

# Curso de Estatística Básica

Pavel Dodonov

pdodonov@gmail.com

Aula 2b – Trabalhando (mais um pouco) com uma amostra

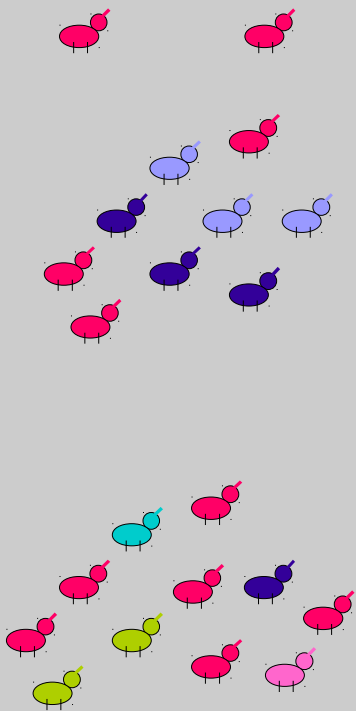
# Estatística

Inferencial

Descritiva

# Estatística inferencial

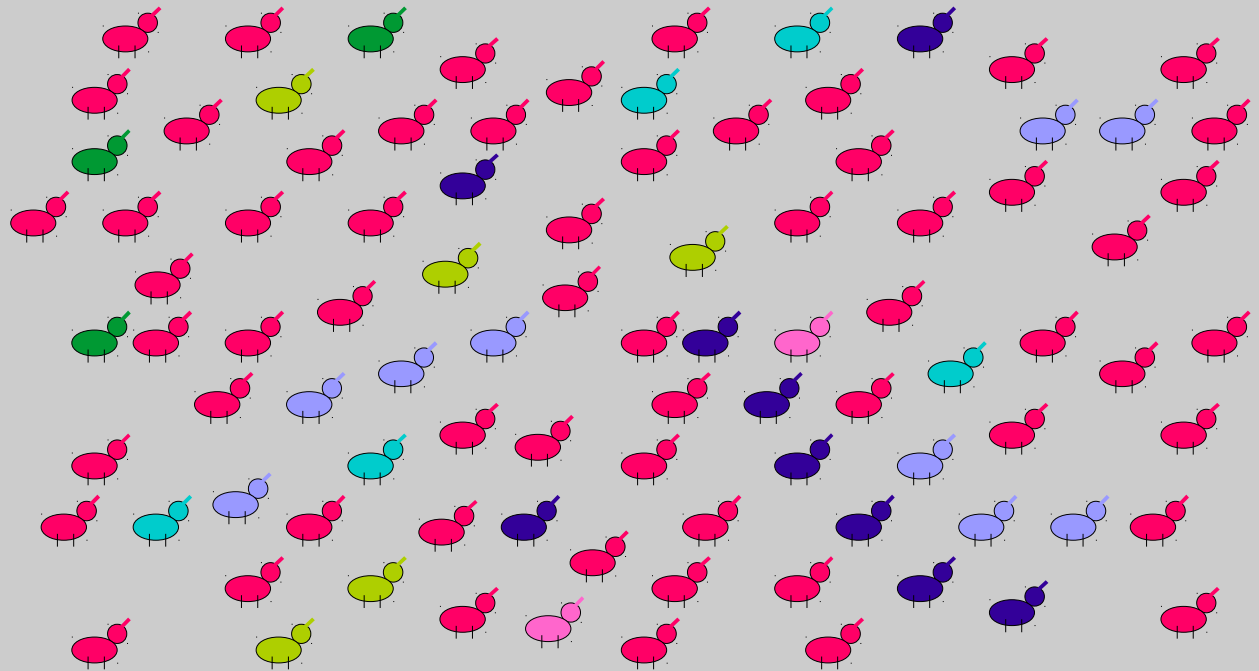
Usualmente, testar hipóteses



Amostra

Estatística

População

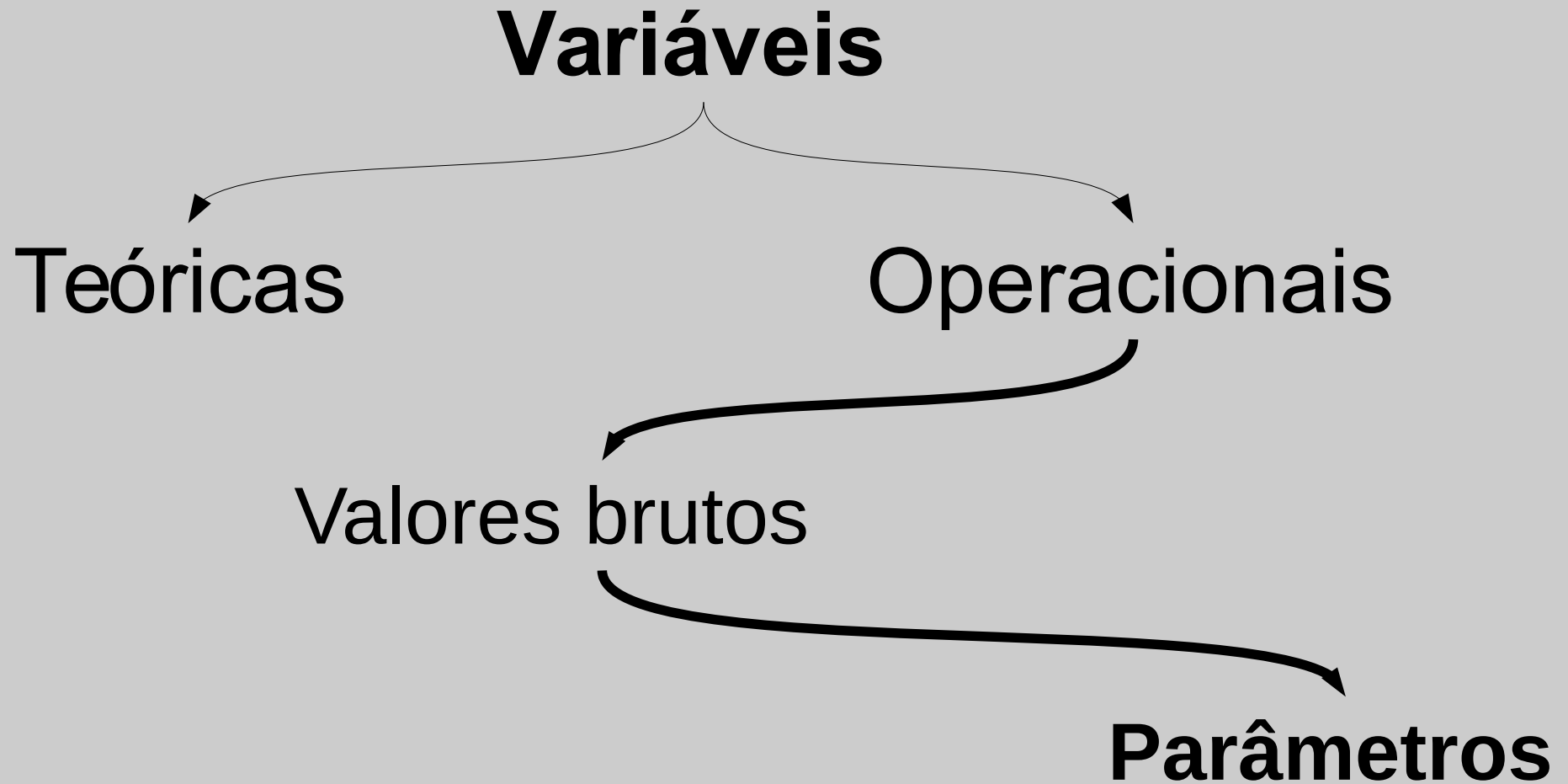


# Estatística descritiva

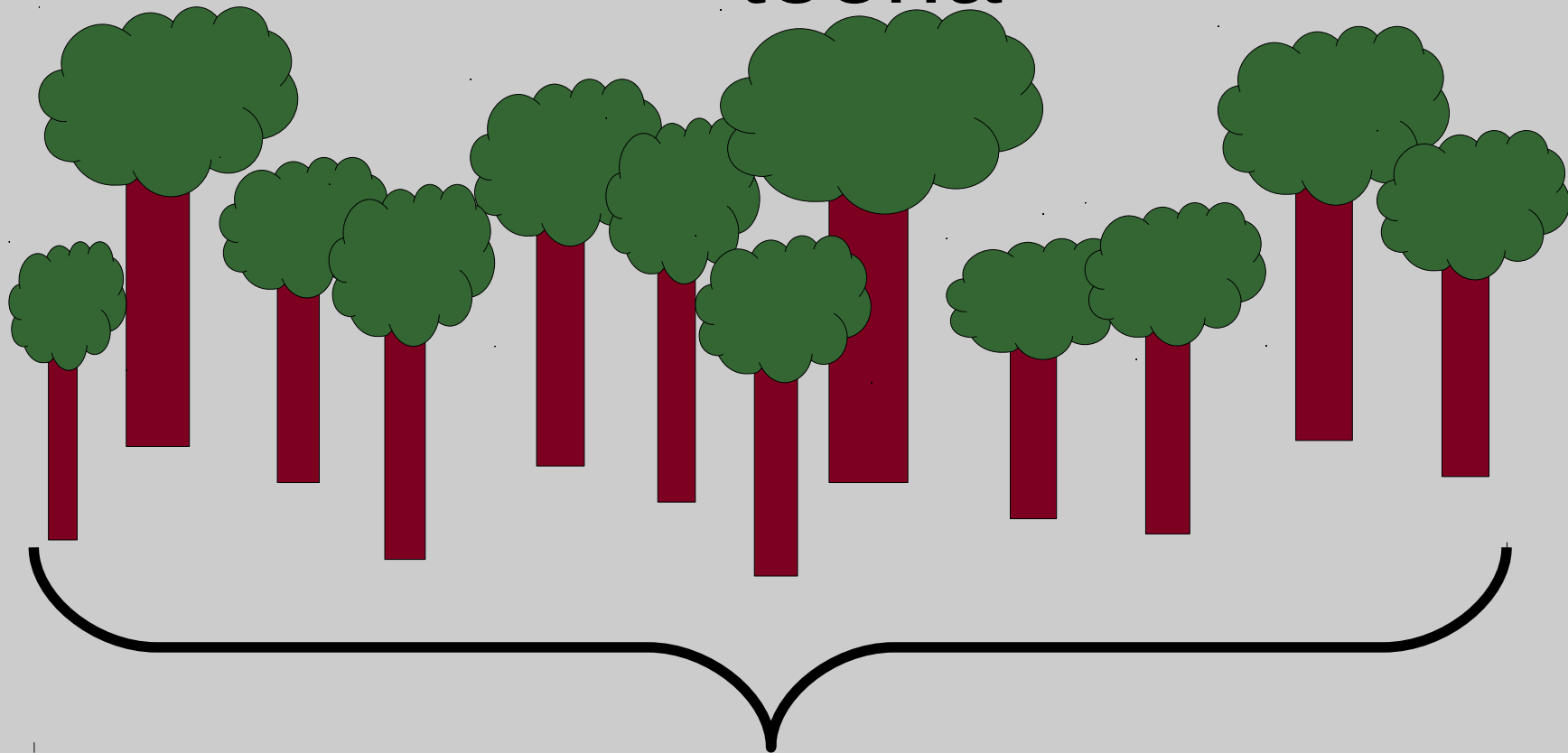
