Capítulo 5

Distribuições de Probabilidades Contínuas

<u>Teorema Central do Limite</u>

Texto baseado no livro:

Estatística Aplicada - Larson / Farber – Editora Pearson – 2010

Seção 5.4: Distribuições amostrais e o Teorema do Limite Central

Objetivos:

Encontrar distribuições amostrais e verificar suas propriedades

Interpretar o Teorema do Limite Central

Aplicar o Teorema do Limite Central para encontrar a probabilidade de uma média da amostra

Distribuições amostrais

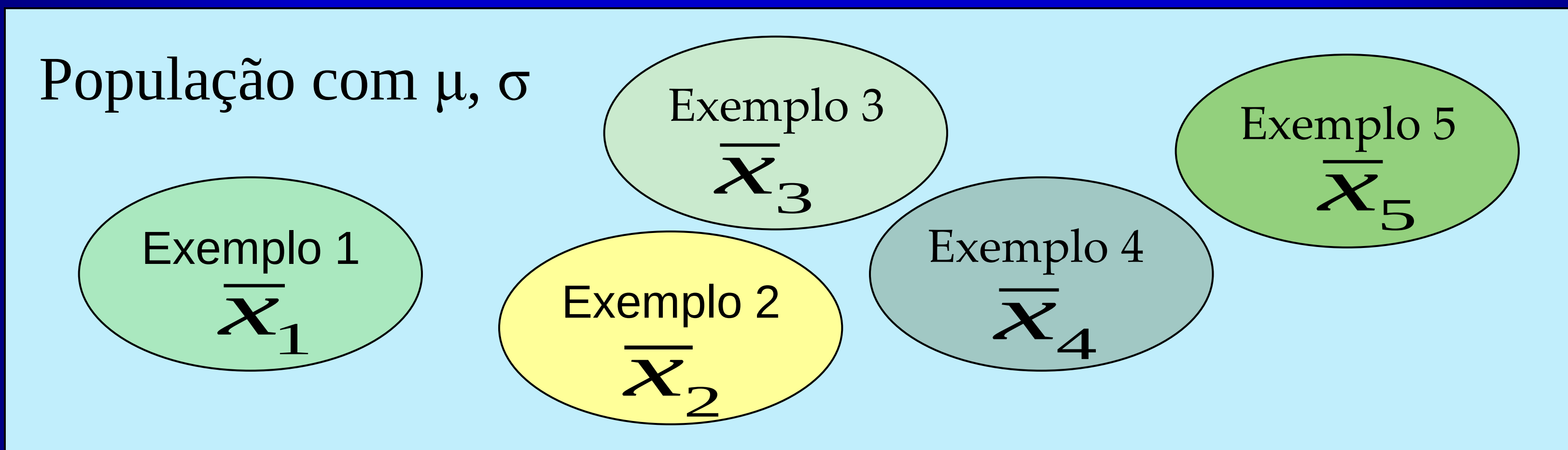
Distribuições amostrais

A distribuição de probabilidades de uma estatística de amostragem

São Formadas quando amostras de tamanho *n* são repetidamente tomadas de uma população

Ex.: distribuição de amostras de médias amostrais

Distribuições de amostras de médias amostrais



A distribuição da amostragem consiste dos valores das médias amostrais, $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$, $\bar{x}_5$, . . .

Propriedades de distribuições de amostras de médias amostrais

1. A média das médias amostrais, $\mu_{\bar{x}}$, é igual à média populacional μ.

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

2. O desvio padrão das médias amostrais, $\sigma_{\bar{x}}$, é igual ao desvio padrão da população σ dividido pela raiz quadrada do tamanho da amostragem n.

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

Também Chamado de **erro padrão da média**

Exemplo: distribuições de amostras de médias amostrais

Os valores populacionais {1, 3, 5, 7} são escritos em pedaços de papel e postos em uma caixa. Dois pedaços de papel são aleatoriamente selecionados, sendo recolocados na caixa após cada seleção.

a) Encontre a média, a variancia e o desvio padrão da população.

Solução:

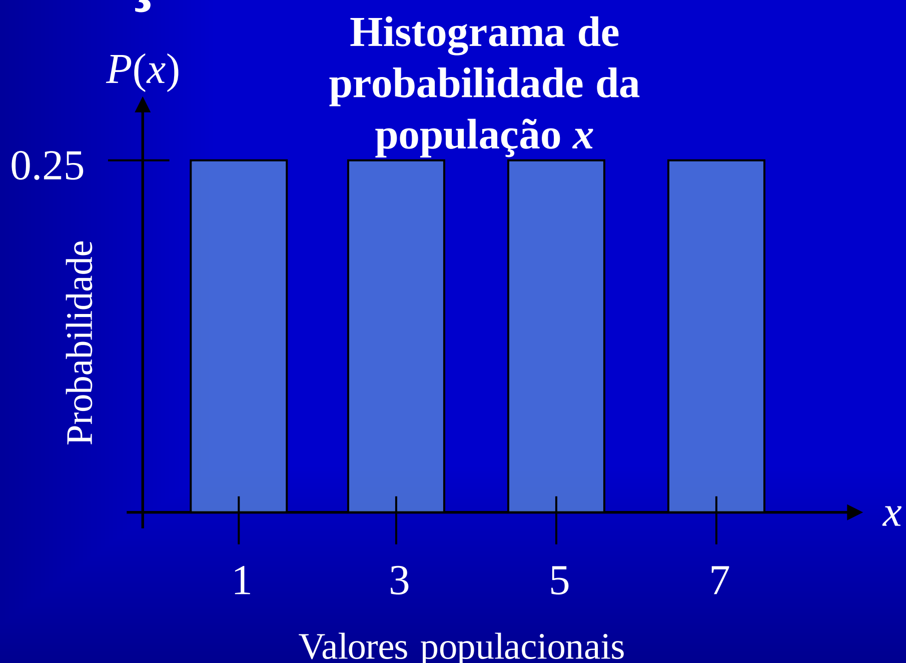
Média: $$\mu = \frac{\Sigma x}{N} = 4$$

Variância: $$\sigma^2 = \frac{\Sigma (x-\mu)^2}{N} = 5$$

Desvio padrã0: $$\sigma = \sqrt{5} \approx 2.236$$

b) Faça o gráfico do histograma de probabilidade dos valores populacionais.

Solução:



Todos os valores têm a mesma probabilidade de serem selecionados (distribuição uniforme)

c) Liste todas as amostragens possíveis de tamanho $n = 2$ e calcule a média de cada amostragem.

Solução:

Amostragem	$\bar{x}$
1, 1	1
1, 3	2
1, 5	3
1, 7	4
3, 1	2
3, 3	3
3, 5	4
3, 7	5

Amostragem	$\bar{x}$
5, 1	3
5, 3	4
5, 5	5
5, 7	6
7, 1	4
7, 3	5
7, 5	6
7, 7	7

Essas médias formam a distribuição amostral das médias amostrais

d) Construa a distribuição de probabilidade das médias amostrais.

Solução:

$\overline{x}$	f	Probabilidade
1	1	0,0625
2	2	0,1250
3	3	0,1875
4	4	0,2500
5	3	0,1875
6	2	0,1250
7	1	0,0625

e) Encontre a média, a variância e o desvio padrão da distribuição amostral das médias amostrais.

