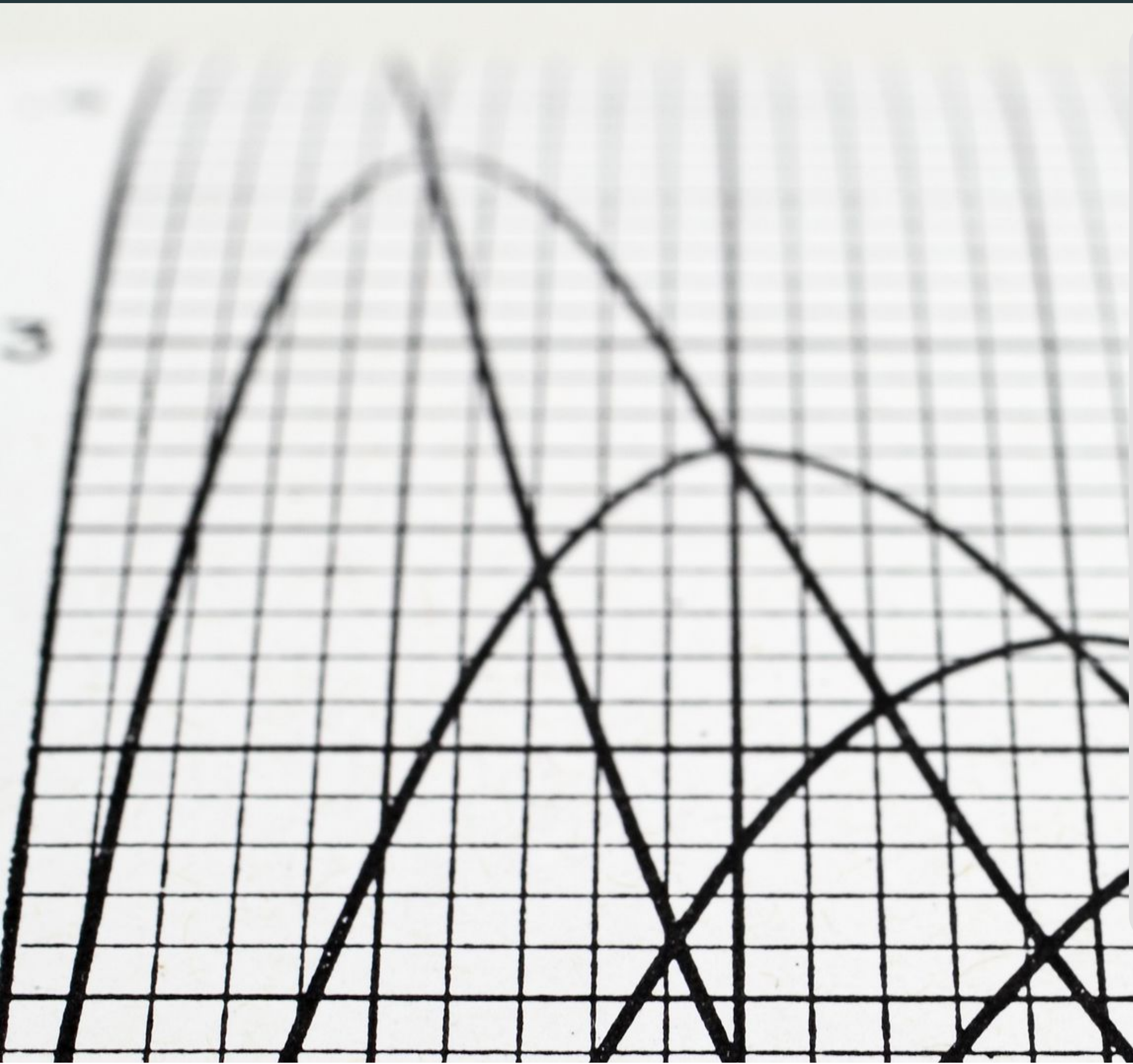
Inferência Estatística e Teste de Hipóteses

O modelo da distribuição normal

Fabio Cop (fabiocopf@gmail.com)
Instituto do Mar - UNIFESP
Última atualização em 18 de março de 2022

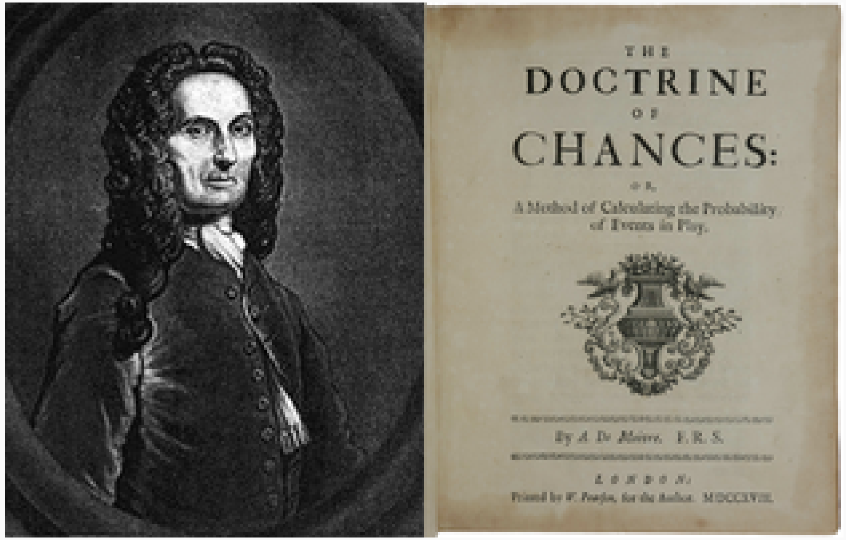
Conteúdo da aula



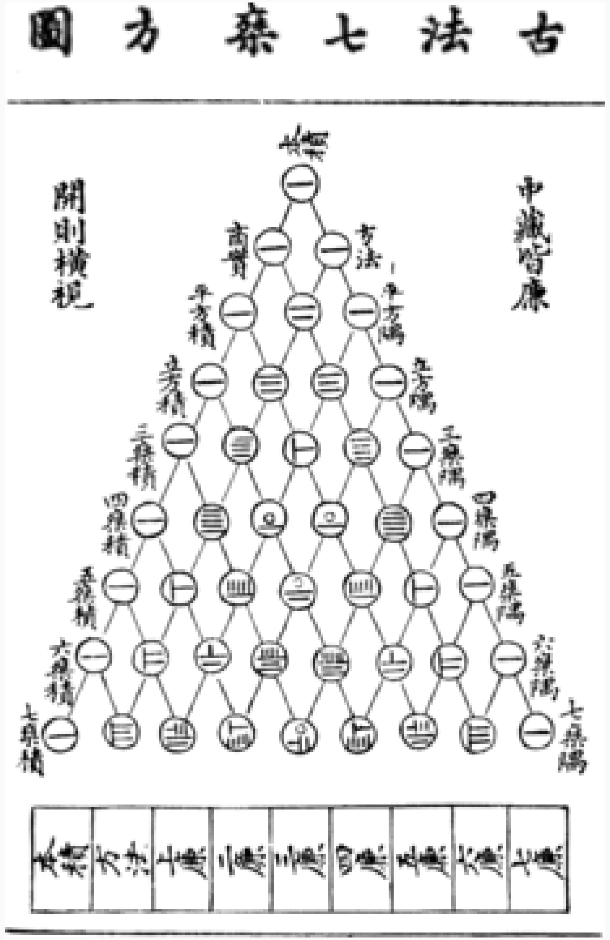
1. Um pouco de história
2. A importância do modelo normal em estatística
3. O modelo normal de probabilidades
4. Entendendo a função normal de densidade de probabilidade
5. Cálculo de probabilidade com a função normal de densidade
6. A distribuição normal padronizada
7. Probabilidades em uma distribuição normal padronizada

