MAE116 – Noções de Estatística

Grupos B e D – II Semestre de 2020

Lista de exercícios 5 – Distribuição normal – CASA (gabarito)

Exercício 1.

A durabilidade de um tipo de pneu da marca Duramais é descrita por uma variável aleatória Normal de média 60000 km e desvio padrão de 8300 km.

Solução:

Para modelar a durabilidade do pneu, defina a seguinte variável aleatória:

X : "Durabilidade, em km, de um tipo de pneu da marca Duramais".

Pelo enunciado, temos que

$$X \sim \mathrm{N}(60000, 8300^2).$$

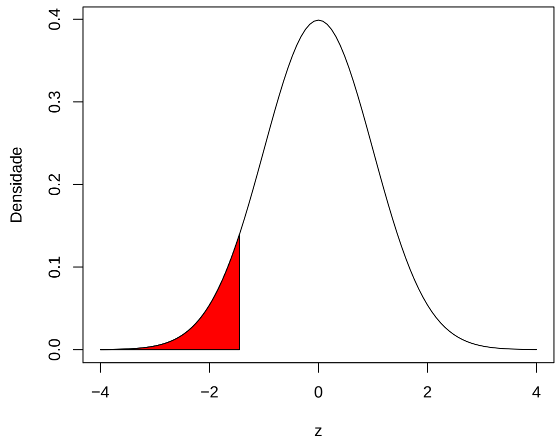
(a) Se a Duramais garante os pneus pelos primeiros 48000 km, qual é a proporção de pneus que deverão ser trocados na garantia?

A proporção (probabilidade) desejada é representada pela probabilidade do pneu durar menos do que os 48000 km, isto é, $\mathrm{P}(X < 48000)$. Fazendo as contas, utilizando a transformação para a distribuição normal padrão, temos que

$$\mathrm{P}(X < 48000) = \mathrm{P}\left(\frac{X - 60000}{8300} < \frac{48000 - 60000}{8300}\right) = \mathrm{P}(Z < -1,45),$$

em que $Z \sim \mathrm{N}(0, 1)$.

Graficamente, a área abaixo da curva da distribuição normal padrão desejada pode ser visualizada abaixo:



Utilizando a tabela disponibilizada no e-disciplinas, percebemos que ela fornece a área acumulada à esquerda na distribuição normal padrão $\mathrm{A}(z) = \mathrm{P}(Z \leq z)$ de apenas quantis z positivos. Entretanto, como a distribuição normal é simétrica, temos que

$$\begin{aligned}
\mathrm{P}(Z < -1,45) &= \mathrm{P}(Z > 1,45) \\
&= 1 - \mathrm{P}(Z \leq 1,45) = 1 - \mathrm{A}(1,45) \\
&= 1 - 0,9265 = 0,0735.
\end{aligned}$$

Logo, a proporção de pneus que deverão ser trocados na garantia é de aproximadamente **0,0735**.

(b) Qual deveria ser a garantia (em km) de tal forma a assegurar que o fabricante trocaria sob garantia no máximo 3% dos pneus?

Desejamos encontrar o valor do quantil x de modo que

$$\mathrm{P}(X < x) = 0,03.$$

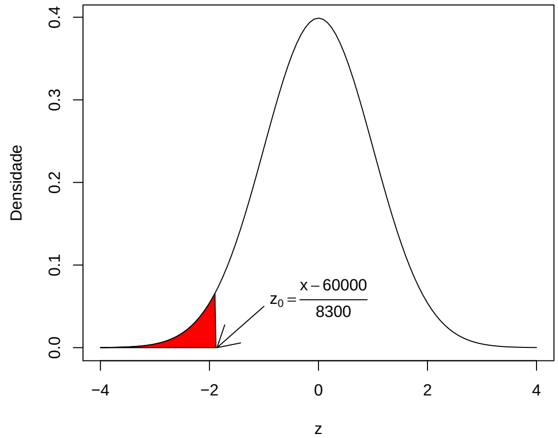
Fazendo a transformação para a distribuição normal padrão, temos que

$$\mathrm{P}(X < x) = 0,03 \Rightarrow \mathrm{P}\left(\frac{X - 60000}{8300} < \frac{x - 60000}{8300}\right) = 0,03 \Rightarrow \mathrm{P}\left(Z < \frac{x - 60000}{8300}\right) = 0,03.$$

Denote

$$z_0 = \frac{x - 60000}{8300}.$$

Graficamente, a área destacada em vermelho representa a probabilidade acumulada à esquerda de 0,03 na distribuição normal padrão.