Descrever a nossa  
amostra

Envolve decidir:  
Quais aspectos dela  
merecem atenção?

# Operacionalizando [mais ainda] a teoria



# Operacionalizando [mais ainda] a teoria



DAP = (10, 40, 20, 19, 25, 24, 30, 60, 28, 30, 42, 20) cm



“Eu medi doze árvores,  
com diâmetros de 10,  
40, 20, 19, 25, 24, 30,  
60, 28, 30, 42 e 20  
cm.”



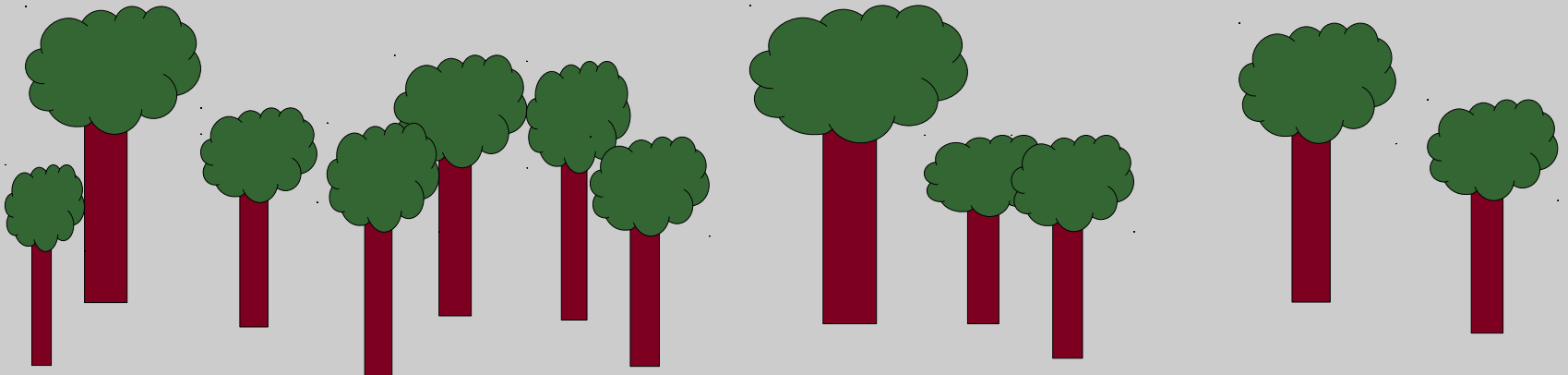
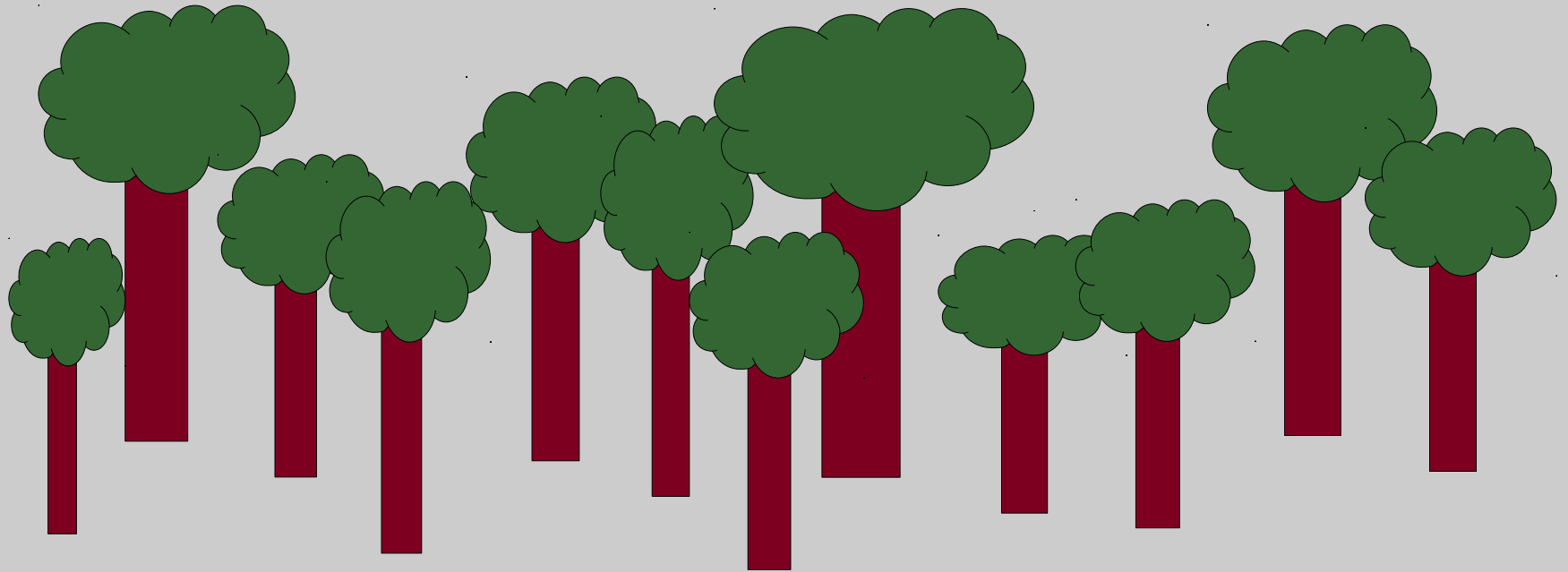
O que importa