1. Um pouco de história

Alguns atribuem a proposição deste modelo a Abraham de Moivre, um matemático Francês que chegou a a distribuição normal como uma aproximação a distribuição binomial em seu livro The Doctrine of Chances em 1718.



Abraham de Moivre (1667 - 1754)



Triângulo de Yang Hui's de 1261

1. Um pouco de história

A distribuição normal também é conhecida como **distribuição gaussiana**. Carl Friedrich Gauss lidou com este modelo para lidar com a distribuição dos erros de observação de um fenômeno no contexto do Método dos Mínimos Quadrados em 1823.

1. Um pouco de história

Pierre-Simon Laplace fez contribuições importantes ao modelo e aplicação da distribuição normal. Calcular a integral $\int e^{-t^2} = \sqrt{\pi}$ em 1782, o que permitiu a **padronização de uma distribuição normal** e, em 1810, apresentou o Teorema Central do Limite no contexto da agregação de várias observações independentes, o que foi fundamental para o desenvolvimento da estatística experimental ao longo do século XX.

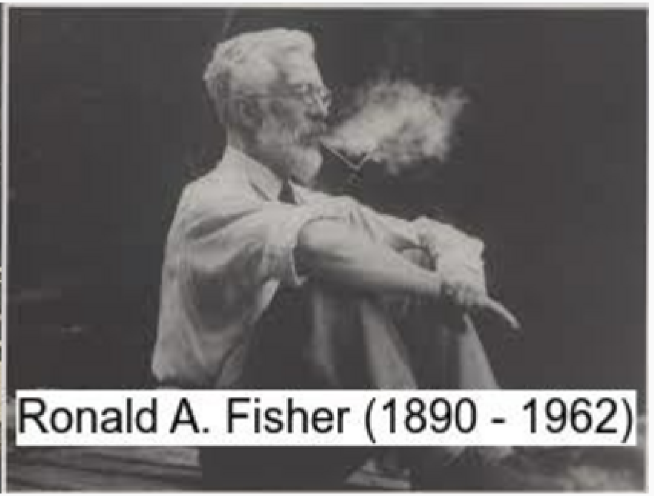
Pierre-Simon Laplace (1748 - 1827)

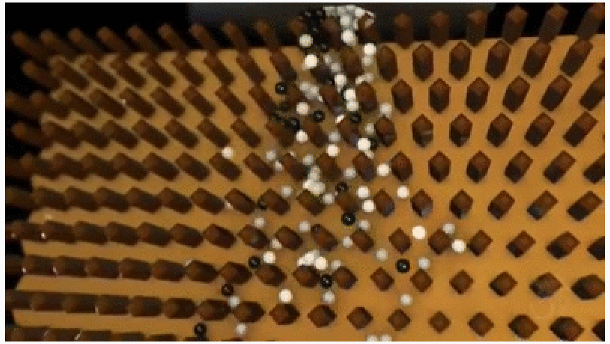
1. Um pouco de história

Até o final do início do século (XX) acurva era conhecida pelo termo **modelo de Gauss-Lalpace**, quando começou a ser referida como **distribuição normal** por Karl Pearson é vários outros autores. Pearson foi também o primeiro a escrever o modelo em termos de seu **desvio padrão** σ. Logo em seguida, em 1915 Ronald Fisher adicionou o parâmetro de posição (m), isto é a média, expressando o modelo como fazemos atualmente.

$$df = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-m)^2}{2\sigma}} dx$$

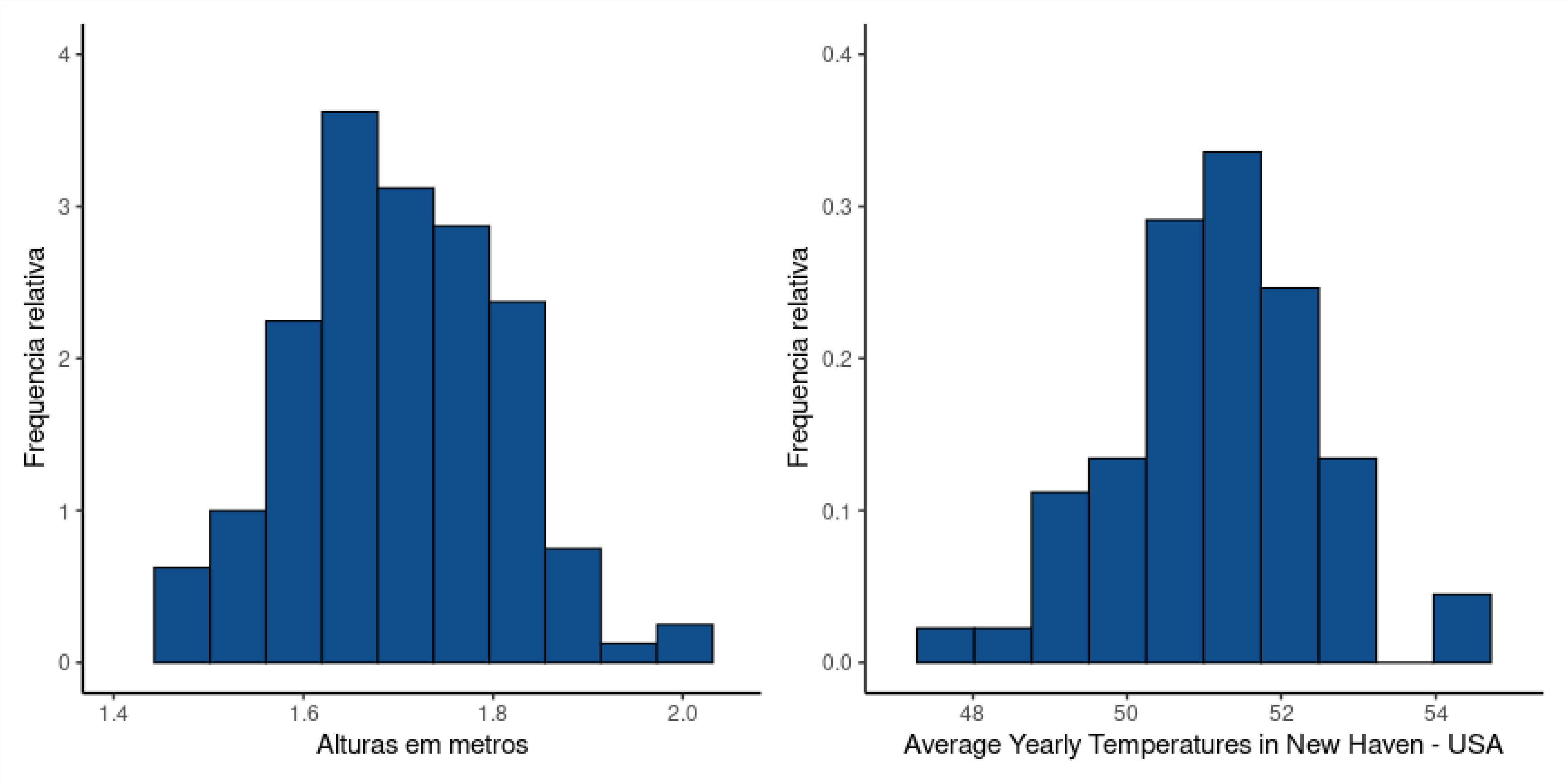
Finalmente, a **distribuição normal padrão**, com média (0) e desvio padrão (1) começou a apracerer nos livros texto por volta de 1950.



