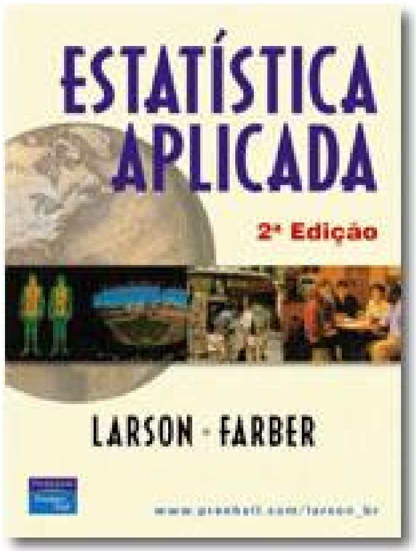
5 Distribuição normal de probabilidade

Estatística Aplicada

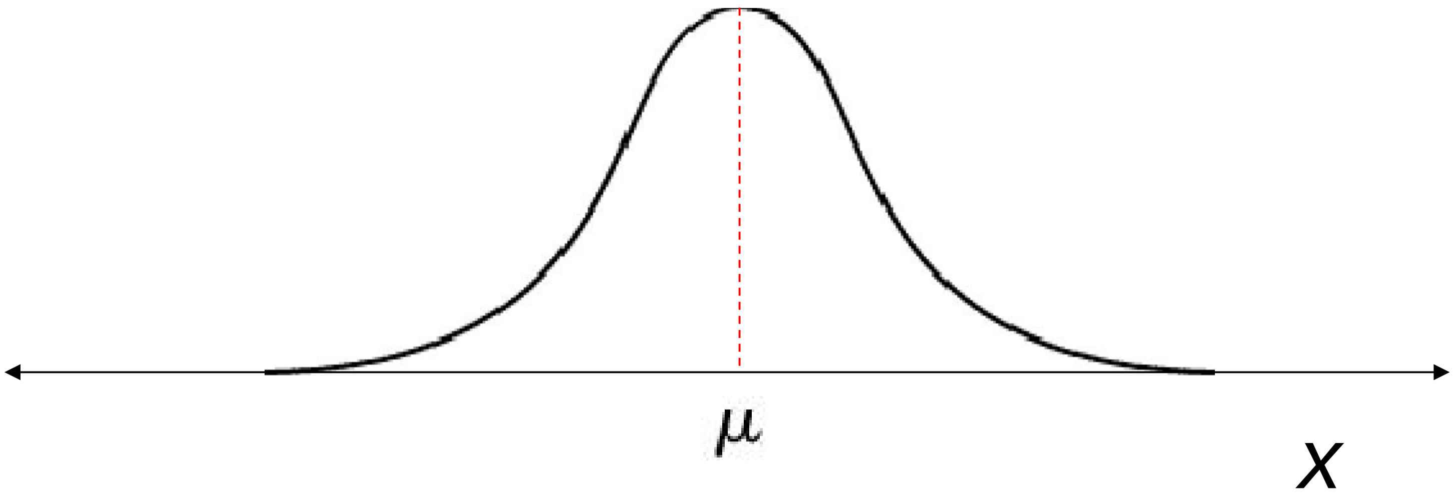
Larson Farber

Seção 5.1

Introdução às distribuições normais



Propriedades de uma distribuição normal

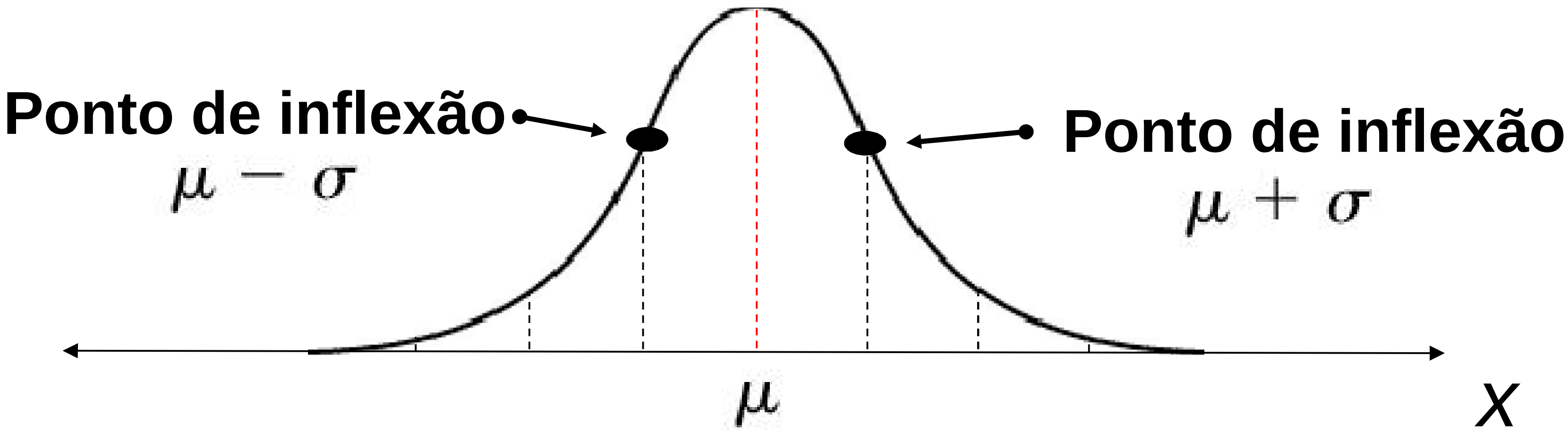


- Suas média, mediana e moda são iguais.

- Tem forma de sino e é simétrica em torno da média.

- A área total sob a curva é de 100%.

Propriedades de uma distribuição normal

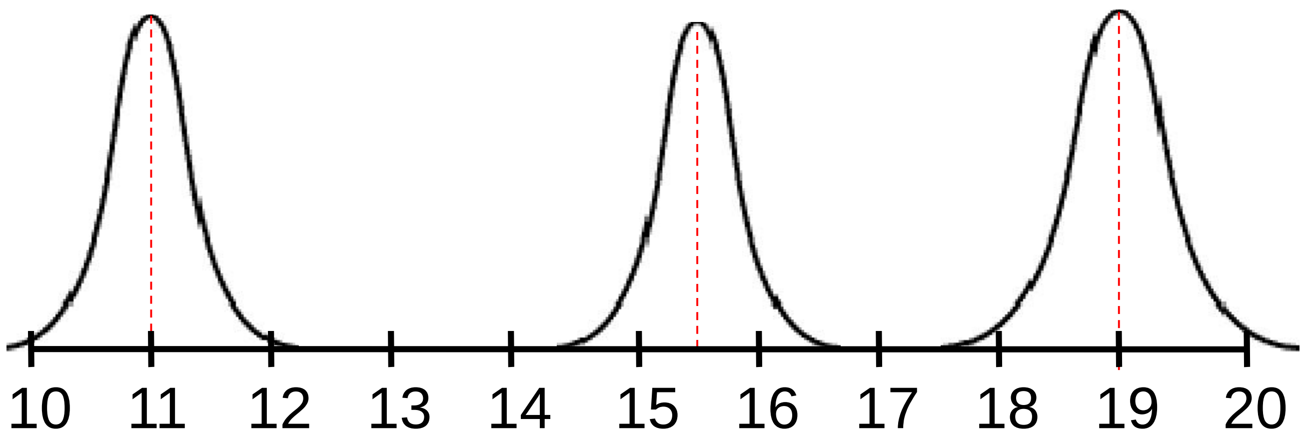


- À medida que a curva se afasta da média, aproxima-se cada vez mais do eixo *x*, mas nunca o toca.

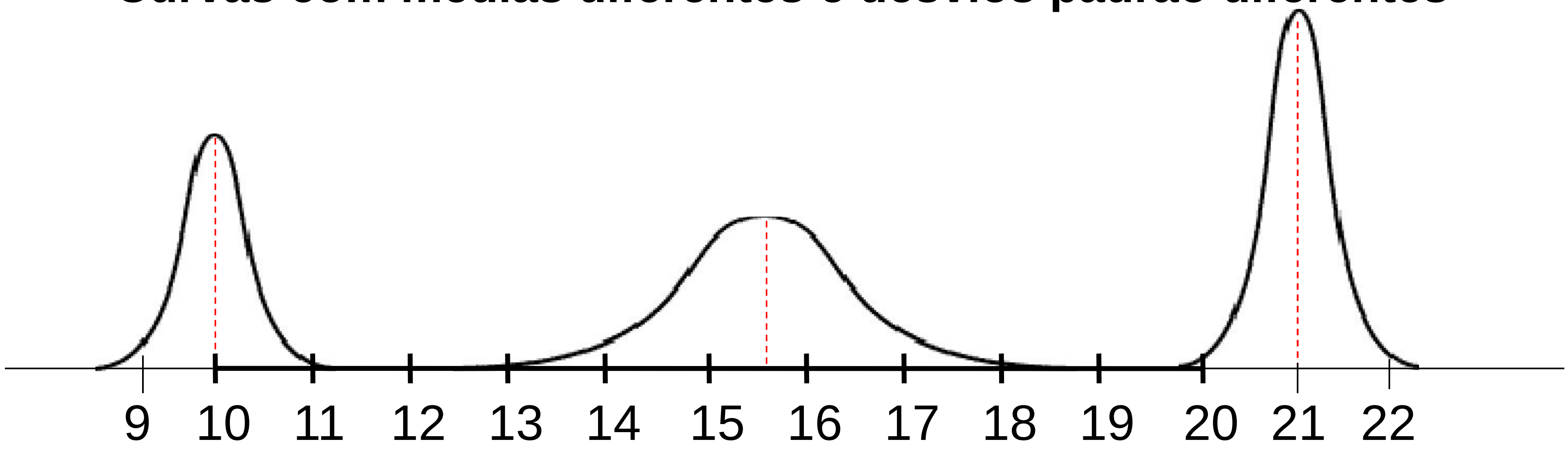
- Os pontos em que a curvatura muda são chamados pontos de inflexão. O gráfico curva-se para baixo entre os pontos de inflexão e, para cima, à esquerda e à direita deles.

