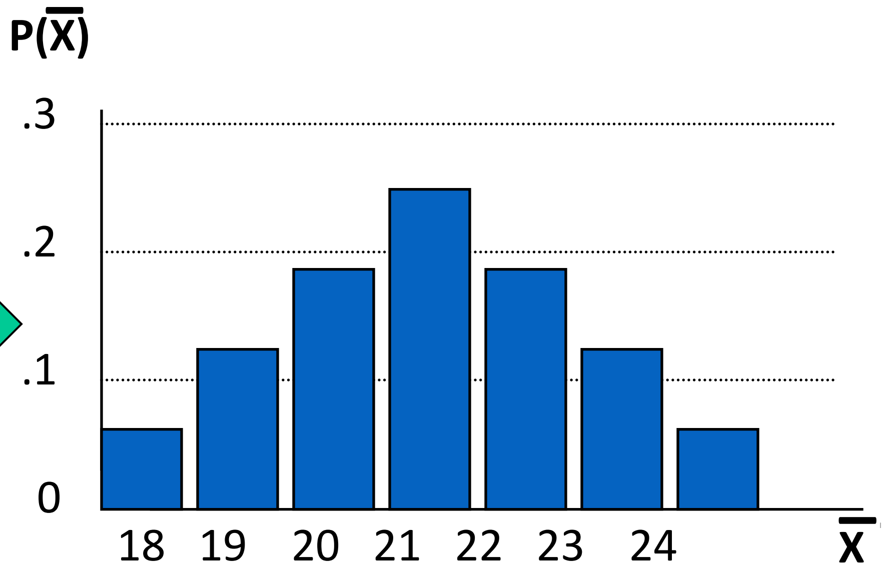
(cont.)

Distribuição amostral de todo as médias de amostras

16 Médias de amostras

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribuição da Média amostral



(não é mais uniforme)

Desenvolvendo uma distribuição de amostragem

(cont.)

Resumo das medidas para a distribuição amostral :

$$E(\overline{X}) = \frac{\sum \overline{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21 = \mu$$

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum (\overline{X}_i - \mu)^2}{N}}$$

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

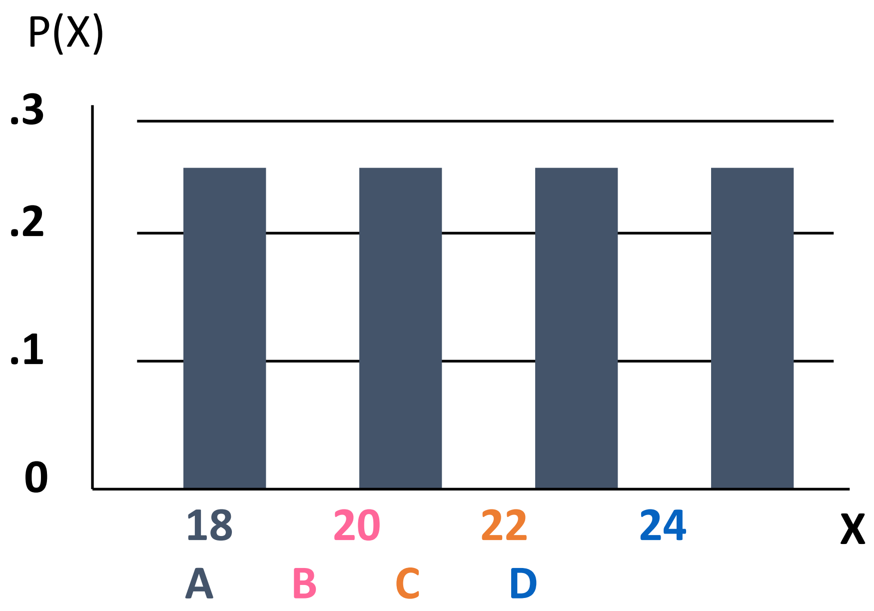
Comparando a Distribuição Populacional com Amostral

