

TRABALHO FINAL



Estatística e Probabilidade

AUTOR

Juno Takano

DATA DE PUBLICAÇÃO

30 de junho de 2025

INTRODUÇÃO

Este trabalho foi elaborado para a disciplina de Estatística e Probabilidade, sob orientação do Prof. Dr. Cleverson Pinheiro, como parte curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de São Paulo, Campus Jacareí. Ele foi desenvolvido utilizando a linguagem de programação [R](#), a ferramenta de publicação [Quarto](#) e o conjunto de pacotes [tidyverse](#).

O código fonte correspondente a este documento está também disponível no GitHub, no repositório [jultty/ESPR-final](#). Ele pode ser visualizado em sua versão como página web em [jultty.github.io/ESPR-final](#).

- Softwares utilizados e versões:
 - R 4.2.2, de 10 de novembro de 2022
 - Quarto 1.7.32, de 16 de junho de 2025
 - tidyverse 2.0.0, de 22 de fevereiro de 2023, com atualizações de pacotes individuais

Os pacotes utilizados podem ser visualizadas na relação abaixo, onde são de fato carregados no ambiente de desenvolvimento.

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(scales)
```

Blocos de código como o exibido acima são utilizados ao longo deste documento. Em algumas ocasiões, eles estão ocultos por terem menor importância para a demonstração da solução – por exemplo, código relacionado estritamente à plotagem de gráficos e não à definição de um cálculo ou algoritmo que de fato informa o gráfico.

Nestes casos, o texto “► Código” aparece onde o código está oculto e

Nestes casos, o texto “Código” aparece onde o código está oculto e apenas seu resultado é exibido, seja ele um número, texto, conjunto de dados ou gráfico. Clicando sobre este texto, é possível expandir e visualizar o código na íntegra:

► Código

```
[1] 0+1.224647e-16i
```

OBSERVAÇÕES

Para manter a consistência com as saídas computacionais, neste trabalho utilizou-se exclusivamente o ponto como marcador de casas decimais.

FUNÇÕES UTILITÁRIAS

Primeiro, podemos definir algumas funções auxiliares que serão úteis ao longo de todo o trabalho.

```
## Usando notação Latex, mostra um número com seis casas decimais
## a porcentagem equivalente com duas casas decimais
ppercent <- function(raw) {
  sprintf("$$P = %.6f \\\approx %.2f\%$$", raw, raw * 100) |
}

## Arredonda um número para 4 casas decimais
r4 <- function(float) {
  round(float, digits = 4)
}

## Arredonda um número para 2 casas decimais
r2 <- function(float) {
  round(float, digits = 2)
}
```

EXERCÍCIO 1

Dada a oportunidade na questão elementar do primeiro exercício, podemos representar as curvas *A*, *B* e *C* usando valores aproximados ao que vemos no gráfico, demonstrando a estabelecendo alguns padrões que se repetirão ao longo deste trabalho.

Para isso, criaremos uma *tibble*, uma estrutura de dados similar a um *dataframe*, que armazena dados em formato tabular. Desta forma podemos relacionar cada curva à sua média e desvio padrão.

```
curves <- tibble(  
  curve = c("A", "B", "C"), # Curva  
  mu = c(45, 60, 45),       # Média  
  sigma = c(2, 3.5, 5),     # Desvio padrão  
)
```

Iremos construir também um vetor *x_vector* para os valores de *x* contendo 1000 valores entre 0 e 100.

```
x_vector <- seq(0, 100, length.out = 1000)
```

Neste momento, temos as informações necessárias para obter os valores correspondentes no eixo *y*.

Criaremos uma segunda *tibble*, *density*, que terá 3000 linhas, mil para cada curva, com respectivos *x* e *y*. Embora mais didático pelo menor nível de abstração, esta implementação não seria tão preferível em larga escala, onde o pacote *purrr* oferece resultados mais eficientes.

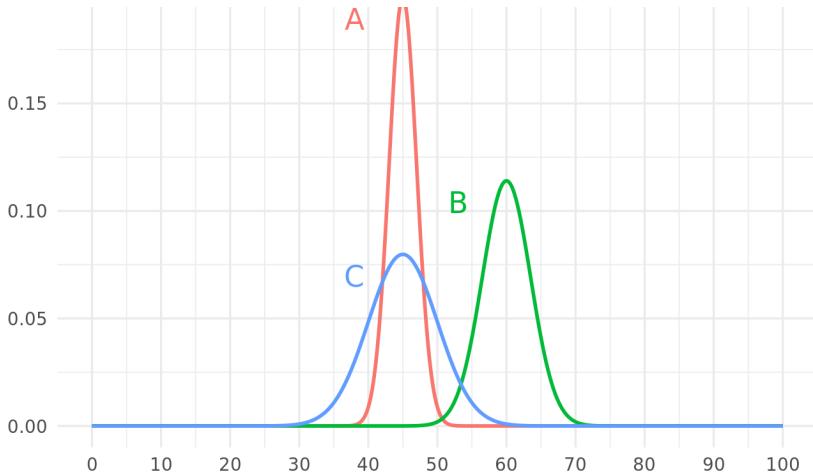
```
density <- curves |>  
  # Para estruturar o dataset, repetimos os atributos de cada  
  slice(rep(seq_len(n()), each = length(x_vector))) |>  
  # Adicionamos uma nova coluna, que insere os valores de x p  
  mutate(x = rep(x_vector, times = nrow(curves))) |>  
  # Finalmente, obtemos o y correspondente na FDP em outra nc  
  mutate(y = dnorm(x, mean = mu, sd = sigma)) |>  
  # Precisamos apenas das colunas curve (A, B ou C), x e y  
  select(curve, x, y)
```

O resultado do que estabelecemos acima pode ser observado na plotagem dos dados:

► Código

Esboço das Curvas Normais A, B e C





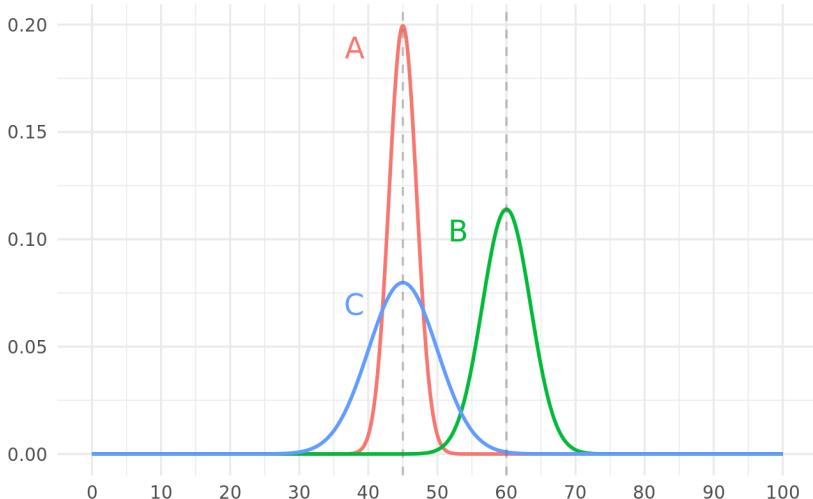
Munidos do gráfico, podemos responder às perguntas do enunciado:

QUESTÃO A

A média maior é a da curva *B*, pois ela está mais deslocada à direita. Sua média fica sobre o ponto 60 do eixo *x*.

► Código

Esboço das Curvas Normais A, B e C com linhas de simetria



As linhas de simetria nos ajudam a verificar que a média maior está localizada no ponto 60 do eixo x , a média da curva B . Não há diferença significativa entre os eixos de simetria das curvas A e C , ambas com aparente média 45, mas visivelmente à esquerda da média de B .

QUESTÃO B

O desvio padrão maior é o da curva C , pois seus limites mínimo e máximo atingem um desvio muito mais amplo em relação à média. Isso também indica que o desvio padrão de C deve ser o maior. Observando o esboço, podemos estimar um desvio padrão em torno de 5.

EXERCÍCIO 2

O exercício 2 nos apresenta um problema similar, que também se debruça sobre a observação de um esboço. Sabendo que a criação de curvas normais será essencial por todo o trabalho, vamos criar um par de funções que nos permita esboçar curvas normais como a que criamos anteriormente através apenas de seus parâmetros:

► Código

```
draft <- function(name, mu, sigma, x_min, x_max, x_step, y_ma

  curve <- tibble(
    curve = name,
    mu = mu,
    sigma = sigma,
  )

  x_vector <- seq(x_min, x_max, length.out = 1000)

  density <- curve |>
    slice(rep(seq_len(n()), each = length(x_vector))) |>
    mutate(x = x_vector) |>
    mutate(y = dnorm(x, mean = mu, sd = sigma)) |>
    select(curve, x, y)

  graph_properties <- list(
    name = name,
    x_min = x_min,
```

```

        x_max = x_max,
        x_step = x_step,
        y_max_multiplier = y_max_multiplier
    )

    plot_draft(density, curve, graph_properties)
}

```

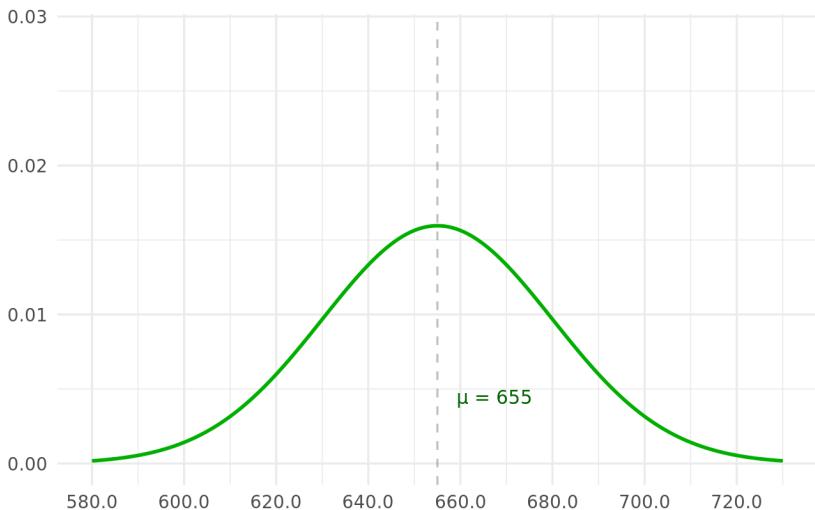
Nossa função pode ser usada da seguinte forma:

```

draft(
  name = "Notas",
  mu = 655,
  sigma = 25,
  x_min = 580,
  x_max = 730,
  x_step = 20,
  y_max_multiplier = 1.8
)

```

Esboço da Curva Normal "Notas"



Podemos responder, pela observação do gráfico, que a nota média é de 655.

Ainda observando o gráfico, podemos ver aproximadamente os pontos

Ainda observando o gráfico, podemos ver aproximadamente os pontos de inflexão em 630 e 680, o que sugere um desvio padrão de $\sigma = 25$.

EXERCÍCIO 3

Os exercícios 3 a 6 apresentam problemas relacionados à área sob a curva normal padrão dado um escore-z.

A linguagem R já nos fornece a função `pnorm`, que responde de forma direta à pergunta sobre qual é a área sob a curva dado um z qualquer.

Para visualizarmos estas informações, iremos criar uma função `z_plot` para obter ainda um gráfico representando a área.

Nossa função receberá apenas um valor `z` e dois parâmetros opcionais:

- `right`, que, se fornecido com o valor `TRUE`, retornará um gráfico com a área à direita de z . Caso o valor seja `FALSE` ou não seja fornecido, ela por padrão retornará um gráfico com a área à esquerda de z
- `opposite`, que, se fornecido, representa um segundo valor oposto ao primeiro que delimita o fim da área, ou seja, permite encontrar uma região entre dois valores

► Código

QUESTÃO A

A função `pnorm` responde à pergunta sobre a área acumulada sob a curva normal padrão à esquerda de -2.19 , que é igual a 0.01426212 ou aproximadamente 1.43% .

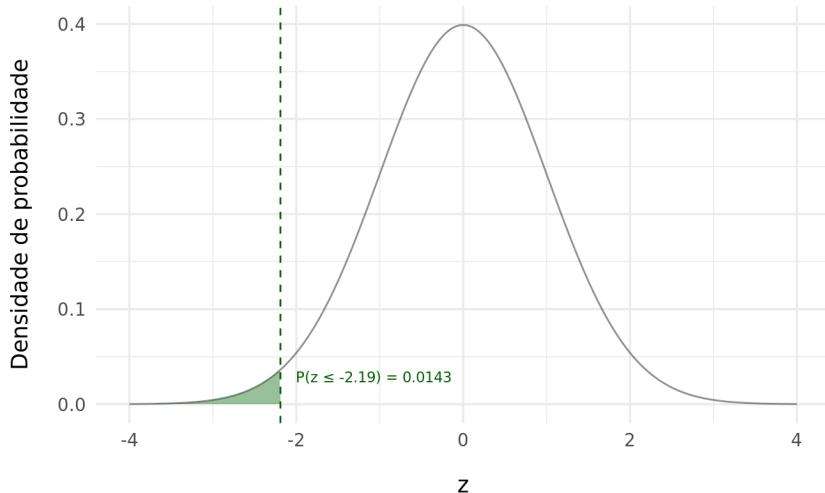
```
pnorm(-2.19) |> ppercent()
```

$$P = 0.014262 \approx 1.43\%$$

Podemos também representar esta área visualmente usando a função `z_plot` que acabamos de definir:

```
z_plot(-2.19)
```

Área acumulada para o escore-z -2.19



QUESTÃO B

Para 2.17, temos:

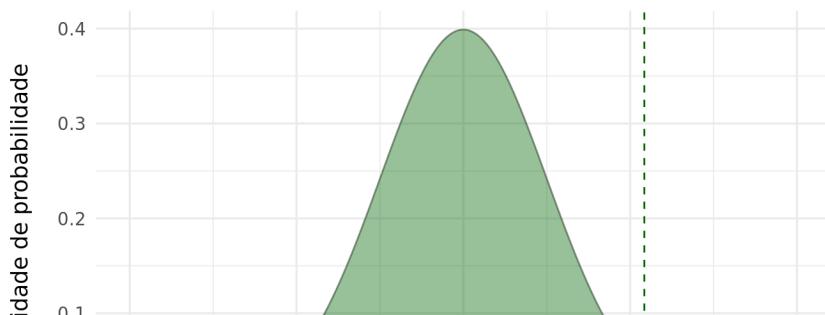
```
pnorm(2.17) |> ppercent()
```

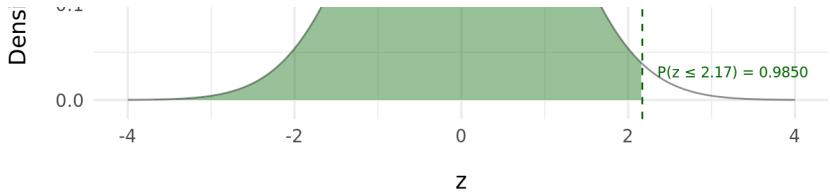
$$P = 0.984997 \approx 98.50\%$$

Que pode ser representado como:

```
z_plot(2.17)
```

Área acumulada para o escore-z 2.17



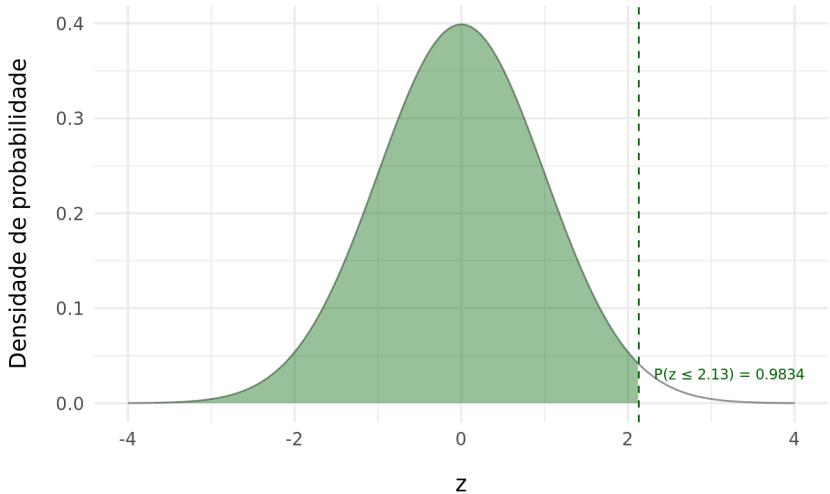


EXERCÍCIO 4

QUESTÃO A

```
z_plot(2.13)
```

Área acumulada para o escore- z 2.13



QUESTÃO B

```
pnorm(2.13) |> ppercent()
```

$$P = 0.983414 \approx 98.34\%$$

EXERCÍCIO 5

QUESTÃO A

Para encontrar a área à direita de um determinado ponto podemos subtrair o mesmo resultado obtido com `pnorm` de 1:

```
1 - pnorm(-2.16)
```

```
[1] 0.9846137
```

Ou podemos usar a opção `lower.tail = FALSE`, fornecida pela própria função:

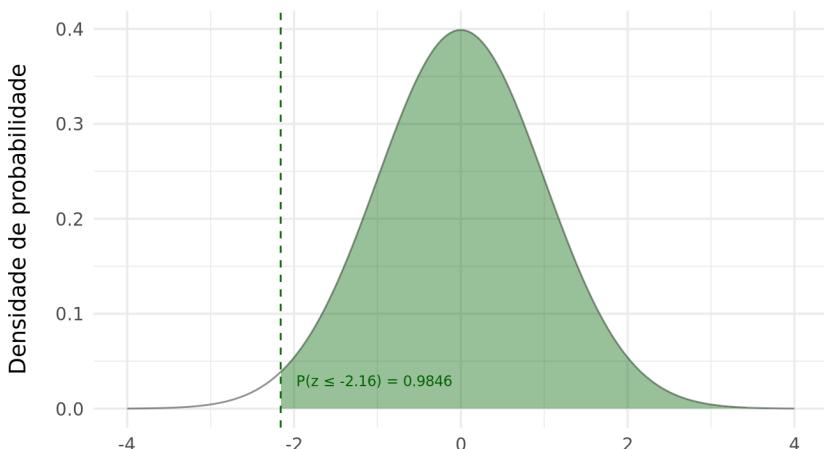
```
pnorm(-2.16, lower.tail = FALSE) |> ppercent()
```

$$P = 0.984614 \approx 98.46\%$$

A curva normal padrão com a área sob a curva à direita de $z = -2.16$ também pode ser obtida com a função que definimos, `z_plot`. Desta vez usamos o parâmetro opcional `right` com o valor `TRUE`, o que retorna o gráfico com a área à direita de z em destaque:

```
z_plot(-2.16, right = TRUE)
```

Área acumulada para o escore-z -2.16



z

QUESTÃO B

```
pnorm(-2.16) |> ppercent()
```

$$P = 0.015386 \approx 1.54\%$$

QUESTÃO C

```
1 - pnorm(-2.16)
```

```
[1] 0.9846137
```

EXERCÍCIO 6

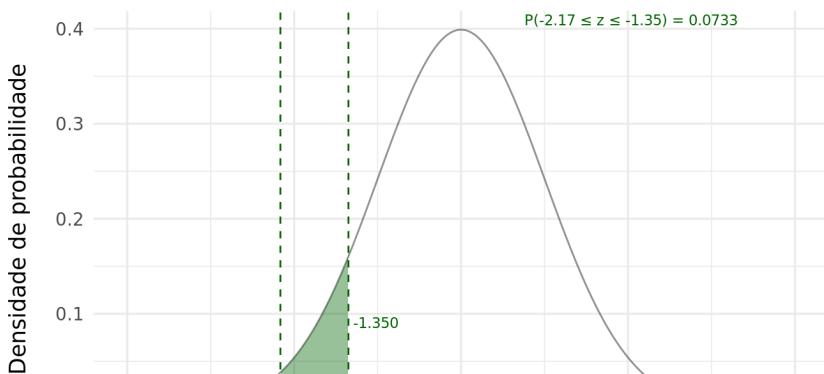
No exercício 6, precisamos encontrar a área *entre* $z = -2.165$ e $z = -1.35$, isto é:

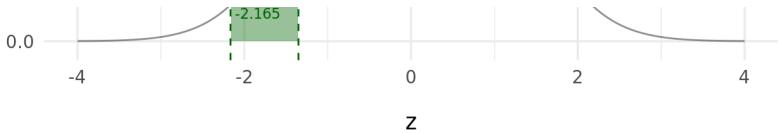
$$P(-2.165 < z < -1.35)$$

QUESTÃO A

```
z_plot(-2.165, right = TRUE, opposite = -1.35)
```

Área entre os escores-z -2.165 e -1.35





QUESTÃO B

```
pnorm(-1.35)
```

[1] 0.08850799

QUESTÃO C

```
pnorm(-2.165)
```

[1] 0.01519384

QUESTÃO D

```
pnorm(-1.35) - pnorm(-2.165)
```

[1] 0.07331415

QUESTÃO E

Seja P a função de distribuição acumulada,

$$z_1 = -2.165$$

$$z_2 = -1.35$$

$$\because z_1 < z_2 \therefore P(z_1 < z < z_2) = P(z_2) - P(z_1)$$

$$P(z_1) = 0.0151938$$

$$P(z_2) = 0.088508$$

$$P = 0.088508 - 0.0151938$$

$$P = 0.073314 \approx 7.33\%$$

Interpretação: A probabilidade de que a média encontre-se entre -2.165 e -1.35 é de aproximadamente 7.33%.

EXERCÍCIO 7

$$\mu = 67$$

$$\sigma = 3.5$$

$$S = P(x > 70)$$

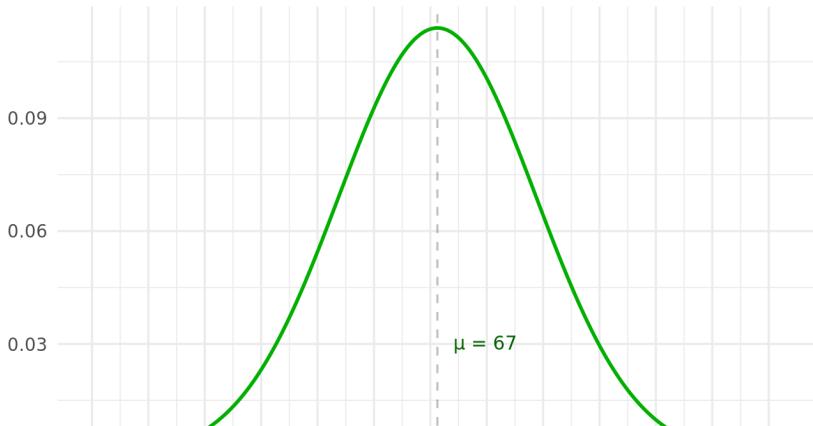
Podemos agregar estes valores em uma lista para acesso posterior:

```
e7 <- list(  
  mu = 67,  
  sigma = 3.5,  
  x = 70  
)
```

QUESTÃO A

```
draft(  
  name = "Velocidade",  
  mu = e7$mu,  
  sigma = e7$sigma,  
  x_min = e7$mu - e7$sigma * 3.5,  
  x_max = e7$mu + e7$sigma * 3.5,  
  x_step = ceiling(e7$sigma / 2),  
)
```

Esboço da Curva Normal "Velocidade"





QUESTÃO B

```
z <- (x - mu) / sigma  
z
```

[1] 0.8571429

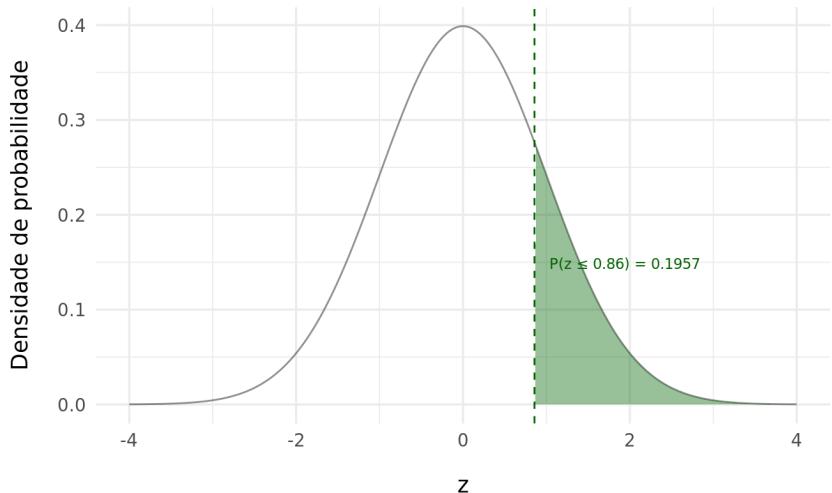
QUESTÃO C