Solução:

A média, a variância, e o desvio padrão de 16 amostras são:

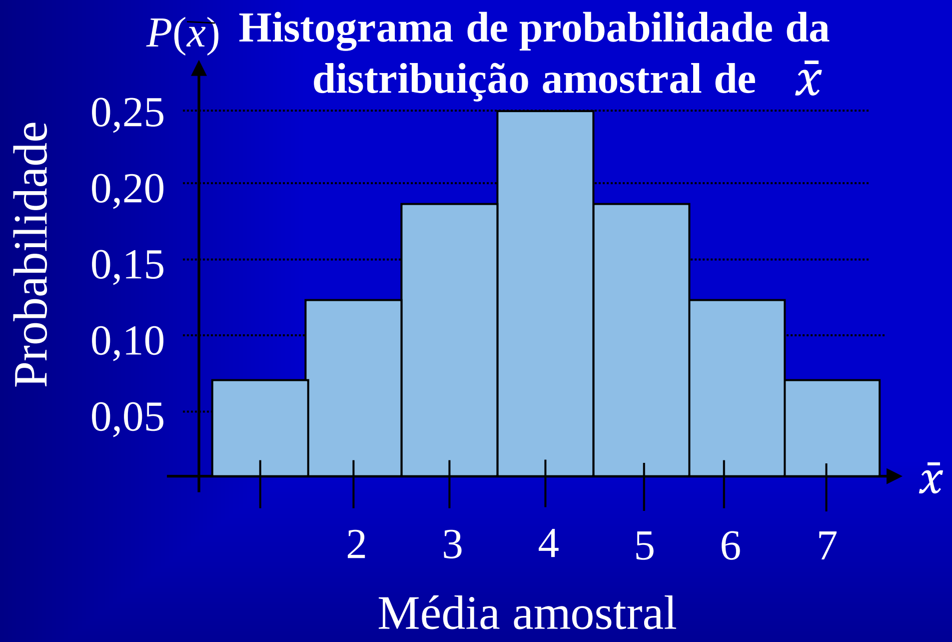
$$\mu_{\bar{x}} = 4 \qquad \sigma_{\bar{x}}^2 = \frac{5}{2} = 2.5 \qquad \sigma_{\bar{x}} = \sqrt{2.5} \approx 1.581$$

Esses resultados satisfazem as propriedades de distribuições de amostras de médias amostrais.

$$\mu_{\bar{x}} = \mu = 4 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{5}}{\sqrt{2}} \approx \frac{2.236}{\sqrt{2}} \approx 1.581$$

f) Faça o gráfico do histograma de probabilidade das médias amostrais.

Solução:



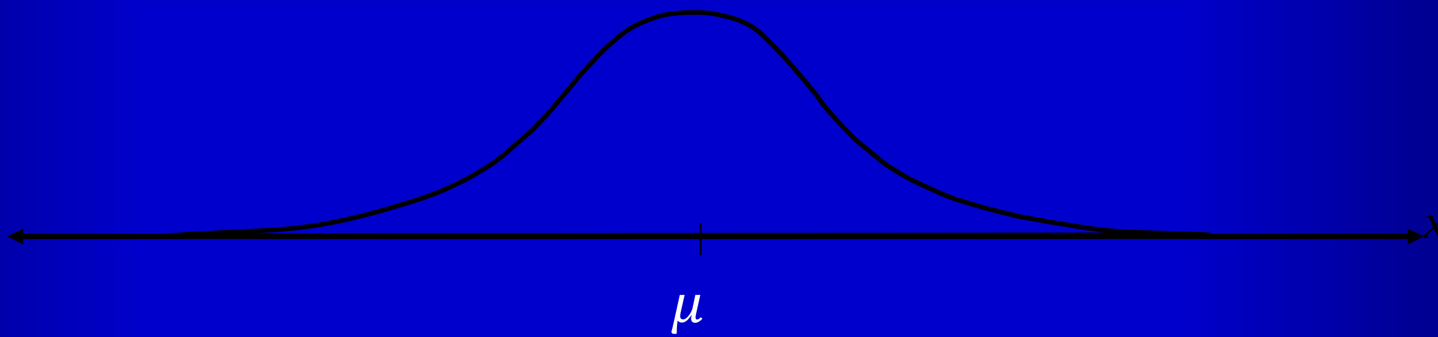
Note que gráfico é simétrico e em formato de sino aproximando-se de uma distribuição normal

O Teorema do Limite Central

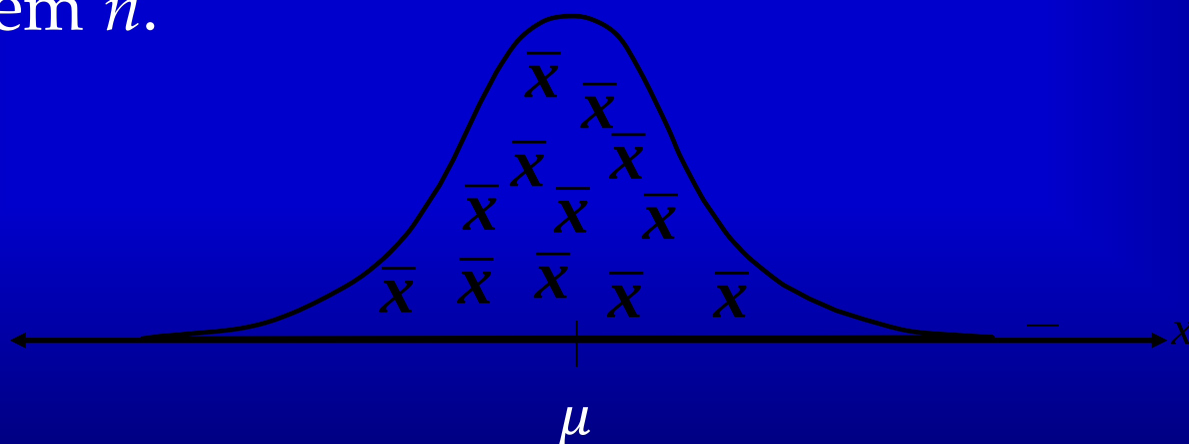
Se amostragens de tamanho $n \geq 30$ são tiradas de qualquer população de média = μ e desvio padrão = σ, então a distribuição amostral da média $\bar{x}$ se aproxima de uma distribuição normal com media $\mu\ e\ \ variância\ \frac{\sigma^2}{n}$

E quanto maior for o tamanho da amostra, melhor será a aproximação

2. Se a própria população é normalmente distribuída,



a distribuição de amostras das médias amostrais é normalmente distribuída para ***qualquer*** tamanho de amostragem n.



Em ambos os casos, a distribuição de amostras de médias amostrais tem uma média igual à média da população.