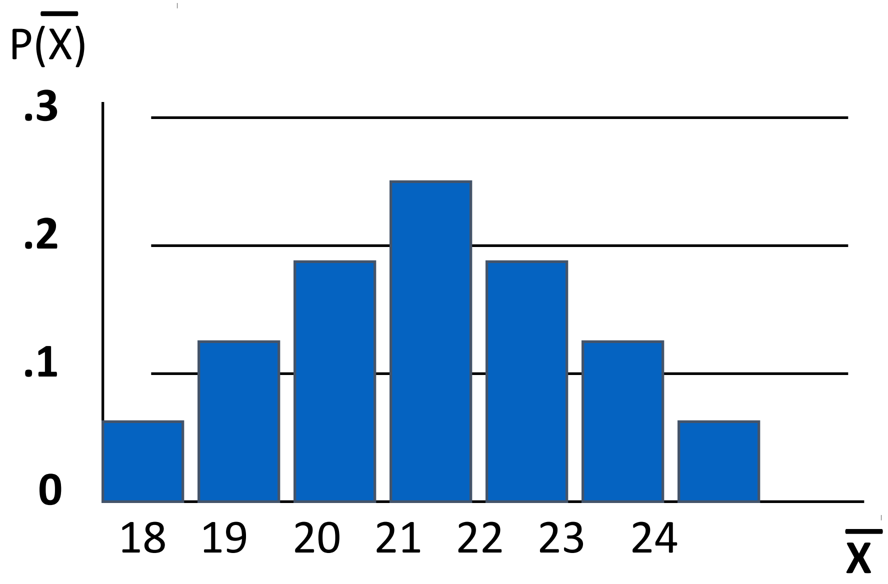
População
N = 4

$\mu = 21 \qquad \sigma = 2.236$

Distribuição da Média
Amostral n = 2

$\mu_{\overline{X}} = 21 \qquad \sigma_{\overline{X}} = 1.58$





Valor Esperado da Média Amostral

- Seja X_1, X_2, . . . X_n uma amostra aleatória da população
- O valor da média amostral destas observações é definida como

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Desvio Padrão da Média (Erro Padrão)

- Amostras diferentes do mesmo tamanho e da mesma população produzirão médias diferentes
- Uma medida da variabilidade na média de amostra para amostra é dada pelo Desvio Padrão da Média:

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

- Observe que o erro padrão da média diminui à medida que o tamanho da amostra aumenta

Se os valores da amostra não forem independentes

(cont.)

- Se o tamanho da amostra n não for uma fração pequena do tamanho da população N, então os membros da amostra individuais não são distribuídos independentemente um do outro
- Assim, as observações não são selecionadas independentemente
- Uma correção é feita para levar isso em conta:

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$$

or

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Se a População tem distribuição Normal

- Se a população é normalmente distribuída com média μ e desvio-padrão σ, a distribuição amostral, $\bar{X}$, também é normalmente distribuída com:

$$\boxed{\mu_{\bar{X}} = \mu}$$

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

- Se o tamanho da amostra n é grande relativamente ao tamanho da população N, então

$$\boxed{\mu_{\bar{X}} = \mu}$$

e

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Lembre que qualquer variavel X com Distribuição Normal pode ser transformada numa Normal Padrão: com média zero e mesma variância.

→ transformar qualquer distribuição normal em uma distribuição normal padronizada

Se X$\sim N(\mu, \sigma^2)$, então Z$= \frac{x-\mu}{\sigma} \sim N(0,1)$.

Z-valor para a Distribuição Amostral da Média

- Z-valor para a distribuição amostral de: $\bar{X}$

$$\boxed{Z = \frac{(\bar{X} - \mu)}{\sigma_{\bar{X}}}}$$

onde:

$\bar{X}$ = média da amostra

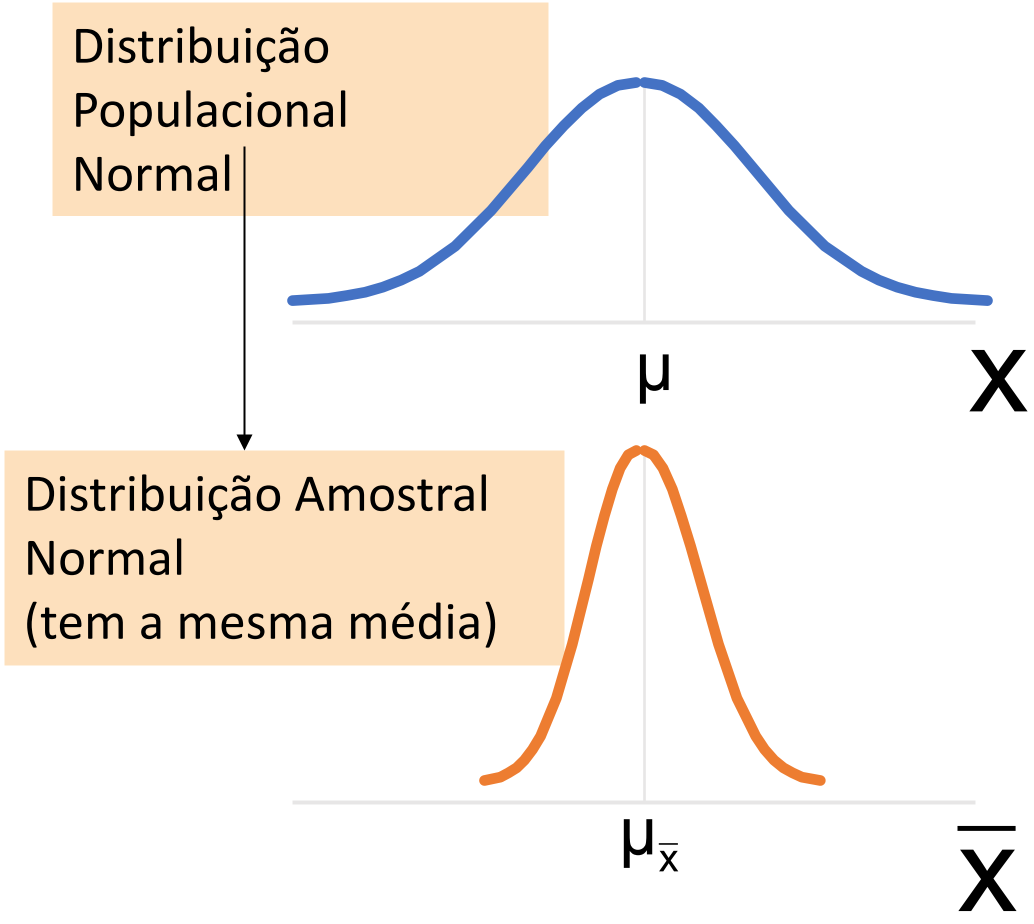
μ = média da população

$\sigma_{\bar{x}}$ = erro padrão da média

Propriedades da Distribuição Amostral

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

$\bar{x}$ (é não viciado, sem viés)