# Estatística descritiva – uma amostra

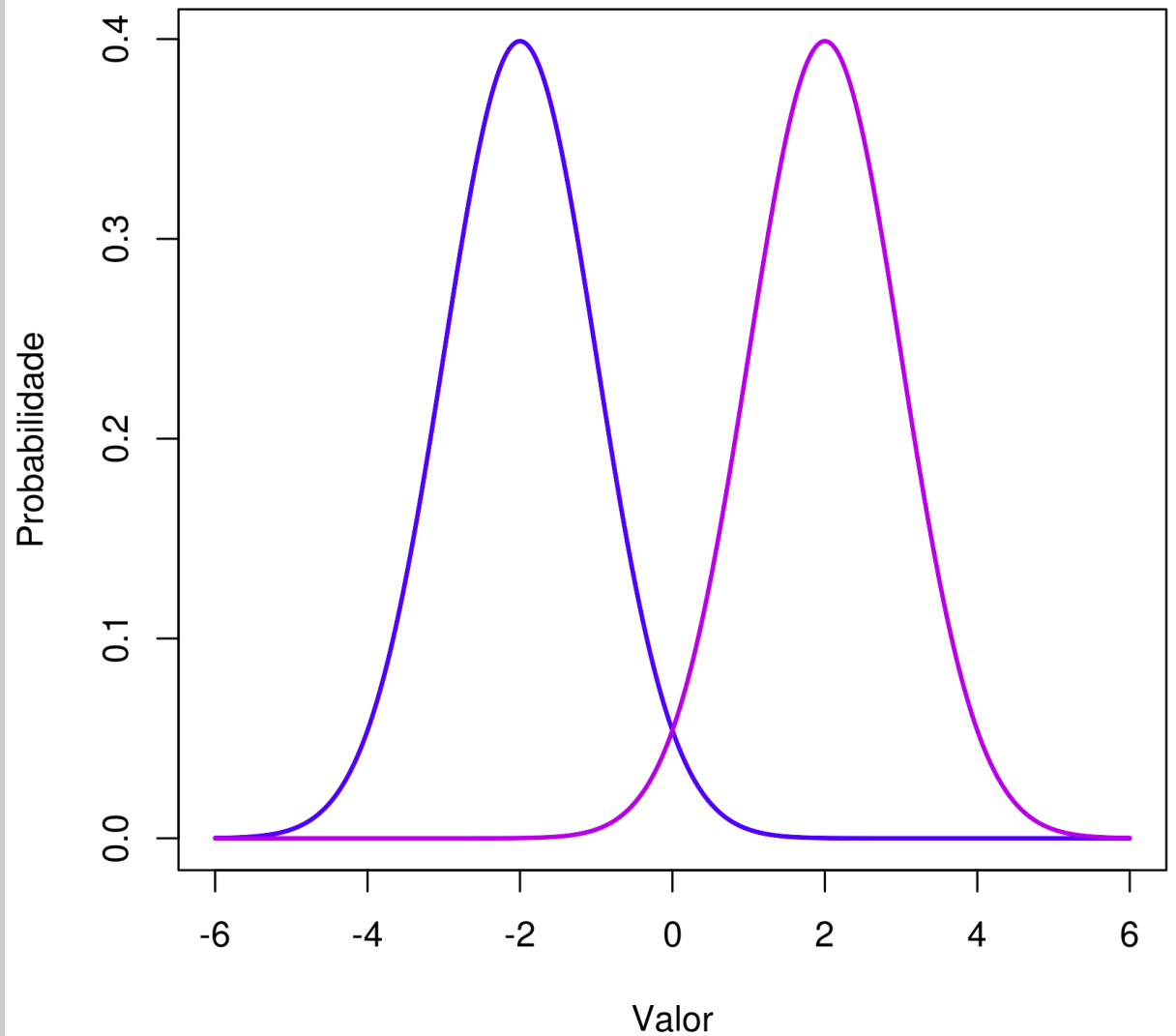
- Tendência central
- Variação
- Forma da distribuição
- Outras

# Tendência central



# Tendência central

Onde está o  
“centro” da  
distribuição?



# Tendência central

- Média

# Aplicação prática: Fator de Impacto

$$IF = \frac{\textit{número de citações}}{\textit{número de artigos}}$$

Muito influenciada por valores extremos

Diferenças menores que uma ou duas unidades são irrelevantes na prática

# Situação hipotética...

Revista Aleatória de  
Coisas (IF=2.37)

# citações	# artigos
0	2
1	6
2	10
3	6
4	3
5	3

Revista de Coisas  
Aleatórias (IF=1.80)

# citações	# artigos
0	4
1	12
2	9
3	8
4	1
5	1

# Situação hipotética...

Revista Aleatória de  
Coisas (IF=2.37)

# citações	# artigos
0	2
1	6
2	10
3	6
4	3
5	3

Revista de Coisas  
Aleatórias (IF=4.53)

# citações	# artigos
0	4
1	12
2	9
3	8
4	1
5	1
100	1

# Aplicação prática: Fator de Impacto

Para saber mais:

<https://dynamicecology.wordpress.com/2016/06/21/impact-factors-are-means-and-therefore-very-noisy/>