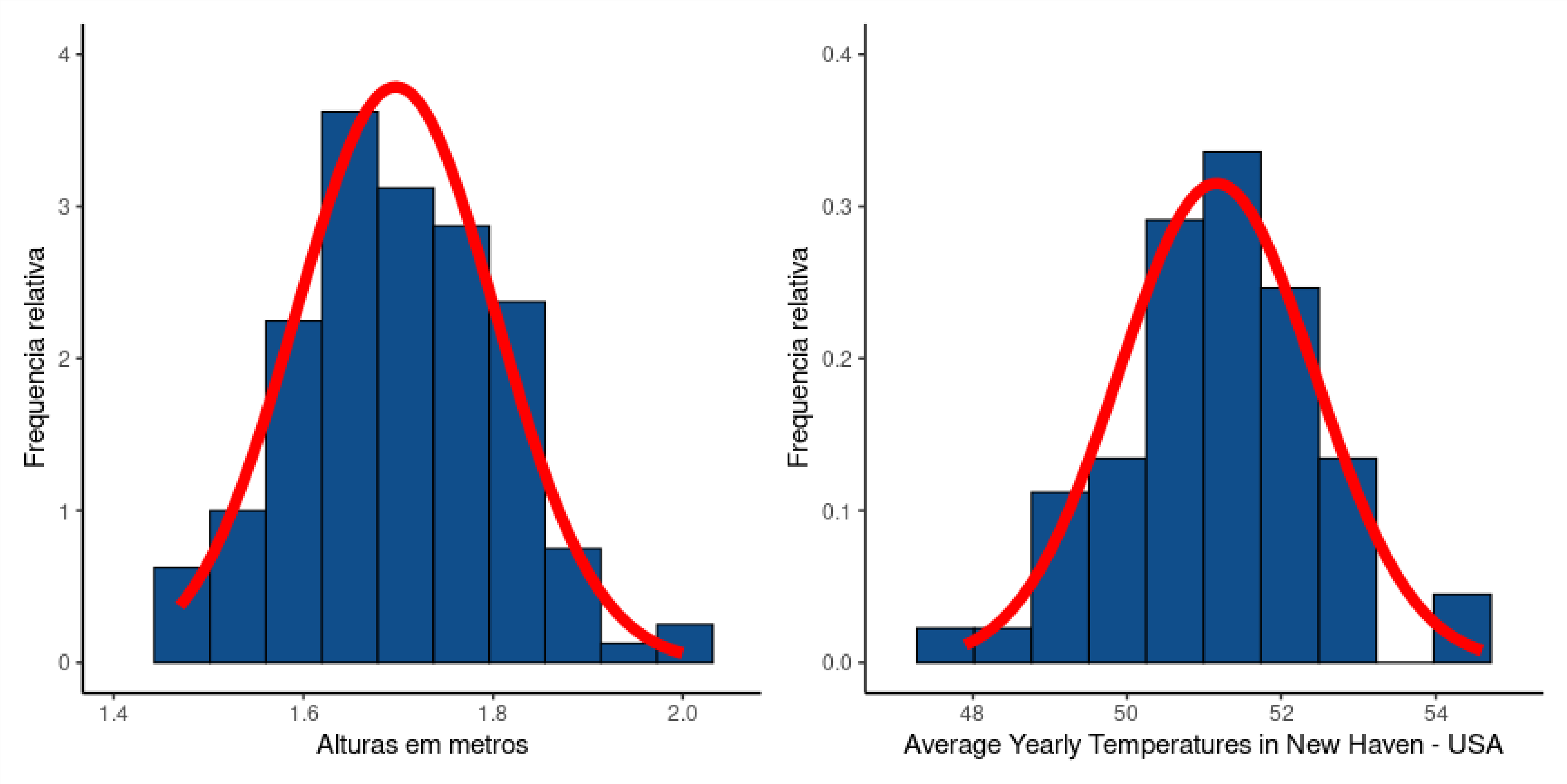


2. A importância do modelo normal em estatística

- A descrição de fenômenos e predições com o modelo normal.