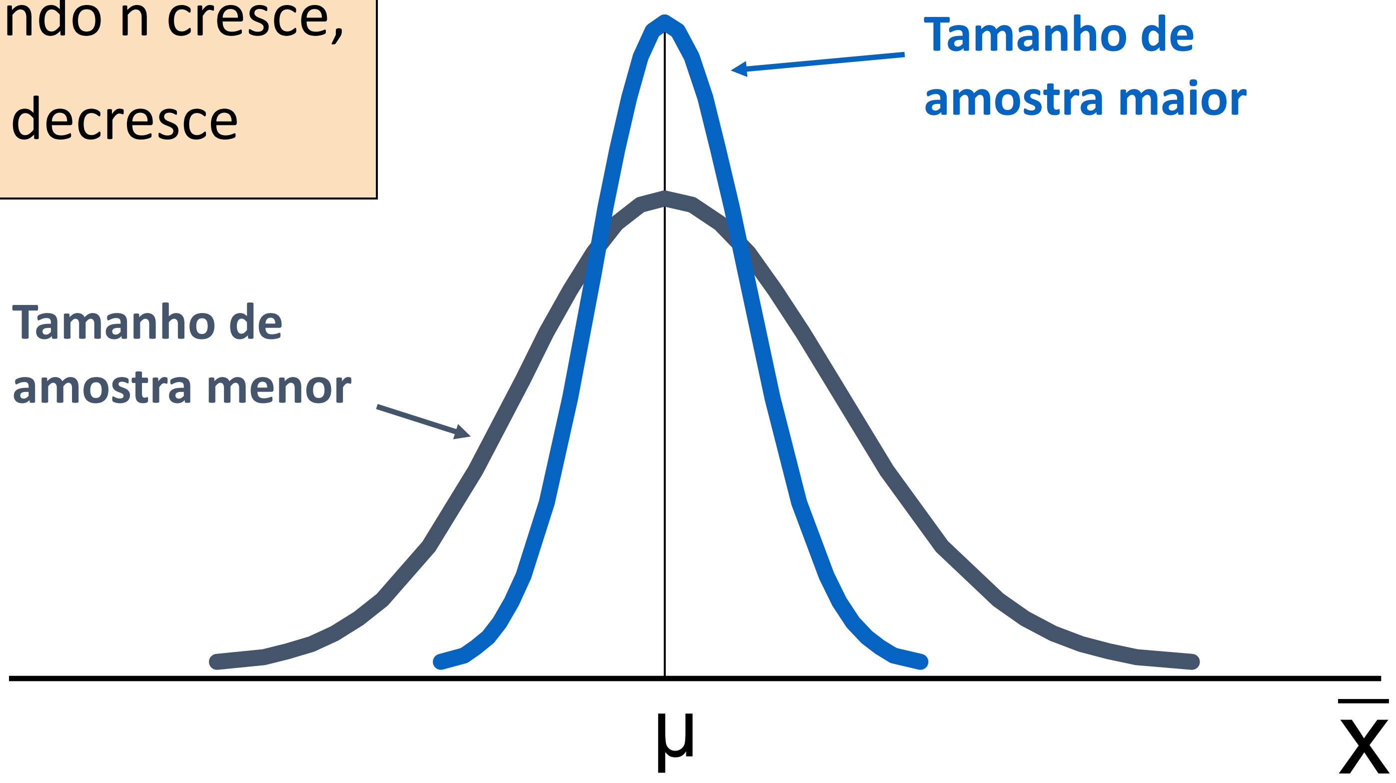


Propriedades da Distribuição Amostral

(cont.)

- Para amostras com reposição:

Quando n cresce,
$\sigma_{\bar{x}}$ decresce



Se a População não for Normalmente distribuída

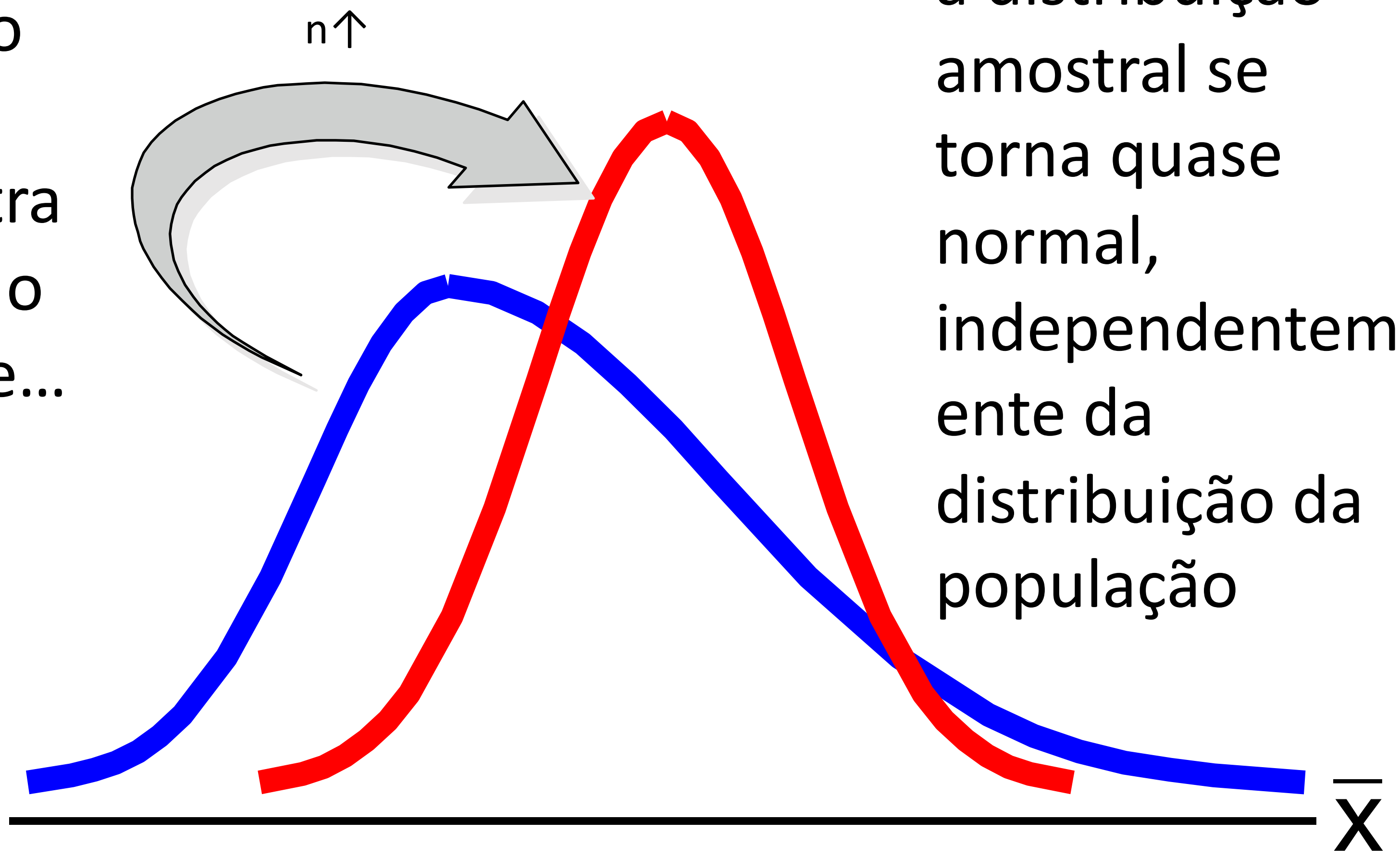
- Podemos aplicar o Theorema Central do Limite :
  - Ainda que a População não seja Normalmente distribuída,
  - …as medias amostrais da população se aproximam de uma normal desde que o tamanho da amostra seja grande o suficiente.

Propriedades da distribuição amostral:

$$\boxed{\mu_{\bar{x}} = \mu}$$ e $$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Teorema Central do Limite

Quando o tamanho da amostra é grande o suficiente…



a distribuição amostral se torna quase normal, independentemente da distribuição da população

Se a População não for Normalmente distribuída

(cont.)