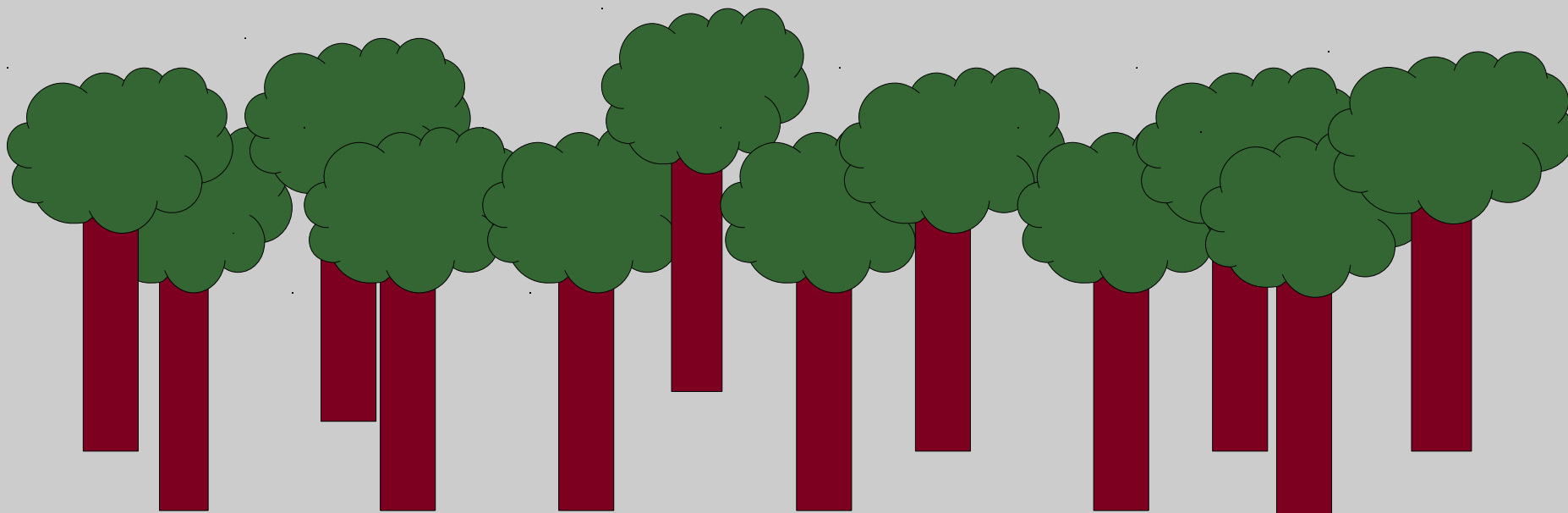
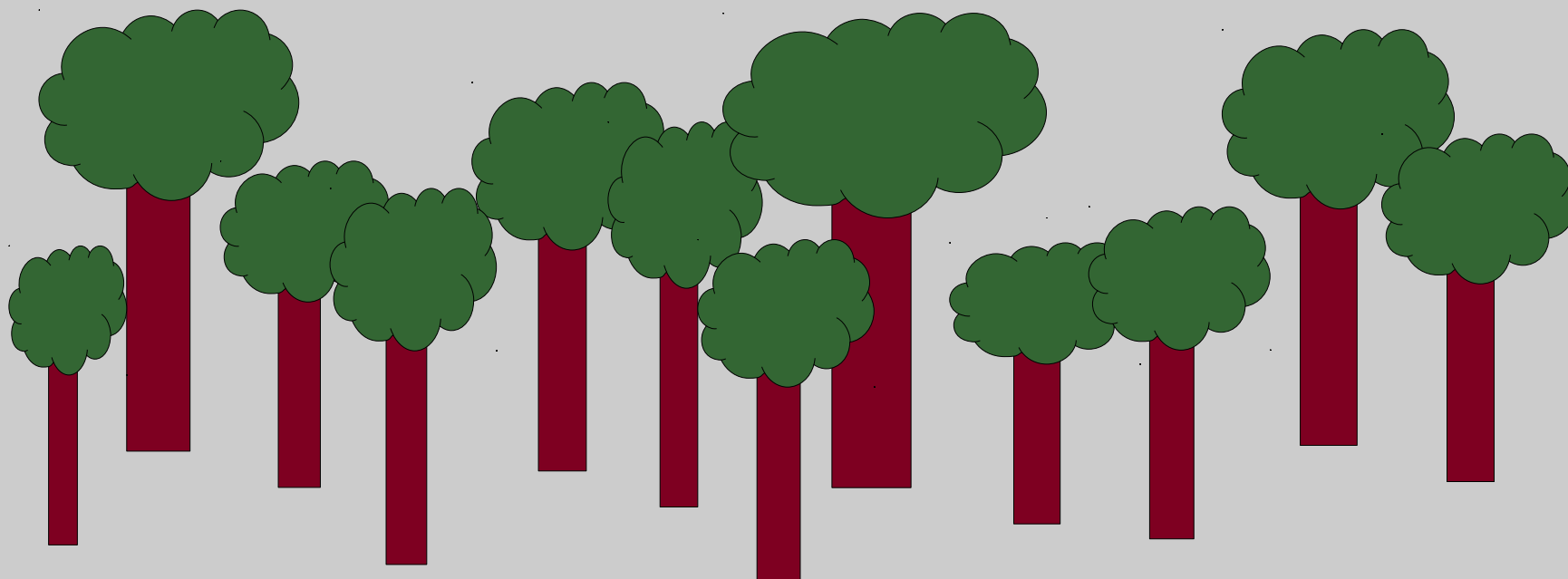
# Tendência central