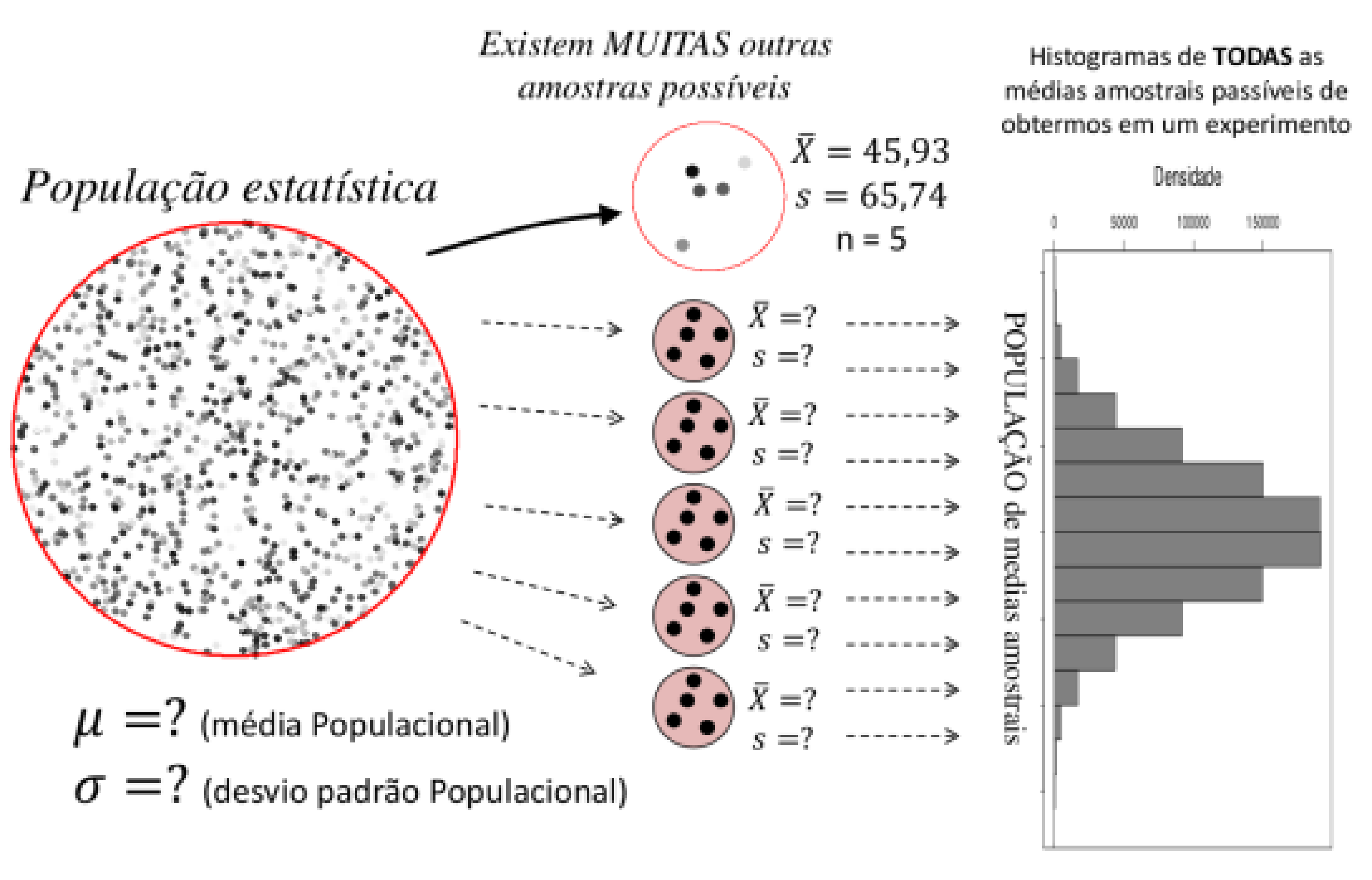


- A descrição de fenômenos e predições com o modelo normal.



2. A importância do modelo normal em estatística

- O modelo normal com resultado do processo de amostragem



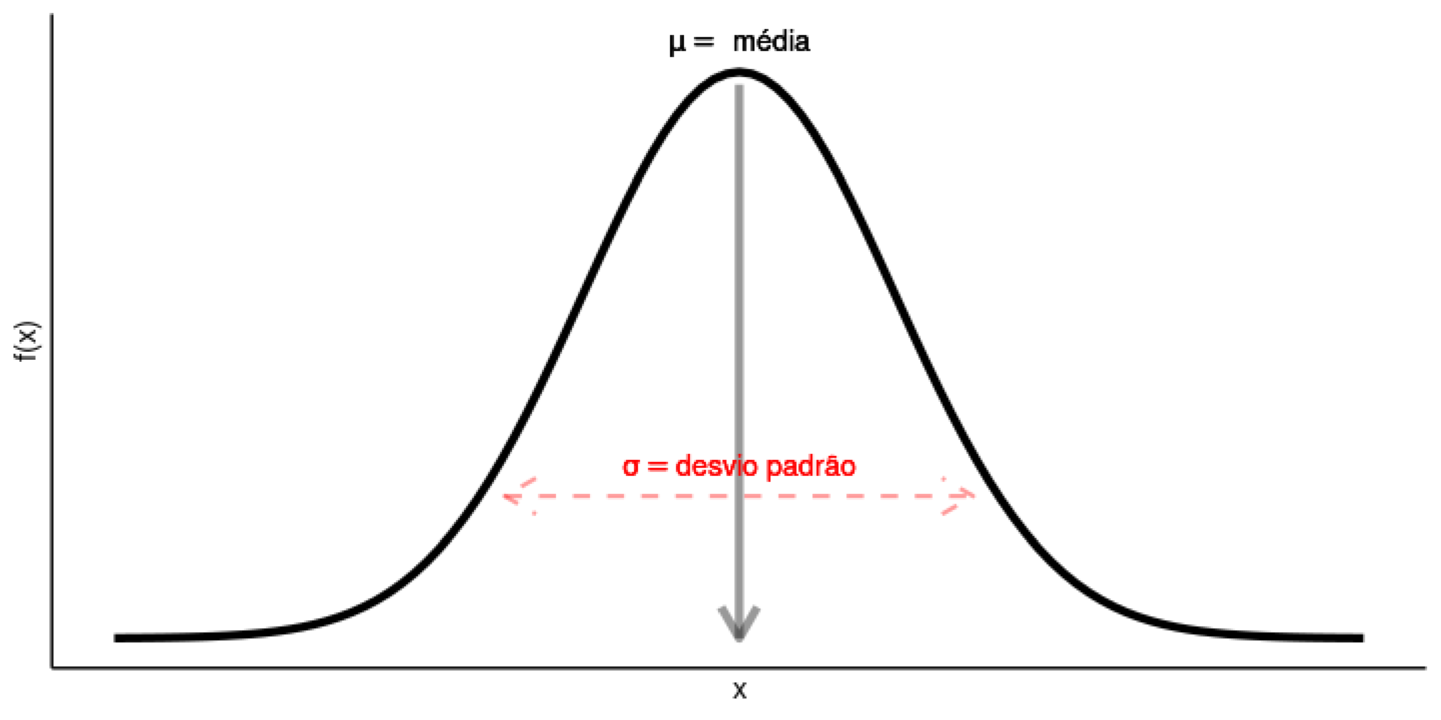
3. O modelo normal de probabilidades

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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