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

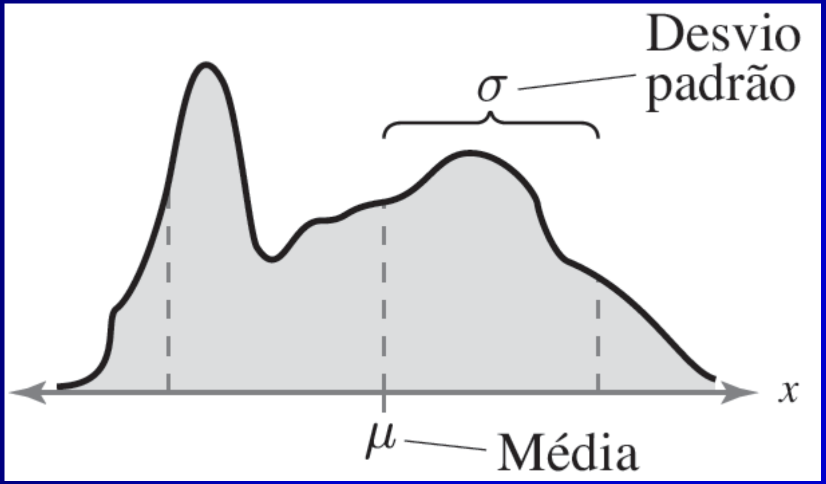
A distribuição da amostra de médias amostrais tem uma variância igual a 1/n vez a variância da população e um desvio padrão igual ao desvio padrão da população dividido pela raiz quadrada de n.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$ Variância

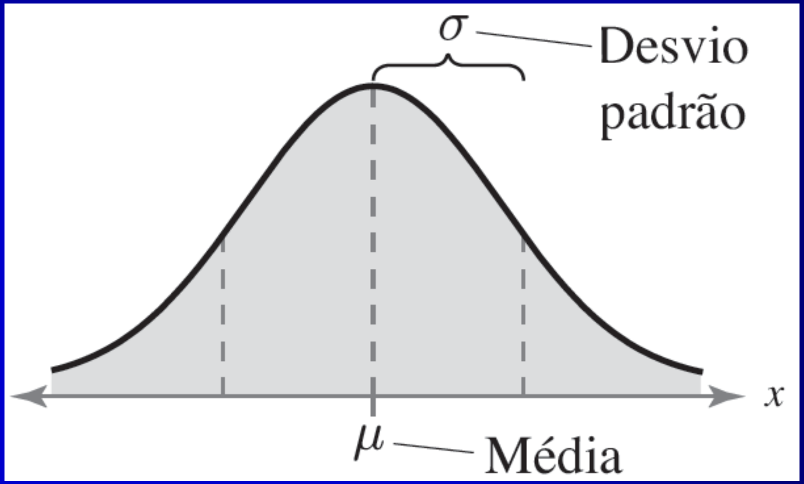
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$ Desvio padrão (**erro padrão da média**)

Em linguagem comum, o teorema do limite central expressa o fato de que a soma de muitas variáveis aleatórias independentes e com mesma distribuição de probabilidade tendem à distribuição normal ("cada vez mais normal à medida que o número de variáveis aumenta")

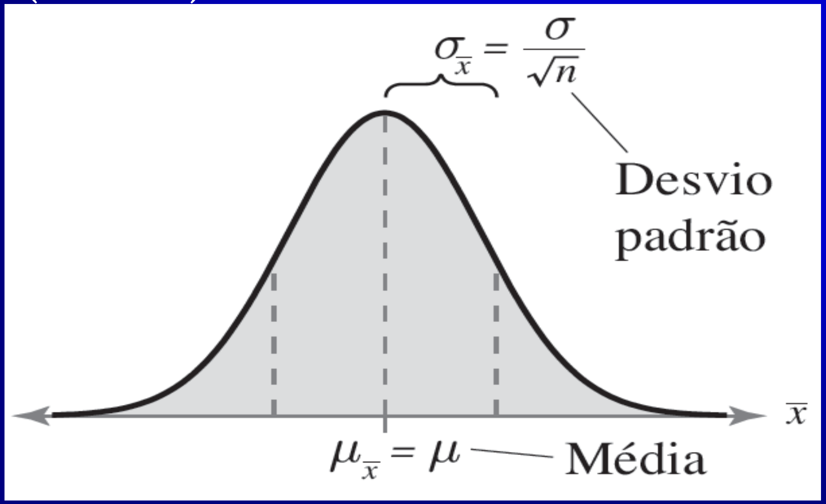
1. Distribuição populacional qualquer



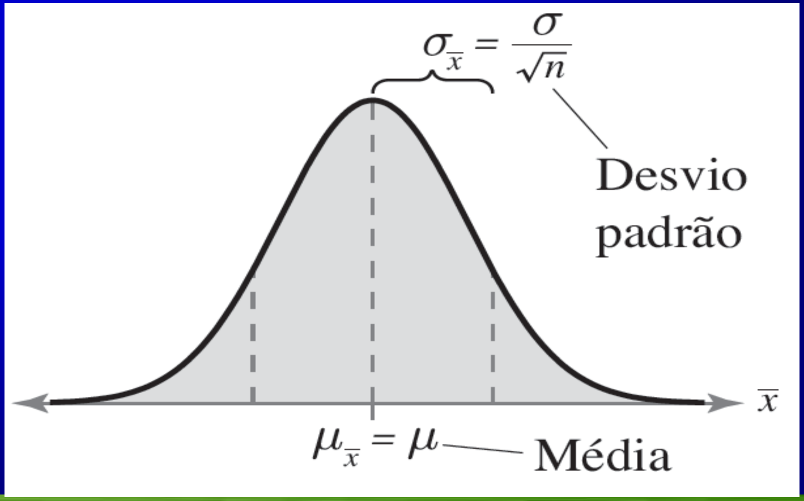
2. Distribuição populacional normal



Distribuição das médias amostrais
($n \geq 30$)



Distribuição das médias amostrais
(n qualquer)



Exemplo: interpretando o Teorema do Limite Central

As contas dos telefones dos habitantes de uma cidade têm uma média de R$ 64 e um desvio padrão de R$ 9. Amostragens aleatórias de 36 contas de telefone são tiradas dessa população e a média de cada amostragem é determinada. Encontre a média e o erro padrão da média da distribuição amostral. Então esboce um gráfico da distribuição amostral das médias amostrais.

Solução: interpretando o Teorema do Limite Central

A média da distribuição de amostras é igual à média da população

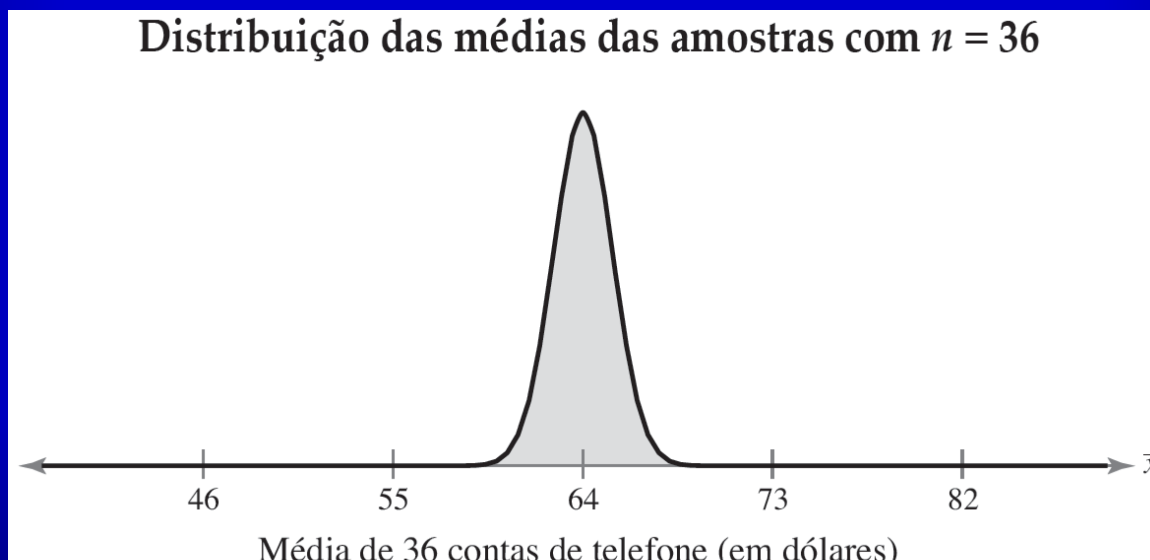
$$\mu_{\bar{x}} = \mu = 64$$

O erro padrão da média é igual ao desvio padrão populacional dividido pela raiz quadrada de *n*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{36}} = 1.5$$

Já que o tamanho da amostragem é maior que 30, a distribuição das amostras pode ser aproximada por uma distribuição normal com

$$\mu_{\bar{x}} = 64 \qquad \sigma_{\bar{x}} = 1.5$$



Exemplo: interpretando o Teorema do Limite Central

As alturas das árvores de carvalho branco adultas são normalmente distribuídas, com uma média de 90 pés ($\approx 27\ m$) e um desvio padrão de 3,5 pés ($\approx 1{,}05\ m$).