```
(1 - pnorm(z)) |> ppercent()
```

$$P = 0.195683 \approx 19.57\%$$

```
z_plot(z, right = TRUE)
```

Área acumulada para o escore-z 0.8571



QUESTÃO D

QUESTÃO 8

Interpretação: A probabilidade de que um veículo aleatório ultrapasse a velocidade máxima de 70 milhas por hora é de aproximadamente 7.33%

EXERCÍCIO 8

$$\mu = 45$$

$$\sigma = 12$$

$$S = P(33 \leq x \leq 60)$$

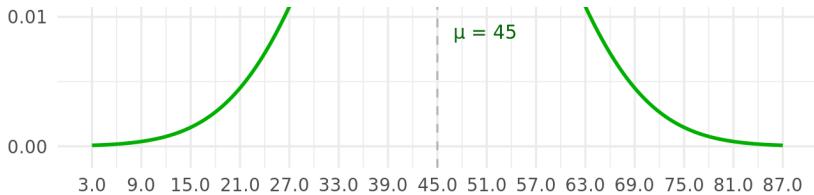
```
e8 <- list(  
  mu = 45,  
  sigma = 12,  
  x1 = 33,  
  x2 = 60  
)
```

QUESTÃO A

```
draft(  
  name = "Permanência",  
  mu = e8$mu,  
  sigma = e8$sigma,  
  x_min = e8$mu - e8$sigma * 3.5,  
  x_max = e8$mu + e8$sigma * 3.5,  
  x_step = ceiling(e8$sigma / 2),  
)
```

Esboço da Curva Normal "Permanência"





QUESTÃO B

```
z1 <- (e8$x1 - e8$mu) / e8$sigma
z1
```

[1] -1

```
z2 <- (e8$x2 - e8$mu) / e8$sigma
z2
```

[1] 1.25

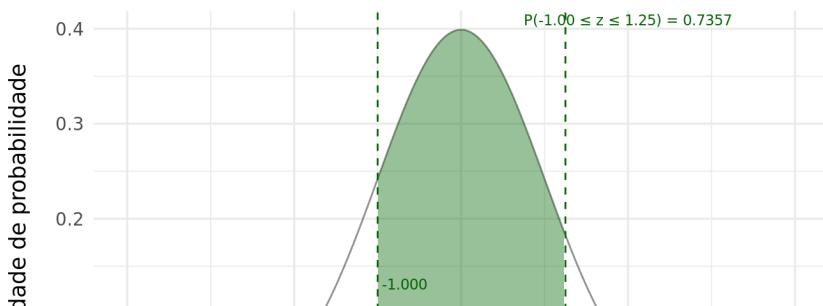
QUESTÃO C

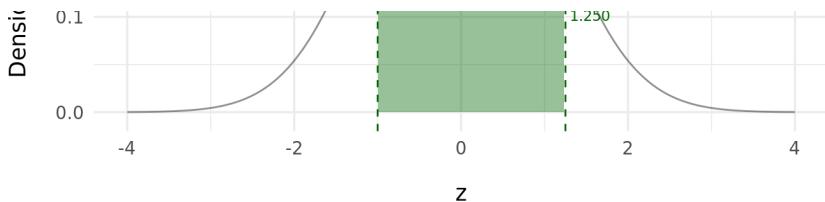
```
p <- abs(pnorm(z2) - pnorm(z1))
p |> ppercent()
```

$$P = 0.735695 \approx 73.57\%$$

```
z_plot(z2, opposite = z1)
```

Área entre os escores-z 1.25 e -1





QUESTÃO D

Interpretação: A probabilidade p de que um consumidor fique na loja entre 33 e 60 minutos é de aproximadamente 0.735695, ou 73.57%. Para 150 consumidores aleatórios, espera-se que aproximadamente $150 \cdot p = 111$ consumidores permaneceriam na loja entre 33 e 60 minutos.

EXERCÍCIO 9

Não há informação suficiente no enunciado para obter a probabilidade de que este evento ocorra.

EXERCÍCIO 10

QUESTÃO 10.1

A função `qnorm` pode fornecer um escore-z correspondente a uma dada probabilidade.

```
z <- qnorm(0.9616)
z
```

[1] 1.769563

QUESTÃO 10.2

Seja α o valor restante, de 0.05, podemos obter o escore-z da diferença entre 1 e $\frac{\alpha}{2}$:

```
alpha <- 0.05
z95 <- qnorm(1 - alpha/2)
-z95
```

z95

```
[1] 1.959964
```

QUESTÃO 10.2 A

```
pnorm(z95)
```

```
[1] 0.975
```

A área acumulada é de 0.975.

QUESTÃO 10.2 B

```
qnorm(pnorm(z95))
```

```
[1] 1.959964
```

A área corresponde a 1.959964 na tabela padrão.

QUESTÃO 10.2 C

O escore-z correspondente é `qnorm(1 - alpha/2)`, 1.959964.

EXERCÍCIO 11

QUESTÕES A A C

Podemos obter cada percentil pela porcentagem ainda através da função `qnorm`.

```
p10 <- qnorm(0.1)
p20 <- qnorm(0.2)
p99 <- qnorm(0.99)
```

```
p10
```

```
[1] -1.281552
```

```
p20
```

```
[1] -0.8416212
```

p99

[1] 2.326348

Temos, portanto:

$$P_{10} = -1.2815516$$

$$P_{20} = -0.8416212$$

$$P_{99} = 2.3263479$$

EXERCÍCIO 12

$$\mu = 52$$

$$\sigma = 15$$

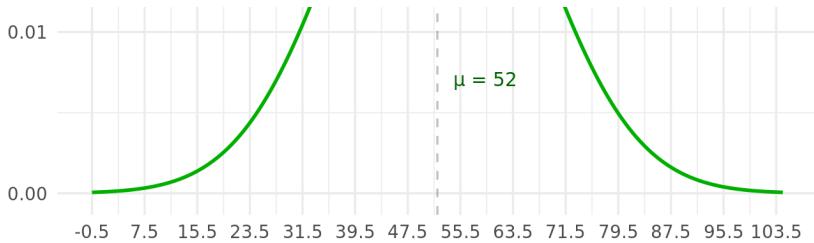
QUESTÃO A

```
e12 <- list(  
  mu = 52,  
  sigma = 15,  
  x1 = NULL,  
  x2 = NULL,  
  x2 = NULL,  
  z1 = -2.33,  
  z2 = 3.1,  
  z3 = 0.58  
)
```

► Código

Esboço da Curva Normal "Pesos"





QUESTÃO B

Podemos inverter a fórmula para obter um escore-z e teremos:

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

Portanto, para cada z_n em $\{-2.33, 3.1, 0.58\}$, podemos obter um x_n :

```
e12$x1 <- e12$mu + e12$sigma * e12$z1
e12$x1
```

[1] 17.05

```
e12$x2 <- e12$mu + e12$sigma * e12$z2
e12$x2
```

[1] 98.5

```
e12$x3 <- e12$mu + e12$sigma * e12$z3
e12$x3
```

[1] 60.7

QUESTÃO C

Os pesos correspondentes são:

$$z_1 = -2.33 \rightarrow x_1 = 17.05$$

$$z_2 = 3.1 \rightarrow x_2 = 98.5$$

$$z_3 = 0.58 \rightarrow x_3 = 60.7$$

EXERCÍCIO 13

$$\mu = 129$$

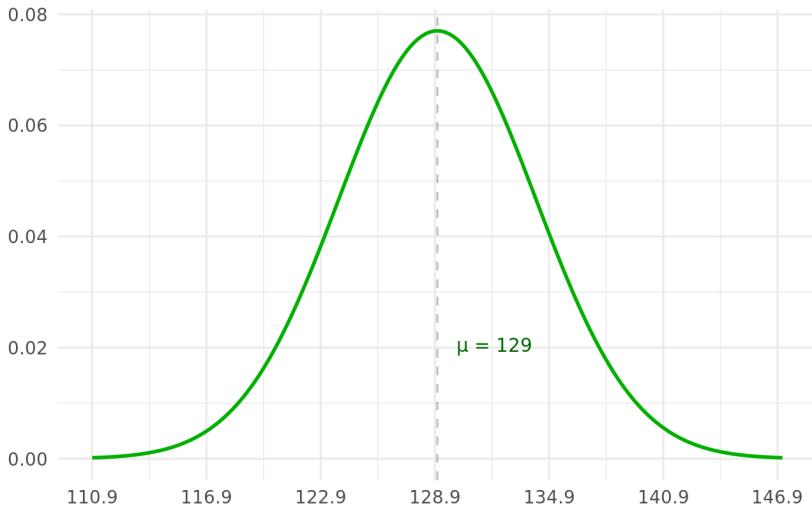
$$\sigma = 5.18$$

QUESTÃO A

```
e13 <- list(  
  mu = 129,  
  sigma = 5.18,  
  x = NULL,  
  z = NULL  
)
```

► Código

Esboço da Curva Normal "Frenagem"



QUESTÃO B

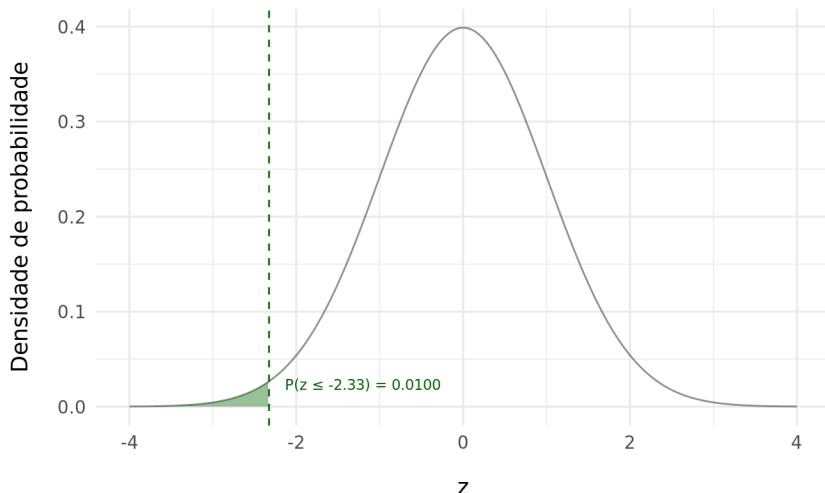
Podemos encontrar o escore-z para 0.01 também com a função `qnorm`:

```
e13$z <- qnorm(0.01)  
e13$z
```

```
[1] -2.326348
```

```
z_plot(e13$z)
```

Área acumulada para o escore-z -2.3263



QUESTÃO C

Mais uma vez, podemos usar:

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

Teremos:

$$x = 129 + 5.18 \cdot -2.3263479$$

```
e13$x <- e13$mu + e13$sigma * e13$z  
e13$x
```

```
[1] 116.9495
```

QUESTÃO D

Interpretação: A maior distância de frenagem que carro um aleatório poderia ter e ainda estar no grupo dos 1% mais baixos é de

aproximadamente 116.95 pés.

EXERCÍCIO 14