Médias e desvios padrão

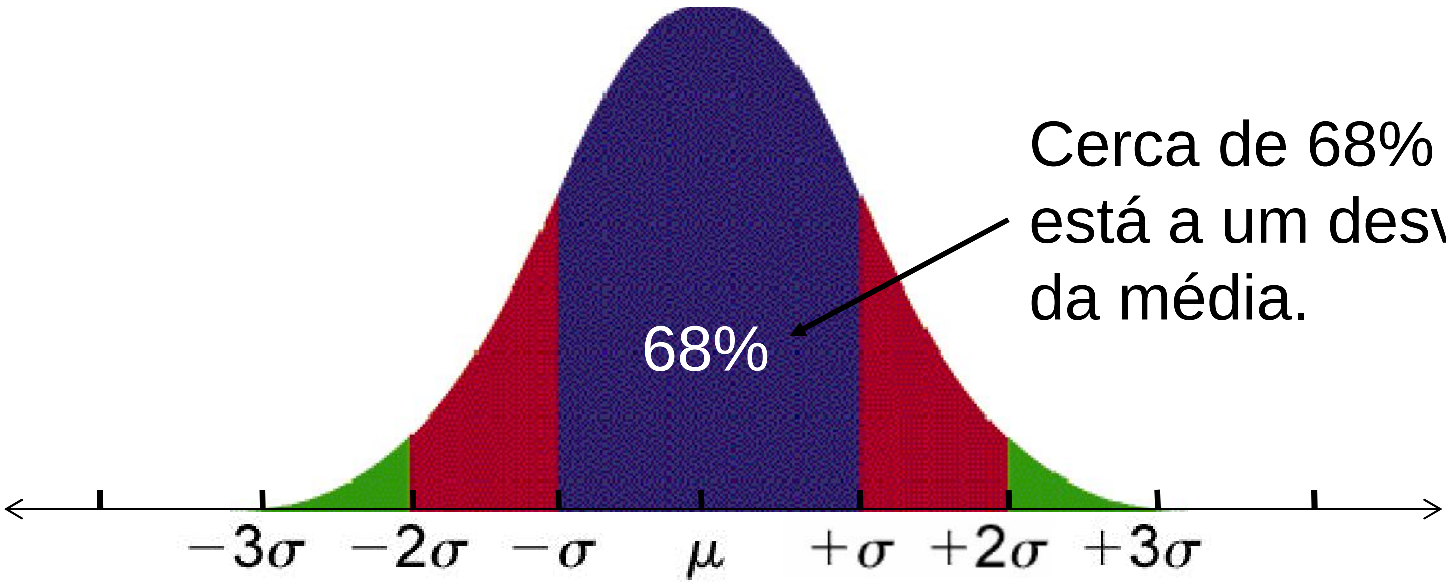
Curvas com médias diferentes e o mesmo desvio padrão



Curvas com médias diferentes e desvios padrão diferentes



Regra Empírica

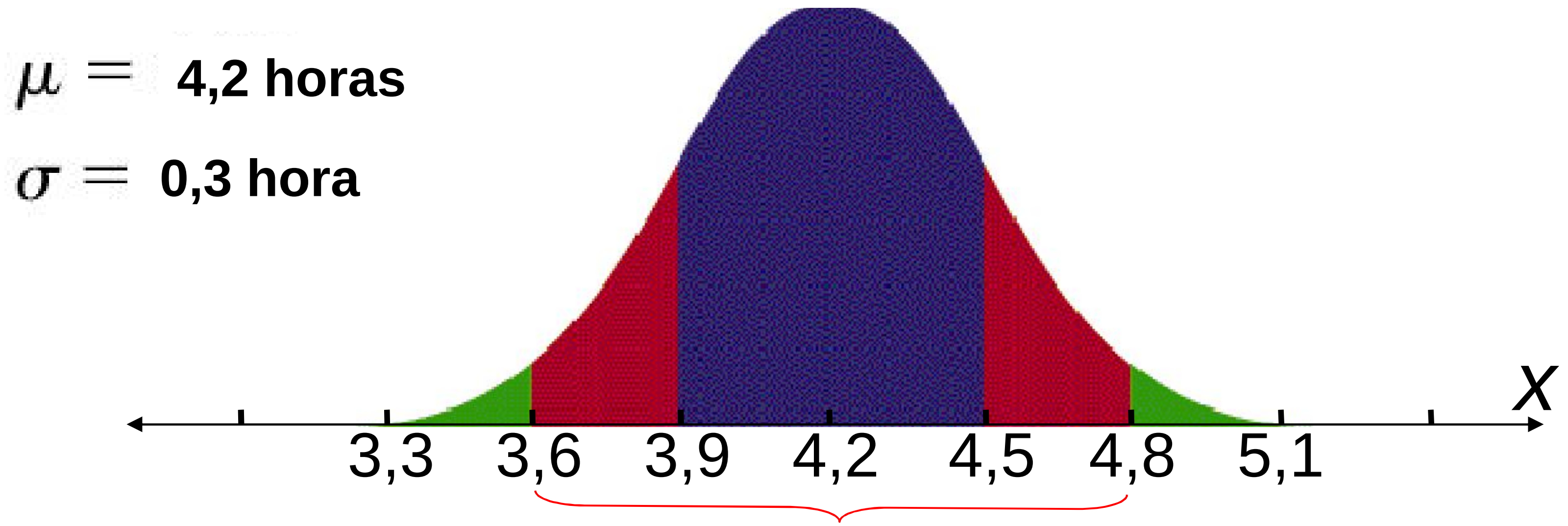


Cerca de 68% da área está a um desvio padrão da média.

Cerca de 95% da área está a dois desvios padrão.

Cerca de 99,7% da área está a três desvios padrão da média.

Como determinar intervalos



Segundo o manual de instruções, o tempo de montagem de certo produto é normalmente distribuído, com uma média de 4,2 horas e um desvio padrão de 0,3 hora. Determine o intervalo no qual caem 95% dos tempos de montagem.

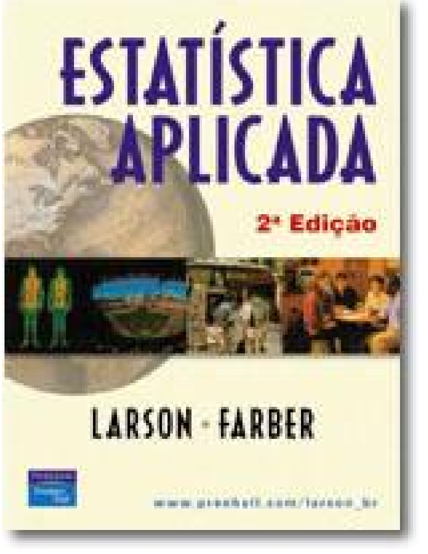
95% dos dados caem a até dois desvios padrão da média.

4,2 – 2 (0,3) = 3,6 e 4,2 + 2 (0,3) = 4,8.

95% dos tempos de montagem estarão entre 3,6 e 4,8 horas.

Seção 5.2

A distribuição normal padrão



O escore padrão

O **escore padrão**, ou **escore *z***, representa o número de desvios padrão que separa uma variável aleatória *x* da média.

$$z = \frac{\textbf{valor – média}}{\textbf{desvio padrão}} = \frac{x - \mu}{\sigma}$$

As pontuações em um concurso público estão normalmente distribuídas, com média de 152 e desvio padrão de 7. Determine o escore *z* para um candidato com pontuação de:

(a) 161 **(b) 148** **(c) 152**

(a)

$$z = \frac{161 - 152}{7}$$

$$z = \mathbf{1{,}29}$$

(b)

$$z = \frac{148 - 152}{7}$$

$$z = -\mathbf{0{,}57}$$

(c)

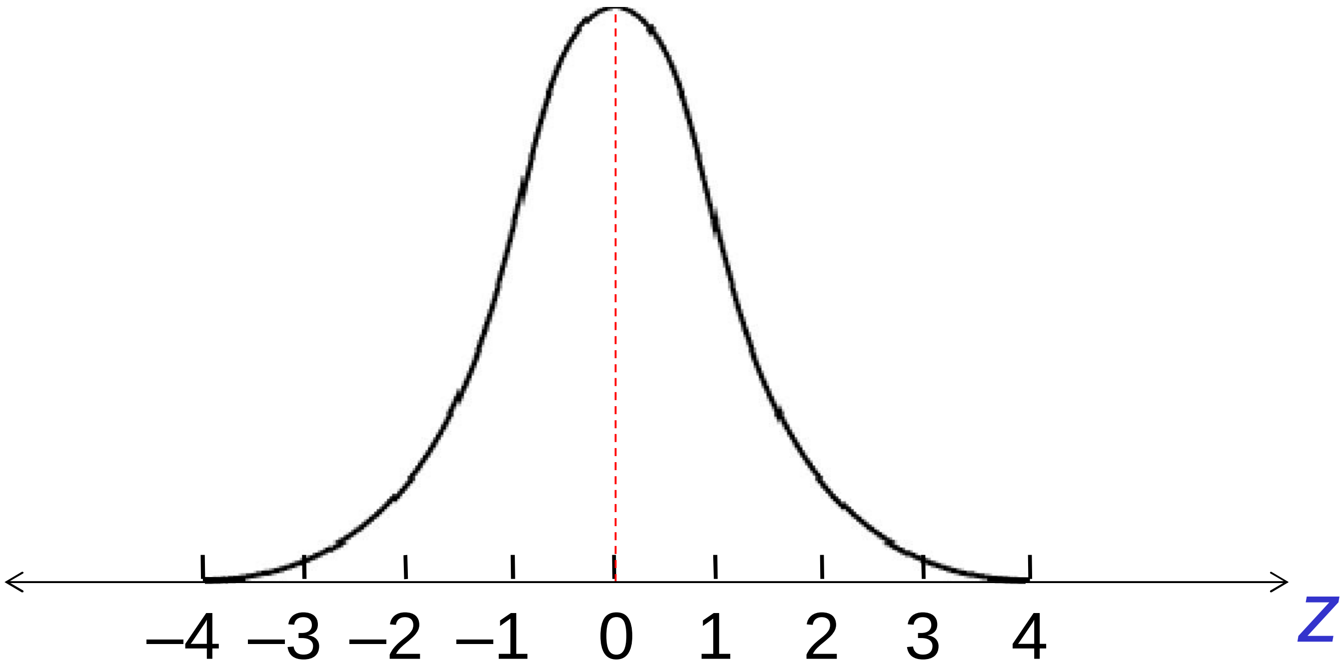
$$z = \frac{152 - 152}{7}$$

$$z = 0$$

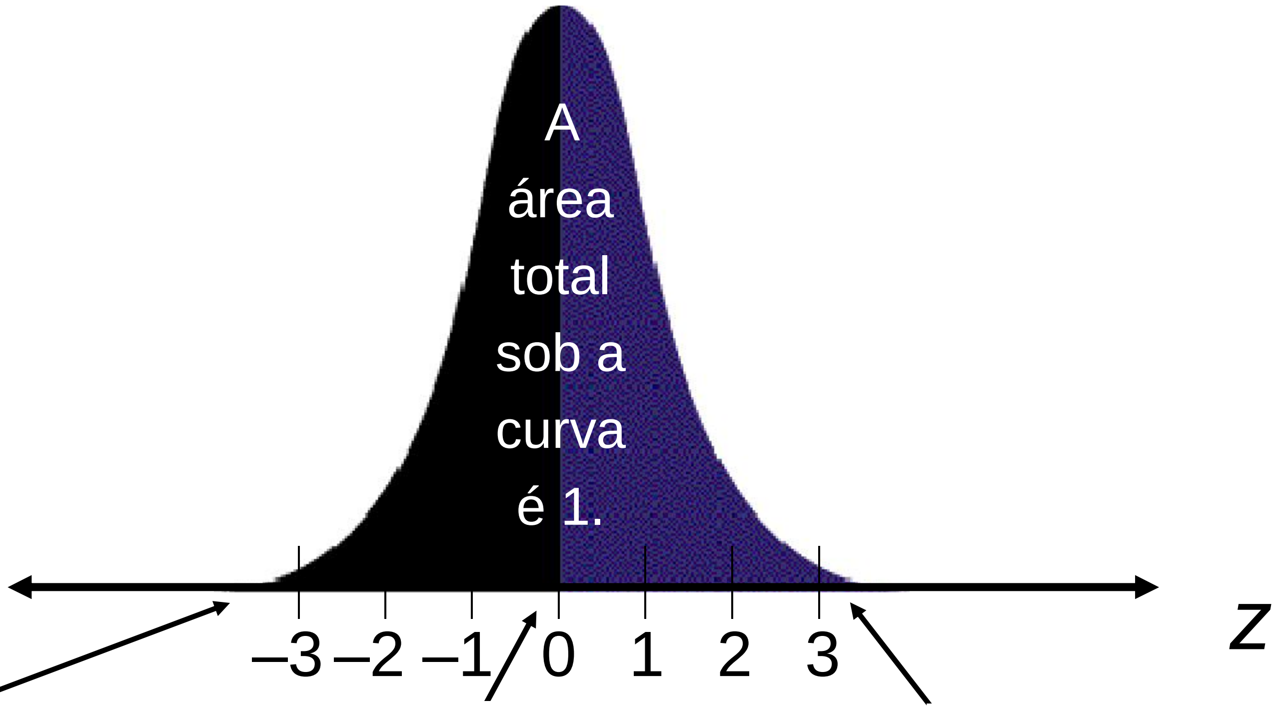
A distribuição normal padrão

A distribuição normal padrão tem média 0 e **desvio padrão de 1.**

Se usar escores *z*, você pode transformar qualquer distribuição normal numa distribuição normal padrão.