- Mediana
  - Metade de valores acima, metade de valores abaixo

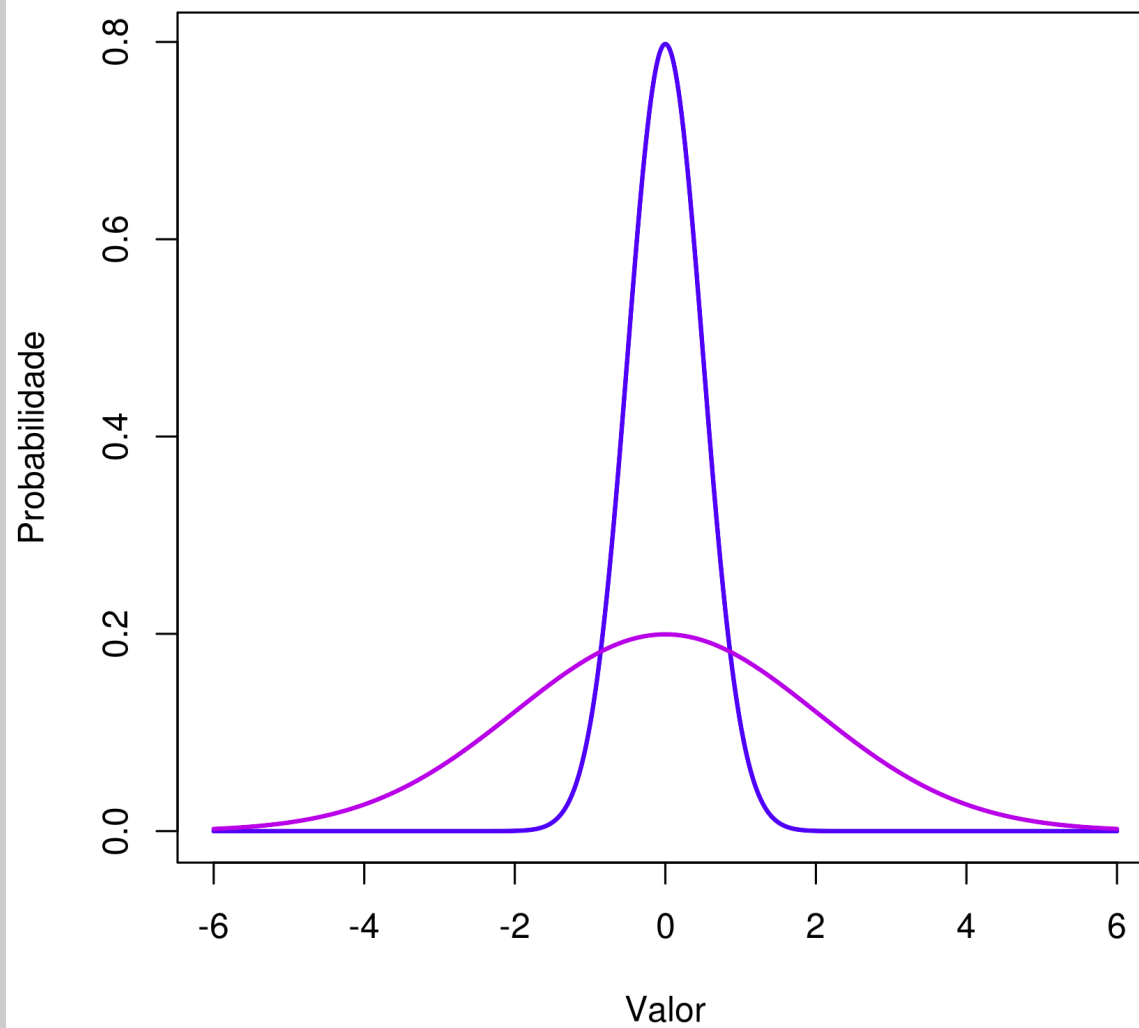
# Tendência central

- Moda
  - O valor mais comum

# Variação



# Variação



# Variação

- Amplitude
  - Diferença entre valor mínimo e máximo

# Variação

- Variância

# Variação

- Desvio-padrão

# Variação

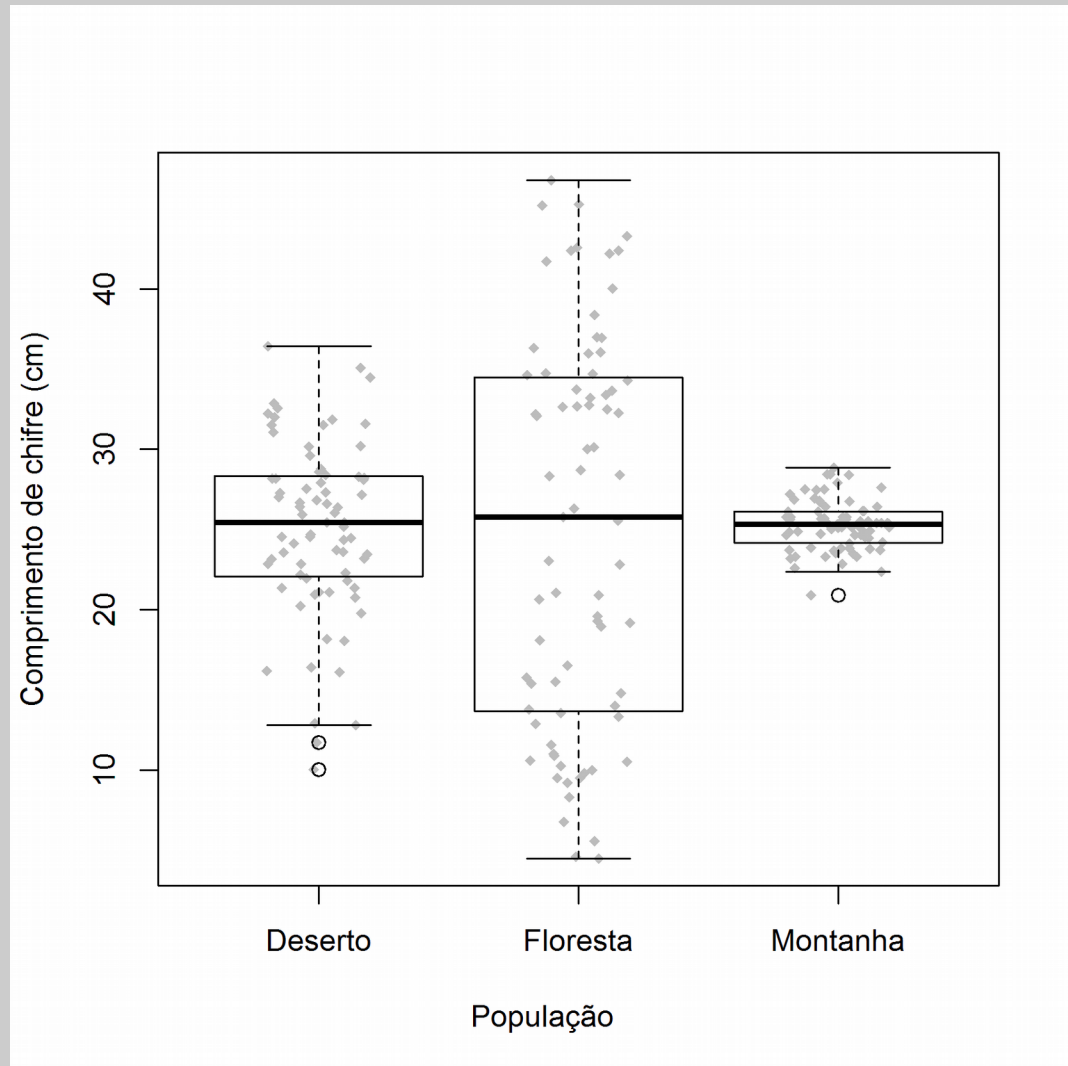
- Coeficiente de variação

$$CV = \frac{\textit{média}}{\textit{desvio padrão}}$$



# Variação

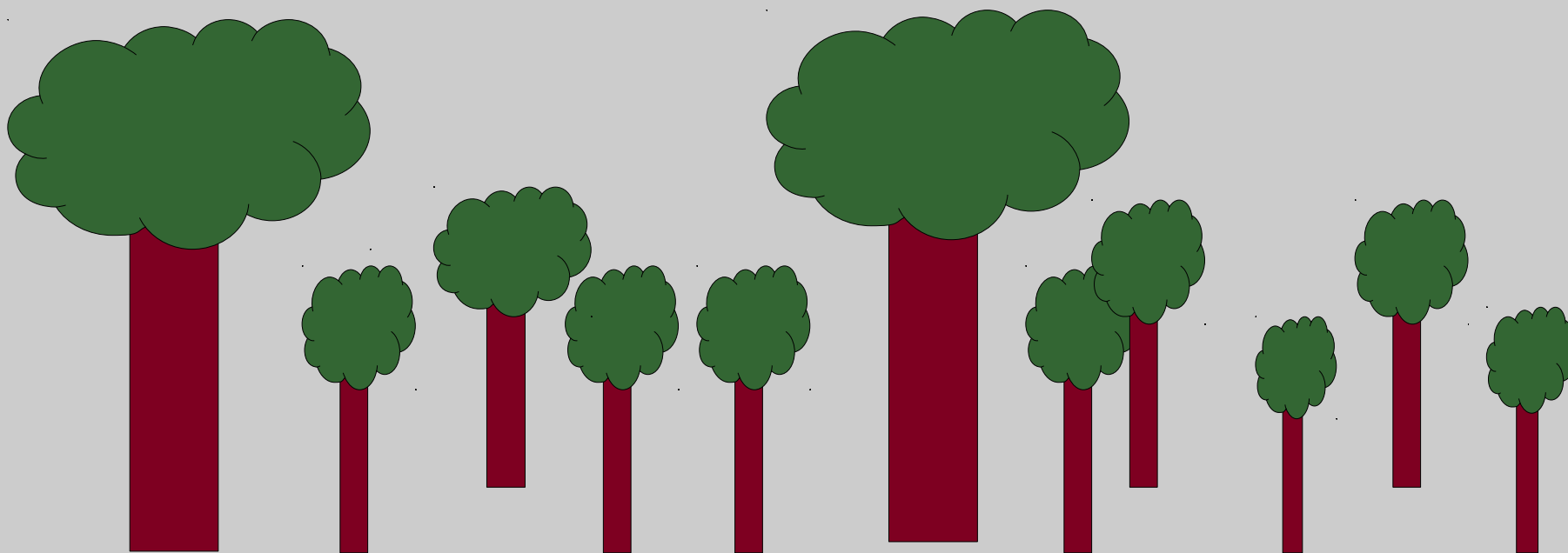
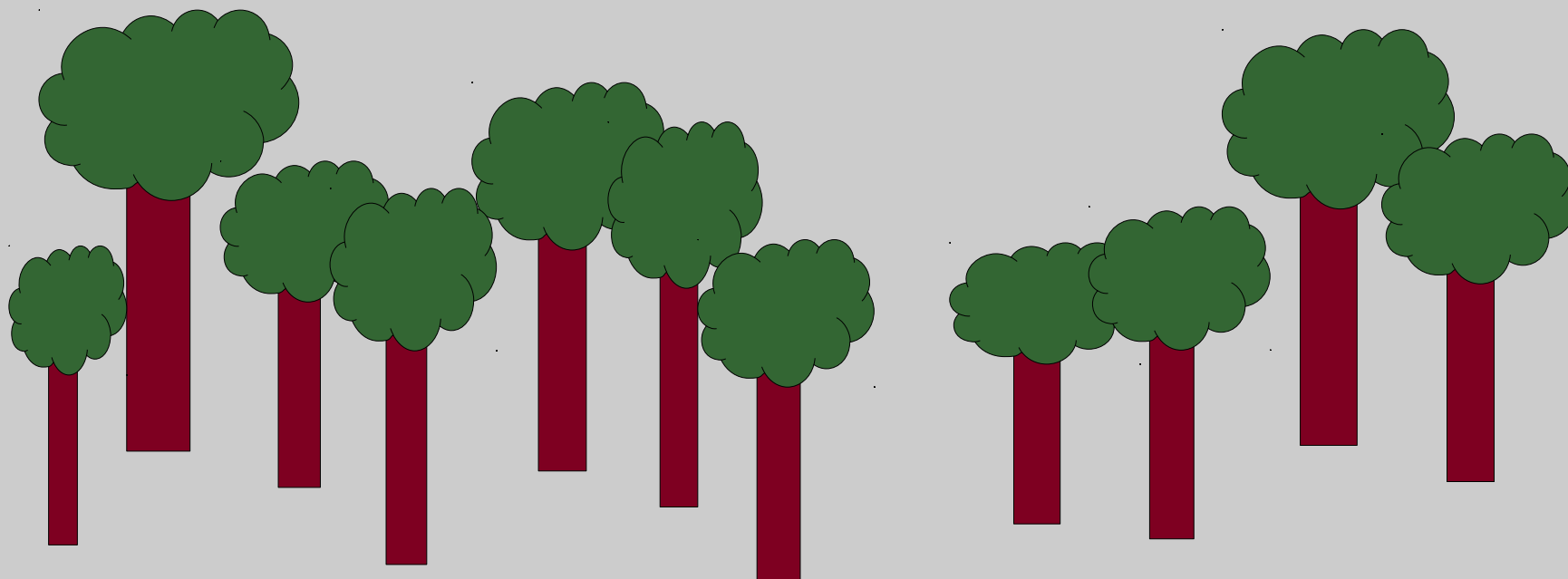
- Distância inter-quantis