```
e14 <- list(  
  mu = 11.2,  
  sigma = 2.1,  
  p = 0.1,  
  x = NULL,  
  z = NULL  
)
```

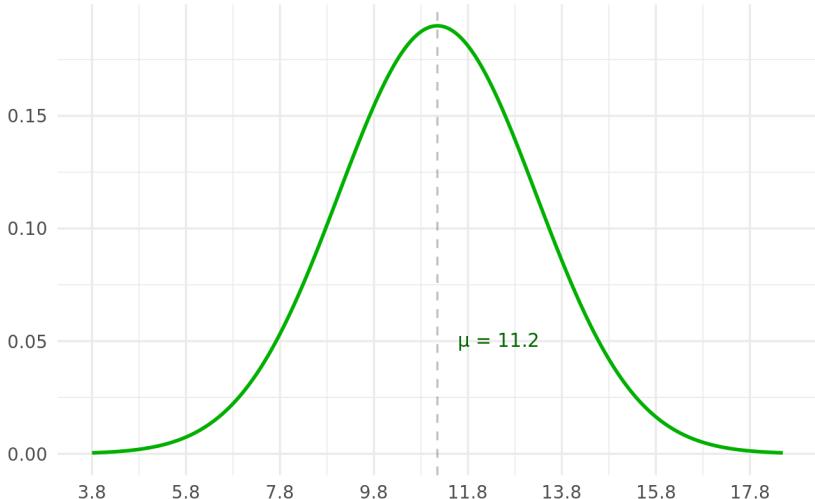
QUESTÃO A

$$\mu = 11.2$$

$$\sigma = 2.1$$

► Código

Esboço da Curva Normal "Tempo de Trabalho"



QUESTÃO B

A função `pnorm` novamente pode nos ajudar a encontrar o score- z .

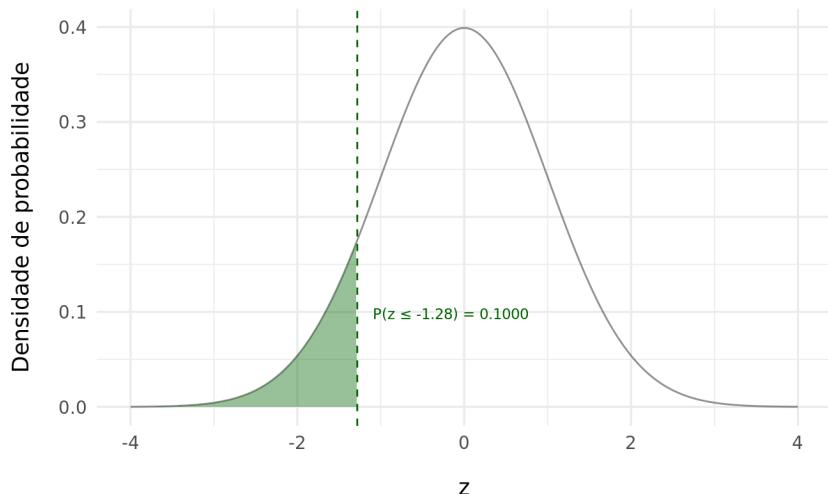
A função `qnorm` novamente pode nos ajudar a encontrar o escore-z para 0.1:

```
e14$z <- qnorm(e14$p)
e14$z
```

```
[1] -1.281552
```

```
z_plot(e14$z)
```

Área acumulada para o escore-z -1.2816



QUESTÃO C

Novamente, pela fórmula:

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

Teremos:

$$x = 11.2 + 2.1 \cdot -1.2815516$$

```
e14$x <- e14$mu + e14$sigma * e14$z
e14$x
```

QUESTÃO D

Interpretação: O tempo máximo que um funcionário pode ter trabalhado na empresa e ainda assim ser demitido é de aproximadamente 8 anos e 6 meses.

EXERCÍCIO 15

► Código

$$\mu = 47 = \bar{x}$$

$$\sigma = 9$$

$$n = 64$$

Para o desvio padrão amostral, temos:

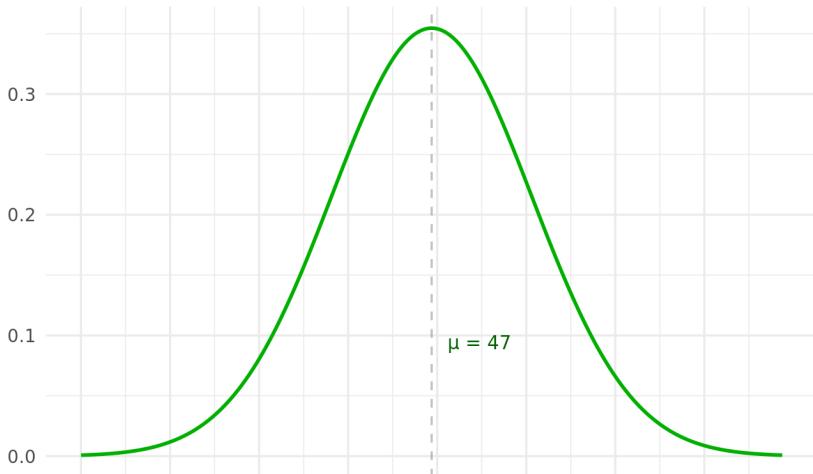
```
e15$s <- e15$sigma / sqrt(e15$n)
```

$$s = 1.125$$

$$\therefore n \geq 30 \therefore \bar{X} \sim N(47, 1.125^2)$$

► Código

Esboço da Curva Normal "Contas de Telefone (Amostral)"



43.1 44.1 45.1 46.1 47.1 48.1 49.1 50.1

EXERCÍCIO 16

- Código

$$\mu = 3.5 = \bar{x}$$

$$\sigma = 0.2$$

$$n = 16$$

Quanto ao erro padrão da distribuição amostral (desvio padrão amostral):

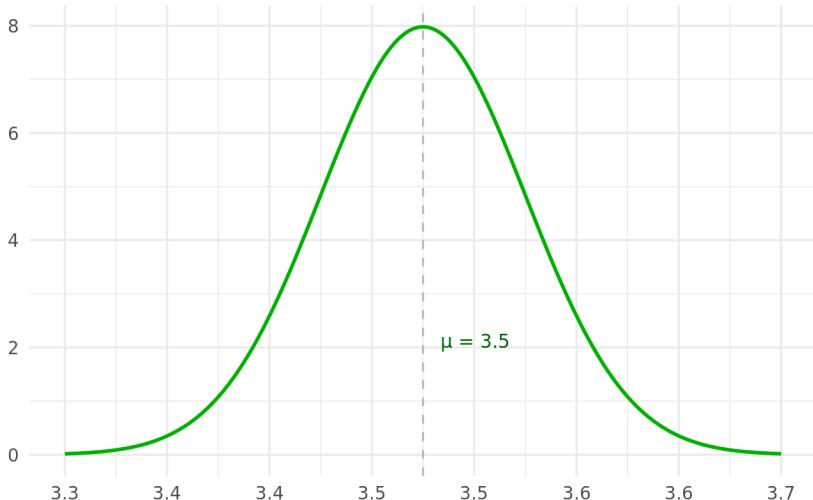
```
e16$s <- e16$sigma / sqrt(e16$n)
```

$$s = 0.05$$

A amostra não é grande o suficiente, mas a população é normalmente distribuída. Logo, podemos usar a distribuição normal:

- Código

Esboço da Curva Normal "Diâmetros (Amostral)"



EXERCÍCIO 17

► Código

Nota: O valor da amostra repete-se no enunciado, como 50 e 100. Foi presumido o segundo valor, 100, por sua compatibilidade com a pergunta.

QUESTÃO A

$$\mu = 25 = \bar{x}$$

$$\sigma = 1.5$$

$$n = 100$$

Desvio padrão amostral:

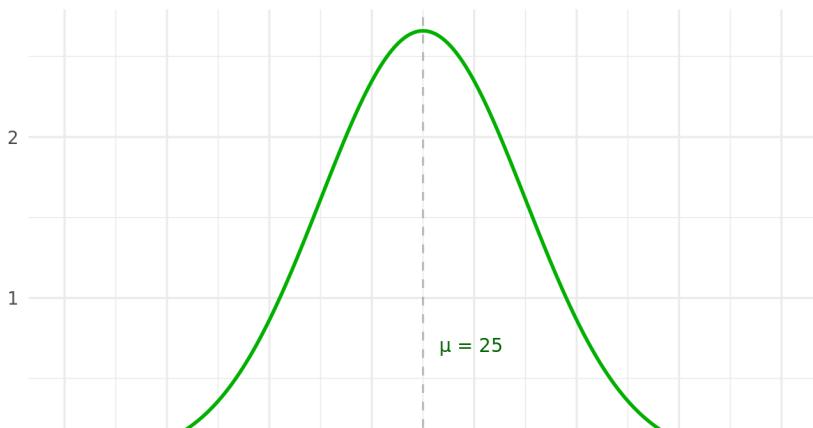
```
e17$s <- e17$sigma / sqrt(e17$n)
```

$$s = 0.15$$

$$\therefore n \geq 30 \therefore \bar{X} \sim N(25, 0.15^2)$$

► Código

Esboço da Curva Normal "Tempo Dirigindo (Amostral)"





QUESTÃO B E C

Ainda precisamos determinar $P(24.7 < x < 25.5)$.

Podemos obter P através de:

$$P = P(25.5) - P(24.7)$$

Podemos usar `pnorm` novamente para expressar o mesmo em R:

```
(pnorm(e17$x2, mean = e17$mu, sd = e17$s) -  
pnorm(e17$x1, mean = e17$mu, sd = e17$s)) |>  
ppercent()
```

$$P = 0.976821 \approx 97.68\%$$

QUESTÃO D

Interpretação: A probabilidade de que o tempo médio que 100 motoristas aleatórios passem entre 24.7 e 25.5 minutos dirigindo é de aproximadamente 97.68%.

EXERCÍCIO 18

► Código

QUESTÃO A

$$\mu = 1.768 \times 10^5 = \bar{x}$$

$$\sigma = 5 \times 10^4$$

$$n = 12$$

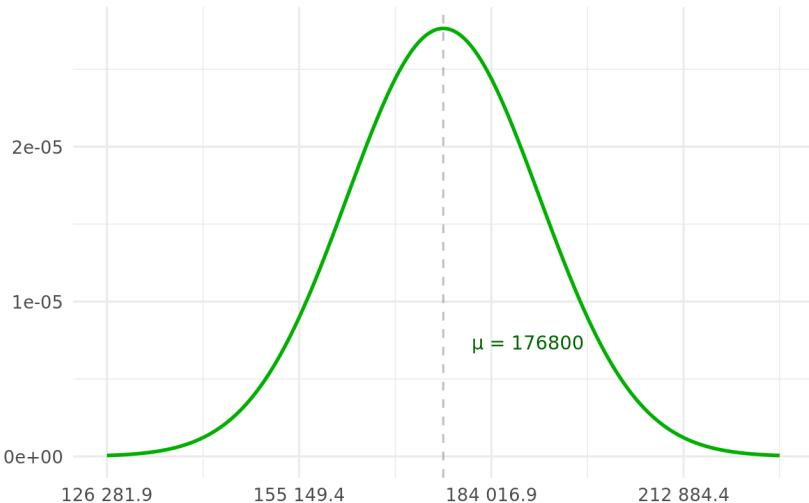
Desvio padrão amostral:

```
e18$s <- e18$sigma / sqrt(e18$n)
```

$$s = 1.4433757 \times 10^4$$

► Código

Esboço da Curva Normal "Residências (Amostral)"

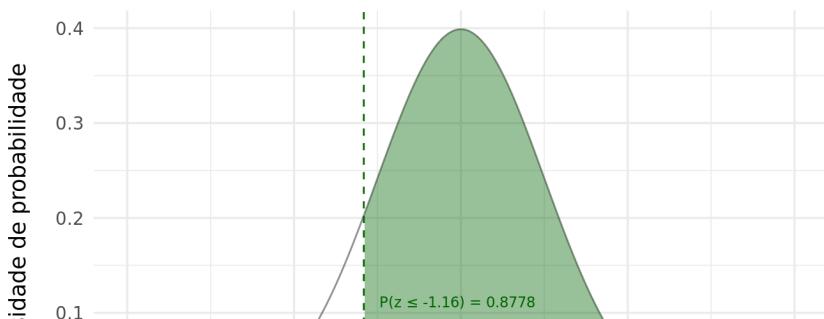


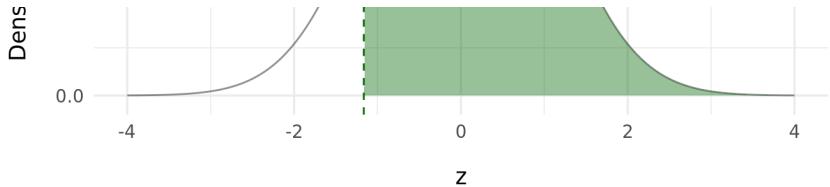
QUESTÃO B

```
e18$z <- (e18$x - e18$mu) / e18$s
```

```
z_plot(e18$z, right = TRUE)
```

Área acumulada para o escore-z -1.1639





QUESTÃO C