Amostras aleatórias de tamanho 4 são tiradas dessa população, e a média de cada amostra é determinada. Encontre a média e o erro padrão da média da distribuição amostral. Então esboce um gráfico da distribuição amostral das médias amostrais.

Solução: interpretando o Teorema do Limite Central

A média da distribuição amostral é igual à média populacional

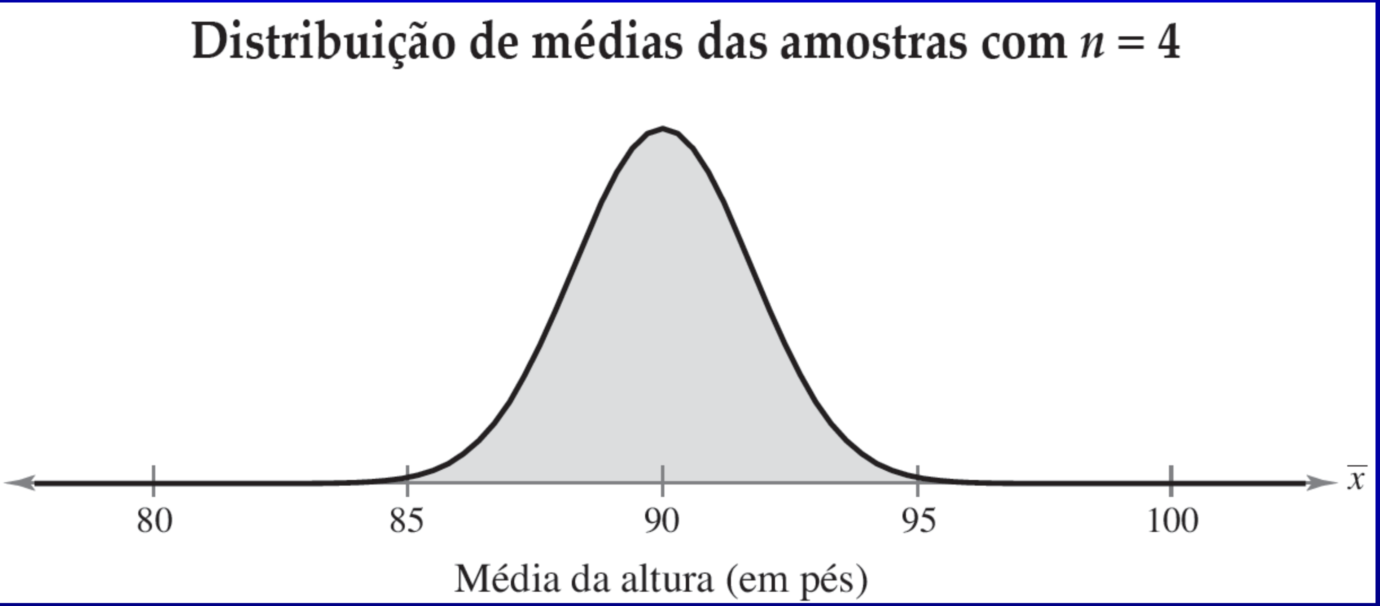
$$\mu_{\bar{x}} = \mu = 90$$

O erro padrão da média é igual ao desvio padrão da população dividido pela raiz quadrada de n.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{4}} = 1.75$$

Já que a população é normalmente distribuída, a distribuição amostral da média amostral também é normalmente distribuída.

$$\mu_{\bar{x}} = 90 \qquad \sigma_{\bar{x}} = 1.75$$



Probabilidade e o Teorema do Limite Central

Para transformer $\bar{x}$ em um escore z

$$z = \frac{\text{Valor} - \text{Média}}{\text{Desvio padrão}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}.$$

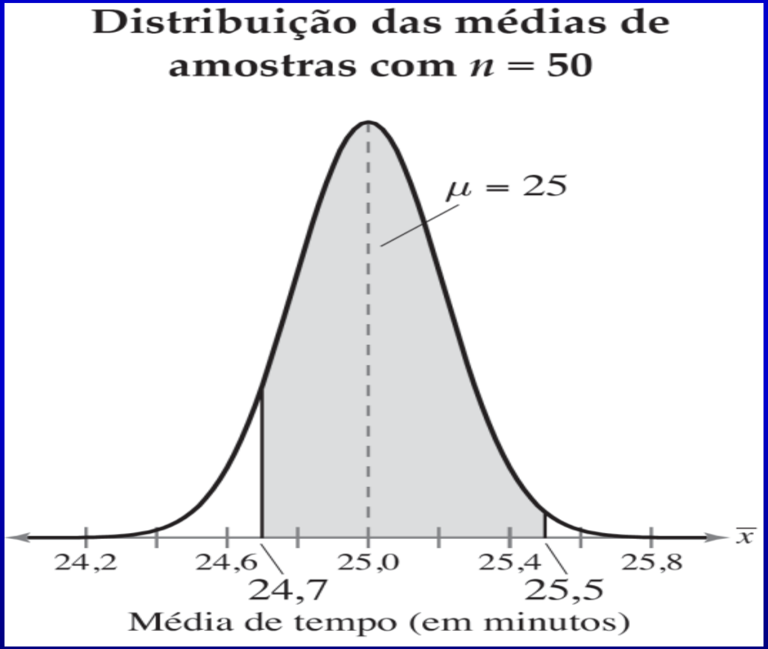
Exemplo 1: probabilidades para distribuições amostrais

O gráfico mostra o tempo gasto pelas pessoas dirigindo a cada dia. Você seleciona aleatoriamente 50 motoristas de 15 até 19 anos. Qual é a probabilidade de que o tempo médio que eles gastem dirigindo diariamente esteja entre 24,7 e 25,5 minutos? Assuma que σ = 1,5 minutos.

Solução: probabilidades para distribuições amostrais

A partir do Teorema do Limite Central (tamanho de amostragem é maior que 30), a distribuição amostral das médias amostrais é aproximadamente normal com

$$\mu_{\bar{x}} = \mu = 25 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{50}} \approx 0.21213$$



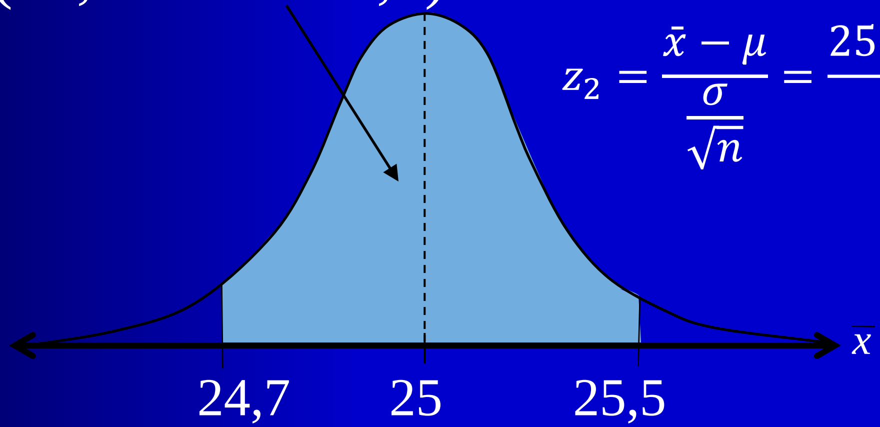
Distribuição normal
$\mu = 25 \quad \sigma = 0{,}21213$