Mais uma vez é necessário recorrer a simetria da distribuição normal para encontrar o valor do quantil z pela tabela. Neste caso, encontrar o valor z_0 tal que $\mathrm{A}(z_0) = 0,03$ (note que $z_0 < 0$) é equivalente a encontrar o valor $-z_0$ de modo que $\mathrm{A}(-z_0) = 0,97$. Olhando agora para o corpo da tabela, obtemos que $-z_0 = 1,88$. Logo,

$$z_0 = -1,88$$

e, portanto,

$$-1,88 = \frac{x - 60000}{8300} \Rightarrow x = 60000 - 1,88 \times 8300 = 44396.$$

Assim, a garantia de tal forma a assegurar que o fabricante trocaria sob garantia no máximo 3% dos pneus deve ser de **44396 km**.

Exercício 2.

Para indivíduos portadores de doença cardíaca coronariana, o nível sérico de colesterol tem distribuição normal com média 244 mg/100 mL e desvio padrão 51 mg/100 mL e para não portadores, distribuição normal com média 219 mg/100 mL e desvio padrão de 41 mg/100 mL. A presença de nível de colesterol maior ou igual a 250 mg/100 mL é usada como indicador preliminar da doença.

Solução:

Defina as seguintes variáveis aleatórias:

X : “Nível sérico de colesterol, em mg/100 mL, de portadores de doença cardíaca coronariana”.

Y : “Nível sérico de colesterol, em mg/100 mL, de não portadores de doença cardíaca coronariana”.

Pelo enunciado, temos que

$$X \sim \mathrm{N}(244, 51^2) \quad \text{e} \quad Y \sim \mathrm{N}(219, 41^2).$$

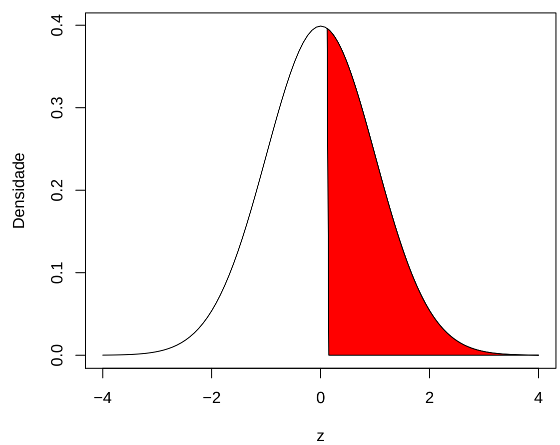
(a) Qual é a probabilidade de se diagnosticar a presença da doença para uma pessoa que realmente seja portadora?

A probabilidade desejada é representada por $\mathrm{P}(X \geq 250)$. Calculando esta probabilidade por meio da transformação para a distribuição normal padrão obtemos que

$$\mathrm{P}(X \geq 250) = \mathrm{P}\left(\frac{X-244}{51} \geq \frac{250-244}{51}\right) = \mathrm{P}(Z \geq 0,12),$$

em que $Z \sim \mathrm{N}(0,1)$.

Graficamente, a área abaixo da curva da densidade da distribuição normal padrão a ser calculada pode ser vista abaixo.



Utilizando a tabela da distribuição normal padrão, temos que

$$\begin{aligned}\mathrm{P}(Z \geq 0,12) &= 1 - \mathrm{P}(Z \leq 0,12)\\ &= 1 - \mathrm{A}(0,12)\\ &= 1 - 0,5478 = 0,4522.\end{aligned}$$

Portanto, a probabilidade de se diagnosticar a presença da doença para uma pessoa que realmente seja portadora da doença coronariana é de aproximadamente **0,4522**.