# Quantis

- Quantil  $x\%$ :  $x\%$  dos dados está abaixo deste valor
- Mediana é o quantil 50% :-)

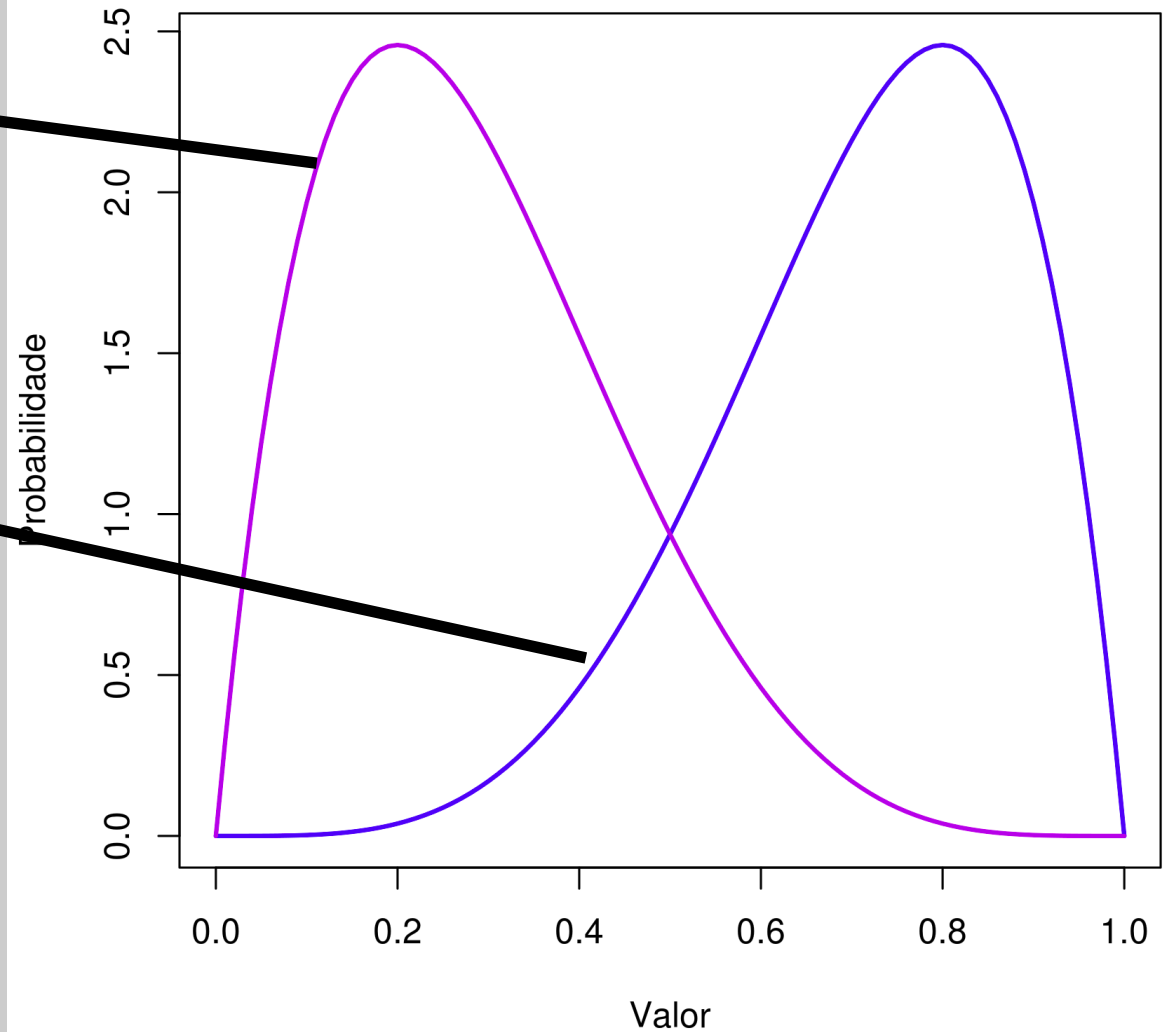
# Forma da distribuição



# Assimetria

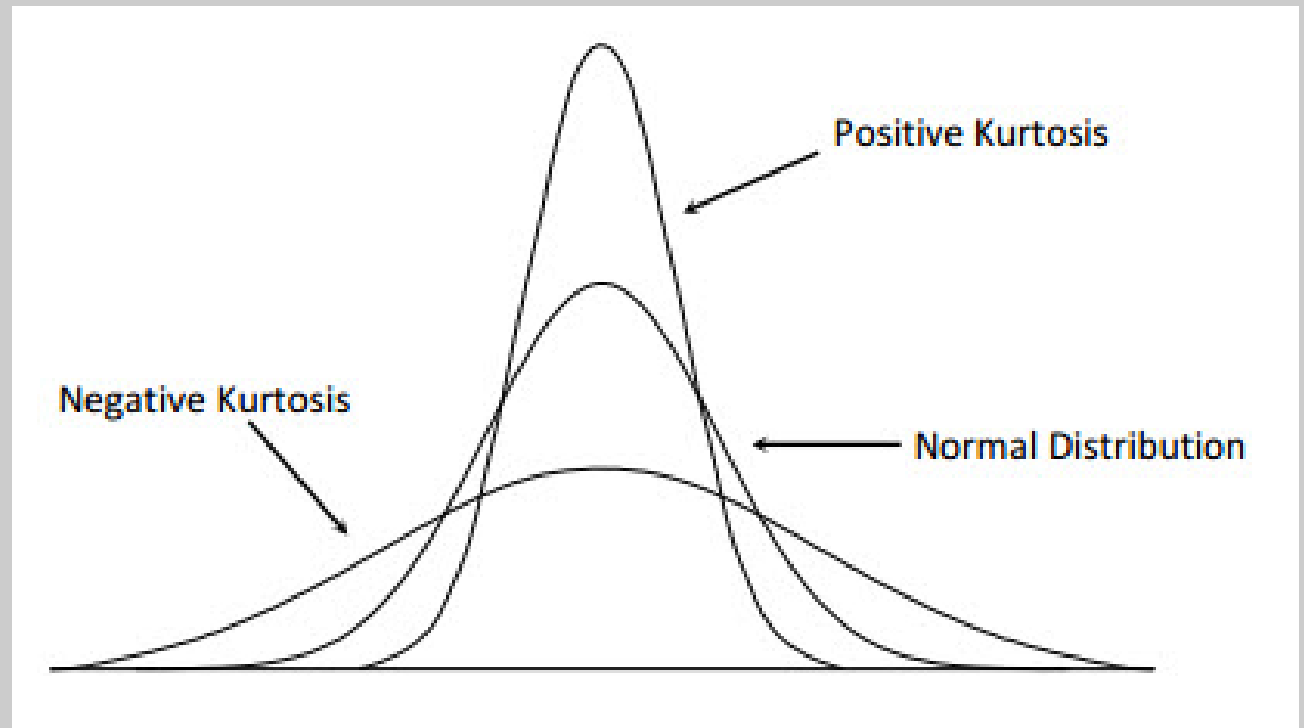
Assimetria positiva  
(cauda para a direita)