Distribuição normal padrão
$\mu = 0$ e $\sigma = 1$

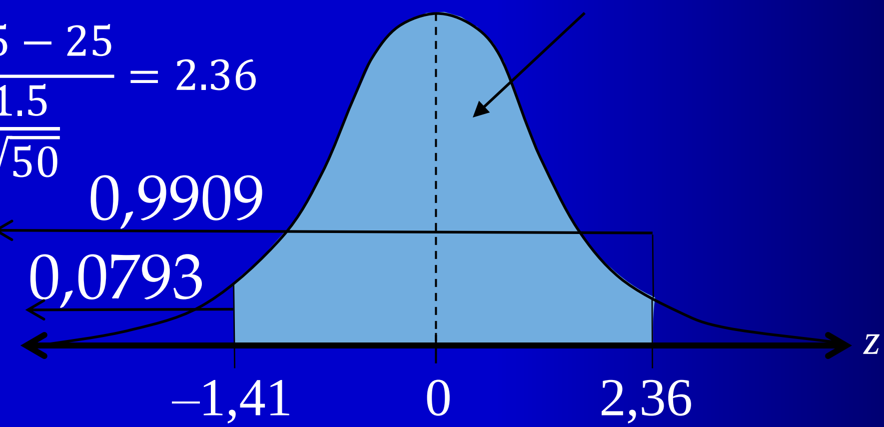
$$z_1 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{24.7 - 25}{\frac{1.5}{\sqrt{50}}} = -1.41$$

$$z_2 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{25.5 - 25}{\frac{1.5}{\sqrt{50}}} = 2.36$$

$P(24{,}7 < \bar{x} < 25{,}5)$



$P(-1{,}41 < z < 2{,}36)$



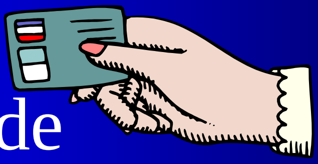
$P(24 < \bar{x} < 54) = P(-1{,}41 < z < 2{,}36)$
$= 0{,}9909 - 0{,}0793 = \mathbf{0{,}9116}$

Exemplo 2:

Um auditor de um banco afirma que os balanços dos cartões de crédito são normalmente distribuídos com uma média de R$ 2.870 e um desvio padrão de R$ 900.

1. Qual é a probabilidade de que um portador de cartão de crédito aleatoriamente selecionado tenha um balanço menor que R$ 2.500?

Solução:



Foi pedido que encontrássemos a probabilidade associada com um certo valor da variável aleatória *x*.

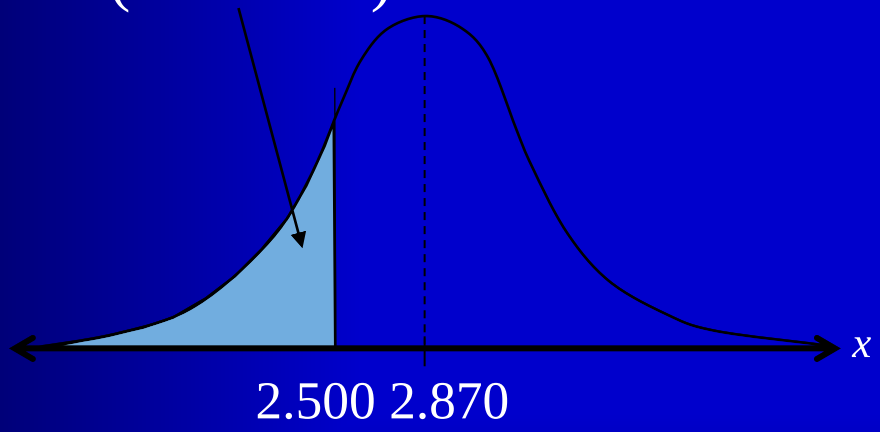
Solução: probabilidades para x e $\bar{x}$

Distribuição normal
$\mu = 2.870 \quad \sigma = 900$

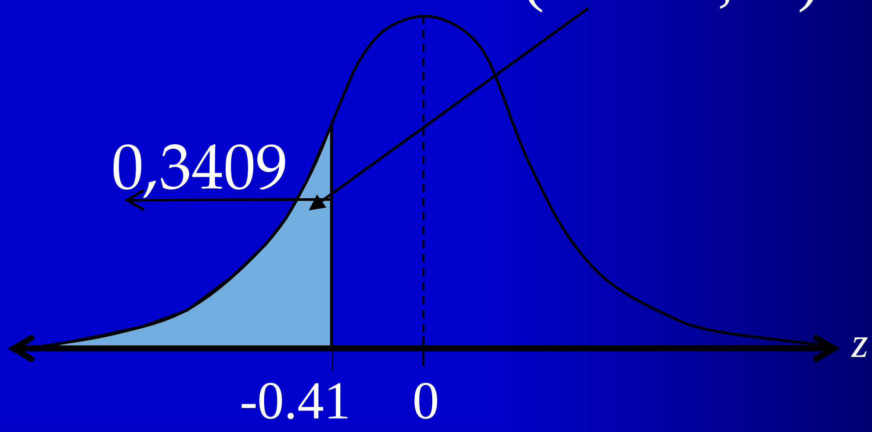
Distribuição normal padrão
$\mu = 0 \quad \sigma = 1$

$$z = \frac{x - \mu}{\sigma} = \frac{2500 - 2870}{900} \approx -0.41$$

$P(x < 2.500)$



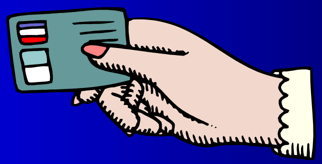
$P(z < -0{,}41)$



$P(x < 2.500) = P(z < -0{,}41) = \mathbf{0{,}3409}$

Exemplo: probabilidades para x e $\bar{x}$

2. Você seleciona aleatoriamente 25 portadores de cartão de crédito. Qual é a probabilidade de que a média dos balanços dos seus cartões de crédito seja menor que R$ 2.500?



Solução:

Foi pedido que encontrássemos a probabilidade associada com uma média amostral $\bar{x}$.

$$\mu_{\bar{x}} = \mu = 2870 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{900}{\sqrt{25}} = 180$$