(b) Qual é a probabilidade de se diagnosticar a presença da doença para um indivíduo não portador da doença coronariana (resultado falso positivo)?

Análogo ao item anterior, mas modelando agora o nível sérico de colesterol de **não portadores** de doença cardíaca coronariana, a solução segue.

Agora, a probabilidade desejada é representada por $\mathrm{P}(Y \geq 250)$. Assim, temos que

$$\mathrm{P}(Y \geq 250) = \mathrm{P}\left(\frac{Y - 219}{41} \geq \frac{250 - 219}{41}\right) = \mathrm{P}(Z \geq 0,76),$$

em que $Z \sim \mathrm{N}(0,1)$.

Utilizando a tabela da distribuição normal padrão, temos que

$$\begin{aligned} \mathrm{P}(Z \geq 0,76) &= 1 - \mathrm{P}(Z \leq 0,76) \\ &= 1 - \mathrm{A}(0,76) \\ &= 1 - 0,7764 = 0,2236. \end{aligned}$$

Logo, a probabilidade de se diagnosticar a presença da doença para um indivíduo não portador da doença coronariana é de aproximadamente **0,2236**.

(c) Determine a probabilidade de não haver indicação preliminar da doença para um portador da doença coronariana (resultado falso negativo).

Para não haver indicação preliminar da doença, o nível sérico de colesterol deve ser menor do que 250 mg/100 mL. Calculando esta probabilidade para os indivíduos portadores da doenã coronariana, a probabilidade desejada é representada por $\mathrm{P}(X < 250)$. Pelo item (a), calculamos diretamente que

$$\mathrm{P}(X < 250) = 1 - \mathrm{P}(X \geq 250) = 1 - 0,4522 = 0,5478.$$

Portanto, a probabilidade de não haver indicação preliminar da doença para um portador da doença coronariana é igual a **0,5478**.

(d) Se numa população o número de portadores de doença coronariana é 1/10 do número de não portadores, qual é a probabilidade de uma pessoa, aleatoriamente selecionada dessa população, tenha nível sérico de colesterol maior ou igual a 250 mg/100 mL?

Defina os seguintes eventos:

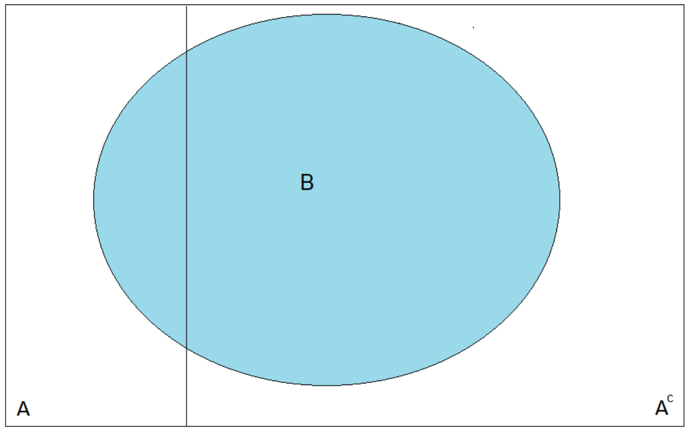
A : "O indivíduo selecionado é portador da doença coronariana."

B : "O indivíduo selecionado tem nível sérico de colesterol superior ou igual a 250 mg/100 mL."