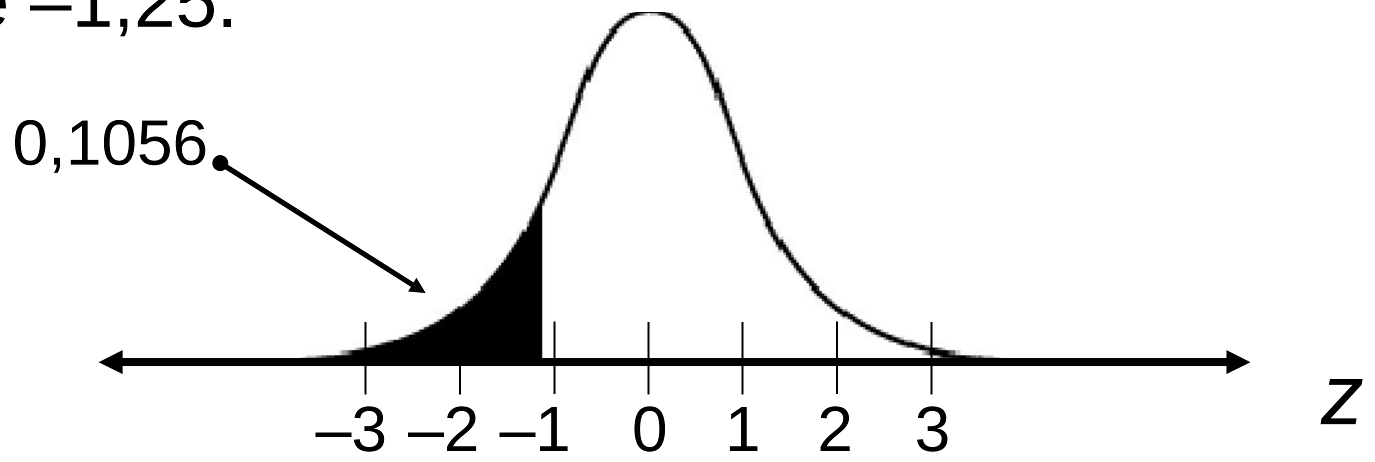
Áreas acumuladas



- **A área acumulada está próxima de 0 para escores *z* próximos de –3,49.**
- **A área acumulada para *z* = 0 é 0,5000.**
- **A área acumulada está próxima de 1 para escores *z* próximos de 3,49.**

Áreas acumuladas

Determine a área acumulada para um escore *z* de –1,25.



Percorra a coluna *z*, à esquerda, até *z* = –1,25; depois siga na transversal até a coluna sob o número 0,05. O valor da célula, 0,1056, corresponde à área acumulada.

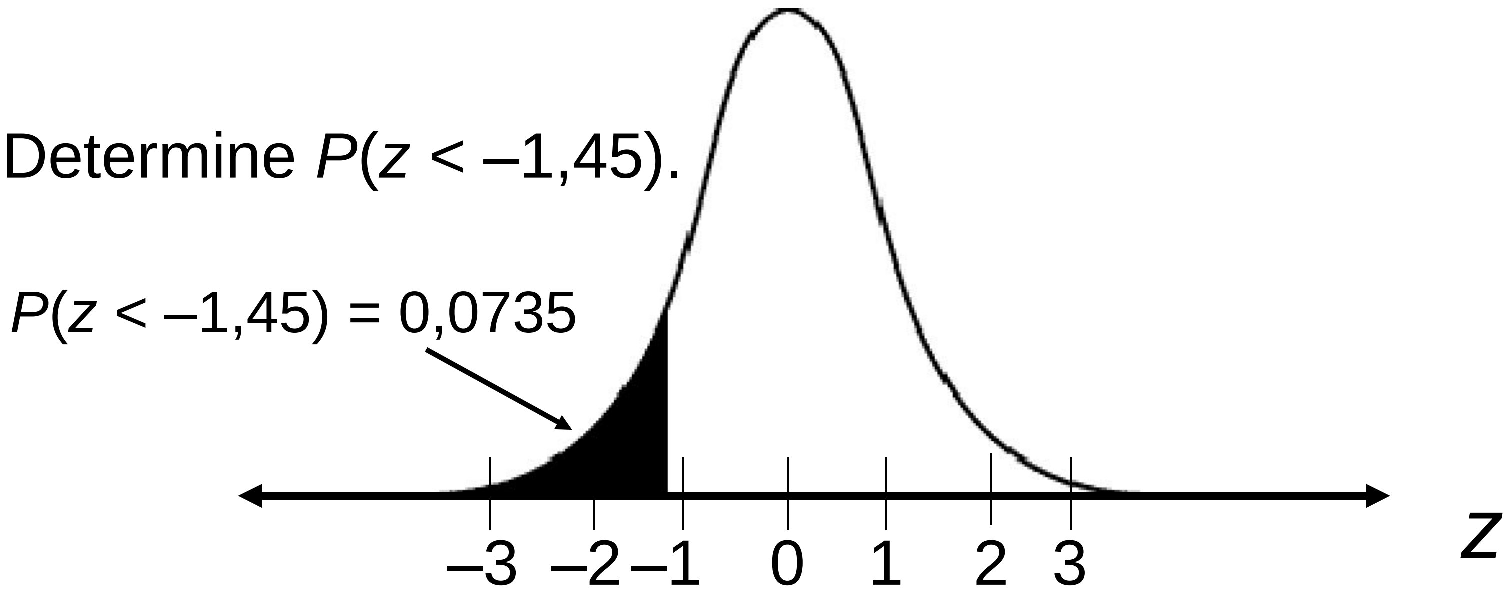
A ***probabilidade*** de que *z* esteja no máximo até –1,25 é de 0,1056.

$$P(z \leq -1{,}25) = 0{,}1056$$

Como determinar probabilidades

Para determinar a probabilidade de *z* ser inferior a um valor dado, encontre a área acumulada na tabela de acordo com o correspondente escore *z*.

Determine $P(z < -1{,}45)$.

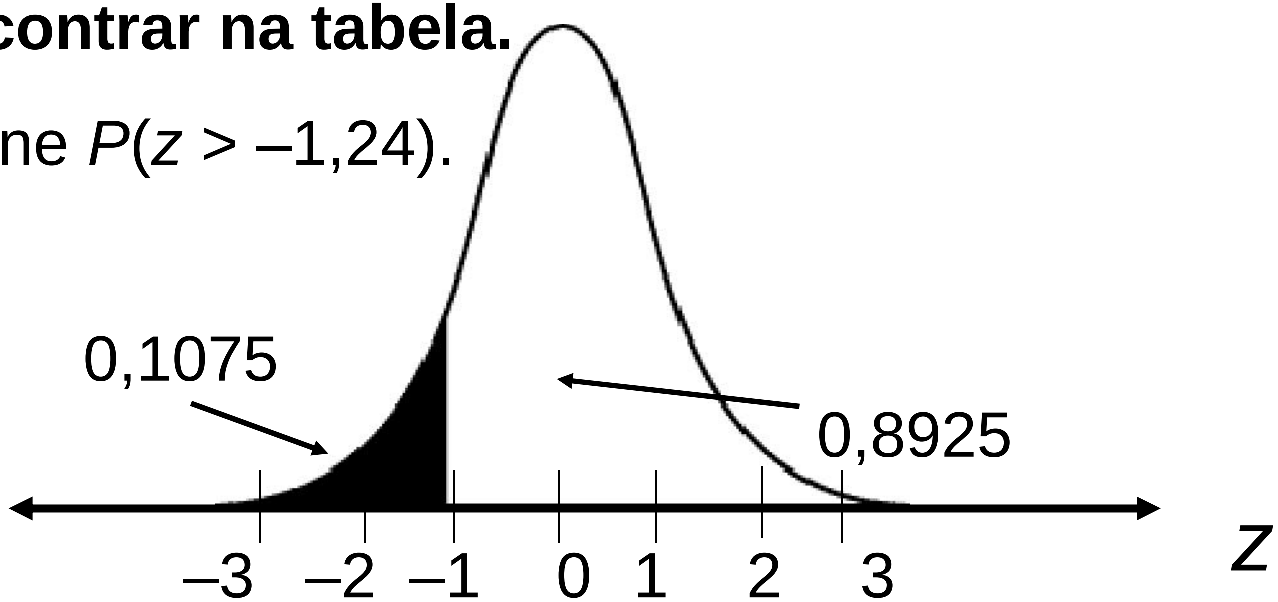


Percorra a coluna *z* até –1,4; depois, vá na transversal até 0,05. A área acumulada é 0,0735.

Como determinar probabilidades

Para determinar a probabilidade de *z* ser superior a um valor dado, subtraia de 1 a área acumulada que você encontrar na tabela.

Determine $P(z > –1,24)$.



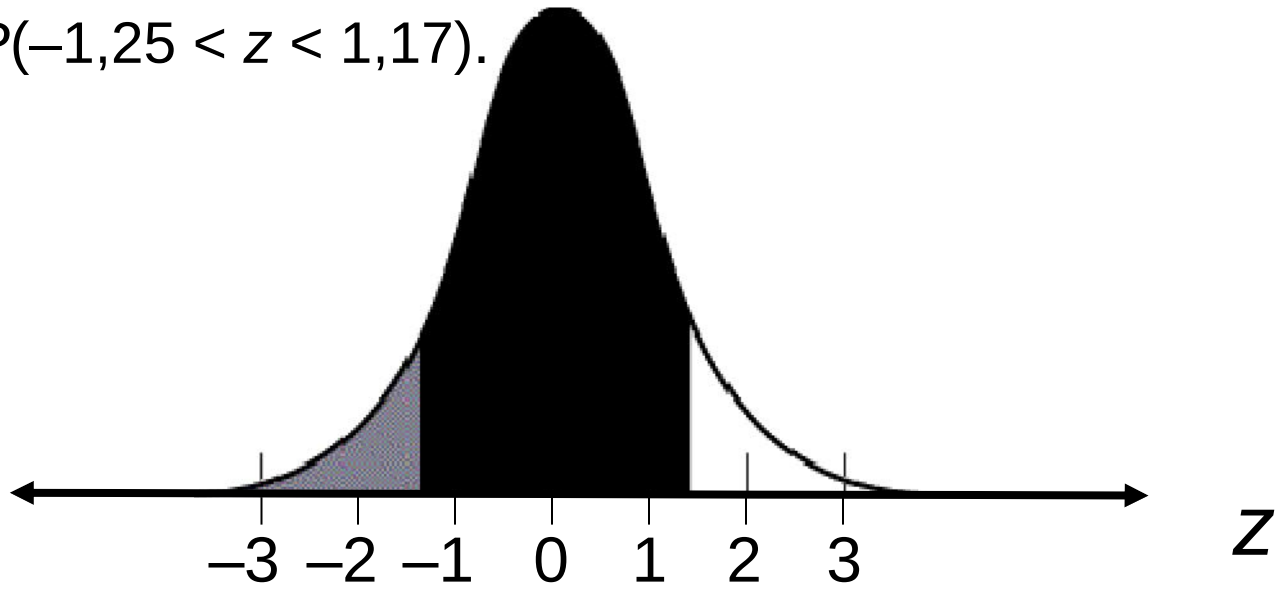
A área acumulada (área à esquerda) é de 0,1075. Logo, a área à direita é: 1 – 0,1075 = 0,8925.

