Code

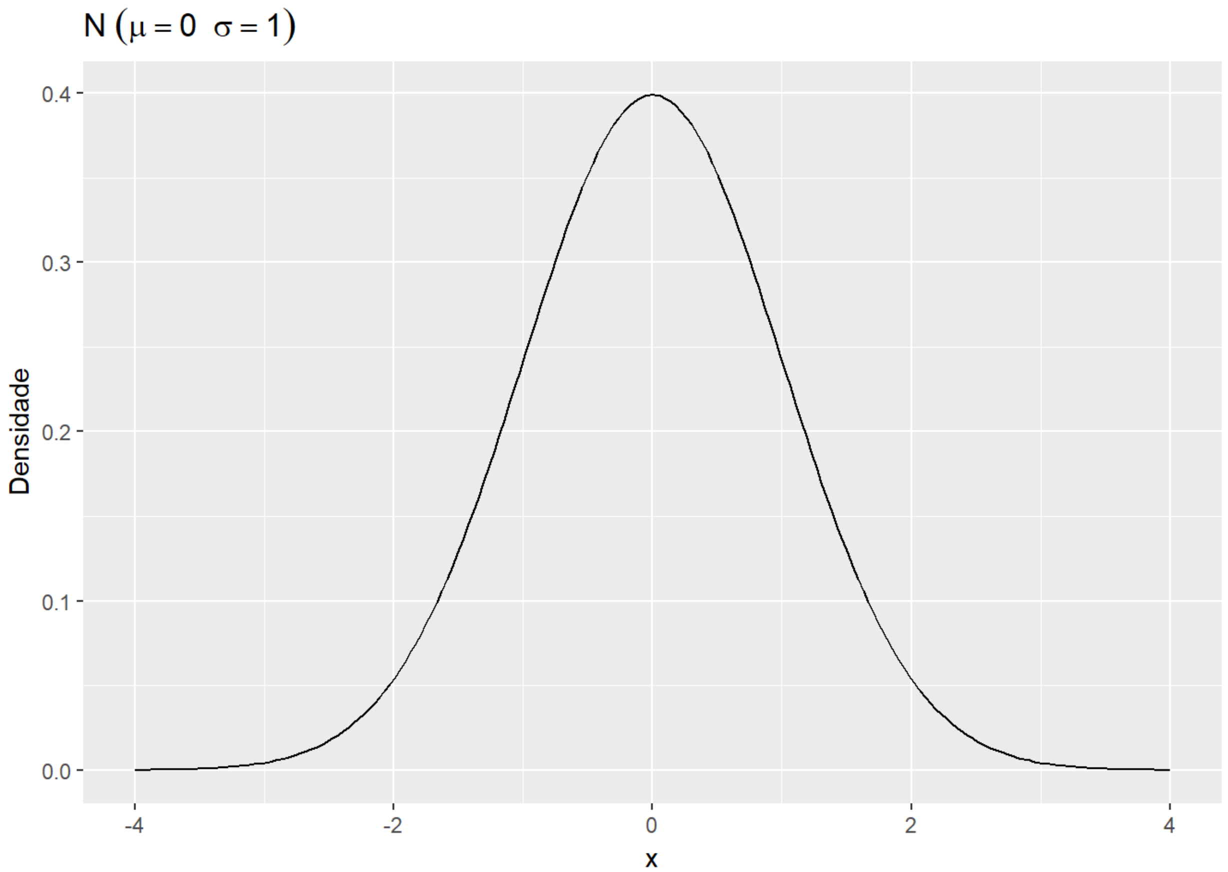
Distribuição Normal (Gaussiana)

Distribuição De-Moivre-Laplace-Gauss

Nos séculos dezoito e dezenove, alguns matemáticos e físicos desenvolveram uma função densidade de probabilidade que descrevia os erros experimentais obtidos em medidas físicas Caire, 2012 (http://repositorio.unesp.br/bitstream/handle/11449/91024/caire_e_me_rcla.pdf?sequence=1). De certa forma todo e qualquer processo de mensuração está sujeito a um erro de medida. Esse erro pode ter diferentes fontes, desde a variação de tempertura, tempo, entre inúmeras outras características não identificáveis. Essa **função densidade de probabilidade** é curva conhecida como distribuição **normal** ou **gaussina**.