A probabilidade desejada é

$$\mathrm{P}(B).$$

Entretanto, note que o evento B depende de qual população o indivíduo pertence (i.e., depende o evento A). O gráfico abaixo ilustra os eventos



Pela regra da probabilidade total,

$$\mathrm{P}(B) = \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A^c) = \mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c).$$

Com base nos itens (a) e (b), respectivamente, calculamos as probabilidades

$$P(B|A) = 0,4522 \quad \text{e} \quad P(B|A^c) = 0,2236.$$

Considere agora o cálculo de $\mathrm{P}(A)$ e $\mathrm{P}(A^c)$. Sejam N_1 e N_2 o tamanho da população de portadores e não portadores da doença coronariana, respectivamente, e seja $N = N_1 + N_2$ o tamanho da população geral. Do enunciado, obtemos que

$$N_1 = \frac{N_2}{10},$$

logo, podemos calcular que

$$N_1 + N_2 = N \quad \Rightarrow \quad N_1 + 10N_1 = N \quad \Rightarrow \quad 11N_1 = N \quad \Rightarrow \quad \frac{N_1}{N} = \frac{1}{11}.$$

Ou seja, a proporção $p_1 = N_1/N$ de indivíduos portadores da doença coronariana na população geral é igual a $p_1 = 0,0909$ e consequentemente, a proporção $p_2 = N_2/N$ de indivíduos não portadores da doença coronariana na população geral é igual a $p_2 = 1 - 0,0909 = 0,9091$.

Assim, calculamos que

$$\mathrm{P}(A) = 0,0909 \quad \text{e} \quad \mathrm{P}(A^c) = 0,9091.$$

Portanto, a probabilidade de uma pessoa, aleatoriamente selecionada dessa população, tenha nível sérico de colesterol maior ou igual a 250 mg/100 mL é

$$\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c) = 0,0909 \times 0,4522 + 0,9091 \times 0,2236 = \mathbf{0,2444}.$$

(e) Qual é a probabilidade de uma pessoa, aleatoriamente selecionada da população acima, seja portadora da doença se seu nível sérico é maior ou igual a 220 mg/100 mL?

Recorde que definimos no item anterior o seguinte evento:

A : "O indivíduo selecionado é portador da doença coronariana."

Agora, defina

C : "O indivíduo selecionado possui o nível sérico de colesterol superior ou igual a 220 mg/100 mL".

A probabilidade desejada é representada por

$$\mathrm{P}(A|C) = \frac{\mathrm{P}(A \cap C)}{\mathrm{P}(C)}.$$

Inicialmente, iremos calcular $\mathrm{P}(C)$. Note que o evento C depende de qual população o indivíduo pertence (i.e., depende o evento A). Análogo ao item anterior,

$$\mathrm{P}(C) = \mathrm{P}(A)\mathrm{P}(C|A) + \mathrm{P}(A^c)\mathrm{P}(C|A^c).$$

Para calcular $\mathrm{P}(C|A)$ e $\mathrm{P}(C|A^c)$, recorde que com base no enunciado temos as variáveis aleatórias X e Y são distribuídas normalmente com as seguintes especificações

$$X \sim \mathrm{N}(244, 51^2) \quad \text{e} \quad Y \sim \mathrm{N}(219, 41^2).$$

Assim, podemos calcular que

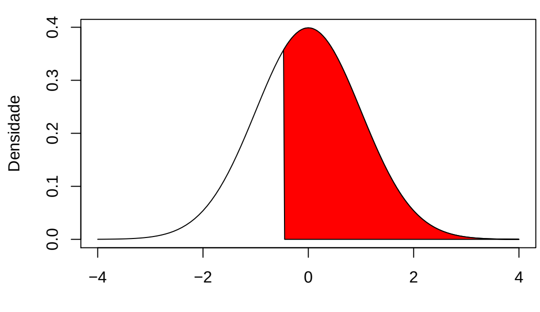
$$\mathrm{P}(C|A) = \mathrm{P}(X \geq 220) = \mathrm{P}\left(\frac{X - 244}{51} \geq \frac{220 - 244}{51}\right) = \mathrm{P}(Z \geq -0,47),$$

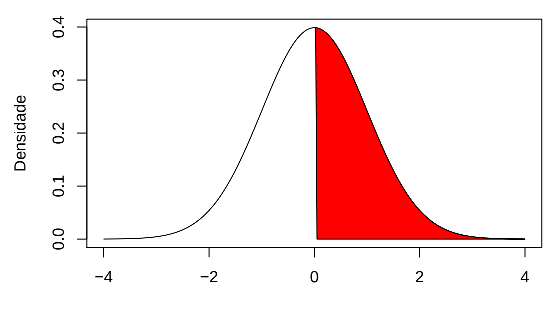
e

$$\mathrm{P}(C|A^c) = \mathrm{P}(Y \geq 220) = \mathrm{P}\left(\frac{Y - 219}{41} \geq \frac{220 - 219}{41}\right) = \mathrm{P}(Z \geq 0,02),$$

em que $Z \sim \mathrm{N}(0, 1)$.