Exemplo: probabilidades para x e $\bar{x}$

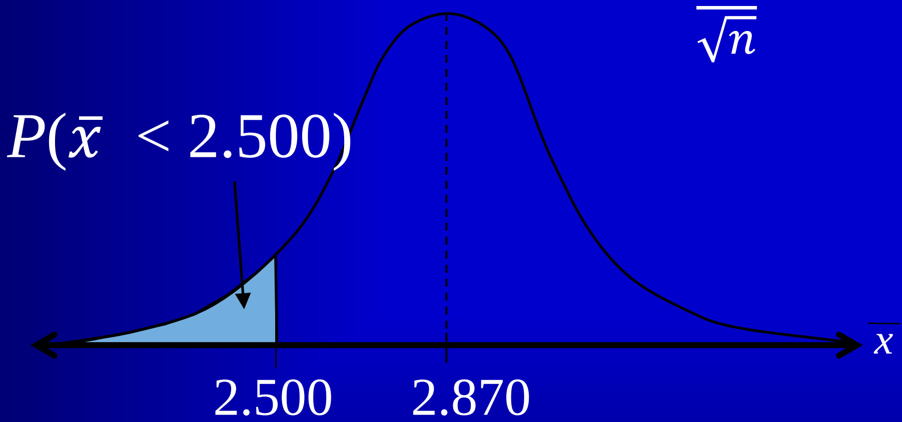
Distribuição normal

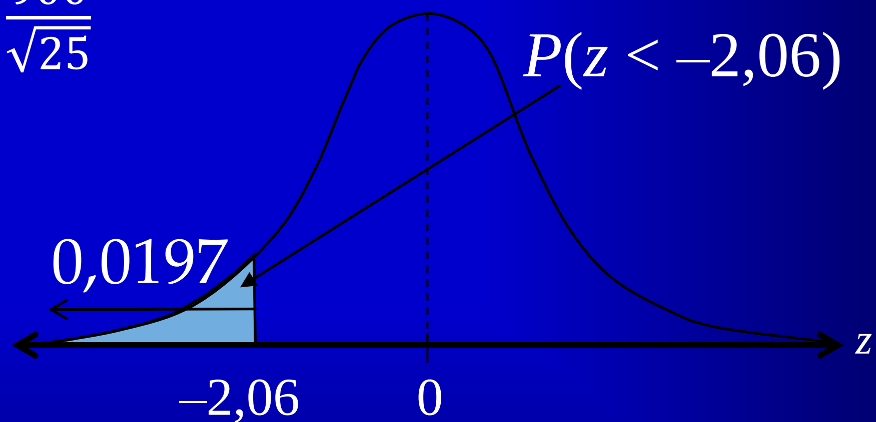
$\mu = 2.870 \quad \sigma = 180$

Distribuição normal padrão

$\mu = 0 \quad \sigma = 1$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2500 - 2870}{\frac{900}{\sqrt{25}}} \approx -2.06$$





$P(\bar{x} < 2.500) = P(z < -2{,}06) = \mathbf{0{,}0197}$

Seção 5.5: Aproximações normais de distribuições binomiais

Objetivos da Seção 5.5

Determinar quando a distribuição normal pode se aproximar da distribuição binomial

Encontrar a correção pela continuidade

Usar a distribuição normal para aproximar probabilidades binomiais

Aproximação normal para uma distribuição binomial

A distribuição normal é usada para aproximar a distribuição binomial quando não seria prático usar a distribuição binomial para encontrar uma probabilidade

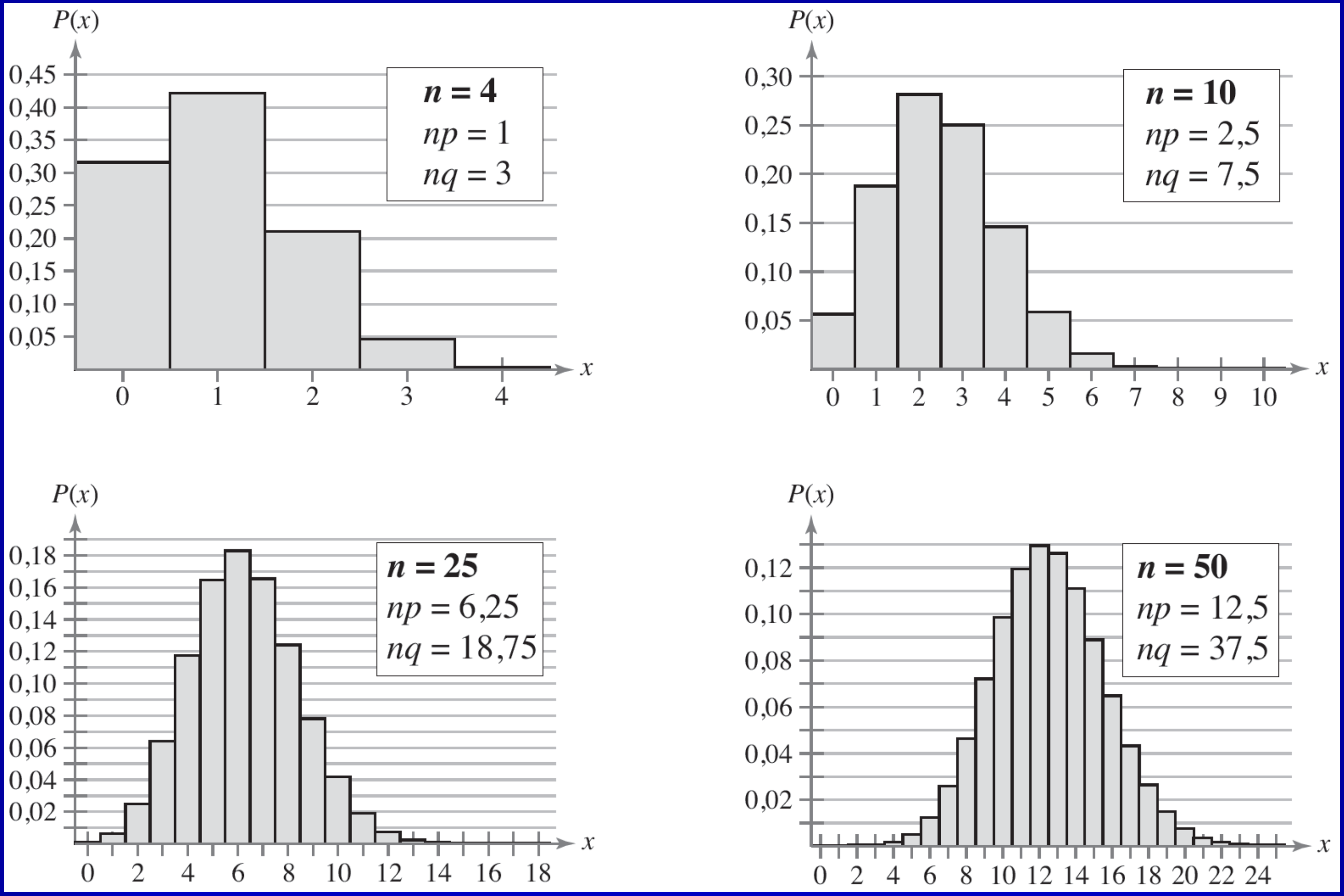
Aproximação normal para uma distribuição binomial

Em geral, se $np \geq 5$ e $nq \geq 5$, então a variável aleatória binomial x é aproximadamente distribuída normalmente com

média $\mu = np$

desvio padrão $\sigma = \sqrt{npq}$

Distribuição binomial: $p = 0{,}25$

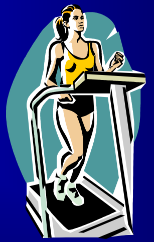


Conforme *n* aumenta, o histograma se aproxima de uma curva normal

Exemplo: aproximando a distribuição binomial

Decida se você pode usar a distribuição normal para aproximar *x*, sendo ele o número de pessoas que responderam sim. Se puder, encontre a média e o desvio padrão.

1. Cinquenta e um por cento dos adultos no Brasil cuja decisão para o ano novo era se exercitar mais alcançaram esse objetivo. Você seleciona aleatoriamente 65 adultos no Brasil cuja resolução era se exercitar mais e pergunta se alcançaram esse objetivo.