A grande utilidade dessa distribuição está associada ao fato de que aproxima de forma bastante satisfatória as curvas de frequências de medidas físicas.

Code



Na época (século XVIII) a sua aplicação inicial era apenas como uma conveniente aproximação da distribuição binomial (http://www.inf.ufsc.br/~andre.zibetti/binomial.html), mais tarde no século XIX a distribuição normal ganhou importância com os trabalhos de Abraham de Moivre (em The Doctrine of Chances (https://archive.org/stream/doctrineofchance00moiv#page/n5/mode/2up)), Pierre Simon Laplace e Carl Friedrich Gauss.

Variável Aleatória Generalizada

Seja X uma variável aleatória contínua com média μ em que $-\infty < x < \infty$, e $\sigma > 0$.

Função Densidade de Probabilidade

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

Podemos dizer que X possui uma **distribuição normal** e escrever $X \sim N(\mu, \sigma^2)$.

Valor Esperado e variância

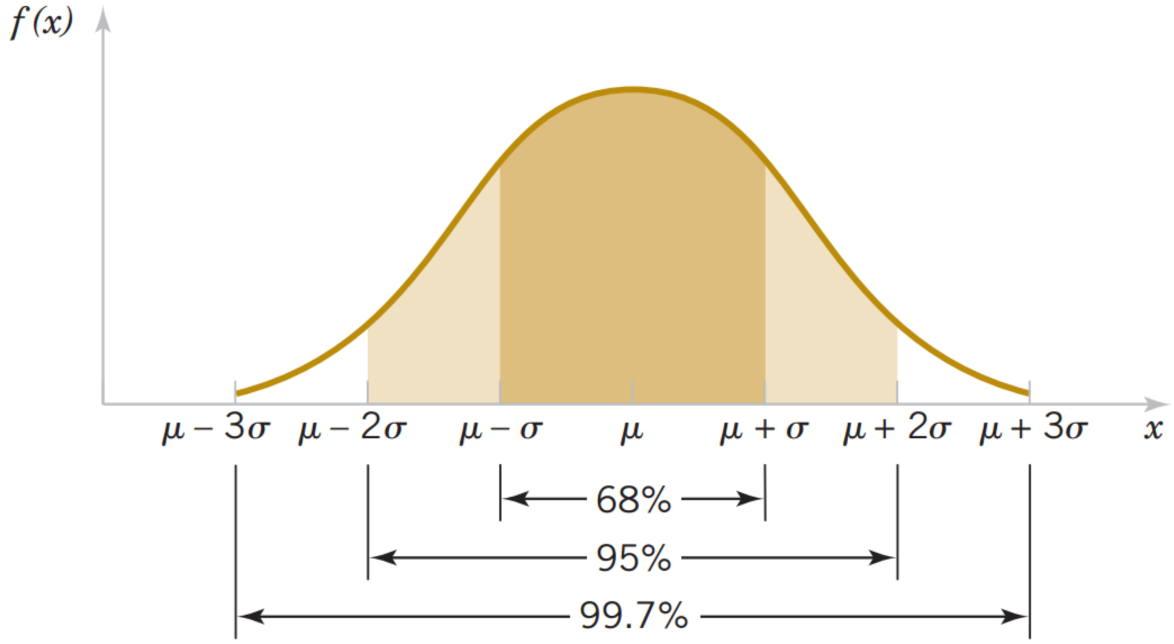
$$E[X] = \mu$$
$$V(X) = \sigma^2$$

Cálculo da Probabilidade

$$\mathbb{P}(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

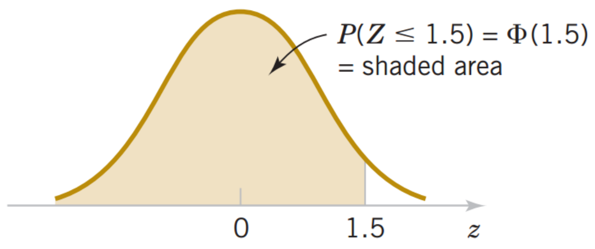
Cabe notar que a integral da função densidade de probabilidade normal não possui solução anaílitica, sendo neste caso o seu cálculo deve ser realizado por um método numérico.