As probabilidades acumuladas podem ser vistas nas respectivas figuras abaixo:





Utilizando a tabela da distribuição normal padrão, temos que

$$\begin{aligned} \mathrm{P}(Z \geq -0,47) &= \mathrm{P}(Z \leq 0,47) \\ &= \mathrm{A}(0,47) \\ &= 0,6808. \end{aligned}$$

e

$$\begin{aligned} \mathrm{P}(Z \geq 0,02) &= 1 - \mathrm{P}(Z \leq 0,02) \\ &= 1 - \mathrm{A}(0,02) \\ &= 1 - 0,5080 = 0,4920. \end{aligned}$$

Logo,

$$P(C|A) = 0,6808 \quad \text{e} \quad P(C|A^c) = 0,4920,$$

e, assim,

$$\begin{aligned} \mathrm{P}(C) &= \mathrm{P}(A)\mathrm{P}(C|A) + \mathrm{P}(A^c)\mathrm{P}(C|A^c) \\ &= 0,0909 \times 0,6808 + 0,9091 \times 0,4920 \\ &= 0,5092. \end{aligned}$$

Por fim, utilizando a regra do produto, calculamos que

$$\mathrm{P}(A \cap C) = \mathrm{P}(A)\mathrm{P}(C|A) = 0,0909 \times 0,6808 = 0,0619.$$

Portanto, a probabilidade de uma pessoa, aleatoriamente selecionada da população geral, seja portadora da doença se seu nível sérico é maior ou igual a 220 mg/100 mL é igual a

$$\mathrm{P}(A|C) = \frac{\mathrm{P}(A \cap C)}{\mathrm{P}(C)} = \frac{0,0619}{0,5092} = \mathbf{0{,}1215}.$$

Exercício 3.

Um bom indicador do nível de intoxicação por benzeno é a quantidade de fenol encontrada na urina. A quantidade de fenol na urina de moradores de uma certa região segue, aproximadamente, uma distribuição normal de média 5 mg/L e desvio padrão 2 mg/L. Considere as seguintes definições em termos da variável quantidade de fenol na urina:

i. Define-se como "valor de referência" a quantidade de fenol tal que 95% da população têm quantidade de fenol maior ou igual a esse valor;

ii. Uma pessoa é considerada "atípica" se a quantidade de fenol em sua urina for inferior a 2 mg/L ou superior a 8 mg/L.

Solução:

Considere a variável aleatória X: quantidade, em mg/L, de fenol encontrada na urina. Então, $X \sim N(5, 2^2)$.

(a) Sorteado um morador ao acaso, qual é a probabilidade de ser "atípico"?

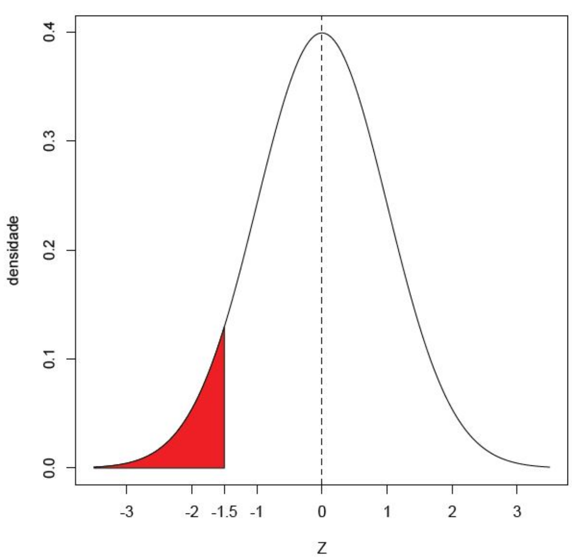
Para ser classificado como "atípico" é necessário que a quantidade de fenol na urina seja inferior a 2 mg/L ou superior a 8 mg/L. Assim, a probabilidade de que o morador sorteado ser atípico é dada por