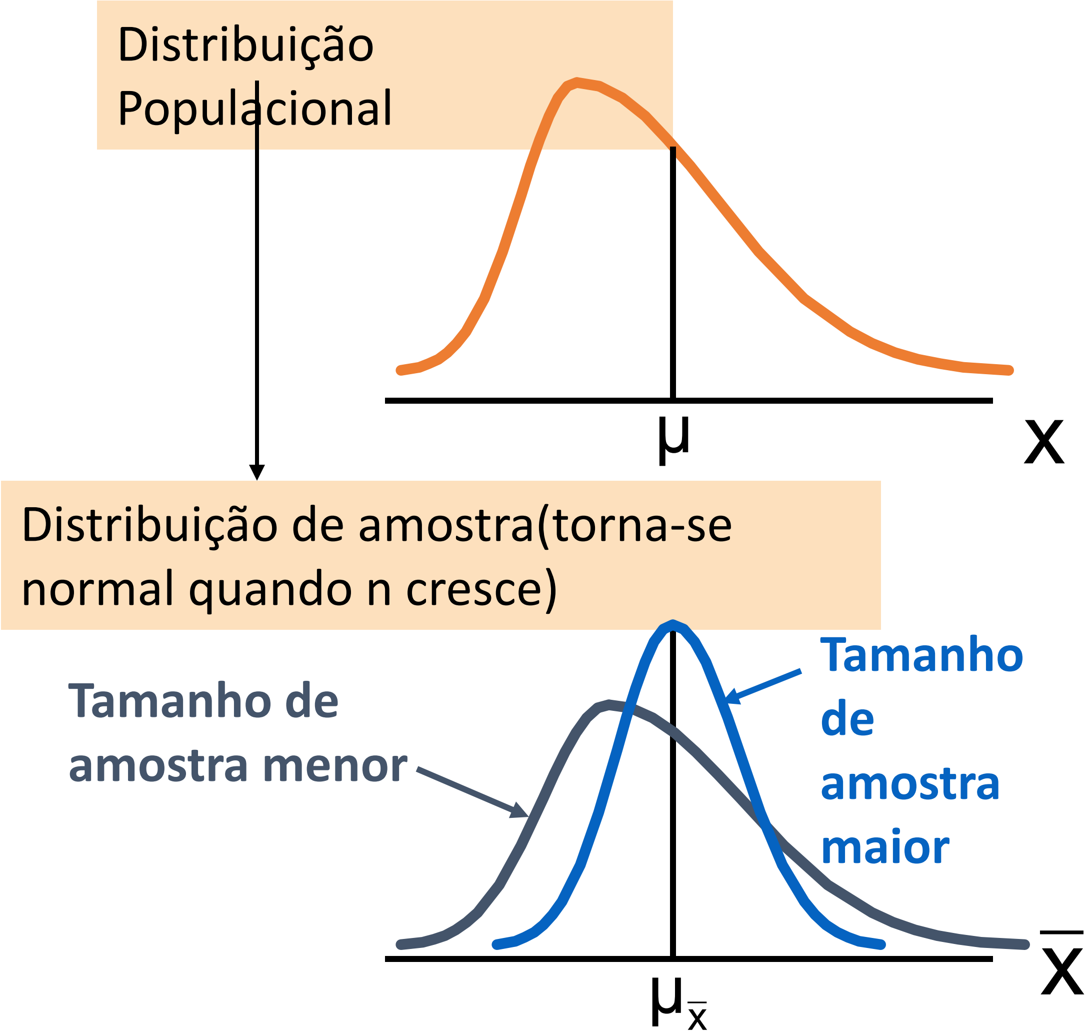
Propriedades de distribuição amostral:

Tendência Central

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

Variação

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



Quão grande é suficiente?

- Para a maioria das distribuições, n> 25 fornecerá uma distribuição de amostra quase normal
- Para distribuições populacionais normais, a distribuição amostral da média é sempre normalmente distribuída

Exemplo

Suponha que uma grande população tenha média $\mu = 8$ e desvio padrão $\sigma = 3$. Suponha que uma amostra aleatória de tamanho $n = 36$ seja selecionada.

- Qual a probabilidade de que a média amostral esteja entre 7.8 e 8.2?

Exemplo

(cont.)

Solução:

- Mesmo que a população não esteja normalmente distribuída, o teorema central do limite pode ser usado (n> 25)
- … então a distribuição amostral de $\bar{X}$ é aproximadamente normal
- … com média = 8 $\mu_{\bar{x}}$
- …e desvio padrão

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$$

Exemplo

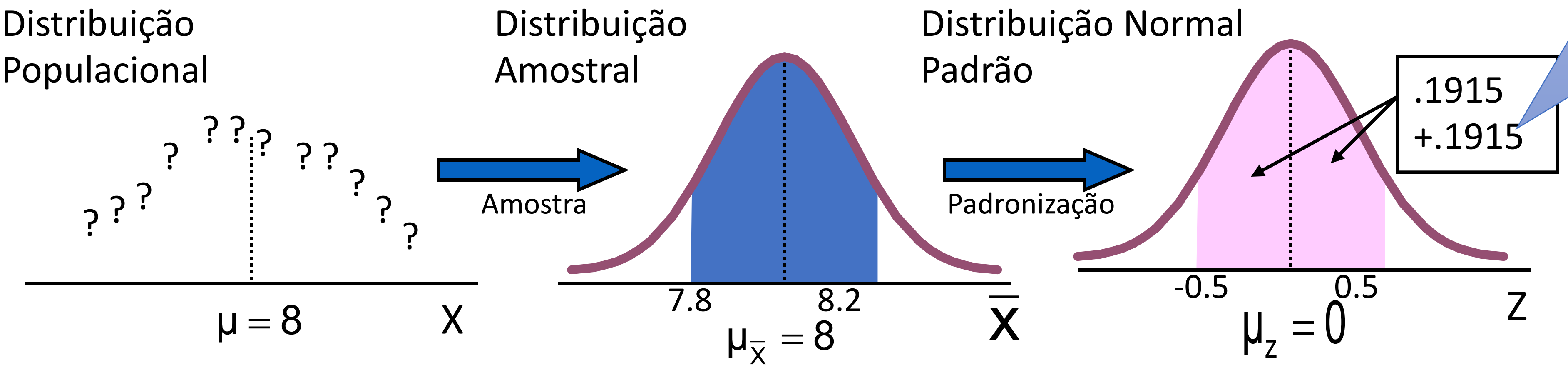
(cont.)