$P(z > –1,24) = 0,8925$

Como determinar probabilidades

Para determinar a probabilidade de *z* estar entre dois valores dados, determine as áreas acumuladas para cada valor e, depois, subtraia a menor da maior.

Determine $P(-1{,}25 < z < 1{,}17)$.



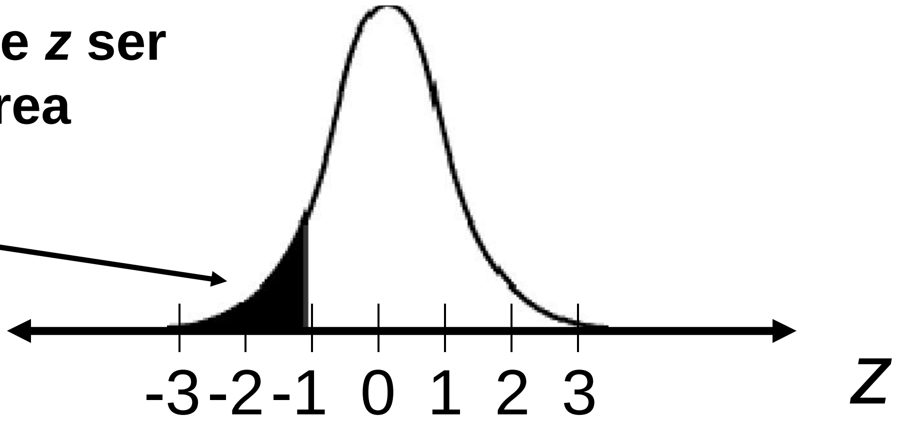
1. $P(z < 1{,}17) = 0{,}8790$

2. $P(z < -1{,}25) = 0{,}1056$

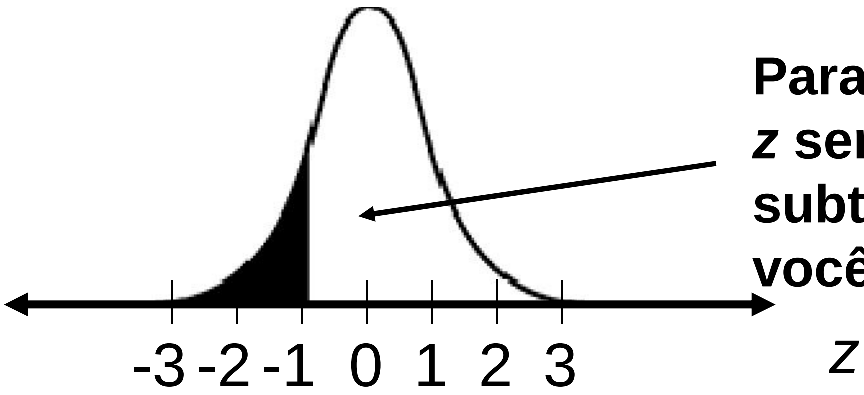
3. $P(-1{,}25 < z < 1{,}17) = 0{,}8790 - 0{,}1056 = 0{,}7734$

Resumo

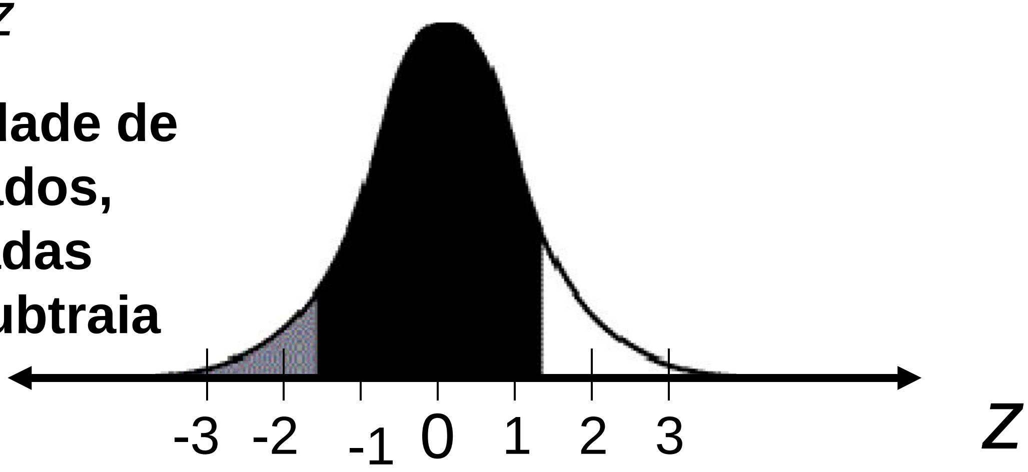
Para determinar a probabilidade de *z* ser inferior a dado valor, encontre a área acumulada correspondente.



Para determinar a probabilidade de *z* ser superior a dado valor, subtraia de 1 a área acumulada que você encontrou na tabela.

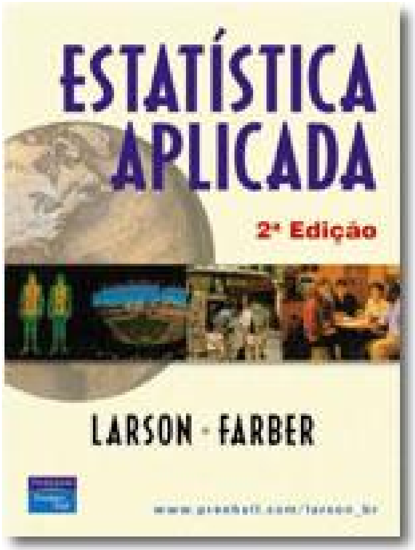


Para determinar a probabilidade de *z* estar entre dois valores dados, determine as áreas acumuladas para cada valor e, depois, subtraia a menor da maior.



Seção 5.3

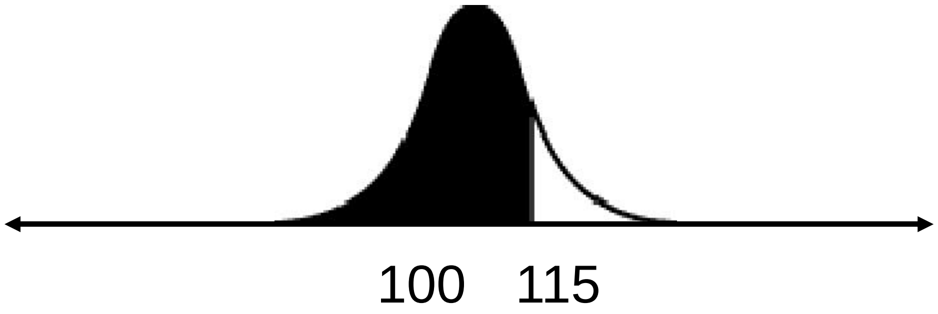
Distribuições normais: determinando probabilidades



Probabilidades e distribuições normais

Se uma variável aleatória *x* é normalmente distribuída, a probabilidade de que ela esteja dentro de dado intervalo é igual à área sob a curva nesse intervalo.

Pontuações de QI são normalmente distribuídas, com uma média de 100 e um desvio padrão de 15. Determine a probabilidade de que uma pessoa selecionada aleatoriamente tenha uma pontuação de QI inferior a 115.



Para determinar a área nesse intervalo, primeiro encontre o escore *z* correspondente a *x* = 115.

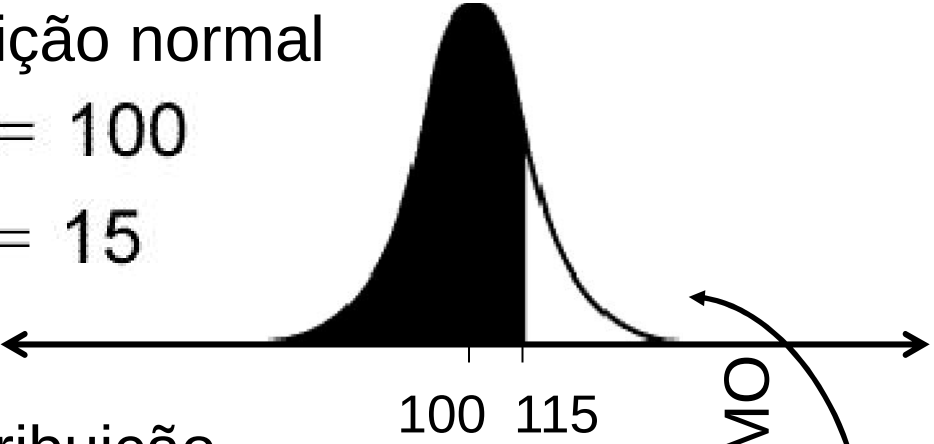
$$z = \frac{150 - 100}{15} = 1$$

Probabilidades e distribuições normais

Distribuição normal

$\mu = 100$

$\sigma = 15$



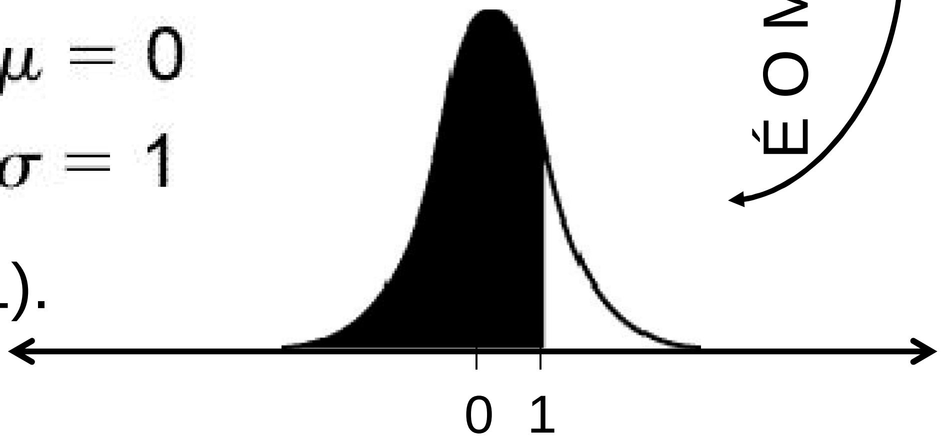
Determine $P(x < 115)$.

É O MESMO

Distribuição normal padrão

$\mu = 0$

$\sigma = 1$



Determine $P(z < 1)$.

É O MESMO

$P(z < 1) = 0{,}8413$, logo $P(x < 115) = 0{,}8413$

Aplicação

As contas mensais de serviços públicos em determinada cidade são normalmente distribuídas, com média de US$ 100 e desvio padrão de US$ 12. Uma conta é escolhida aleatoriamente. Determine a probabilidade de ela estar entre US$ 80 e US$ 115.