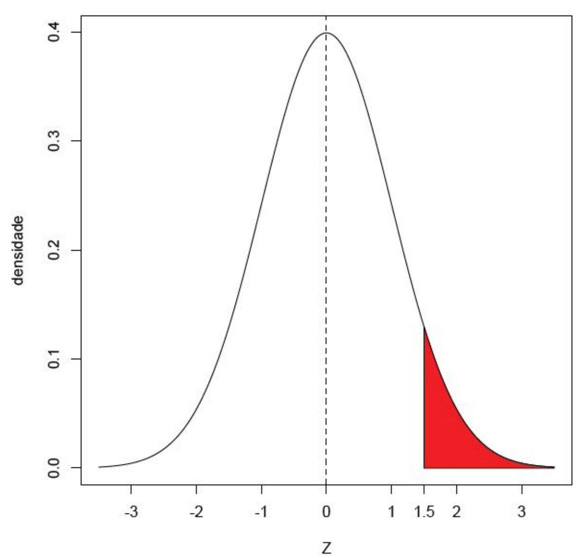
$$P(X \leq 2) + P(X \geq 8).$$

$$P(X \leq 2) = P\left(\frac{X-5}{2} \leq \frac{2-5}{2}\right) = P(Z \leq -1,5) = 1 - P(Z \leq 1,5) = 1 - 0,9332 = 0,0668$$

$$P(X \geq 8) = P\left(\frac{X-5}{2} \geq \frac{8-5}{2}\right) = P(Z \geq 1,5) = 1 - P(Z \leq 1,5) = 1 - A(1,5) = 1 - 0,9332 = 0,0668$$

Portanto, a probabilidade do morador ser classificado como atípico é 0,0668 + 0,0668 = 0,1336.



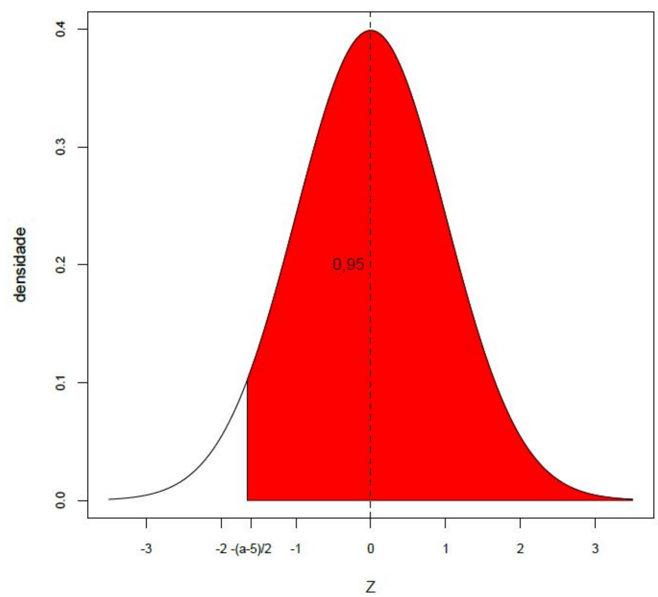


(b) Qual é o valor de referência da população?

Queremos obter o valor a tal que $P(X \geq a) = 0,95$.

$$P(X \geq a) = 0,95 \Rightarrow P\left(\frac{X-5}{2} \geq \frac{a-5}{2}\right) = 0,95 \Rightarrow P\left(Z \geq \frac{a-5}{2}\right) = 0,95 \Rightarrow P\left(Z \leq -\frac{a-5}{2}\right) = 0,95$$

$$\Rightarrow A\left(-\frac{(a-5)}{2}\right) = 0,95 \Rightarrow -\frac{(a-5)}{2} = 1,65 \Rightarrow = 5 - 2 \times 1,65 = 1,7$$



O valor de referência da população é 1,7 mg/L.