```
pnorm(e18$z, lower.tail = FALSE) |> ppercent()
```

$$P = 0.877775 \approx 87.78\%$$

QUESTÃO D

Interpretação: Há uma probabilidade de aproximadamente 87.78% que o preço de vendas médio seja maior que US\$ 160 000.

EXERCÍCIO 19

- Código

$$\mu = 190 = \bar{x}$$

$$\sigma = 48$$

$$n = 10$$

Desvio padrão amostral:

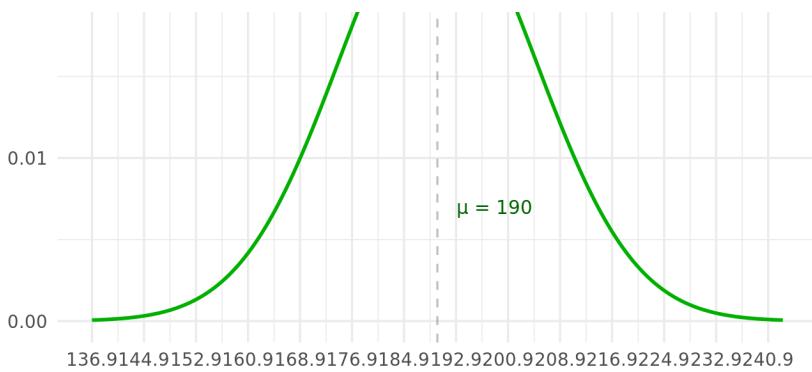
```
e19$s <- e19$sigma / sqrt(e19$n)
```

$$s = 15.1789328$$

- Código

Esboço da Curva Normal "Monitores (Amostral)"





```
e19$z <- (e19$x - e19$mu) / e19$sigma
e19$z_barra <- (e19$x - e19$mu) / e19$s
```

QUESTÃO B

```
pnorm(e19$z) |> ppercent()
```

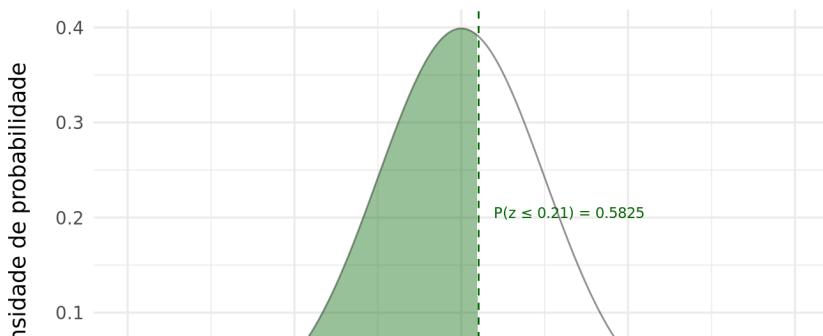
$$P = 0.582516 \approx 58.25\%$$

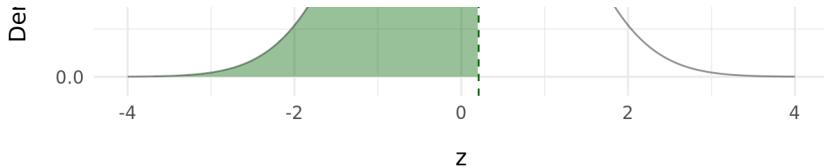
```
pnorm(e19$z_barra) |> ppercent()
```

$$P = 0.744990 \approx 74.50\%$$

```
z_plot(e19$z)
```

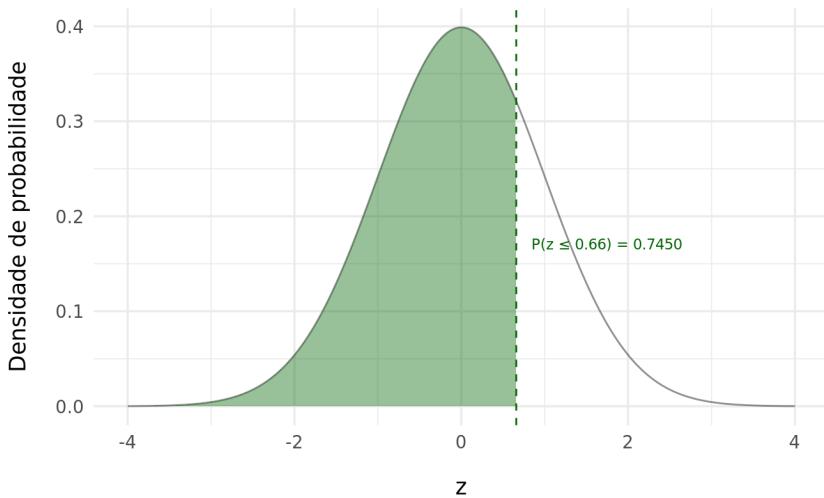
Área acumulada para o escore-z 0.2083





```
z_plot(e19$z_barra)
```

Área acumulada para o escore-z 0.6588



QUESTÃO C

A amostra apresenta uma probabilidade maior. Isto é esperado, já que ela restringe a quantidade de casos extremos.

EXERCÍCIO 20

► Código

QUESTÃO A

```
e20$q <- 1 - e20$p
```

$$n = 100$$

$$p = 0.34$$

$$q = 0.66$$

QUESTÃO B

```
e20$np <- e20$n * e20$p  
e20$nq <- e20$n * e20$q
```

$$np = 34$$

$$nq = 66$$

QUESTÃO C

Podemos usar uma distribuição normal para aproximar, pois $np \geq 5 \wedge nq \geq 5$.

QUESTÃO D

```
mu_b <- e20$np  
sigma_b <- sqrt(e20$n * e20$p * e20$q)
```

$$\mu = 34$$

$$\sigma = 4.7371$$

EXERCÍCIO 21

► Código

```
e21$q <- 1 - e21$p  
e21$np <- e21$n * e21$p  
e21$nq <- e21$n * e21$q
```

$$n = 100$$

$$p = 0.34$$

$$q = 0.66$$

$$np = 34$$

$$nq = 66$$

QUESTÃO A

Novamente podemos usar uma distribuição normal, dado que $np \geq 5 \wedge nq \geq 5$.

QUESTÃO B

```
e21$mu <- e21$np  
e21$sigma <- sqrt(e21$n * e21$p * e21$q)
```

$$\mu = 34$$

$$\sigma = 4.7371$$

QUESTÃO C

Para corrigir $P(x > 30)$, usaremos $x \geq 30.5$.

```
e21$a <- 30.5
```

QUESTÃO D

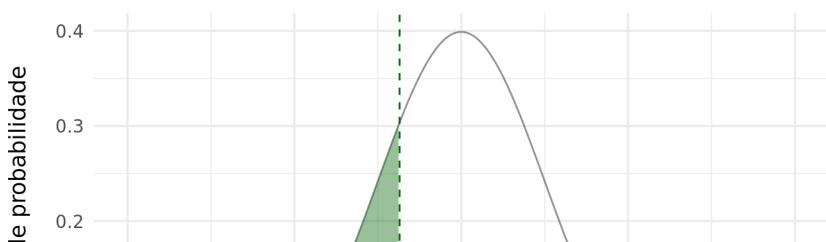
```
e21$z <- (e21$a - e21$mu) / e21$sigma  
e21$ps <- pnorm(e21$z, lower.tail = FALSE)
```

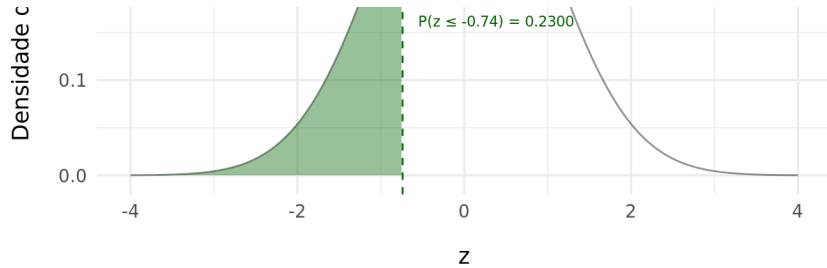
$$z = -0.7388506$$

$$P(X > 30) \approx P(Z > z) = 0.77$$

```
z_plot(e21$z)
```

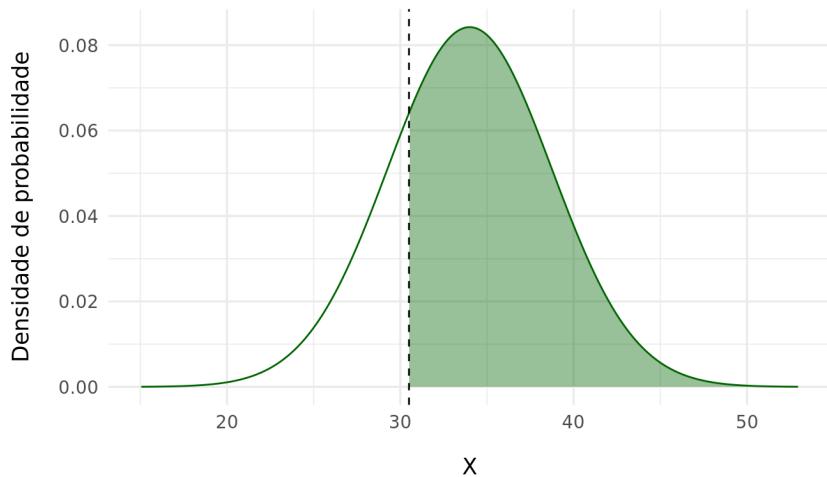
Área acumulada para o escore-z -0.7389





► Código

Aproximação Normal para Binomial, $n = 100$ $p = 0.34$
Em destaque, $P(X > 30)$ com correção de continuidade



EXERCÍCIO 22

► Código

```
e22$q <- 1 - e22$p
e22$np <- e22$n * e22$p
e22$nq <- e22$n * e22$q
```

$$n = 100$$

$$p = 0.58$$

$$q = 0.42$$

$$np = 58$$

$$nq = 42$$

QUESTÃO A

Mais uma vez podemos usar uma distribuição normal, dado que $np \geq 5 \wedge nq \geq 5$.

QUESTÃO B

```
e22$mu <- e22$np  
e22$sigma <- sqrt(e22$n * e22$p * e22$q)
```

$$\mu = 58$$

$$\sigma = 4.9356$$

EXERCÍCIO 23

► Código

```
e23$q <- 1 - e23$p  
e23$np <- e23$n * e23$p  
e23$nq <- e23$n * e23$q
```

$$n = 75$$

$$p = 0.32$$

$$q = 0.68$$

$$np = 24$$

$$nq = 51$$

QUESTÃO A

Ainda podemos usar uma distribuição normal, dado que $np \geq 5 \wedge nq \geq 5$.

QUESTÃO B

```
e23$mu <- e23$np  
e23$sigma <- sqrt(e23$n * e23$p * e23$q)
```

$$\mu = 24$$

$$\sigma = 4.0398$$

QUESTÃO C

```
e23$xi <- e23$x - 0.5  
e23$xs <- e23$x + 0.5  
  
e23$zi <- (e23$xi - e23$mu) / e23$sigma  
e23$zs <- (e23$xs - e23$mu) / e23$sigma
```

$$x_i = 14.5$$

$$x_s = 15.5$$

$$z_i = -2.3516004$$

$$z_s = -2.1040635$$

QUESTÃO D

Para aproximar $P(X = 15)$ para a distribuição normal:

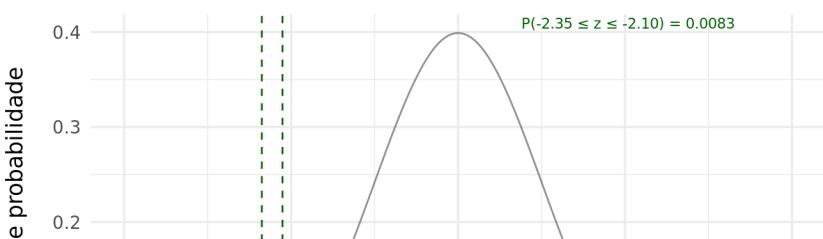
```
pnorm(e23$zs) - pnorm(e23$zi) |> percent()
```

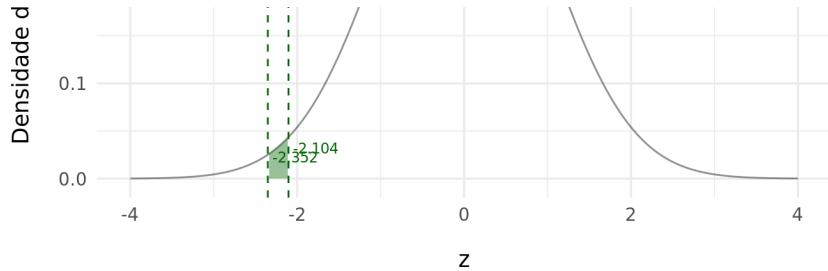
$$P = 0.009346 \approx 0.93\%$$