Solução (cont.):

$$P(7.8 < \mu_{\bar{X}} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\mu_{\bar{X}}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right)$$

$$= P(-0.5 < Z < 0.5) = \boxed{0.3830}$$

Ver Tabela Normal Padrão



Aceitação de Intervalo – Primeiras idéias sobre Intervalo de Confiança

- Objetivo: determine um interval dentro do qual a média amostral é provável de ocorrer, dados a média e variância populacional

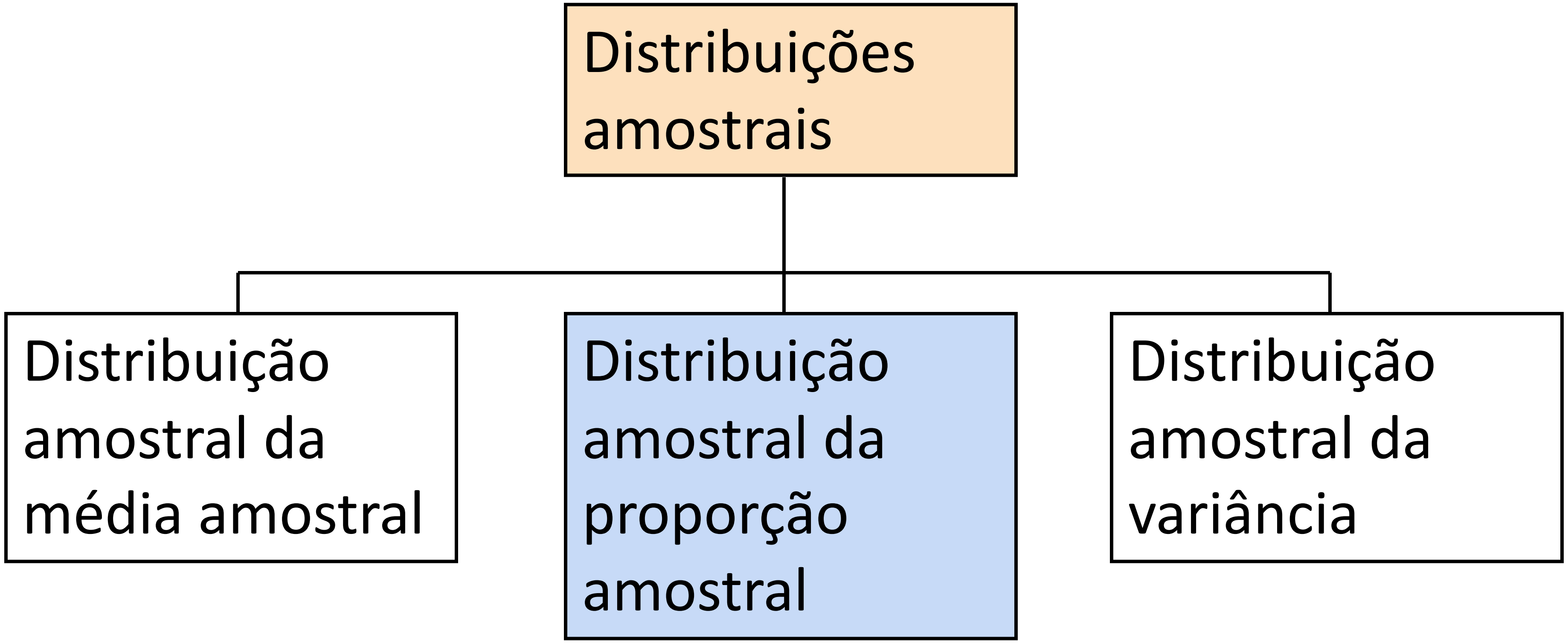
  - Pelo Teorema Central do Limite, sabemos que a distribuição de $\overline{X}$ é aproximadamente normal se n é grande o suficiente, com média μ e desvio padrão $\sigma_{\overline{X}}$