(c) Sorteadas 5 pessoas ao acaso, qual é a probabilidade de se ter no mínimo 4 "atípicas"?

Defina a variável aleatória Y: número de pessoas classificadas como "atípicas" entre as 5 pessoas sorteadas ao acaso. Pelo item (a) sabemos que a probabilidade de uma pessoa sorteada ser "atípica" é 0,1336. Assim, $Y \sim b(5; 0,1336)$. Os valores da distribuição $b(5; 0,1336)$ são dados na tabela a seguir.

y	$P(Y=y)$
0	0,48819380
1	0,37640050
2	0,11608290
3	0,01790013
4	0,00138011
5	0,00004256

A probabilidade de obtermos no mínimo 4 “atípicas” é obtida calculando $P(Y \geq 4)$.

$$P(Y \geq 4) = P(Y = 4) + P(Y = 5) = 0,0014$$

(d) Sabendo que uma pessoa não é atípica, qual é a probabilidade de ter quantidade de fenol no intervalo 4mg/L a 6 mg/L?

Queremos calcular a probabilidade condicional

$$P(4 \leq X \leq 6|2 \leq X \leq 8) = \frac{P(4 \leq X \leq 6, 2 \leq X \leq 8)}{P(2 \leq X \leq 8)} = \frac{P(4 \leq X \leq 6)}{P(2 \leq X \leq 8)}$$

$$\begin{aligned} P(4 \leq X \leq 6) &= P\left(\frac{4-5}{2} \leq \frac{X-5}{2} \leq \frac{6-5}{2}\right) = P(-0,5 \leq Z \leq 0,5) = 2P(0 \leq Z \leq 0,5) \\ &= 2(A(0,5) - 0,5) = 2(0,6915 - 0,5) = 0,3830 \end{aligned}$$

Pelo item (a) temos que

$$P(2 \leq X \leq 8) = 1 - 0,1336 = 0,8664$$

Então,

$$P(4 \leq X \leq 6|2 \leq 8) = \frac{0,3830}{0,8664} = 0,4421$$

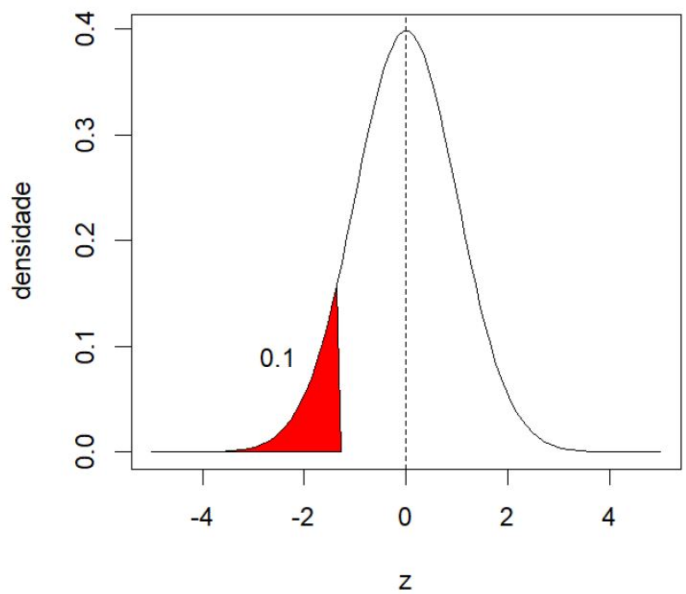
Sabendo que uma pessoa não é atípica, a probabilidade de ter quantidade de fenol no intervalo 4 mg/L a 6 mg/L é 0,4421.

Exercício 4.

Uma máquina de empacotar um determinado produto o faz segundo uma distribuição normal com média μ e desvio padrão 10 g. Com a máquina assim regulada,

(a) Em quanto deve ser fixado o peso médio para que apenas 10% dos pacotes tenham menos de 500 g?