Distribuição normal

$\mu = 100$

$\sigma = 12$

P(80 < *x* < 115)

$$z = \frac{80 - 100}{12} = -\mathbf{1{,}67}$$

$$z = \frac{115 - 100}{12} = \mathbf{1{,}25}$$

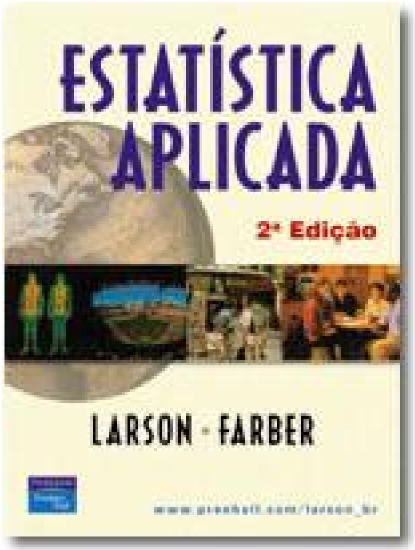
***P*(–1,67 < *z* < 1,25)**

0,8944 – 0,0475 = 0,8469

A probabilidade de uma conta estar entre US$ 80 e US$ 115 é 0,8469.

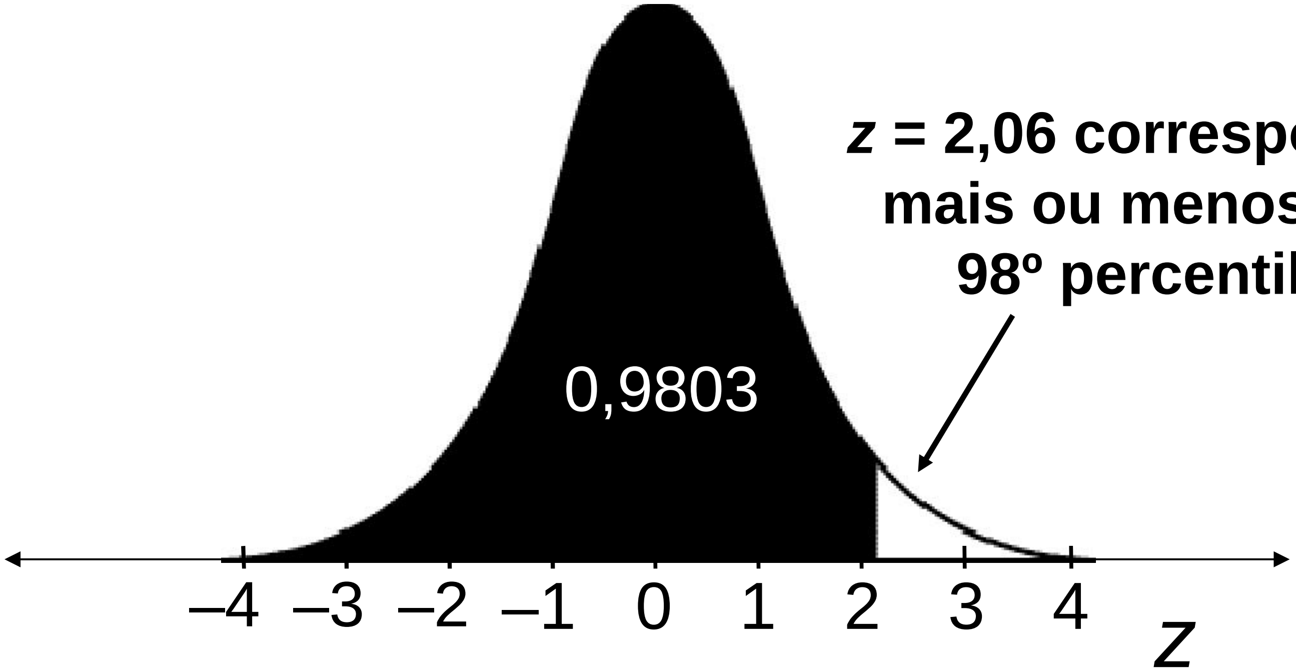
Seção 5.4

Distribuições normais:
obtendo valores



Da área ao escore *z*

Determine o escore *z* correspondente a uma área acumulada de 0,9803.

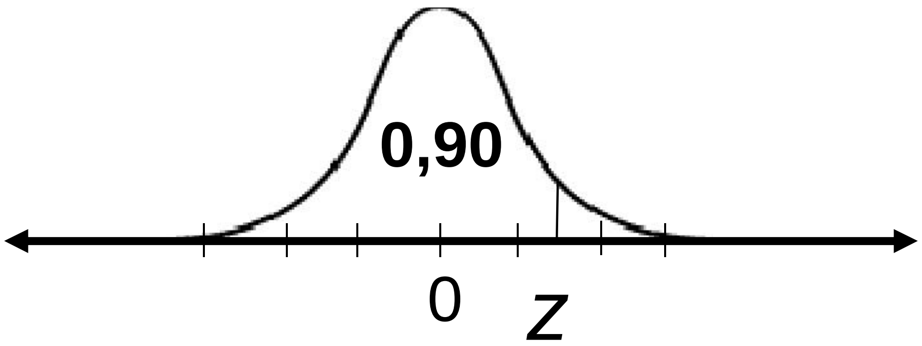


***z* = 2,06 corresponde mais ou menos ao 98º percentil.**

Localize 0,9803 na tabela. Leia os valores no início da linha e no alto da coluna correspondentes. O *escore z* será 2,06.

Determinando escores *z* a partir de áreas

Determine o escore *z* correspondente ao 90º percentil.

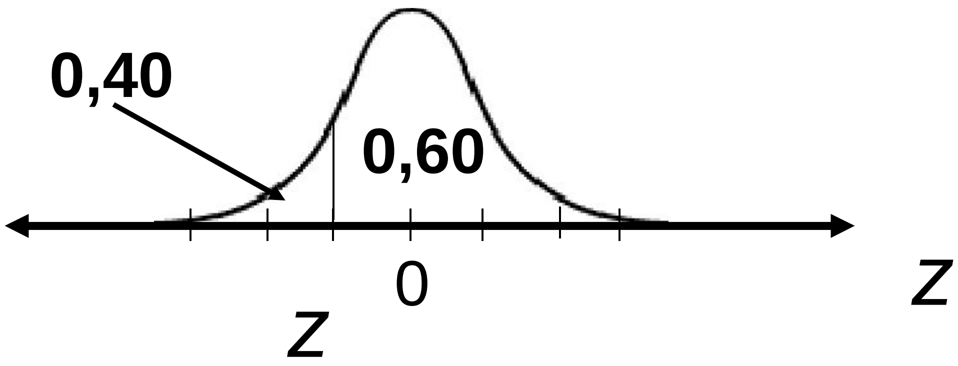


Na tabela, o valor mais próximo é 0,8997. O início da linha é 1,2 e o topo da coluna é 0,08. Isso corresponde a *z* = 1,28.

Um escore *z* de 1,28 corresponde ao 90º percentil.

Determinando escores *z* a partir de áreas

Determine um escore *z* que tenha uma área de 0,60 à sua direita.

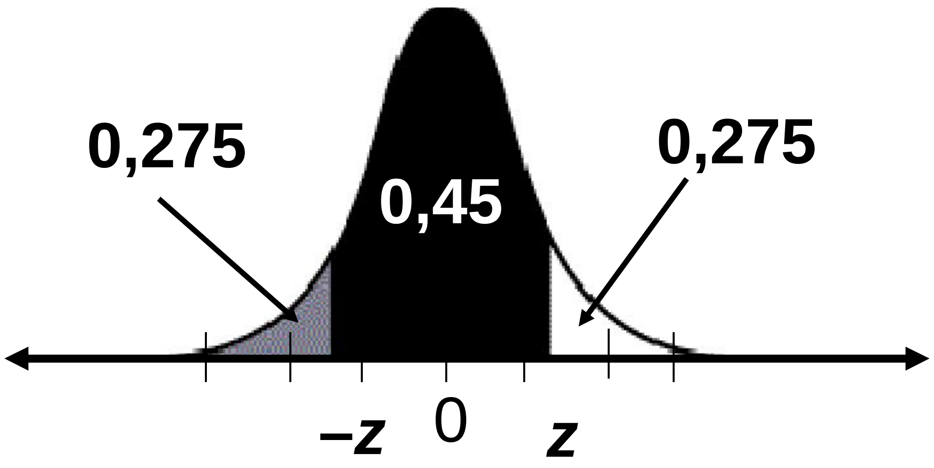


Com 0,60 à direita, a área acumulada é de 0,40. O valor mais próximo é de 0,4013. O início da linha é 0,2 e o topo da coluna é 0,05. Logo, o escore *z* é 0,25.

Um escore *z* de 0,25 tem uma área de 0,60 à sua direita. Isso corresponde ao 40º percentil.

Determinando escores *z* a partir de áreas

Determine um escore *z* tal que 45% da área sob a curva fique entre –*z* e *z*.



A área restante nas pontas é de 0,55. Metade dessa área está em cada ponta; logo, 0,55/2 = 0,275 é a área acumulada para o valor negativo de *z*, e 0,275 + 0,45 = 0,725 é a área acumulada para o *z* positivo. O valor mais próximo na tabela é de 0,2743 e, assim, o escore *z* é 0,60. O escore *z* positivo é 0,60.

De escores *z* a escores brutos

Para determinar um valor *x* a partir de um escore *z*:

$$x = \mu + z\sigma$$

As pontuações em um concurso público estão normalmente distribuídas, com média de 152 e desvio padrão de 7. Determine a pontuação de um candidato com escore *z*:

(a) 2,33 (b) –1,75 (c) 0

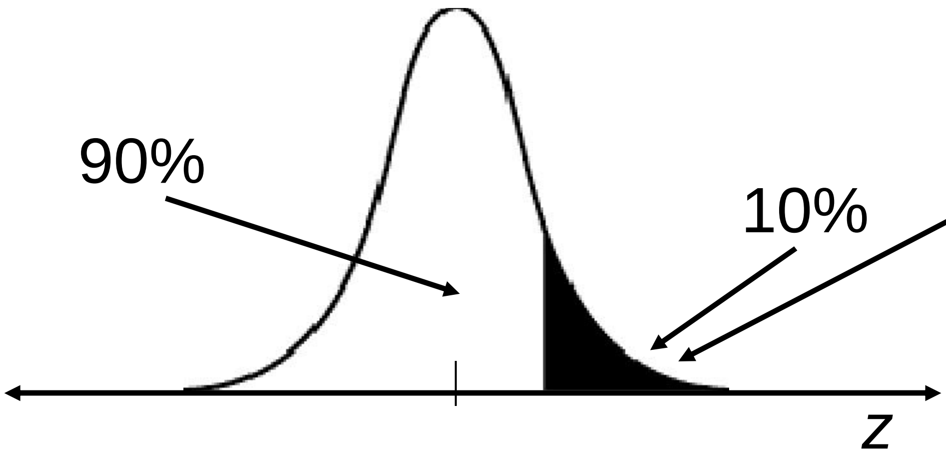
(a) $x = 152 + (2{,}33)(7) = 168{,}31$

(b) $x = 152 + (-1{,}75)(7) = 139{,}75$

(c) $x = 152 + (0)(7) = 152$

Determinando percentis ou valores de corte

As contas mensais de serviços públicos em determinada cidade são normalmente distribuídas, com média de US$ 100 e desvio padrão de US$ 12. Qual é o valor mais baixo entre os 10% mais altos?



US$ 115,36 é o valor mais baixo entre os 10% mais altos.