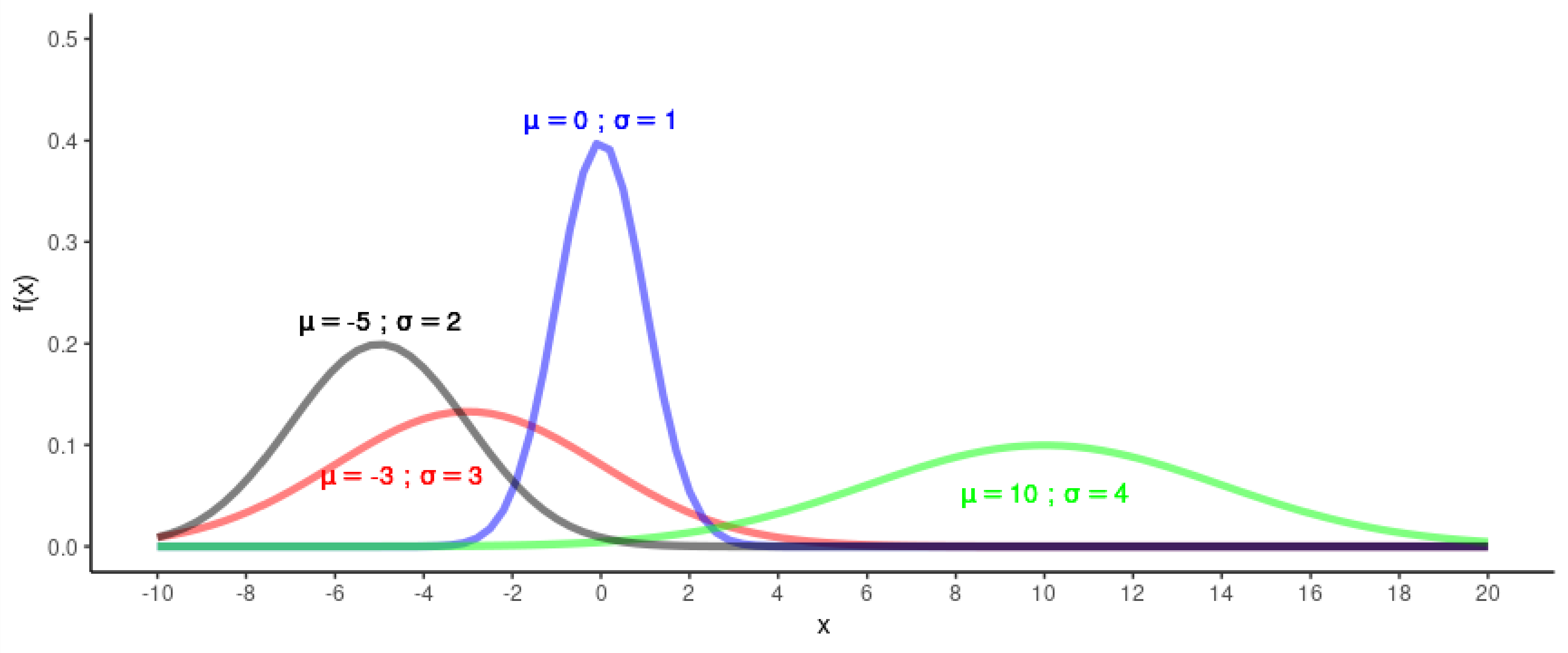
Lê-se:

X é uma variável aleatória com distribuição normal, média μ e desvio padrão σ.



$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Podemos alterar o formato da distribuição normal alterando seu parâmetos de posição (média - μ) e de dispersão (desvio padrão - σ).



4. Entendendo a função normal de densidade de probabilidade

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Encontrando $f(x)$ para pontos específicos.

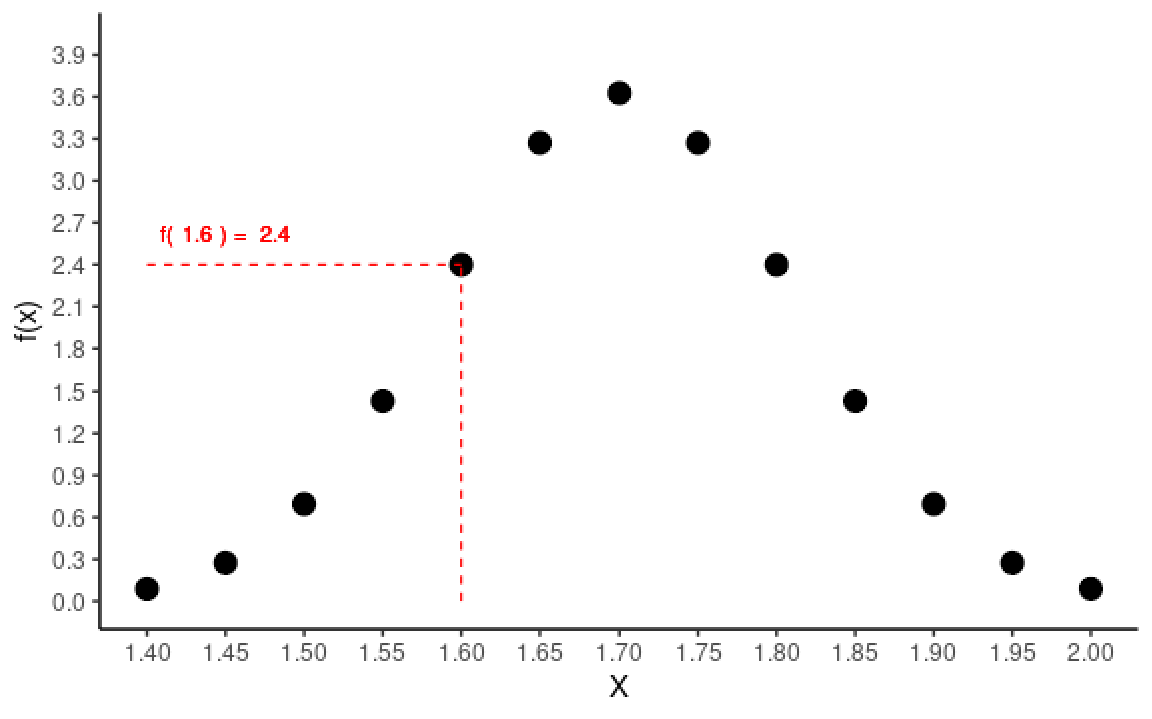
Seja:

$\mu = 1.7$; $\sigma = 0.11$

O ponto $x = 1.6$ tem:

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}(\frac{1.6-\mu}{\sigma})^2}$$

$$f(x) = 2.4$$



4. Entendendo a função normal de densidade de probabilidade

Calculando de $f(x)$ no R: a função `dnorm()`

Seja $\mu = 1.7$ e $\sigma = 0.11$

Para $x = 1.5$ fazemos:

```
mu <- 1.7
dp <- 0.11
dnorm(1.5, mean = mu, sd = dp)
```

```
## [1] 0.6945048
```

E para múltiplos valores de x:

```
x <- c(1.4, 1.5, 1.6, 1.7)
dnorm(x, mean = mu, sd = dp)
```