Para sanar tal dificuldade a função pode ser padronizada com a substituição dos parâmetros por $\mu = 0$ e $\sigma^2 = 1$. Essa abordagem é dada pela definição de uma nova variável aleatória Z, chamada de **variável aleatória normal padronizada**.



Padronizando uma variável aleatória Normal
Se X for uma variável aleatória normal com $E[X] = \mu$ e $V(X) = \sigma^2$, a variável aleatória $$Z = \frac{X - \mu}{\sigma}$$ será uma variável aleatória normal, com $E[Z] = 0$ e V(Z)=1. Ou seja, Z é uma variável aleatória normal padrão.

Dessa forma, é possível obter a área sob a curva da normal padrão de forma analítica. Assim uma vez que a v.a. X é padronizada em Z é possível obter a área entre dois pontos sob a curva diretamente com o uso de uma tabela, e essa área representa uma probabilidade.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Exemplo - Normal 1

Os dados de uma pesquisa mostram algumas informações sobre o tempo de cirurgias para recostrução ACL em hospitais com alto volume de cirurgia. A partir dos dados foram calculados, o tempo médio de 129 minutos com um desvio padrão de 14 minutos.

a. **Qual é a probabilidade de uma cirurgia ACL, em um hospital com alto volume de cirurgias, requerer um tempo maior do que dois desvios-padrão acima da média?**
b. **Qual é a probabilidade de uma cirurgia ACL, em um hospital com alto volume de cirurgias ser completada em menos de 100 minutos?**
c. **Em qual tempo a probabilidade de uma cirurgia ACL em um hospital com alto volume de cirurgias é igual a 0.95?**

d. **Se uma cirurgia requer 199 minutos, o que você conclui sobre o volume de tais cirurgias em um hospital? Explique.**

Solução

a. **Qual é a probabilidade de uma cirurgia ACL, em um hospital com alto volume de cirurgias, requerer um tempo maior do que dois desvios-padrão acima da média?**

$$1 - \mathbb{P}(Z < 2) = 1 - \phi(2) = 0.0228$$

Code

```
[1] 0.02275013
```

b. **Qual é a probabilidade de uma cirurgia ACL, em um hospital com alto volume de cirurgias ser completada em menos de 100 minutos?**

Seja X o tempo, onde $X \sim N(129, 14^2)$

$$Z = \frac{X - \mu}{\sigma} = \frac{100 - 129}{14} = -2.0714$$

$$\mathbb{P}(X < 100) = \phi(-2.0714) = 0.01916072$$

Code

```
[1] 0.01916072
```

c. **Em qual tempo a probabilidade de uma cirurgia ACL em um hospital com alto volume de cirurgias é igual a 0.95?**