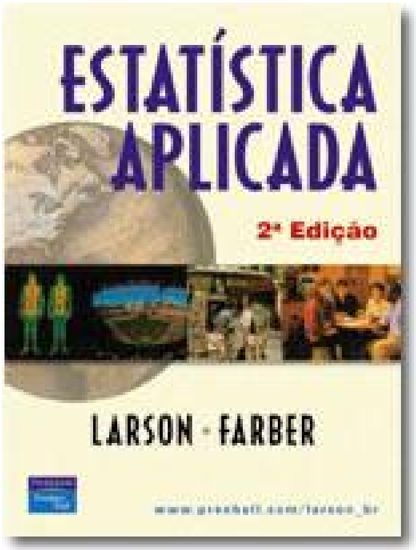
Determine, na tabela, a área acumulada mais próxima a 0,9000 (o 90º percentil). A área 0,8997 corresponde a um escore *z* de 1,28.

Para determinar o valor *x* correspondente, use: $x = \mu + z\sigma$

$$x = 100 + 1{,}28(12) = 115{,}36.$$

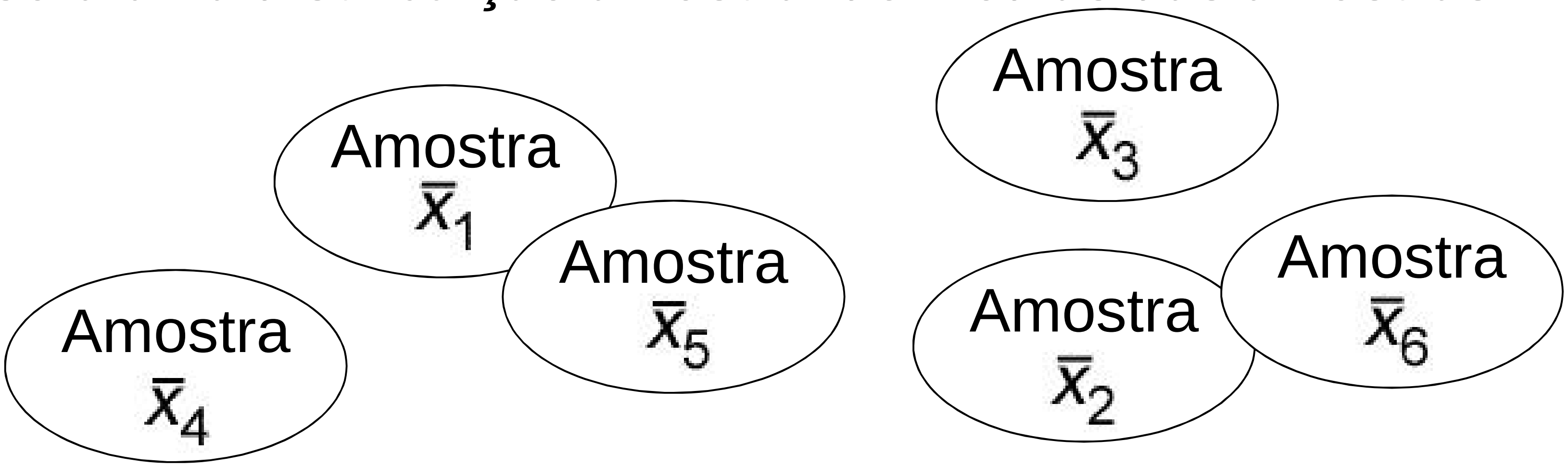
Seção 5.5

Teorema do
Limite Central



Distribuições amostrais

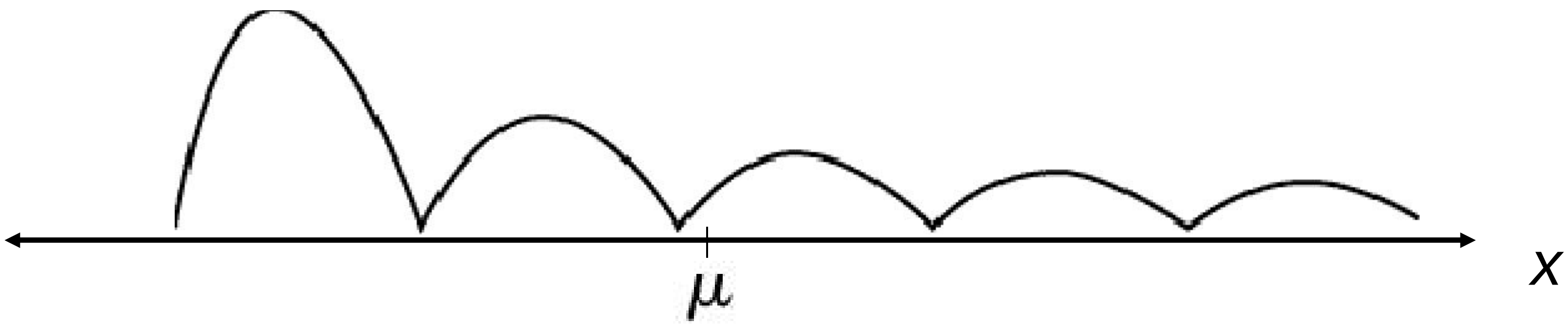
Uma distribuição amostral é a distribuição de probabilidade de uma estatística da amostra formada quando amostras de tamanho *n* são colhidas várias vezes de uma população. Se a estatística da amostra for a sua média simples, a distribuição será uma *distribuição amostral de médias das amostras*.



A distribuição amostral consiste nos valores das médias da amostra, $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \ldots$

O Teorema do Limite Central

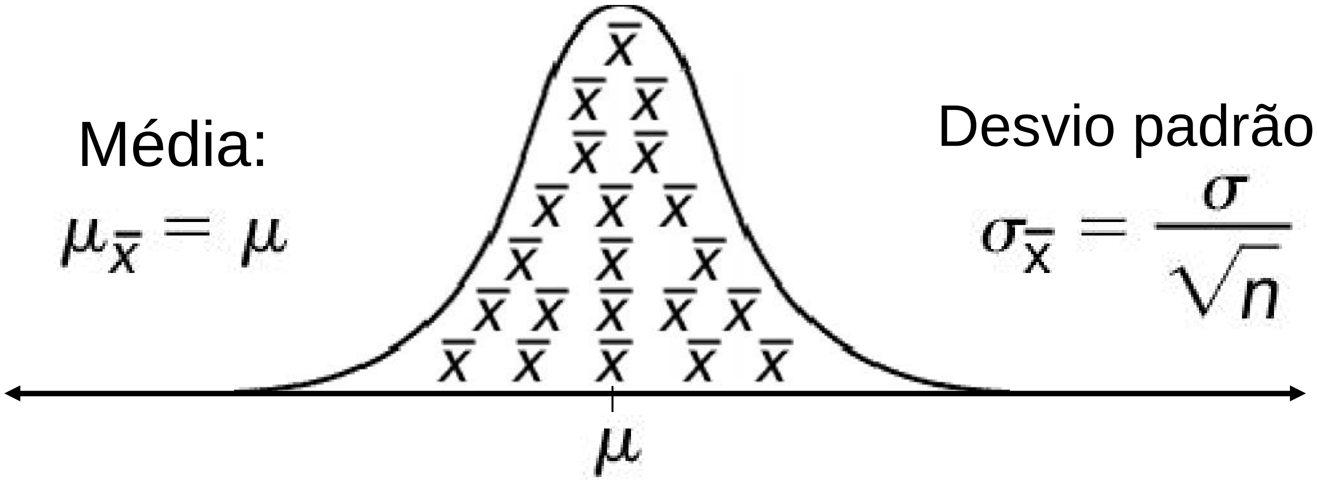
Se uma amostra $n \leq 30$ for tirada de uma população com ***qualquer tipo de distribuição***, média = μ e desvio padrão = σ



as ***médias da amostra*** terão **distribuição normal**.

Média:

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

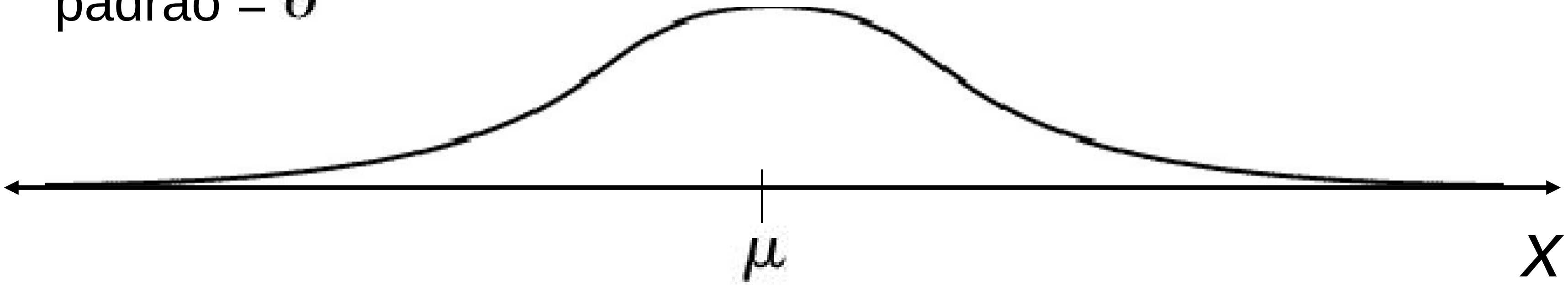


Desvio padrão:

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

O Teorema do Limite Central

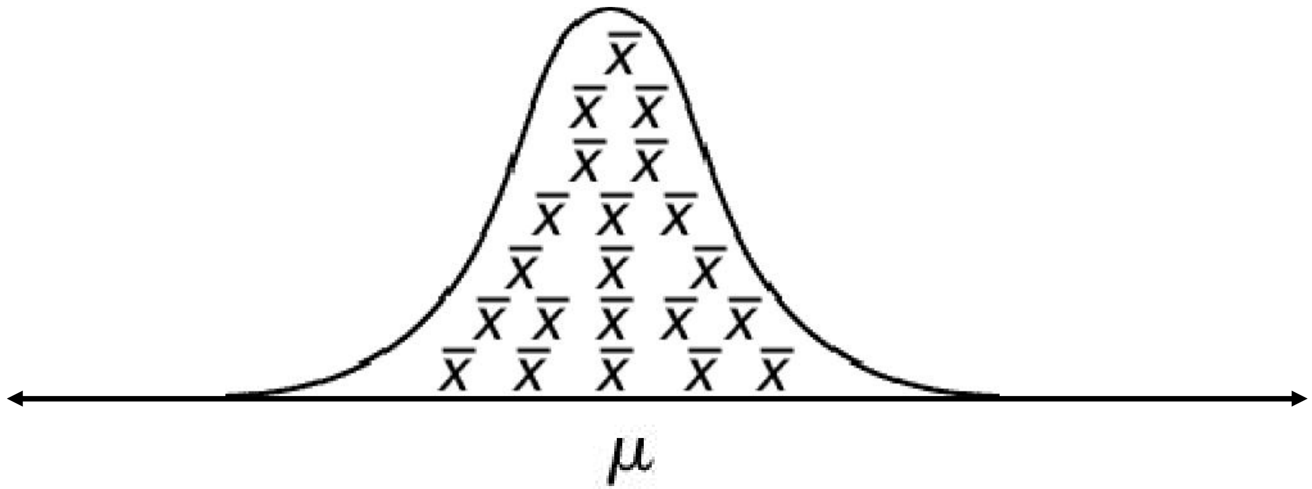
Se uma amostra de qualquer tamanho for tirada de uma população com ***distribuição normal***, média = μ e desvio padrão = σ



a distribuição das médias da amostra de tamanho *n* será **normal**, com média $\mu_{\overline{x}} = \mu$