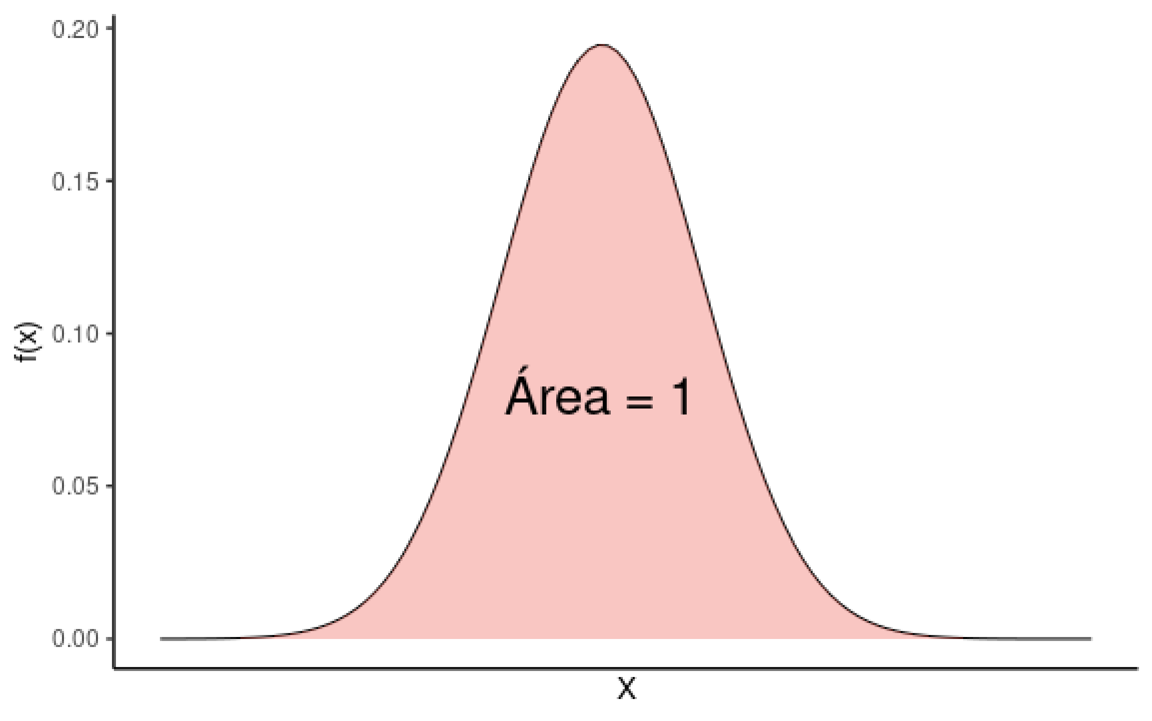
```
## [1] 0.0879777 0.6945048 2.3991470 3.6267480
```

5. Cálculo de probabilidade com a função normal de densidade

A distribuição normal é uma função de **densidade de probabilidade**.

A **área** abaixo de $f(x)$ entre $-\infty$ e $+\infty$ é igual a 1.

$$P(-\infty \leq X \leq +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1$$

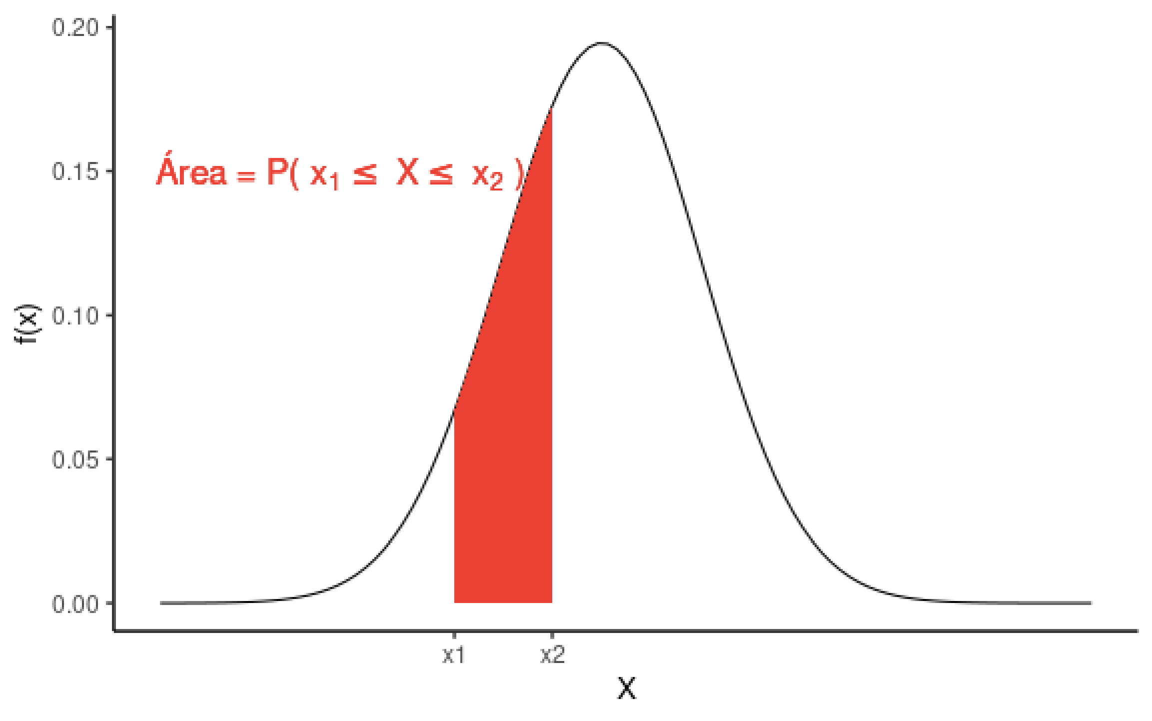


5. Cálculo de probabilidade com a função normal de densidade

A distribuição normal é uma função de **densidade de probabilidade**.

Utilizamos a distribuição normal para encontrar a probabilidade de uma variável aleatória X estar entre x_1 e x_2, o que corresponde a encontrar a **área** abaixo de $f(x)$.

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$



5. Cálculo de probabilidade com a função normal de densidade

Calculando de $P(X)$ no R: a função `dnorm()`

Seja $\mu = 1.7$ e $\sigma = 0.11$

$P(X < 1.5)$:

```
mu <- 1.7
dp <- 0.11
pnorm(q = 1.5, mean = mu, sd = dp, lower.tail = TRUE
```

```
## [1] 0.03451817
```

$P(X > 1.5)$ --> `lower.tail == FALSE`:

```
mu <- 1.7
dp <- 0.11
pnorm(q = 1.5, mean = mu, sd = dp, lower.tail = FALSE
```

```
## [1] 0.9654818
```

Os argumentos da função `pnorm` (veja o menu de ajuda digitando `?pnorm` no Console do R):

> **q**: o valor de x
>
> **mean**: média μ da função normal
>
> **sd**: desvio padrão σ da função normal
>
> **lower.tail**: se a função irá retornar a probabilidade abaixo (TRUE) ou acima (FALSE) de q

5. Cálculo de probabilidade com a função normal de densidade

Calculando de $P(X)$ no R: a função `dnorm()`

Seja $\mu = 1.7$ e $\sigma = 0.11$

$P(X < 1.5)$:

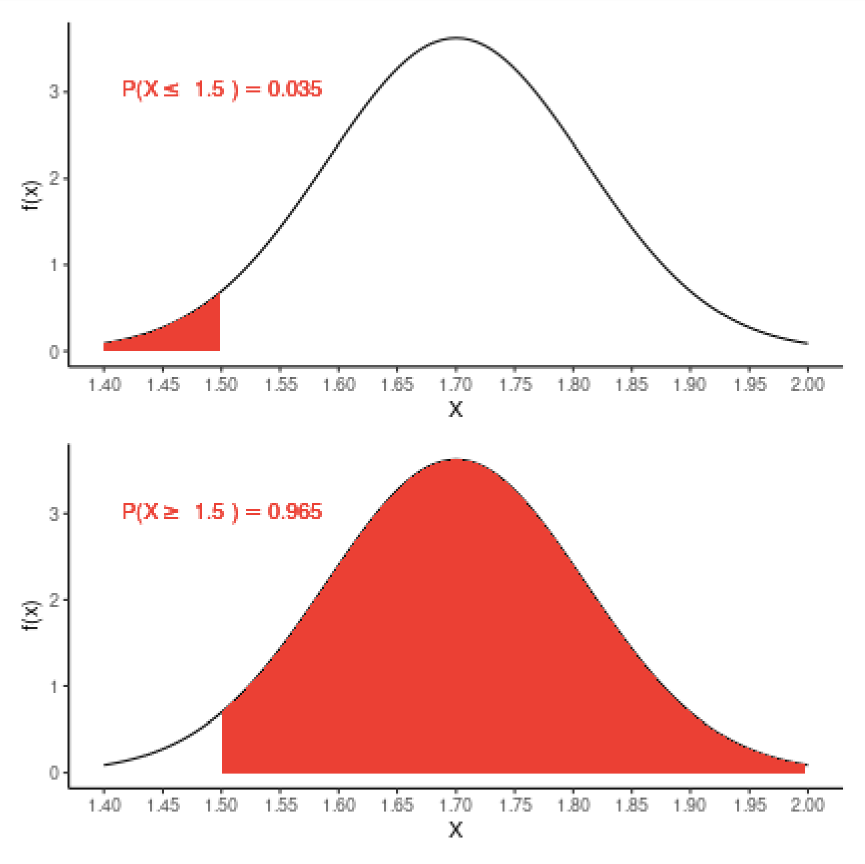
```
mu <- 1.7
dp <- 0.11
pnorm(q = 1.5, mean = mu, sd = dp, lower.tail = TRUE
```

```
## [1] 0.03451817
```

$P(X > 1.5)$ --> `lower.tail == FALSE`:

```
mu <- 1.7
dp <- 0.11
pnorm(q = 1.5, mean = mu, sd = dp, lower.tail = FALS
```

```
## [1] 0.9654818
```



5. Cálculo de probabilidade com a função normal de densidade

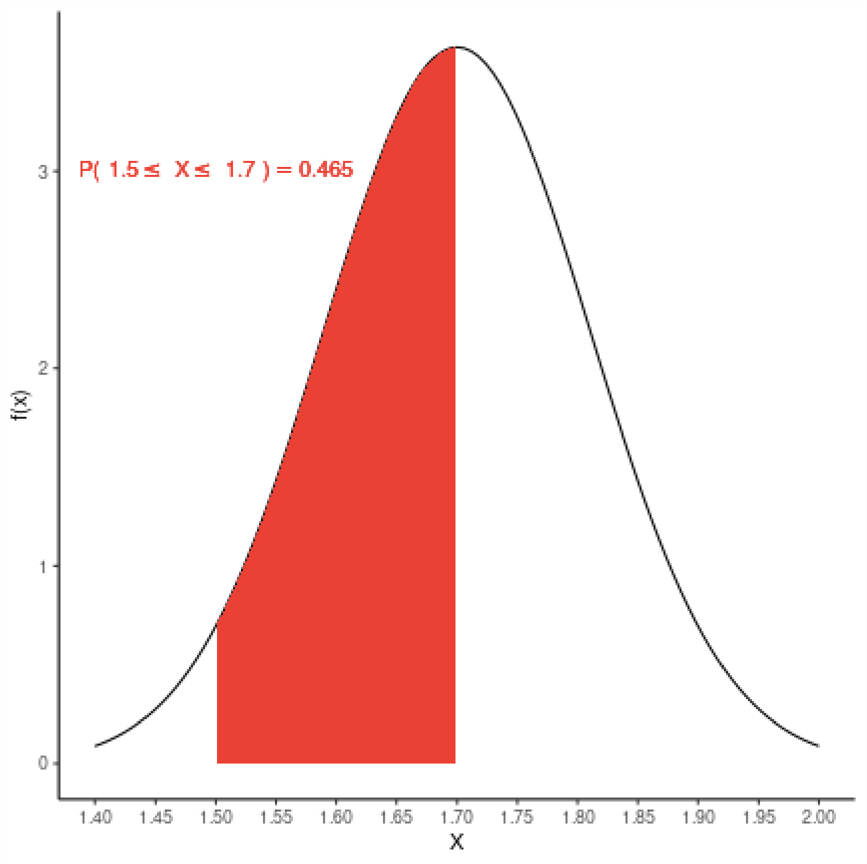
Calculando de $P(X)$ no R: a função `dnorm()`

Seja $\mu = 1.7$ e $\sigma = 0.11$

$$P(1.5 \leq X \leq 1.7) = P(X \leq 1.7) - P(X \leq 1.5)$$

```
diff(pnorm(q = c(1.5, 1.7),
           mean = mu,
           sd = dp,
           lower.tail = TRUE)
     )
```

```
## [1] 0.4654818
```