```
numeric(0)
```

```
z_plot(e23$zs, opposite = e23$zi)
```

Área entre os escores-z -2.1041 e -2.3516





EXERCÍCIO 24

O problema apresentado neste exercício é mais simples por poder ser reduzido a um vetor tal que:

```
hours <- c(
  26, 25, 32, 31, 28, 28,
  28, 22, 28, 25, 21, 40,
  32, 22, 25, 22, 26, 24,
  46, 20, 35, 22, 32, 48,
  32, 36, 38, 32, 22, 19
)
```

QUESTÕES A E B

Com isso, podemos obter a média amostral, que corresponde também à média populacional:

```
x_bar <- mean(hours)
x_bar
```

[1] 28.9

EXERCÍCIO 25

QUESTÃO A

Usando o mesmo vetor, teremos:

```
n      <- length(hours)      # tamanho da amostra
```

```
sigma <- 7.9                      # desvio populacional
alpha <- 0.05                       # 95% de confiança
z_c   <- qnorm(1 - alpha/2) # z-crítico
```

QUESTÃO B

Podemos obter a margem de erro por $z_c \cdot \sigma / \sqrt{n}$:

```
E = z_c * sigma / sqrt(n)
```

QUESTÃO C

$$n = 30$$

$$\sigma = 7.9$$

$$\alpha = 0.05$$

$$z_c = 1.959964$$

$$E = 2.8269267$$

Interpretação: Com 95% de confiança, a média de horas semanais está no intervalo (26.31; 31.95).

EXERCÍCIO 26

```
n <- length(hours)
s <- sd(hours)
alpha <- 0.05
f <- n - 1 # degrees of freedom
t_c <- qt(1 - alpha/2, f)
```

QUESTÃO A

```
x_bar <- mean(hours)
E <- t_c * s / sqrt(n)
```

$$\bar{x} = 28.9$$

$$E = 2.7480781$$

QUESTÃO B

```
li <- x_bar - E  
ls <- x_bar + E
```

$$l_i = 26.1519219$$

$$l_s = 31.6480781$$

QUESTÃO C

Interpretação: Com 95% de confiança, a média populacional de horas semanais está no intervalo (26.15; 31.65).

EXERCÍCIO 27

Novamente, podemos usar apenas um vetor para os dados:

```
hours <- c(  
  30, 26, 33, 26, 26, 33, 31, 31, 21, 37,  
  27, 20, 34, 35, 30, 24, 38, 34, 39, 31,  
  22, 30, 23, 23, 31, 44, 31, 33, 33, 26,  
  27, 28, 25, 35, 23, 32, 29, 31, 25, 27  
)  
  
sigma <- 7.9  
E <- 2  
alpha <- 0.05  
z_c <- qnorm(1 - alpha / 2)
```

$$\sigma = 7.9$$

$$E = 2$$

$$\alpha = 0.05$$

$$z_c = 1.959964$$

Dadas estas informações, podemos obter:

$$n = (z_c \cdot \sigma / E)^2$$

```
n <- ceiling((z_c * sigma / E)^2)  
n
```

[1] 60

Interpretação: Pelo menos 60 funcionários devem ser incluídos na

amostra para ter 95% de confiança de que a diferença máxima entre a média amostral e a média populacional seja de 2 horas.

EXERCÍCIO 28

QUESTÃO A

Dada uma amostra de $n = 22$, podemos obter o grau de liberdade $f = n - 1$:

```
n <- 22  
f <- n - 1
```

QUESTÃO B

```
c <- 0.90
```

QUESTÃO C

Novamente, usamos a função `qt` para obter o valor t_c .

```
alpha <- 1 - c  
t_c <- qt(1 - alpha / 2, f)
```

QUESTÃO D

O valor crítico t_c para um nível de confiança de 90% com uma amostra de 22 é 1.7207429.

EXERCÍCIO 29

Podemos definir as seguintes variáveis a partir do enunciado:

```
n <- 16  
mu <- 162  
s <- 10  
f <- n - 1  
se <- s / sqrt(n)
```

Com elas, podemos então construir duas listas, cada uma contendo os

limites superiores e inferiores para os intervalos de confiança de 90% e 99%:

```
## 90%
alpha90 <- 0.10
t90 <- qt(1 - alpha90 / 2, f)
e90 <- t90 * se
ci90 <- list(i = x_bar - e90, s = x_bar + e90)

## 99%
alpha99 <- 0.01
t99 <- qt(1 - alpha99 / 2, f)
e99 <- t99 * se
ci99 <- list(i = x_bar - e99, s = x_bar + e99)
```

Os valores obtidos mostram:

- Intervalo de confiança de 90%
 - Limite inferior 24.5174
 - Limite superior 33.2826
- Intervalo de confiança de 99%
 - Limite inferior 21.5332
 - Limite superior 36.2668

EXERCÍCIO 30

Novamente, podemos definir as seguintes variáveis a partir do enunciado:

```
n      <- 36
mu    <- 9.75
s      <- 2.39
f      <- n - 1
se    <- s / sqrt(n)
```

```
## 90%
t90  <- qt(0.95, f)
e90  <- t90 * se
ci90 <- list(i = mu - e90, s = x_bar + e90)

## 95%
t95  <- qt(0.975, f)
```

```
e95 <- t95 * se  
ci95 <- list(i = mu - e95, s = x_bar + e95)
```

Temos, portanto, os intervalos de confiança:

```
ci90
```

```
$i  
[1] 9.076987
```

```
$s  
[1] 29.57301
```

```
ci95
```

```
$i  
[1] 8.94134
```

```
$s  
[1] 29.70866
```

QUESTÃO A

- Intervalo de confiança de 90%
 - $t_c = 1.6896$
 - $E = 0.673$
- Intervalo de confiança de 95%
 - $t_c = 2.0301$
 - $E = 0.8087$

QUESTÃO B

- Intervalo de confiança de 90%
 - Limite inferior 9.077
 - Limite superior 29.573
- Intervalo de confiança de 95%
 - Limite inferior 8.9413
 - Limite superior 29.7087

QUESTÃO C

Podemos comparar as larguras usando uma subtração simples:

- Intervalo de confiança de 90%
 - Largura: 20.496
- Intervalo de confiança de 95%
 - Largura: 20.7673

EXERCÍCIO 31

Como o desvio padrão da população é desconhecido e temos uma amostra pequena, devemos usar a distribuição t com um índice de liberdade de $n - 1$.

Mais uma vez, vamos primeiro designar algumas variáveis com os valores do enunciado:

```
n <- 18
x_bar <- 64
s <- 2.5
f <- n - 1
```

Dado o intervalo de confiança de 90%, temos $\alpha = 0.1$:

```
alpha <- 0.10
t_c <- qt(1 - alpha / 2, f)
se <- s / sqrt(n)
```

E com isso podemos obter a margem de erro:

```
me <- t_c * se
```

Nosso intervalo de confiança estará, portanto, entre:

```
c_i <- x_bar - me
c_s <- x_bar + me
```

Ou seja, entre 62.9749 e 65.0251.

EXERCÍCIO 32

QUESTÃO A

—

Pelo enunciado, temos:

```
x <- 123  
n <- 2462
```

QUESTÃO B

Podemos obter \hat{p} com uma proporção simples de $\frac{x}{n}$.

```
p_h <- x / n
```

$$\hat{p} = 0.0499594$$

EXERCÍCIO 33

33. Use os dados do exercício 32 para construir um intervalo de confiança de 90% para a proporção populacional de professores americanos que disseram que “todas ou quase todas” as informações que eles encontram utilizando ferramentas de busca on-line são corretas ou confiáveis.

QUESTÃO A

```
x      <- 123  
n      <- 2462  
p_h <- x / n  
q_h <- 1 - p_h
```

$$\hat{p} = 0.0499594$$

$$\hat{q} = 0.9500406$$

QUESTÃO B

```
np <- n * p_h  
nq <- n * q_h
```

A distribuição amostral pode ser aproximada por uma distribuição normal pois $np = 123$ $nq = 2339$.

QUESTÃO C

```
z_c <- qnorm(0.95)
e <- z_c * sqrt(p_h * q_h / n)
```

$$z_c = 1.6448536$$

$$E = 0.0072221$$

QUESTÃO D

```
l_i <- p_h - e
l_s <- p_h + e
```

QUESTÃO E

O intervalo de confiança populacional é aproximadamente (0.0435; 0.0565).

Com 90% de confiança, entre 4.35% e 5.65% dos professores dos Estados Unidos responderiam “todas ou quase todas”.

EXERCÍCIO 34

QUESTÃO A

```
n <- 498
p_h <- 0.25
q_h <- 1 - p_h
p_h
```

[1] 0.25

```
q_h
```

[1] 0.75

$$\hat{p} = 0.25$$

$$\hat{q} = 0.75$$

QUESTÃO B

```
no <- n * p_h
```

```
nq <- n * q_h
```

A amostra pode ser aproximada pela distribuição normal dado que \bar{x} np = 124.5 nq = 373.5 \$.

QUESTÃO C

```
z_c <- qnorm(0.995)
e <- z_c * sqrt(p_h * q_h / n)
```

$$z_c = 2.5758293$$

$$E = 0.04999808$$

QUESTÃO D

```
l_i <- p_h - e
l_s <- p_h + e
```

$$l_i = 0.2000192$$

$$l_s = 0.2999808$$

QUESTÃO E

Com 99% de confiança, a proporção de adultos americanos que consideram pessoas acima de 65 anos os motoristas mais perigosos está entre aproximadamente 20% e 30%.

EXERCÍCIO 35

QUESTÃO A

```
p_h1 <- 0.5
q_h1 <- 1 - p_h1

p_h2 <- 0.31
q_h2 <- 1 - p_h2

z_c <- qnorm(0.95)
e <- 0.02
```

QUESTÃO B

A partir de,

$$\frac{n = (z_c^2 \cdot \hat{p} \cdot \hat{q})}{E^2}$$

Temos:

```
n1 <- (z_c^2 * p_h1 * q_h1) / e^2  
n2 <- (z_c^2 * p_h2 * q_h2) / e^2
```

$$n_1 = 1690.9646588$$

$$n_2 = 1446.7893621$$

QUESTÃO C

Dado o domínio da aplicação, precisamos dos valores arredondados:

```
n1 <- ceiling(n1)  
n2 <- ceiling(n2)
```

Sem estimativa prévia, precisa-se de 1691 pessoas Com $\hat{p} = 0.31$,
precisa-se de 1447 pessoas

EXERCÍCIO 36

36. Encontre os valores críticos χ^2_R e χ^2_L para um intervalo de confiança de 90% quando o tamanho da amostra é 30.

QUESTÃO A

```
n <- 30  
f <- n - 1  
conf <- 0.90  
alpha <- 1 - conf  
f
```

[1] 29

```
alpha
```

```
[1] 0.1
```

QUESTÃO B

As áreas são 0.05 à direita e 0.05 à esquerda.

QUESTÃO C

Podemos usar a distribuição chi quadrado. Na linguagem R, a função `qchisq` oferece o que precisamos:

```
c_r <- qchisq(1 - alpha / 2, f)
c_l <- qchisq(alpha / 2, f)
```

$$c_r = 42.5569678$$

$$c_l = 17.7083662$$

QUESTÃO D

Com 90% de confiança, χ^2 está entre aproximadamente 17.71 e aproximadamente 42.56.

EXERCÍCIO 37

QUESTÃO A

```
n <- 30
f <- n - 1
s <- 1.20

conf1 <- 0.90; alpha1 <- 1 - conf1
conf2 <- 0.95; alpha2 <- 1 - conf2

c_r1 <- qchisq(1 - alpha1 / 2, f)
c_l1 <- qchisq(alpha1 / 2, f)

c_r2 <- qchisq(1 - alpha2 / 2, f)
c_l2 <- qchisq(alpha2 / 2, f)
```

QUESTÃO B

```
var_est <- (f * s^2)