e desvio padrão

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

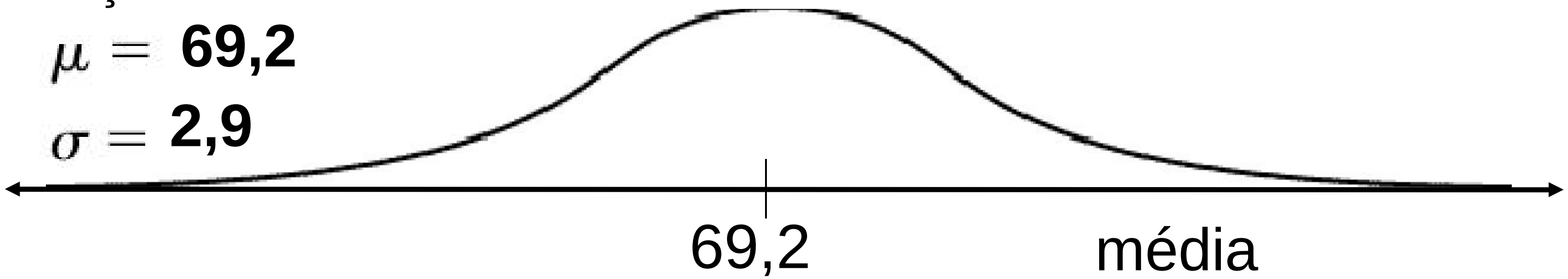


Aplicação

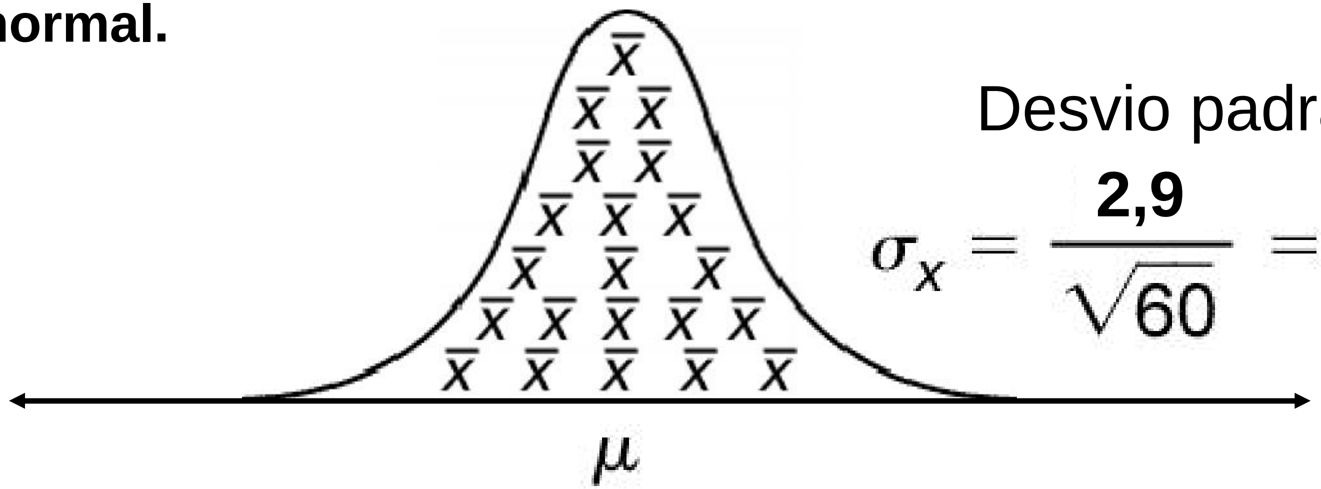
A média de altura dos homens norte-americanos (de 20 a 29 anos) é de $\mu =$ **69,2** **e** $\sigma =$ **2,9** polegadas. Amostras aleatórias de 60 homens são selecionadas. Determine a média e o desvio padrão (erro padrão) da distribuição amostral.

$\mu =$ **69,2**

$\sigma =$ **2,9**



A distribuição de médias da **amostra de tamanho 60**, $\mu_{\bar{x}} = \mu =$ **69,2**, será **normal.**



Desvio padrão

$$\sigma_{\bar{x}} = \frac{2{,}9}{\sqrt{60}} = \mathbf{0{,}3744}$$

Interpretando o Teorema do Limite Central

A média de altura dos homens norte-americanos (de 20 a 29 anos) é μ = 69,2 polegadas. Se uma amostra aleatória de 60 homens nessa faixa etária for selecionada, qual é a probabilidade de que a média de altura na amostra seja superior a 70 polegadas? Admita um desvio padrão de 2,9 polegadas.

Uma vez que *n* > 30, a distribuição amostral de $\bar{x}$ será normal

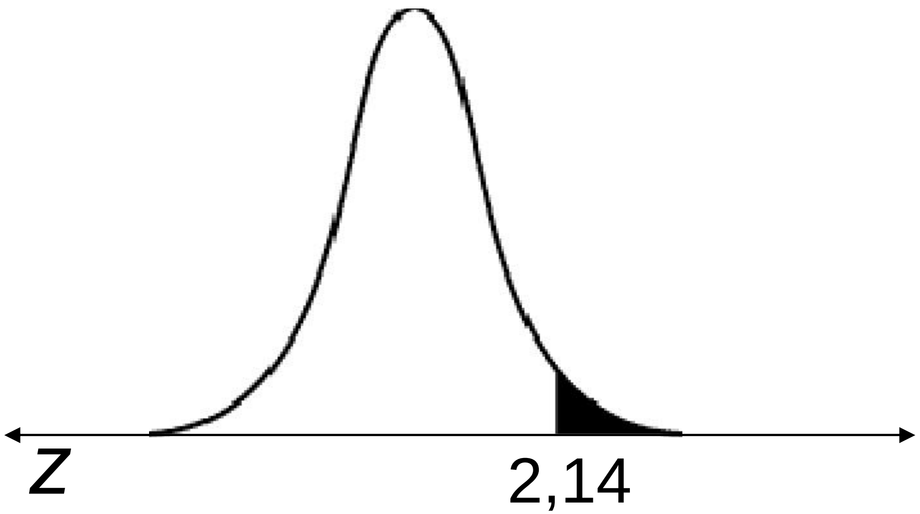
Média: $\mu_{\bar{x}} = \mathbf{69{,}2}$

Desvio padrão: $\sigma_{\bar{x}} = \dfrac{2{,}9}{\sqrt{60}} = \mathbf{0{,}3744}$

Determine o escore *z* para uma média amostral de 70:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{70 - 69{,}2}{0{,}3744} = 2{,}14$$

Interpretando o Teorema do Limite Central



$$P(\bar{x} > 70)$$

$$= P\,(z > \mathbf{2{,}14}\,)$$

$$= 1 - \mathbf{0{,}9838}$$

$$= \mathbf{0{,}0162}$$

Há uma probabilidade de 0,0162 de que uma amostra com 60 homens tenha uma média de altura superior a 70 polegadas.

Aplicando o Teorema do Limite Central

Em certa semana o preço médio da gasolina na Califórnia foi de US$ 1,164 por galão. Qual é a probabilidade de que o preço médio em uma amostra de 38 postos esteja entre US$ 1,169 e US$ 1,179? Admita que o desvio padrão seja de US$ 0,049.

Uma vez que *n* > 30, a distribuição amostral de $\bar{x}$ será normal.

Média: $\mu_{\bar{x}} = \mu_{\bar{x}} = 1{,}164$

Desvio padrão: $$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{0{,}049}{\sqrt{38}} = 0{,}0079$$

Calcule o escore *z* para valores amostrais de US$ 1,169 e US$ 1,179.

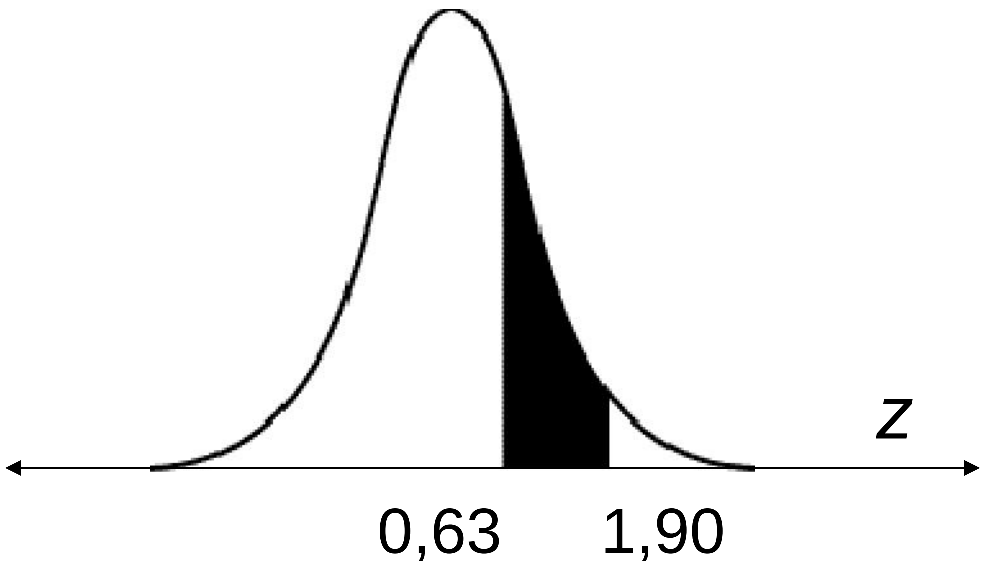
$$z = \frac{1{,}169 - 1{,}164}{0{,}0079} = 0{,}63 \qquad z = \frac{1{,}179 - 1{,}164}{0{,}0079} = 1{,}90$$

Aplicando o Teorema do Limite Central

$P(0{,}63 < z < 1{,}90)$

$= 0{,}9713 - 0{,}7357$

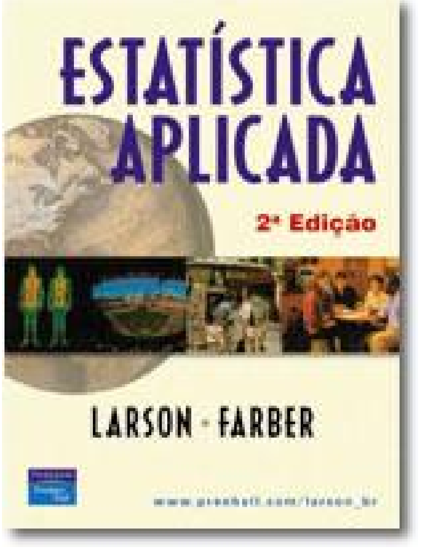
$= 0{,}2356$



A probabilidade de que a média da amostra esteja entre US$ 1,169 e US$ 1,179 é de 0,2356.

Seção 5.6

Aproximações normais para as distribuições binomiais



Características da distribuição binomial

- O número de tentativas independentes (*n*) é fixo.
- Cada tentativa pode ter dois resultados, sucesso *ou* fracasso.
- A probabilidade de sucesso numa única tentativa é *p* e de fracasso é *q*. $p + q = 1$
- É possível determinar a probabilidade de exatamente *x* sucessos em *n* tentativas, sendo $x = 0$ ou 1 ou 2 … *n*.
- *x* é uma variável aleatória ***discreta*** que representa uma contagem do número de sucessos em *n* tentativas.

$$\mu = np \quad \text{e} \quad \sigma = \sqrt{npq}$$

Aplicação

34% dos norte-americanos têm sangue tipo A^+. Se 500 pessoas dessa nacionalidade forem selecionadas aleatoriamente, qual é a probabilidade de ao menos 300 terem sangue tipo A^+?