Code

```
[1] 1.644854
```

Code

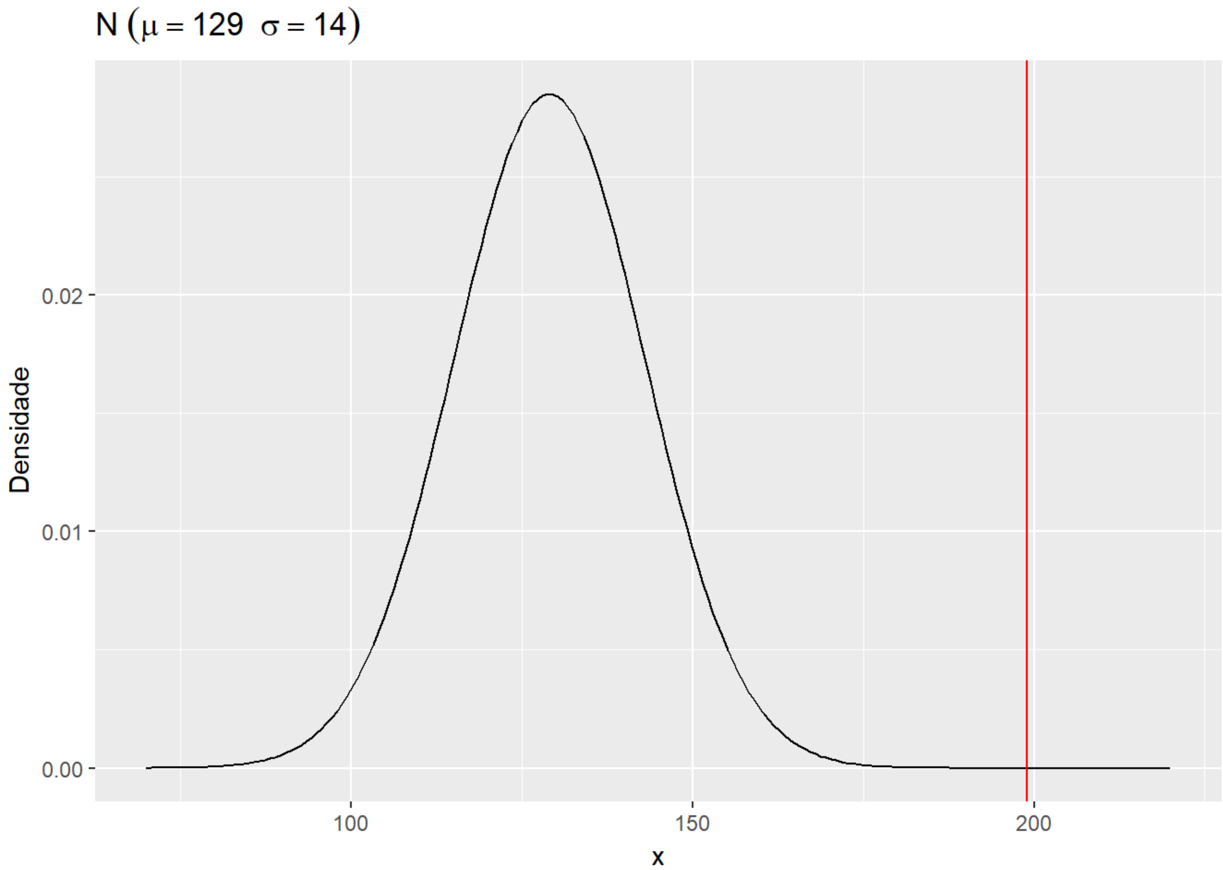
```
[1] 152.028
```

O tempo em que 95 % das cirurgias irão acabaram é dentro de 152.028 minutos.

d. **Se uma cirurgia requer 199 minutos, o que você conclui sobre o volume de tais cirurgias em um hospital? Explique.**

De acordo com a distribuição sabemos que menos de 5 % das cirurgias irão demandar tal quantidade de tempo. Observe a figura abaixo e conclua.

Code



Exemplo - Normal 2

Numa ilha existem três modelos de automóveis disponíveis para compra, um de cada uma das seguintes marcas: A, B e C. Segundo os fabricantes, e tendo em conta as características da rede de estradas local, o consumo dos referidos automóveis (em litros aos cem km) é caracterizado pelas seguintes distribuições:

- $A \sim N(10, 1^2)$
- $B \sim N(11, 1.5^2)$
- $C \sim N(8, 1.25^2)$

Um determinado Senhor, residente da ilha, possui um automóvel da marca C, dois da marca A e três da marca B. O referido Senhor tem por costume escolher na segunda-feira de manhã o automóvel que vai utilizar ao longo da semana.