Assimetria negativa  
(cauda para a esquerda)



# Curtose

- Quão “íngreme” é o pico da distribuição



# Curtose

- Aplicação na vida real



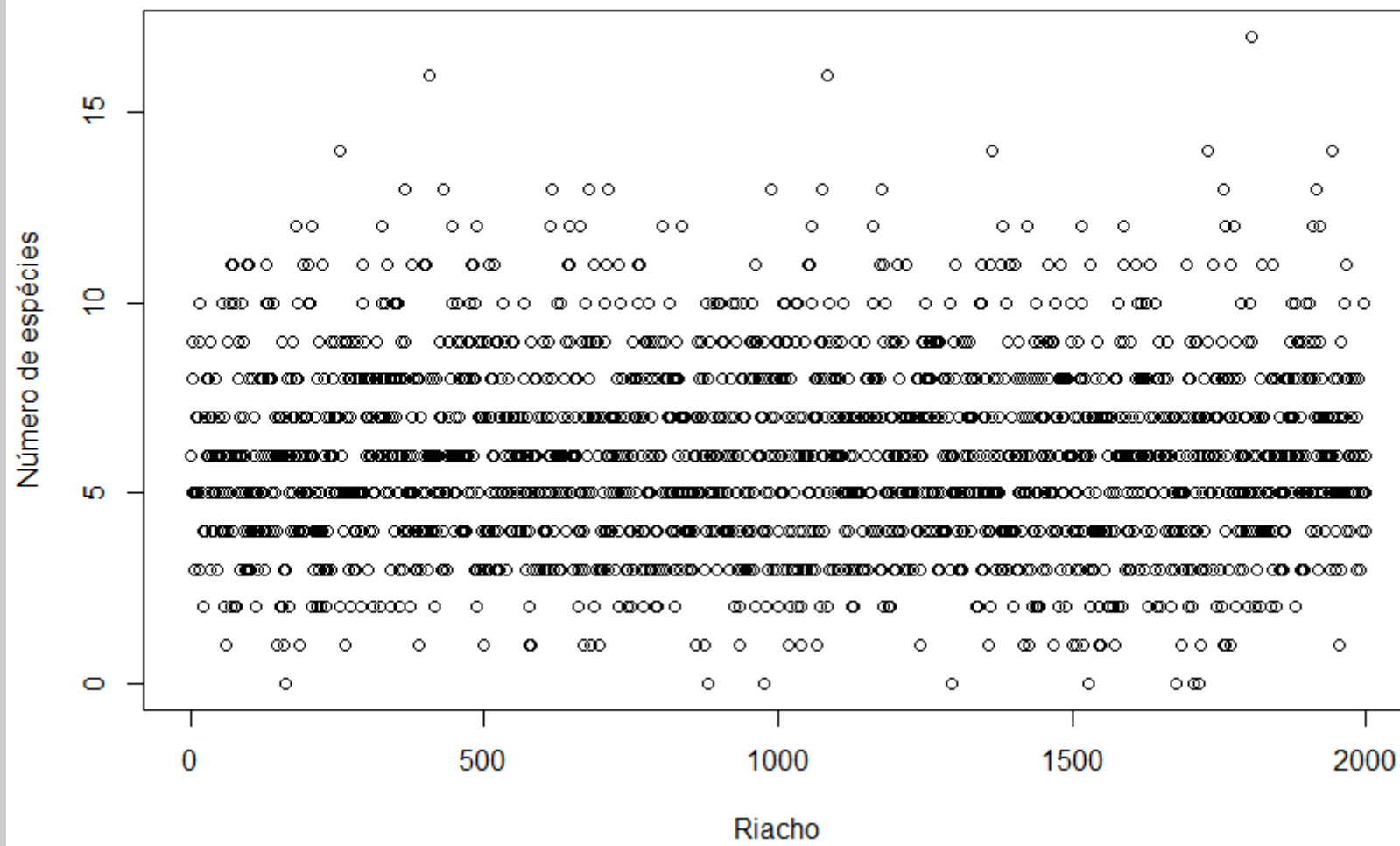
# Outras

- Valores máximo e mínimo

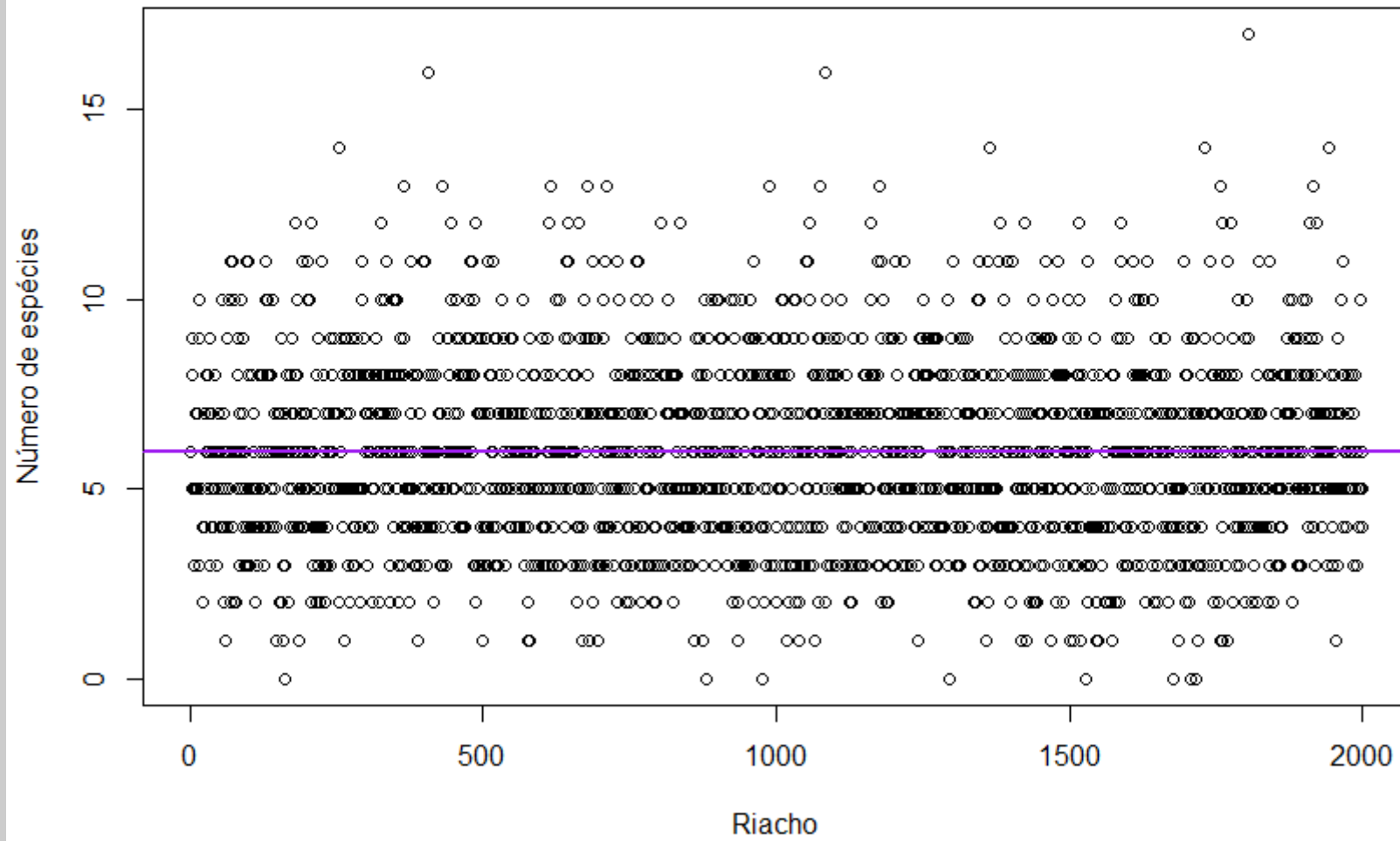