# 90%
v1_i <- var_est / c_r1
v1_s <- var_est / c_l1

# 95%
v2_i <- var_est / c_r2
v2_s <- var_est / c_l2
```

QUESTÃO C

```
sd1_i <- sqrt(v1_i)
sd1_s <- sqrt(v1_s)

sd2_i <- sqrt(v2_i)
sd2_s <- sqrt(v2_s)
```

QUESTÃO D

- Intervalo de confiança de 90%
 - Para a variância: 0.98 a 2.36
 - Para o desvio padrão: 0.99 a 1.54
- Intervalo de confiança de 95%
 - Para a variância: 0.91 a 2.6
 - Para o desvio padrão: 0.96 a 1.61

EXERCÍCIO 38

QUESTÃO A

1. $\mu \neq 74$
2. $\sigma^2 \leq 2.7$
3. $p > 0.24$

QUESTÃO B

1. $\mu = 74$
2. $\sigma^2 > 2.7$
3. $p \geq 0.24$

QUESTÃO C



1. H_0 e H_a para média

- $H_0 : \mu = 74$
- $H_a : \mu \neq 74$ (afirmação)

2. H_0 e H_a para variância

- $H_0 : \sigma^2 \leq 2.7$ (afirmação)
- $H_a : \sigma^2 > 2.7$

3. H_0 e H_a para proporção

- $H_0 : p \leq 0.24$
- $H_a : p > 0.24$ (afirmação)

EXERCÍCIO 39

QUESTÃO A

- $H_0 : p \leq 0.01$
- $H_a : p > 0.01$

QUESTÃO B

- Erro tipo I: rejeitar H_0 quando $p \leq 0.01$, concludendo que a taxa é $> 1\%$ quando na realidade é $\leq 1\%$
- Erro tipo II: não rejeitar H_0 quando $p > 0.01$, concludendo que a taxa é $\leq 1\%$ quando na realidade é $> 1\%$

QUESTÃO C

O erro tipo II é mais sério.

EXERCÍCIO 40

QUESTÃO A

1. Analista de consumo

- $H_0 : \mu = 74$
- $H_a : \mu \neq 74$

2. Corretor de imóveis

- $H_0: p = 0.24$
- $H_a: p > 0.24$

QUESTÃO B

1. Teste bilateral: $\mu \neq 74$

```
z0 <- 2.1  
p_b <- 2 * (1 - pnorm(abs(z0)))  
p_b
```

[1] 0.03572884

2. Teste unilateral à direita: $p > 0.24$

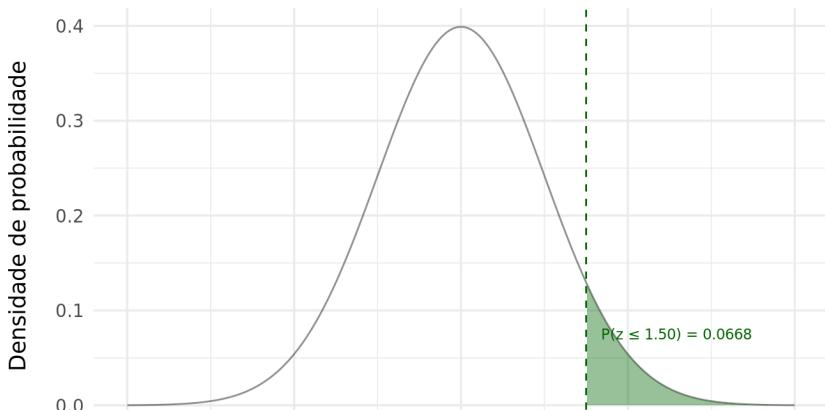
```
z0 <- 1.5  
p_u <- 1 - pnorm(z0)  
p_u
```

[1] 0.0668072

QUESTÃO C

```
z_plot(qnorm(1 - p_u), right = TRUE)
```

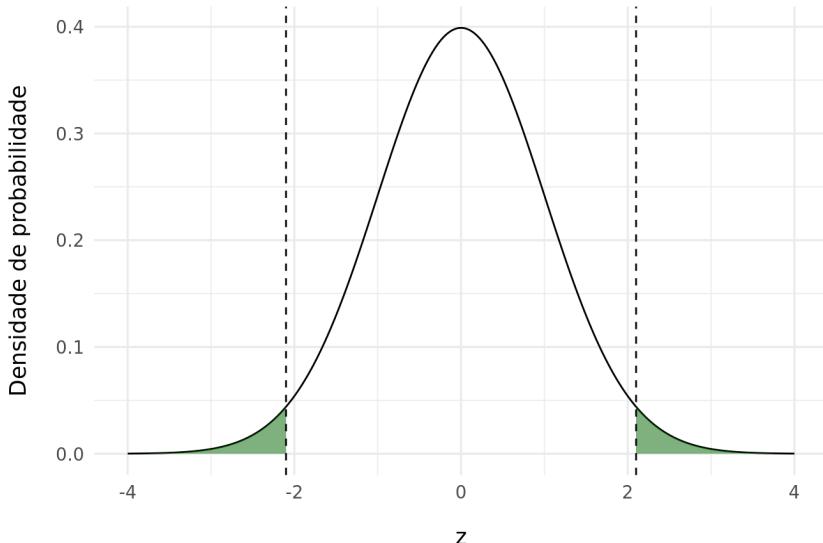
Área acumulada para o escore-z 1.5





► Código

Teste Bilateral



EXERCÍCIO 41

INTERPRETAÇÃO: REJEITAR H_0

- Proporção de estudantes: 61%
 - Há evidência suficiente para afirmar que a proporção de estudantes em atividades extracurriculares não é 0.61
 - Rejeitar $H_0: p = 0.61 \implies p \neq 0.61$
- Tempo médio de troca de óleo é menor que 15 min
 - Há evidência suficiente para concluir que o tempo médio é menor que 15 minutos
 - Rejeitar $H_0: \mu = 15 \implies \mu < 15$

INTERPRETAÇÃO: NÃO REJEITAR H_0

- Proporção de estudantes é 61%
 - Não há evidência suficiente para afirmar que a proporção difere de 0.61
- Tempo médio de troca de óleo é menor que 15 min
 - Não há evidência suficiente para afirmar que o tempo médio é menor que 15 minutos

EXERCÍCIO 42

42. Você realiza um teste de hipótese para cada uma das afirmações.
Como você deveria interpretar sua decisão se rejeitar H_0 ? E se você não rejeitar H_0 ? Ha (afirmação): Um corretor de imóveis divulga que a proporção de proprietários que acham suas casas muito pequenas para suas famílias é mais que 24%.

QUESTÃO A

Se rejeitarmos H_0 , a interpretação seria de que há evidência suficiente para concluir que a proporção de proprietários que acham suas casas muito pequenas é maior que 24%.

QUESTÃO B

Se não rejeitarmos H_0 , a interpretação seria de que não há evidência suficiente para concluir que a proporção de proprietários que acham suas casas muito pequenas é maior que 24%.

EXERCÍCIO 43

43. Encontre o valor p para um teste de hipótese unilateral à esquerda com uma estatística de teste padronizada $z = -1,71$. Decida se rejeita H_0 quando o nível de significância é $\alpha = 0,05$.

QUESTÃO A

```
z <- -1.71  
p <- pnorm(z)  
p
```

11 0.04363294

QUESTÃO B

Este valor é o mesmo obtido na questão A, 0.0436329.

QUESTÃO C

Interpretação: Uma vez que $p = 0.0436329 < \alpha = 0.1$, devemos rejeitar H_0 .

EXERCÍCIO 44

QUESTÃO A

```
z <- 1.64  
p <- pnorm(z)  
p
```

[1] 0.9494974

QUESTÃO B

```
p_2 <- 2 * (1 - pnorm(z))  
p_2
```

[1] 0.1010052

QUESTÃO C

$p = 0.9494974 > \alpha = 0.10 \implies$ Não rejeitar H_0 .

EXERCÍCIO 45

QUESTÃO A

- $H_0: \mu = 35$
- $H_a: \mu > 35$

QUESTÃO B

- $\alpha = 0.05$

QUESTÃO C

```
x_bar <- 36  
mu <- 35  
sigma <- 4  
n <- 100  
  
z <- (x_bar - mu) / (sigma / sqrt(n))  
z
```

[1] 2.5

QUESTÃO D

```
p <- 1 - pnorm(z)  
p
```

[1] 0.006209665

QUESTÃO E

Dado que $p = 0.0062097 < \alpha = 0.05$, devemos rejeitar H_0 .

QUESTÃO F

Interpretação: Há evidência suficiente para concluir que a velocidade média dos veículos ultrapassa 35 milhas por hora.

EXERCÍCIO 46

QUESTÃO A

- Afirmação: “o tempo médio para recuperar o custo é 3 anos”.
- $H_0 : \mu = 3$
- $H_1 : \mu \neq 3$

QUESTÃO B

```
alpha <- 0.01
```

QUESTÃO C

```
x_bar <- 3.3
mu   <- 3
sigma <- 0.5
n    <- 25

z <- (x_bar - mu) / (sigma / sqrt(n))
z
```

[1] 3

QUESTÃO D

```
p <- 2 * (1 - pnorm(abs(z)))
p
```

[1] 0.002699796

QUESTÃO E

Como $p \approx 0.0027 < \alpha = 0.01$, devemos rejeitar H_0 .

QUESTÃO F

Interpretação: Há evidências suficientes de que o tempo médio para recuperar o custo difere de 3 anos. No caso da presente estatística, ele parece ser maior.

EXERCÍCIO 47

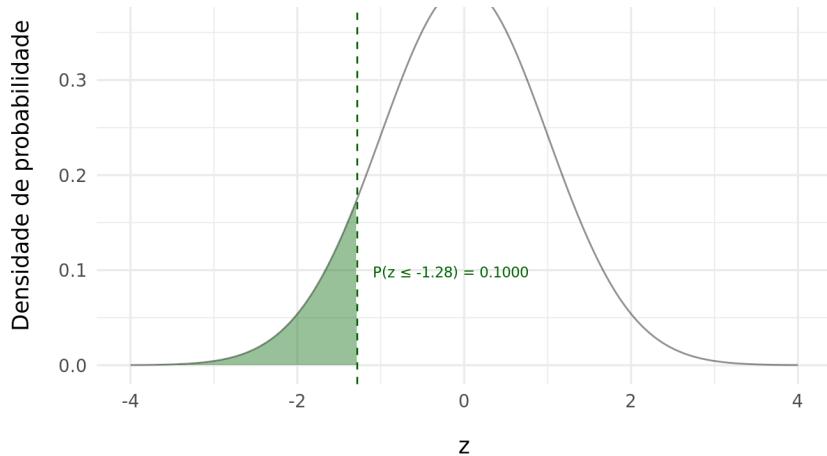
QUESTÃO A

```
alpha <- 0.10
```

```
z_plot(qnorm(alpha))
```

Área acumulada para o escore-z -1.2816





QUESTÃO B

```
z <- -1.28
a <- pnorm(z)
a
```

[1] 0.1002726

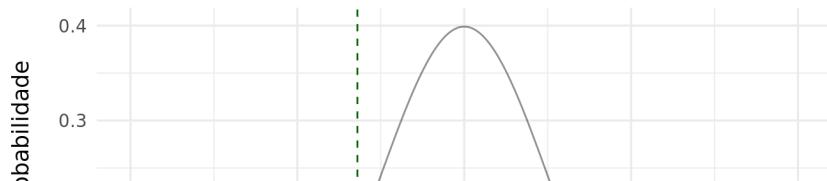
QUESTÃO C

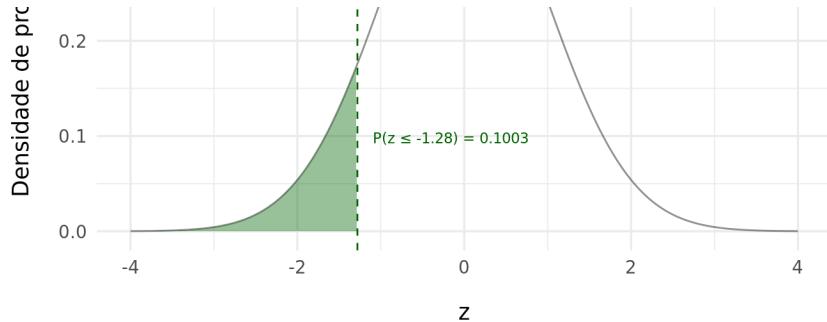
```
z_a <- qnorm(a)
z_a
```

[1] -1.28