Seja X: peso dos pacotes empacotados pela máquina. Do enunciado $X \sim Normal(\mu; 10^2)$. Queremos encontrar o valor μ de modo que $P(X < 500) = 0,1 \Rightarrow P\left(\frac{X-\mu}{10} < \frac{500-\mu}{10}\right) = 0,1$ $\Rightarrow P\left(Z < \frac{500-\mu}{100}\right) = 0,1$. Logo, $z = \frac{500-\mu}{10}$.
Encontrar um valor $z < 0$ de modo que $A(z) = 0,1$ é o mesmo que encontrar um valor $a > 0$ de modo que $A(a) = 0,90$ e $z = -a$.



Pela tabela $a = 1,28$ e $z = -1,28 = \frac{500-\mu}{10} \Rightarrow \mu = 500 + 1,28 \times 10 = 512,8$. Então, para apenas 10% dos pacotes tenham menos de 500g o peso médio deve ser fixado em 512,8g.

(b) Qual é a probabilidade de que o peso de um pacote exceda 550 g?

$$P(X > 550) = P\left(\frac{X - 512,8}{10} > \frac{550 - 512,8}{10}\right) = P(Z > 3,72) \approx 0$$

(c) Determine a proporção de pacotes em que o peso não se afasta da média em mais que dois desvios padrões.

$$
\begin{aligned}
P(-2\sigma < X-\mu < 2\sigma) &= P\left(\frac{-2\sigma}{\sigma} < \frac{X-\mu}{\sigma} < \frac{2\sigma}{\sigma}\right) = P(-2 < Z < 2) = P(Z \le 2) - P(Z \le -2) \\
&= P(Z \le 2) - P(Z \ge 2) = P(Z \le 2) - [1 - P(Z \le 2)] = 2P(Z \le 2) - 1 \\
&= 2A(2) - 1 = 2 \times 0,9772 - 1 = 0,9544
\end{aligned}
$$

(d) Se forem selecionados 12 pacotes enchidos por essa máquina, qual é a probabilidade de que pelo menos dois pacotes tenham peso inferior a 500 g?

Seja Y : número de pacotes com peso inferior a 500 g, dentre os 12 pacotes selecionados. Do Item (a) sabemos que quando $\mu = 512,8 \Rightarrow P(X < 500) = 0,1$. Então, $Y \sim Binomial(12; 0,1)$.
Logo, queremos calcular $P(Y \ge 2) = 1 - P(Y = 1) - P(Y = 0)$.

Podemos utilizar o Rcmdr para calcular essa probabilidade utilizando o caminho no menu: "Distribuições" $\Rightarrow$ "Distribuições Discretas" $\Rightarrow$ "Distribuição Binomial" $\Rightarrow$ "Probabilidades da Binomial". Em "Experimentos da Binomial" coloque 12 e em "Probabilidade de sucesso" coloque 0,10. Como resultado será apresentada a seguinte tabela de probabilidades:

```
Probability
0 0.282429536481
1 0.376572715308
2 0.230127770466
3 0.085232507580
4 0.021308126895
5 0.003788111448
6 0.000491051484
7 0.000046766808
8 0.000003247695
9 0.000000160380
10 0.000000005346
11 0.000000000108
12 0.000000000001
```

Pela tabela de probabilidades temos que $P(Y \ge 2) \approx 1 - 0,3766 - 0,2824 \approx 0,3410$.
Alternativamente, podemos calcular a probabilidade pela probabilidade da cauda superior da distribuição Binomial, ou seja, $P(Y \ge 2) = P(Y > 1)$. Vá em "Distribuições" $\Rightarrow$ "Distribuições Discretas" $\Rightarrow$ "Distribuição Binomial" $\Rightarrow$ "Probabilidades da cauda da Binomial". Em "Valores da Variável" coloque 1, em "Experimentos da Binomial" coloque 12, em "Probabilidade de sucesso" coloque 0,1 e marque a opção "cauda superior". O resultado será idêntico ao anterior.

```
> pbinom(c(1), size=12, prob=0.1, lower.tail=FALSE)
[1] 0.3409977
```