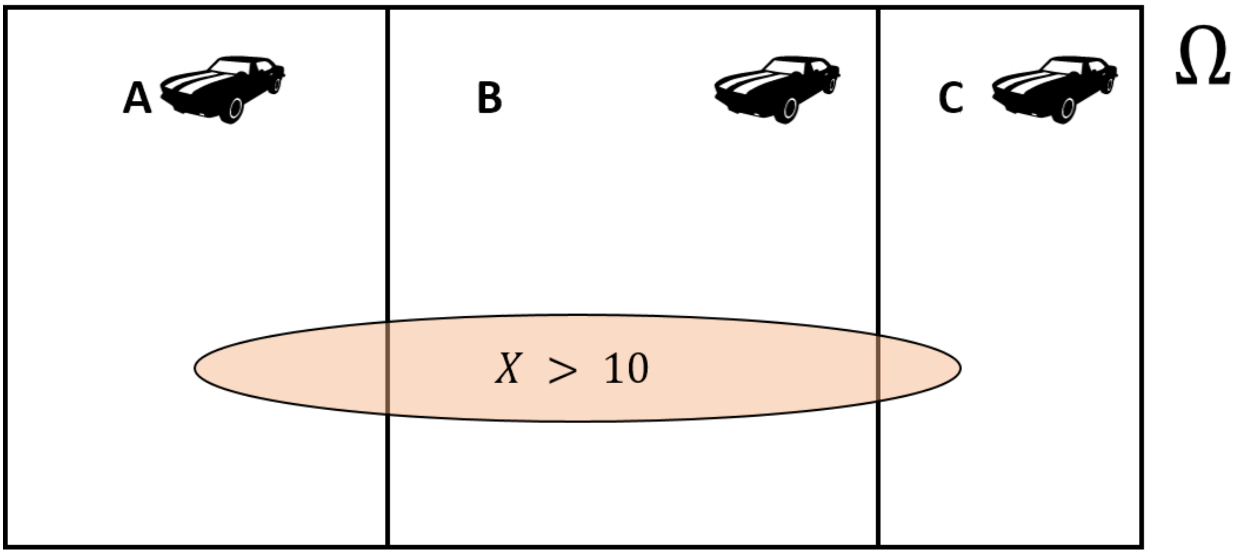
a. **Sabendo que na última semana o automóvel utilizado gastou em média mais de 10 litros (aos cem km), calcule a probabilidade de ter sido um automóvel da marca A o escolhido.**
b. **Qual a probabilidade de o consumo de um automóvel da marca C ser pelo menos 1 litro inferior ao consumo de um automóvel da marca A?**

Solução

a - Sabendo que na última semana o automóvel utilizado gastou em média mais de 10 litros (aos cem km), calcule a probabilidade de ter sido um automóvel da marca A o escolhido.

A questão pode ser respondida utilizando o teorema de Bayes. Imagine o espaço amostral particionado com eventos mutuamente exclusivos (carros), agora suponha o evento de interesse ocorrendo neste espaço amostral ($X > 10$), onde $X = consumo$, v.a. contínua com distribuição normal. Podemos representar esse cenário (partições) de acordo com a figura abaixo e suas probabilidades.

$$\mathbb{P}(A) = \frac{2}{6}$$
$$\mathbb{P}(B) = \frac{3}{6}$$
$$\mathbb{P}(C) = \frac{1}{6}$$