```
z_plot(z_a)
```

Área acumulada para o escore-z -1.28





QUESTÃO D

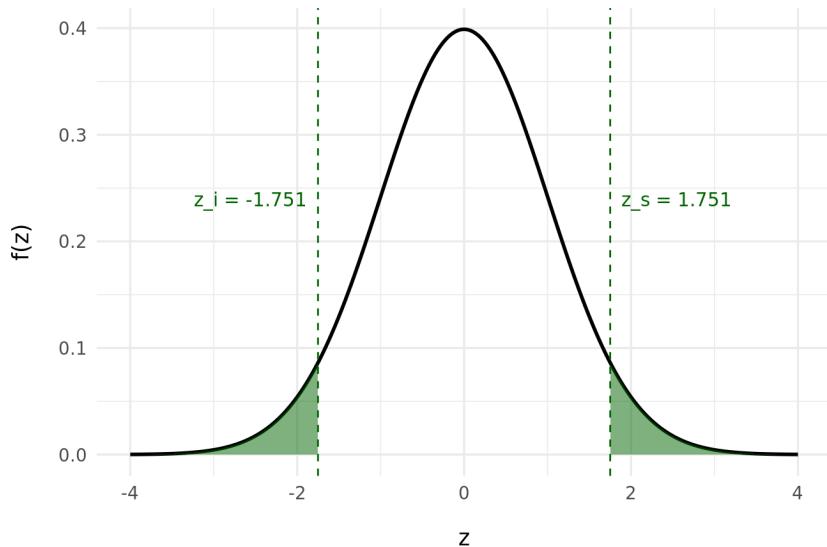
A zona de rejeição está à esquerda de $z = -1.28$, ou seja, devemos rejeitar se $z < -1.28$.

EXERCÍCIO 48

QUESTÃO A

► Código

Curva Normal Padrão: $\alpha/2$ em cada cauda



QUESTÃO B

```
alpha <- 0.08
a_i <- pnorm(qnorm(alpha/2))
a_s <- pnorm(qnorm(1 - alpha/2))
```

$$a_i = 0.04$$

$$a_s = 0.96$$

QUESTÃO C

```
z_i <- qnorm(0.04)
z_s <- qnorm(0.96)
z_i # -1.750686
```

```
[1] -1.750686
```

```
z_s # 1.750686
```

```
[1] 1.750686
```

QUESTÃO D

Rejeitar H_0 se $z < -1.7506861$ ou $z > 1.7506861$.

EXERCÍCIO 49

- Hipóteses
 - $H_0 : \mu = 68000$
 - $H_1 : \mu < 68000$

```
x_bar <- 66900
mu <- 68000
sigma <- 5500
n <- 20

z <- (x_bar - mu) / (sigma / sqrt(n))
p <- pnorm(z)
```

$$z = -0.8944272$$

$$p = 0.1855467$$

$0.1855467 > \alpha = 0.05$, logo, não devemos rejeitar H_0 .

Interpretação: Não há evidência suficiente de que o salário médio seja menor que 68000.

EXERCÍCIO 50

QUESTÃO A

```
alpha1 <- 0.1  
alpha2 <- 0.01
```

QUESTÃO B

```
z_1 <- qnorm(1 - alpha1 / 2)  
z_2 <- qnorm(1 - alpha2 / 2)
```

$$z_1 = 1.6448536$$

$$z_2 = 2.5758293$$

As regiões de rejeição são $z < -z_0$ ou $z > z_0$.

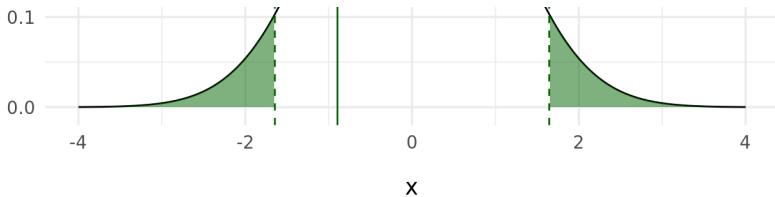
QUESTÃO C

► Código

Teste Z bilateral

$$\alpha = 0.1, z = -0.894$$





QUESTÃO D

Interpretação: Com $\alpha = 0.1$, há evidência suficiente para rejeitar a afirmação de que o custo médio é 13960 (o custo médio parece menor). Com $\alpha = 0.01$, não há evidência suficiente para rejeitar a afirmação.

EXERCÍCIO 51

QUESTÃO A

$$gl = n - 1 = 9 - 1 = 8$$

QUESTÃO B

```
alpha <- 0.10
f      <- 8

t0 <- qt(1 - alpha, f)
```

$$t_0 = 1.3968153$$

EXERCÍCIO 52

52. Encontre os valores críticos $-t_0$ e t_0 para um teste bilateral para $\alpha = 0,05$ e $n = 16$.

QUESTÃO A

$$gl = n - 1 = 16 - 1 = 15$$

QUESTÃO B

```
alpha <- 0.05
```

```
f <- 15  
t_i <- qt(alpha / 2, f)  
t_s <- qt(1 - alpha / 2, f)
```

$$t_i = -2.1314495$$

$$t_s = 2.1314495$$

EXERCÍCIO 53

QUESTÃO A

$$H_0: \mu = 1200$$

$$H_a: \mu < 1200$$

QUESTÃO B

```
alpha <- 0.1  
n <- 7  
f <- n - 1
```

QUESTÃO C

```
t <- qt(alpha, f)
```

$$t = -1.4397557$$

QUESTÃO D

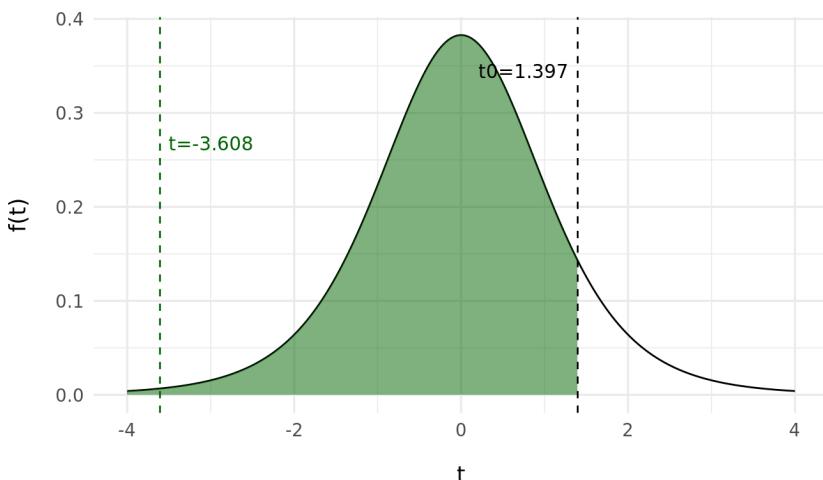
```
x_bar <- 1125  
mu <- 1200  
s <- 55  
n <- 7  
  
t_stat <- (x_bar - mu) / (s / sqrt(n))  
p <- pt(t_stat, f)
```

$$t_{stat} = -3.6078427$$

$$p = 0.00563$$

Distribuição t (df=6)

Área à esquerda de t_0 = região de rejeição



QUESTÃO E

$$p \approx 0.044 < \alpha = 0.1 \implies \text{Rejeita } H_0$$

QUESTÃO F

Interpretação: Há evidência de que o custo médio do seguro é menor que 1200.

EXERCÍCIO 54

54. Uma indústria afirma que o nível médio do pH da água em um rio próximo é de 6,8. Você seleciona aleatoriamente 39 amostras de água e mede o pH de cada uma. A média amostral e o desvio padrão são de 6,7 e 0,35, respectivamente. Há evidência suficiente para rejeitar a afirmação da indústria considerando nível de significância $\alpha = 0,05$?

Esta mesma indústria afirma que a condutividade média do rio é de 1.890 mg/l. A condutividade de uma amostra de água é uma medida do

total de sólidos dissolvidos na amostra. Você seleciona aleatoriamente 39 amostras de água e mede a condutividade de cada uma. A média e o desvio padrão amostral são de 2.350 mg/l e 900 mg/l, respectivamente. Há evidência suficiente para rejeitar a afirmação da indústria considerando o nível de significância $\alpha = 0,01$?

QUESTÃO A

pH

-

$$H_0: \mu_{\text{pH}} \leq 6.8$$

-

$$H_a: \mu_{\text{pH}} > 6.8$$

CONDUTIVIDADE

-

$$H_0: \mu_{\text{cond}} \leq 1890$$

-

$$H_a: \mu_{\text{cond}} > 1890$$

QUESTÃO B

-

$$\text{\texttt{\text{pH: }} } \& \alpha = 0.05, \quad df = 39 - 1 = 38$$

-

$$\text{\texttt{\text{Cond: }} } \& \alpha = 0.01, \quad df = 39 - 1 = 38$$

QUESTÃO C

```
df      <- 38
t_crit05 <- qt(1 - 0.05 / 2, df)
t_crit01 <- qt(1 - 0.01 / 2, df)
t_crit05; -t_crit05
```

```
[1] 2.024394
```

```
[1] -2.024394
```

```
# [1] 2.024394 -2.024394  
t_crit01; -t_crit01
```

```
[1] 2.711558
```

```
[1] -2.711558
```

```
# [1] 2.715158 -2.715158
```

QUESTÃO D

```
# pH  
xbar1 <- 6.7  
mu01 <- 6.8  
s1 <- 0.35  
n1 <- 39  
df <- n1 - 1  
  
t1 <- (xbar1 - mu01) / (s1 / sqrt(n1))  
p1 <- 2 * pt(-abs(t1), df)  
  
# condutividade  
xbar2 <- 2350  
mu02 <- 1890  
s2 <- 900  
n2 <- 39  
  
t2 <- (xbar2 - mu02) / (s2 / sqrt(n2))  
p2 <- 2 * pt(-abs(t2), df)  
  
t1; p1
```

```
[1] -1.784285
```

```
[1] 0.08236385
```

```
# t1 ≈ -1.786, p1 ≈ 0.081
```

```
t2; p2
```

```
[1] 3.191888
```

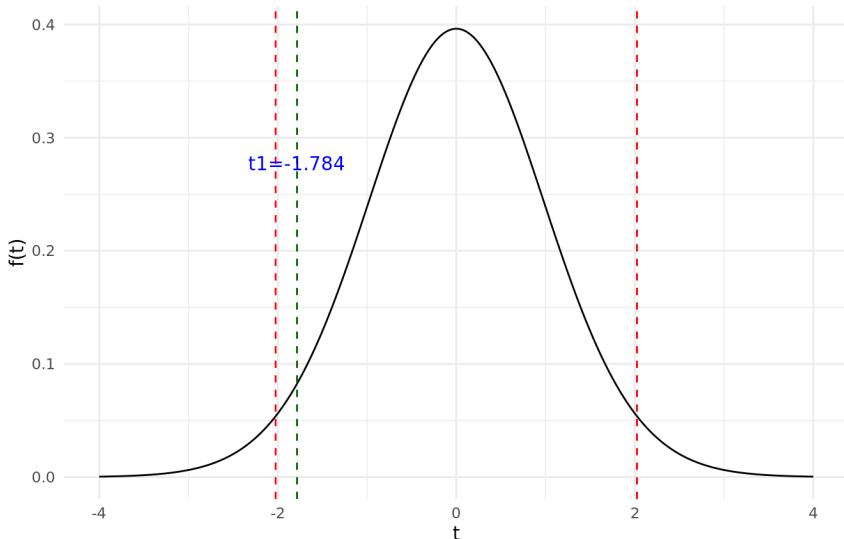
```
[1] 0.002835512
```

```
# t2 ≈ 3.190, p2 ≈ 0.0029
```

► Código

Teste t bilateral para pH

$t_1 = -1.784$ | região de rejeição $|t| > 2.0244$



QUESTÃO E

- pH ($\alpha = 0.05$)
 - $p_1 \approx 0.081 > 0.05 \implies$ Não rejeita H_0
- Condutividade ($\alpha = 0.01$)
 - $p_2 \approx 0.0029 > 0.01 \implies$ Rejeita H_0

QUESTÃO F

Interpretação: - pH: Não há evidência suficiente para rejeitar a afirmação de nível médio 6.8. - Condutividade: Há evidência suficiente para rejeitar a afirmação de média 1890 mg/L.