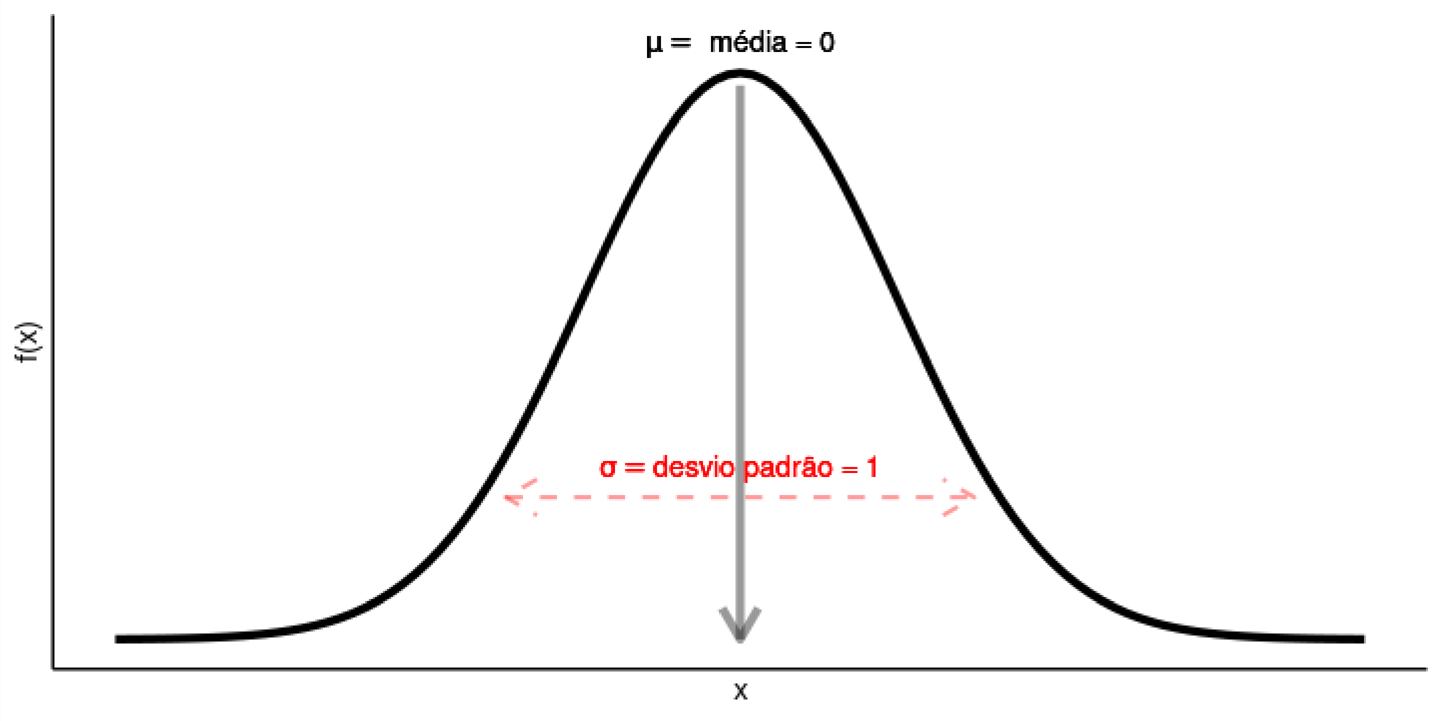


6. A distribuição normal padronizada

$$f(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

$$X \sim \mathcal{N}(0,\, 1)$$

X é uma variável aleatória com distribuição normal **padronizada** se, média $\mu = 0$ e desvio padrão $\sigma = 1$.



6. A distribuição normal padronizada: a transformação Z

$$z_i = \frac{x_i - \mu}{\sigma}$$

Seja $\mu = 1.7$ e $\sigma = 0.11$

A observação $x = 1.5$ corresponte a:

$$z_i = \frac{x_i - \mu}{\sigma} = \frac{1.5 - 1.7}{0.11} = -1.82$$

Em uma distribuição normal com $\mu = 1.7$ e $\sigma = 0.11$, a observação $x = 1.5$ está 1.82 desvios padrões abaixo da média.

Seja $\mu = 1.7$ e $\sigma = 0.11$

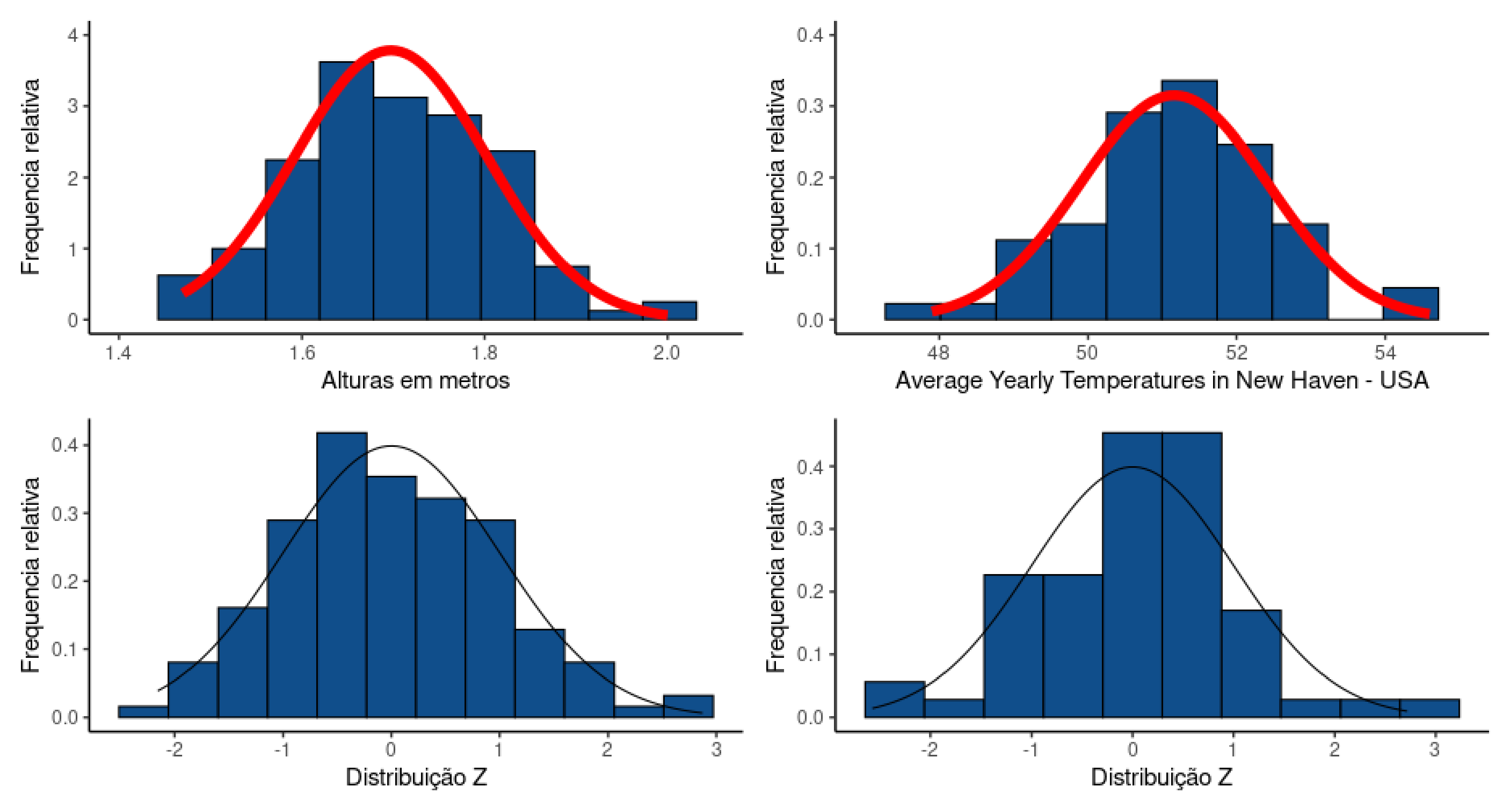
A observação $x = 2.2$ corresponte a:

$$z_i = \frac{x_i - \mu}{\sigma} = \frac{2.2 - 1.7}{0.11} = 4.55$$

Em uma distribuição normal com $\mu = 1.7$ e $\sigma = 0.11$, a observação $x = 2.2$ está 4.55 desvios padrões acima da média.

6. A distribuição normal padronizada

Após a transformação Z nos exemplos sobre altura dos alunos e chuva mensal temos:



6. Probabilidades em uma distribuição normal padronizada

Limites conhecidos de probabilidade