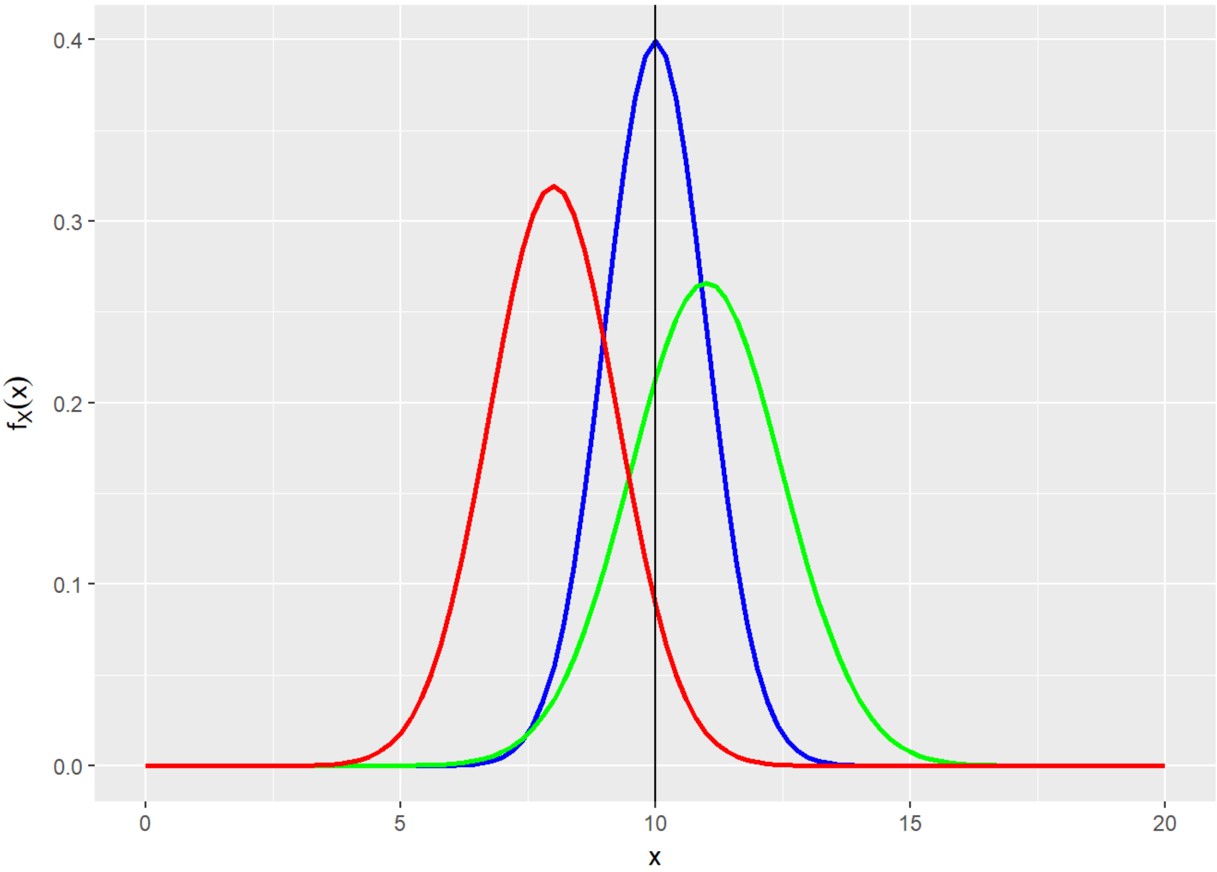


Dado que o consumo foi maior do que 10 litros/100 km, $X > 10$, qual é a probabilidade de ter sido o carro A o escolhido, ou seja,

$$\mathbb{P}(A|X > 10) = \frac{\mathbb{P}(A).\, \mathbb{P}(X > 10|A)}{\mathbb{P}(X > 10)}$$
$$\mathbb{P}(A|X > 10) = \frac{\mathbb{P}(A).\, \mathbb{P}(X > 10|A)}{\mathbb{P}(A).\, \mathbb{P}(X > 10|A) + \mathbb{P}(B).\, \mathbb{P}(X > 10|B) + \mathbb{P}(C).\, \mathbb{P}(X > 10|C)}$$

Para calcularmos as probabilidades condicionais temos,

Code



$$\mathbb{P}(X > 10|A) = 0.50$$

$$\mathbb{P}(X > 10|B) = 0.7475075$$

$$\mathbb{P}(X > 10|C) = 0.05479929$$

Code

```
P(X>10|A) =  0.5
```

Code

```
P(X>10|B) =  0.7475075
```

Code

```
P(X>10|C) =  0.05479929
```

Assim temos que,

$$\mathbb{P}(A|X > 10) = \frac{\mathbb{P}(A).\,\mathbb{P}(X > 10|A)}{\mathbb{P}(X > 10)} = \frac{(2/6)(0.5)}{(2/6)(0.5) + (3/6)(0.748) + (1/6)(0.0548)} = 0.303$$

A probabilidade de ter sido escolhido o carro A, dado que $X > 10$ ocorreu é de 0.303 ou 30.3 %.

b - Qual a probabilidade de o consumo de um automóvel da marca C ser pelo menos 1 litro inferior ao consumo de um automóvel da marca A?

Nesta questão vamos que utilizar a distribuição da diferença entre variáveis aleatórias, ou seja a distribuição da diferença entre o consumo do carro C e A. Como consequência teremos uma nova distribuição normal, sendo C e A normais, e aplicaremos a propriedade da linearidade do valor esperado e da variância para determinar os parâmetros desta nova distribuição.

Assim temos que a nova variável aleatória é D onde,

$$D = C - A$$

Sabemos que D possui distribuição normal, com média e variância / desvio padrão.