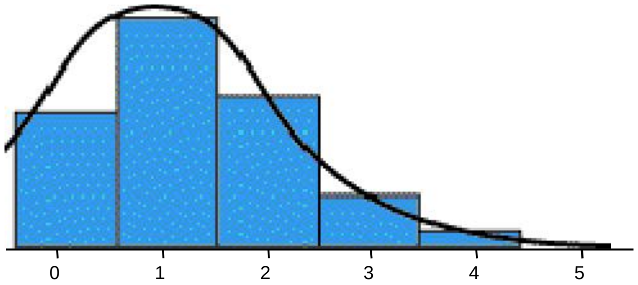
Com as técnicas do Capítulo 4, você poderia calcular a probabilidade de exatamente 300, exatamente 301… exatamente 500 norte-americanos terem sangue tipo A^+ e depois somar as probabilidades.

Ou… você pode usar as probabilidades de curva normal para aproximar as probabilidades binomiais.

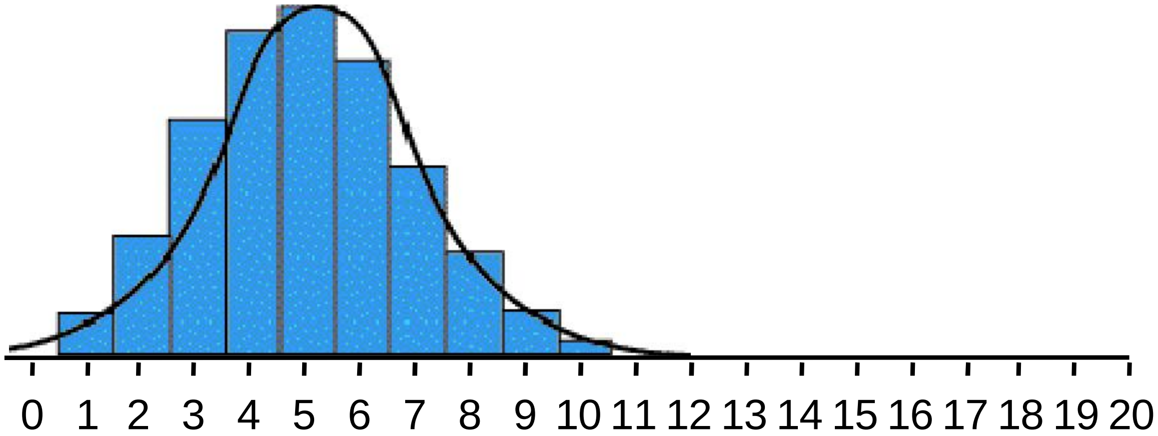
Se $np \geq 5$ e $nq \geq 5$, a variável aleatória binomial x tem distribuição aproximadamente normal com:

$$\mu = np \quad \textbf{e} \quad \sigma = \sqrt{npq}$$

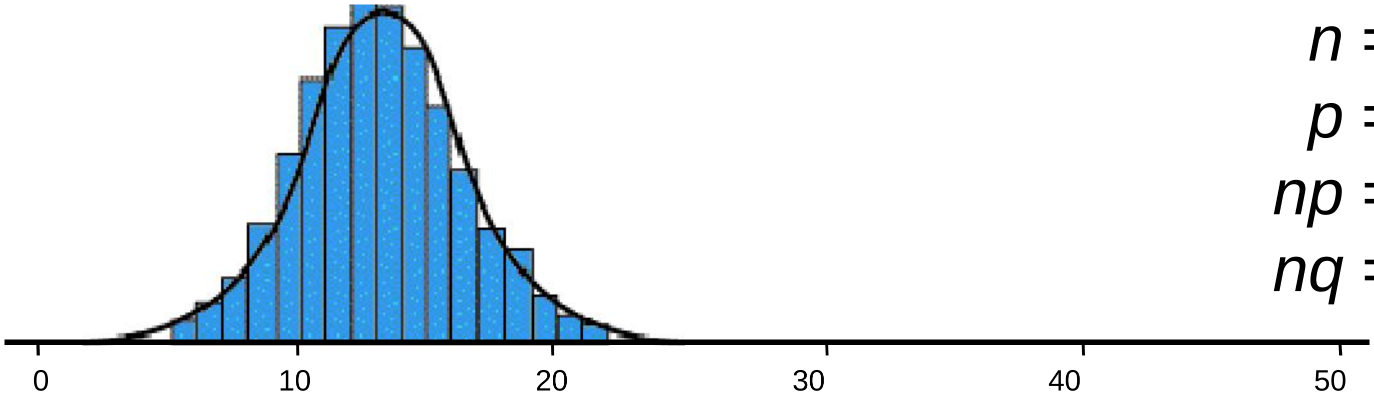
Por que precisamos de *np* $\geq$ 5 e *nq* $\geq$ 5?



$n = 5$
$p = 0{,}25,\ q = 0{,}75$
$np = 1{,}25 \quad nq = 3{,}75$



$n = 20$
$p = 0{,}25$
$np = 5 \quad nq = 15$



$n = 50$
$p = 0{,}25$
$np = 12{,}5$
$nq = 37{,}5$

Probabilidades binomiais

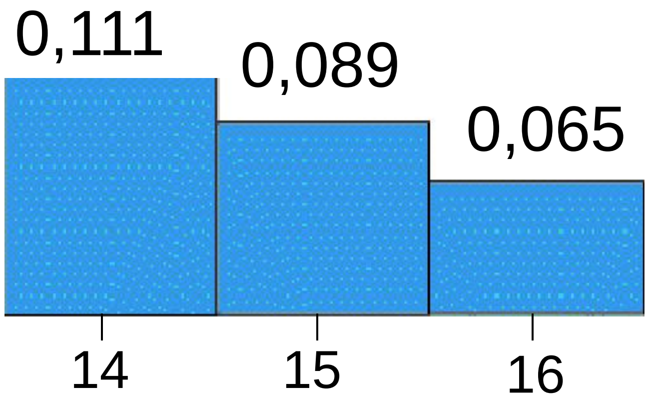
A distribuição binomial é discreta e pode ser representada por um histograma de probabilidade. A probabilidade de que um específico valor de *x* ocorra é igual à área do retângulo com ponto médio *x*.

Se *n* = 50 e *p* = 0,25, determine $P(14 \le x \le 16)$.

Some as áreas dos retângulos com pontos médios em

$x = 14, \quad x = 15, \quad x = 16.$

$0{,}111 + 0{,}089 + 0{,}065 = 0{,}265$

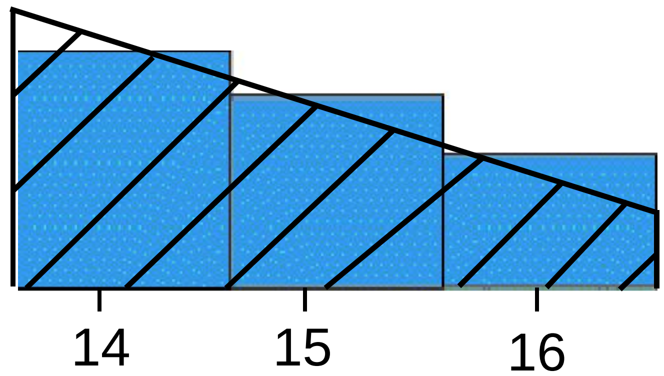


$P(14 \le x \le 16) = 0{,}265$

Correção pela continuidade

Use a aproximação normal para a binomial a fim de determinar $P(14 \leq x \leq 16)$ **se n =** 50 **e** $p =$ **0,25**

Verifique que $np =$ **12,5** ≥ 5 **e** $nq =$ **37,5** $\geq 5.$

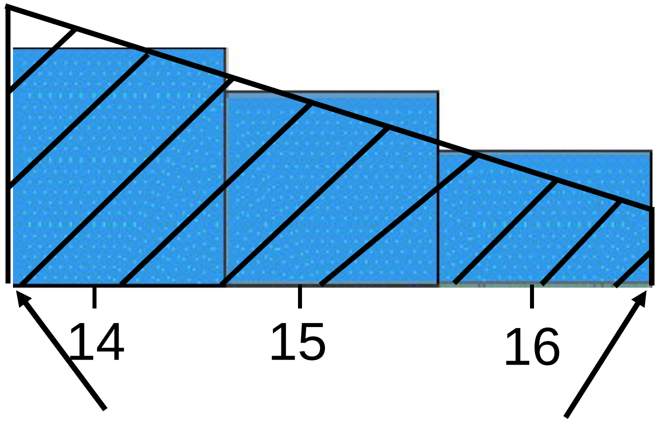


Os valores para a variável aleatória binomial *x* são 14, 15 e 16.

Correção pela continuidade

Use a aproximação normal para a binomial a fim de determinar $P(14 \leq x \leq 16)$ **se n =** 50 **e** $p = $ **0,25**.

Verifique que $np =$ **12,5** ≥ 5 **e** $nq =$ **37,5** $\geq 5.$



O intervalo de valores sob a curva normal é

13,5 $\leq x \leq$ **16,5**

Para garantir que as fronteiras de cada retângulo estejam incluídas no intervalo, subtraia 0,5 das fronteiras à esquerda e some 0,5 às que estão à direita.

Aproximação normal para a binomial

Use a aproximação normal para a binomial a fim de determinar: $P(14 \leq x \leq 16)$ **se** $n = 50$ **e** $p = 0{,}25$.

Com as fórmulas de distribuição binomial, determine a média e o desvio padrão.

$$\mu = np = 50\,(0{,}25) = 12{,}5$$

$$\sigma = \sqrt{npq} = \sqrt{50\,(0{,}50)(0{,}75)} = 3{,}062$$

Ajuste os pontos extremos para corrigir pela continuidade $P\ 13{,}5 \leq x \leq 16{,}5$.

Converta cada ponto extremo em um escore *z*.

$$z = \frac{13{,}5 - 12{,}5}{3{,}062} = 0{,}33 \qquad z = \frac{16{,}5 - 12{,}5}{3{,}062} = 1{,}31$$

$$P(\,0{,}33 \leq z \leq 1{,}31\,) = 0{,}9049 - 0{,}6293 = 0{,}2756$$

Aplicação

Segundo um levantamento entre os usuários da Internet, 75% são a favor de que o governo regulamente o 'lixo eletrônico'. Se 200 internautas forem selecionados aleatoriamente, determine a probabilidade de que menos de 140 sejam a favor da regulação governamental.

Uma vez que $np = 150 \geq 5$ e $nq = 50 \geq 5$, você pode usar a distribuição normal para aproximar a probabilidade binomial.

$$\mu = np = 200\,(0{,}75) = 150$$

$$\sigma = \sqrt{npq} = \sqrt{200\,(0{,}75)(0{,}25)} = \mathbf{6{,}1237}$$

A frase binomial "menos de 140" significa 0, 1, 2, 3…139.

Use a correção pela continuidade para traduzir isso à variável contínua no intervalo $(-\infty,\ \mathbf{139{,}5})$. Determine $P(x < 139{,}5)$.

Aplicação

Segundo um levantamento entre os usuários da Internet, 75% são a favor de que o governo regulamente o ‘lixo eletrônico’. Se 200 internautas forem selecionados aleatoriamente, determine a probabilidade de que menos de 140 sejam a favor da regulação governamental.

Use a correção pela continuidade $P(x < 139{,}5)$.

$$z = \frac{139{,}5 - 150}{6{,}1237} = -1{,}71$$

$P(z < -1{,}71) = 0{,}0436$

A probabilidade de que menos de 140 sejam a favor da regulação governamental é de aproximadamente 0,0436.