Solução: aproximando a distribuição binomial

Você pode usar a aproximação normal

$n = 65$, $p = 0{,}51$, $q = 0{,}49$

$np = (65)(0{,}51) = 33{,}15 \geq 5$

$nq = (65)(0{,}49) = 31{,}85 \geq 5$

Média: $\mu = np = 33{,}15$

Desvio padrão: $\sigma = \sqrt{npq} = \sqrt{65 \cdot 0.51 \cdot 0.49} \approx 4.03$

Exemplo: aproximando a distribuição binomial

2. Quinze por cento dos adultos no Brasil não traçam objetivos de ano novo. Você seleciona aleatoriamente 15 adultos e pergunta a cada um se traçou um objetivo de ano novo.

Você não pode usar a aproximação normal, pois:

$n = 15, p = 0{,}15, q = 0{,}85$

$np = (15)(0{,}15) = 2{,}25 < 5$

$nq = (15)(0{,}85) = 12{,}75 \geq 5$

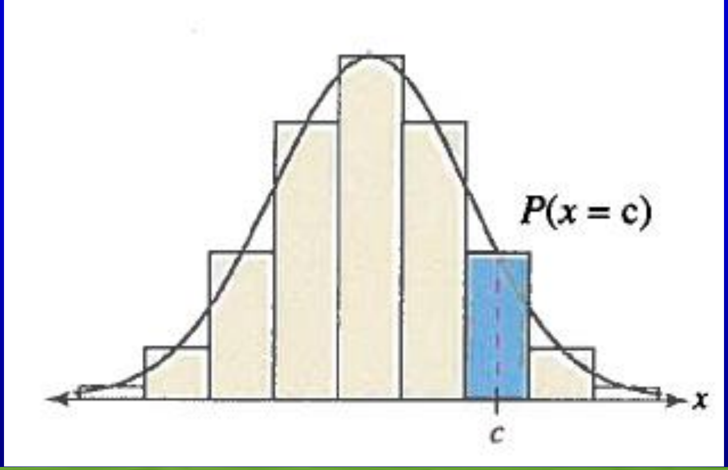
Porque $np < 5$, você não deve usar a distribuição normal para aproximar a distribuição.

Correção pela continuidade

A distribuição binomial é discreta e pode ser representada pelo histograma de probabilidade

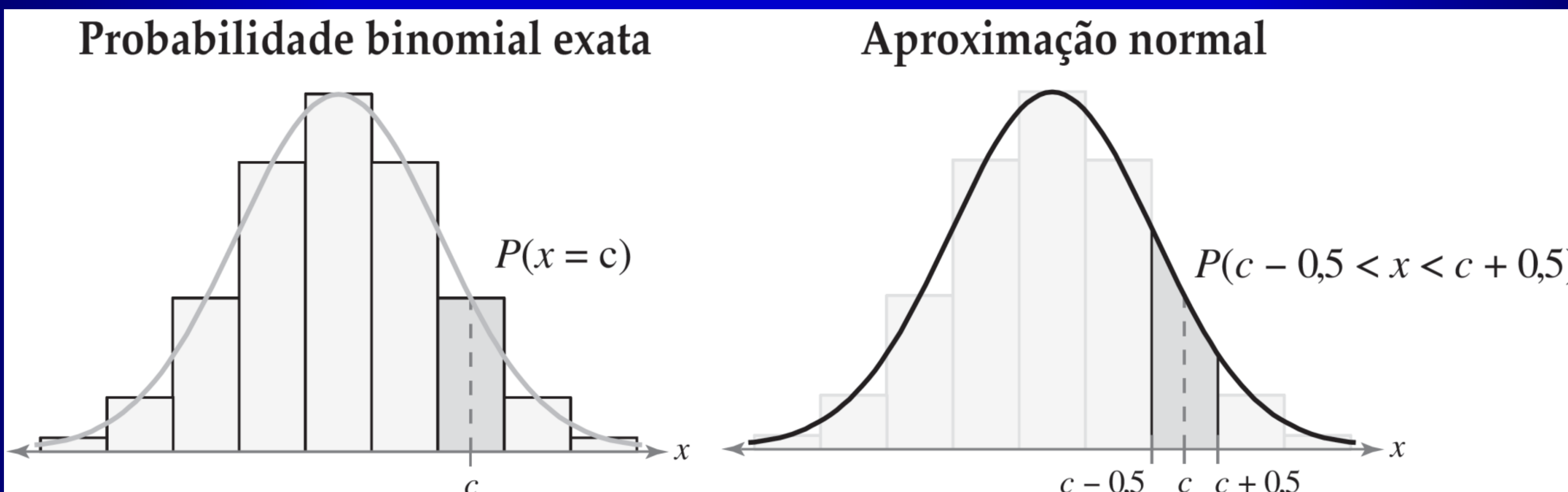
Para calcular as probabilidades binomiais *exatas*, a fórmula binomial é usada para cada valor de x e os resultados são somados

Isso corresponde geometricamente a somar as áreas das barras no histograma de probabilidade



Quando você usa uma distribuição normal que é *contínua* para aproximar uma distribuição binomial que é discreta, você precisa mover 0,5 unidade para a esquerda e para a direita do ponto médio para incluir todos os possíveis valores x no intervalo (**correção pela continuidade usada para melhorar a aproximação**). Ou seja, faremos o seguinte:

$$P(X = K) \approx (K - \frac{1}{2} \leq X \leq K + \frac{1}{2})$$



Exemplo: usando uma correção pela continuidade

Use uma correção pela continuidade para converter os intervalos binomiais no intervalo de distribuição normal.

1. A probabilidade de se obter entre 270 e 310 sucessos.

Solução: Os valores dos pontos médios discretos são: 270, 271, …, 310

O intervalo correspondente para a distribuição normal contínua é $269{,}5 < x < 310{,}5$

Use uma correção pela continuidade para converter os intervalos binomiais no intervalo de distribuição normal.

A probabilidade de se obter pelo menos 158 sucessos.

Solução:
Os valores dos pontos médios discretos são:
158, 159, 160, ….

O intervalo correspondente para a distribuição normal contínua é $x > 157{,}5$

Use uma correção pela continuidade para converter os intervalos binomiais no intervalo de distribuição normal.

3. A probabilidade de se obter menos de 63 sucessos.

Solução:

Os valores dos pontos médios discretos são:
… , 60 , 61 , 62 , 63
O intervalo correspondente para a distribuição contínua normal é $x < 62{,}5$

Usando a distribuição normal para aproximar probabilidades binomiais

Em palavras	***Em símbolos***
1. Verifique se a distribuição binomial se aplica.	Especifique n, p e q.
2. Determine se você pode usar a distribuição binomial para aproximar *x,* a variável binomial.	np é ≥ 5? nq é ≥ 5?
3. Encontre a média μ e o desvio padrão σ para a distribuição.	$\mu = np$ $\sigma = \sqrt{npq}$

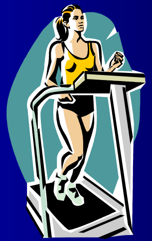
Em palavras	Em símbolos
4. Aplique a correção pela continuidade apropriada. Sombreie a área correspondente sob a curva.	Some ou subtraia 0,5 dos pontos finais.
5. Encontre o(s) escore(s) z correspondente(s).	$z = \dfrac{x - \mu}{\sigma}$
6. Encontre a probabilidade.	Use a tabela normal padrão.

Exemplo: aproximando a probabilidade binomial

Cinquenta e um por cento dos adultos no Brasil cuja resolução de ano novo era se exercitar mais alcançaram essa resolução. Você seleciona aleatoriamente 65 adultos cuja resolução era se exercitar mais e pergunta se eles alcançaram essa resolução. Qual é a probabilidade de que menos de 40 deles responda que sim?