Calculando os parâmetros,

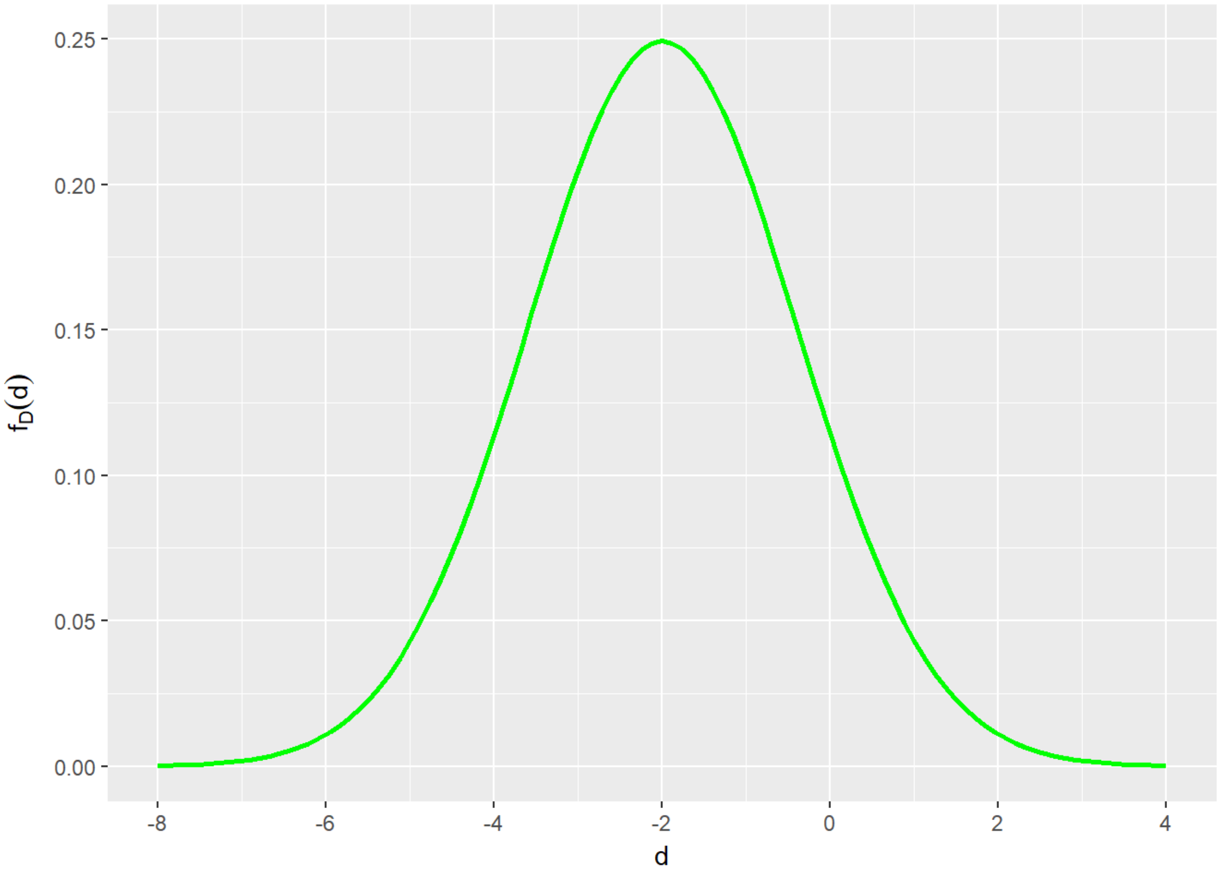
$$E[C - A] = E[C] - E[A]$$
$$V(C - A) = V(C) + V(A)$$

Substituindo,

$$E[D] = E[C] - E[A] = 8 - 10 = -2$$
$$V(D) = V(C) + V(A) = (1)^2 + (1.25)^2 = 2.5625$$
$$DP(D) = 1.60$$

Dessa forma a distribuição da v.a. D representa a diferença no consumo do carro C em relação ao carro A.

Code



Para calcularmos a probabilidade de o consumo de um automóvel da marca C ser pelo menos 1 litro inferior ao consumo de um automóvel da marca A,

$$\mathbb{P}(D < -1) = 0.7340145$$

Intepretação:

Considere o seguinte,

- se $D = 0$ significa que o consumo dos dois automóveis C e A não se diferenciam
- se $D = 1$ significa que o consumo de C é 1 litro superior ao de A
- se $D = -1$ significa que o consumo de C é 1 litro inferior ao de A
- na média o consumo de D é de 2 litros inferior ao de A ($\mu = -2$)

Assim,

Code

```
P(D < -1) =  0.7340145
```

Code