- Seja $z_{\alpha/2}$ o valor z que deixa a área α/2 da calda acima da distribuição normal (i.e., o intervalo entre - $z_{\alpha/2}$ e $z_{\alpha/2}$ inclui média amostral (1 – α)%
- Então

$$\boxed{\mu \pm z_{\alpha/2}\sigma_{\overline{X}}}$$

é o interval que inclui $\overline{X}$ com probability (1 – α)

Distribuições amostrais de proporções amostrais



Distribuições amostrais de proporções amostrais

P = proporção da população que possui alguma característica

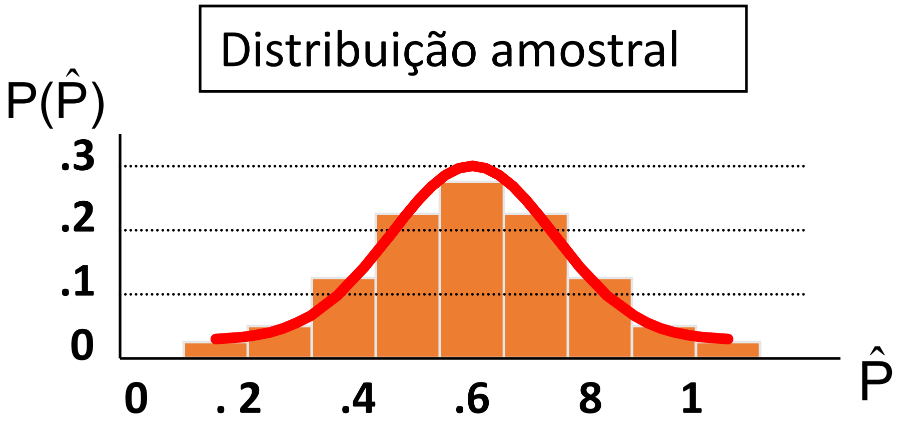
Proporção da amostra ($\hat{p}$) provê uma estimative de P:

$$\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \leq \hat{p} \leq 1$
- $\hat{p}$ tem distribuição binomial, mas pode ser aproximada por uma distribuição normal quando $nP(1 – P) > 5$

Distribuição amostral de $\hat{p}$

- Aproximação da Normal :



Propriedades:

$$\boxed{E(\hat{p}) = P}$$

e

$$\boxed{\sigma_{\hat{p}}^2 = Var\left(\frac{X}{n}\right) = \frac{P(1-P)}{n}}$$

(P = proporção da população)

Z-Valor para Proporções

Padronizando $\hat{p}$ para um Z valor usando a fórmula:

$$Z = \frac{\hat{p} - P}{\sigma_{\hat{p}}} = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Exemplo

- Se a verdadeira proporção de eleitores que apoiam a Proposta A é $P = 0.4$, qual a probabilidade de uma amostra de tamanho 200 resultar a mesma proporção amostral entre 0.40 and 0.45?

- i.e.: **se $P = 0.4$ e $n = 200$, o que é**

  $P(0.40 \leq \hat{p} \leq 0.45)$?

Exemplo

(cont.)

- se P = 0.4 e n = 200, o que é
- $P(0.40 \leq \hat{p} \leq 0.45)$?

Encontre $\sigma_{\hat{p}}$:

$$\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$$

Converta para normal padrão:

$$P(.40 \leq \hat{p} \leq .45) = P\left(\frac{.40-.40}{.03464} \leq Z \leq \frac{.45-.40}{.03464}\right)$$
$$= P(0 \leq Z \leq 1.44)$$

Exemplo

(cont.)

- **se P = 0.4 e n = 200, o que é**
- **$P(0.40 \le \hat{p} \le 0.45)$?**

Use a tabela da Normal Padrão: $P(0 \le Z \le 1.44) =$ $\boxed{0.4251}$