Solução:

Pode usar a aproximação normal



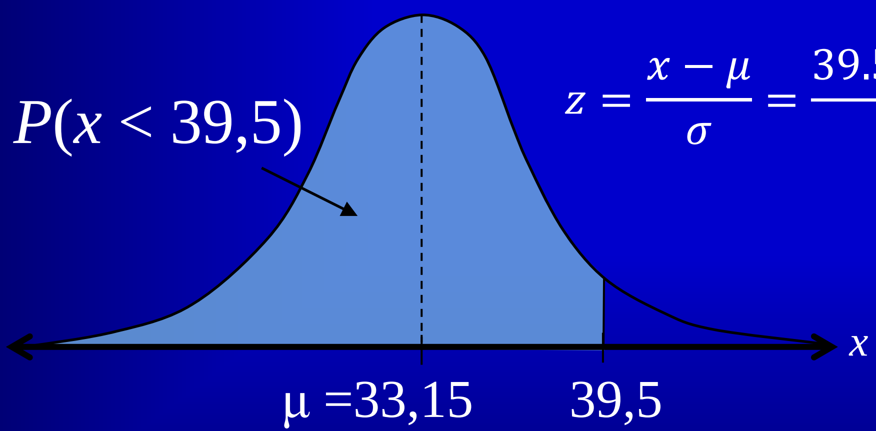
$\mu = 65.0{,}51 = 33{,}15$ $\quad$ $\sigma = \sqrt{65 \cdot 0.51 \cdot 0.49} \approx 4.03$

Solução: aproximando a probabilidade binomial

Aplique a correção pela continuidade:

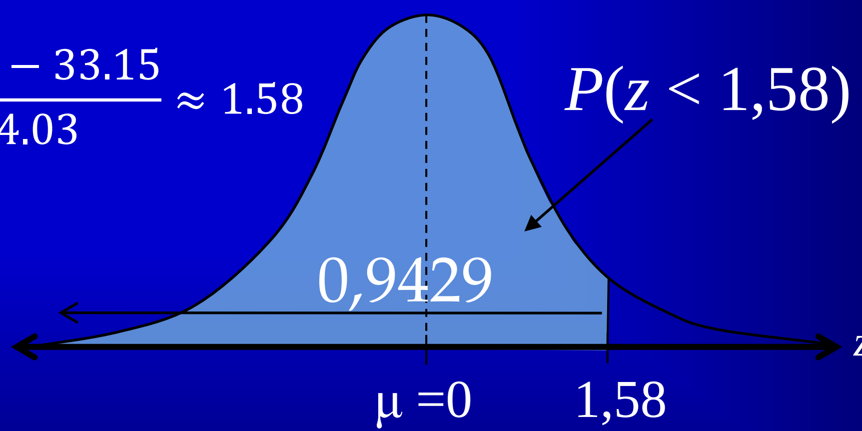
Menos de 40 (…37, 38, 39) corresponde ao intervalo de distribuição contínua normal ***x* < 39,5**

Distribuição normal
$\mu = 33,15 \quad \sigma = 4,03$



$$z = \frac{x-\mu}{\sigma} = \frac{39.5 - 33.15}{4.03} \approx 1.58$$

Normal padrão
$\mu = 0 \quad \sigma = 1$



$P(\bar{x} < 39,5) = P(z < 1,58) = \mathbf{0,9429}$

Exemplo: aproximando a probabilidade binomial

Um pesquisa reporta que 86% dos usuários de internet utilizam o Internet Explorer do Windows como seu navegador. Você seleciona aleatoriamente 200 usuários de internet e pergunta se eles usam o Internet Explorer. Qual é a probabilidade de que exatamente 176 deles digam que sim?

Solução:

Pode usar a aproximação normal

$$np = (200)(0.86) = 172 \geq 5 \quad nq = (200)(0,14) = 28 \geq 5$$

$$\mu = 200 \cdot 0{,}86 = 172 \qquad \sigma = \sqrt{200 \cdot 0.86 \cdot 0.14} \approx 4.91$$

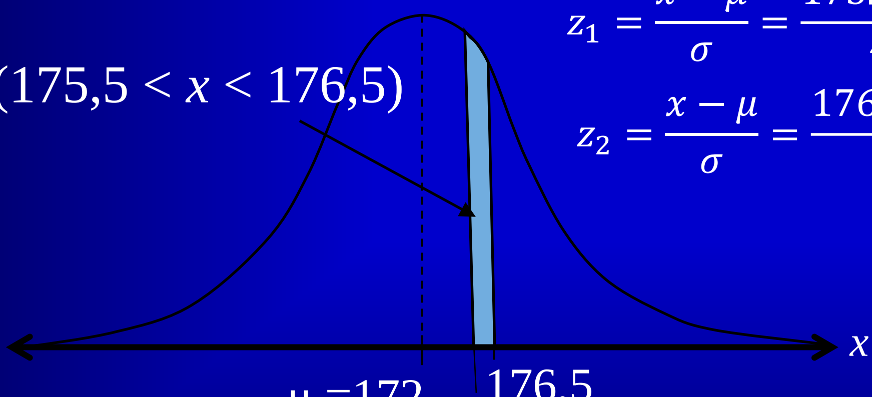
Solução: aproximando a probabilidade binomial

Aplique a correção pela continuidade:

Exatamente 176 corresponde ao intervalo de distribuição contínua normal **175,5 < *x* < 176,5**

Distribuição normal

μ = 172 e σ = 4,91

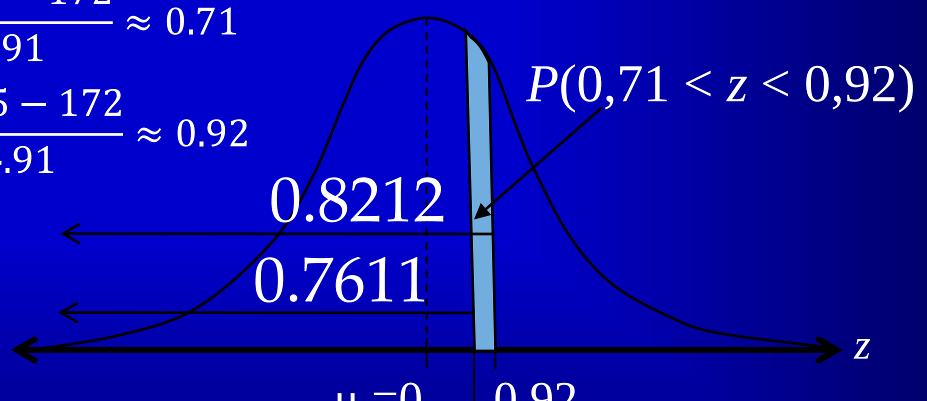


$$z_1 = \frac{x-\mu}{\sigma} = \frac{175.5 - 172}{4.91} \approx 0.71$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{176.5 - 172}{4.91} \approx 0.92$$

Normal padrão

μ = 0 e σ = 1



$P(175{,}5 < x < 176{,}5) = P(0{,}71 < z < 0{,}92) = 0{,}8212 - 0{,}7611 = \mathbf{0{,}0601}$

Exemplos

1. Seja X: N(100, 25). Calcular:

a) $P(100 \leq X \leq 106)$ **Resp: 0,384930**

b) $P(89 \leq X \leq 107)$ **Resp: 0,90534**

c) $P(112 \leq X \leq 116)$ **Resp: 0,007510**

d) $P(X \geq 108)$ **Resp: 0,054799**

2. Sendo X: N(50,16), determine X_a tal que:

a) $P(X \geq X_a) = 0{,}05$ **Resp: X_a = 56,56**

b) $P(X \leq X_a) = 0{,}99$ **Resp: X_a = 59,28**