# Erro-padrão e Intervalo de confiança



# População simulada



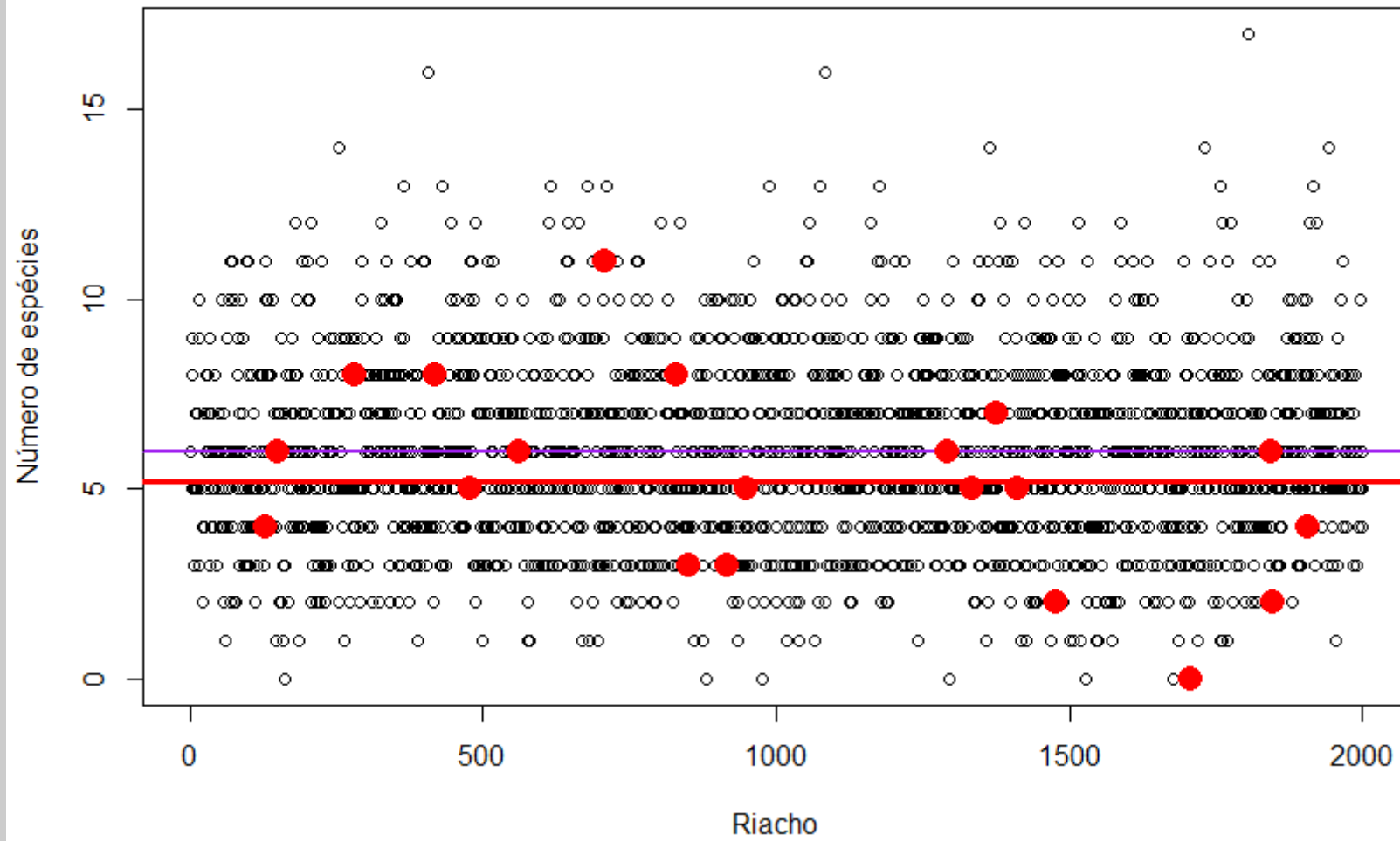
# População simulada



População simulada

Média = 5.97

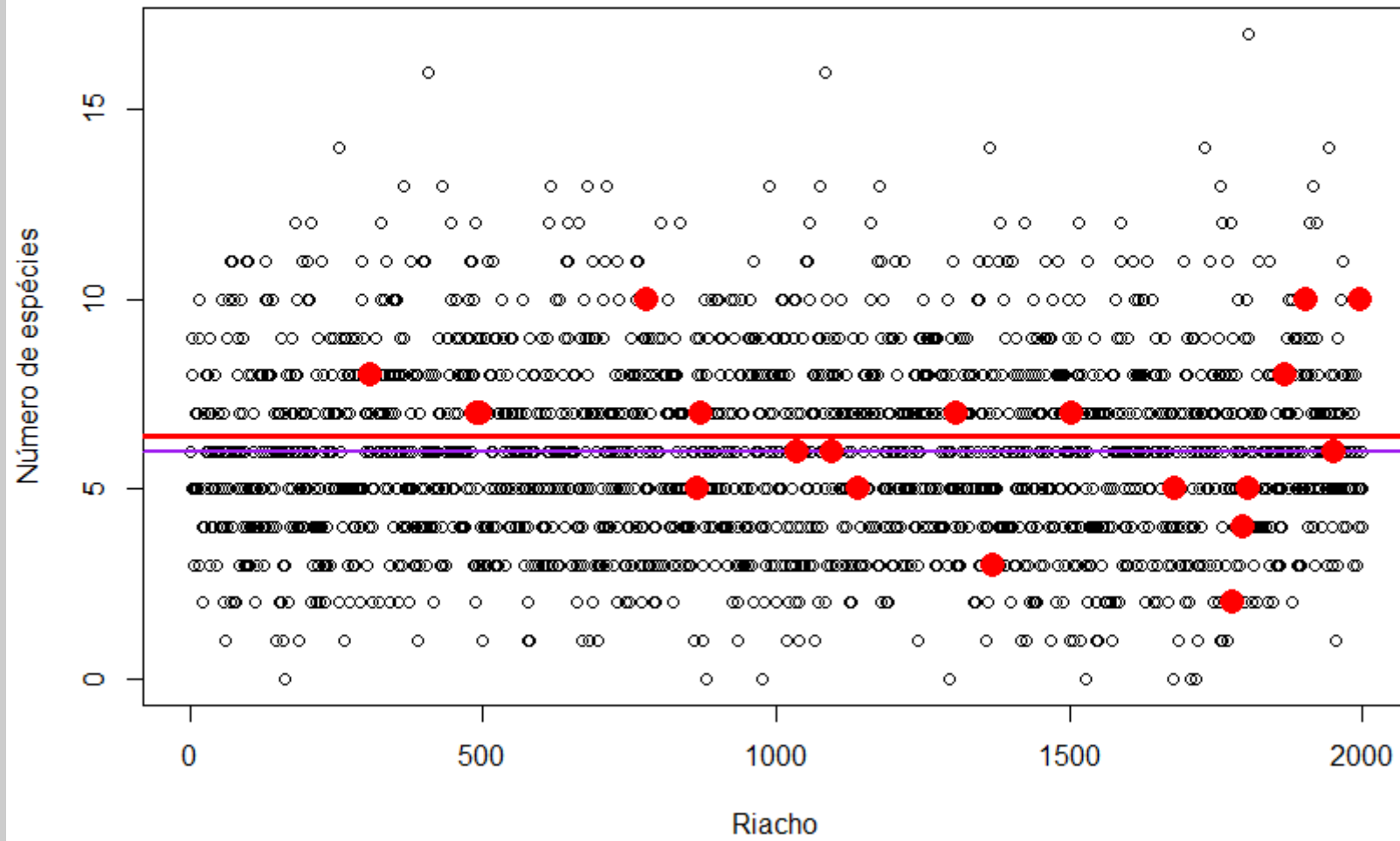
# Amostragem



Amostra 1

Média = 5.2

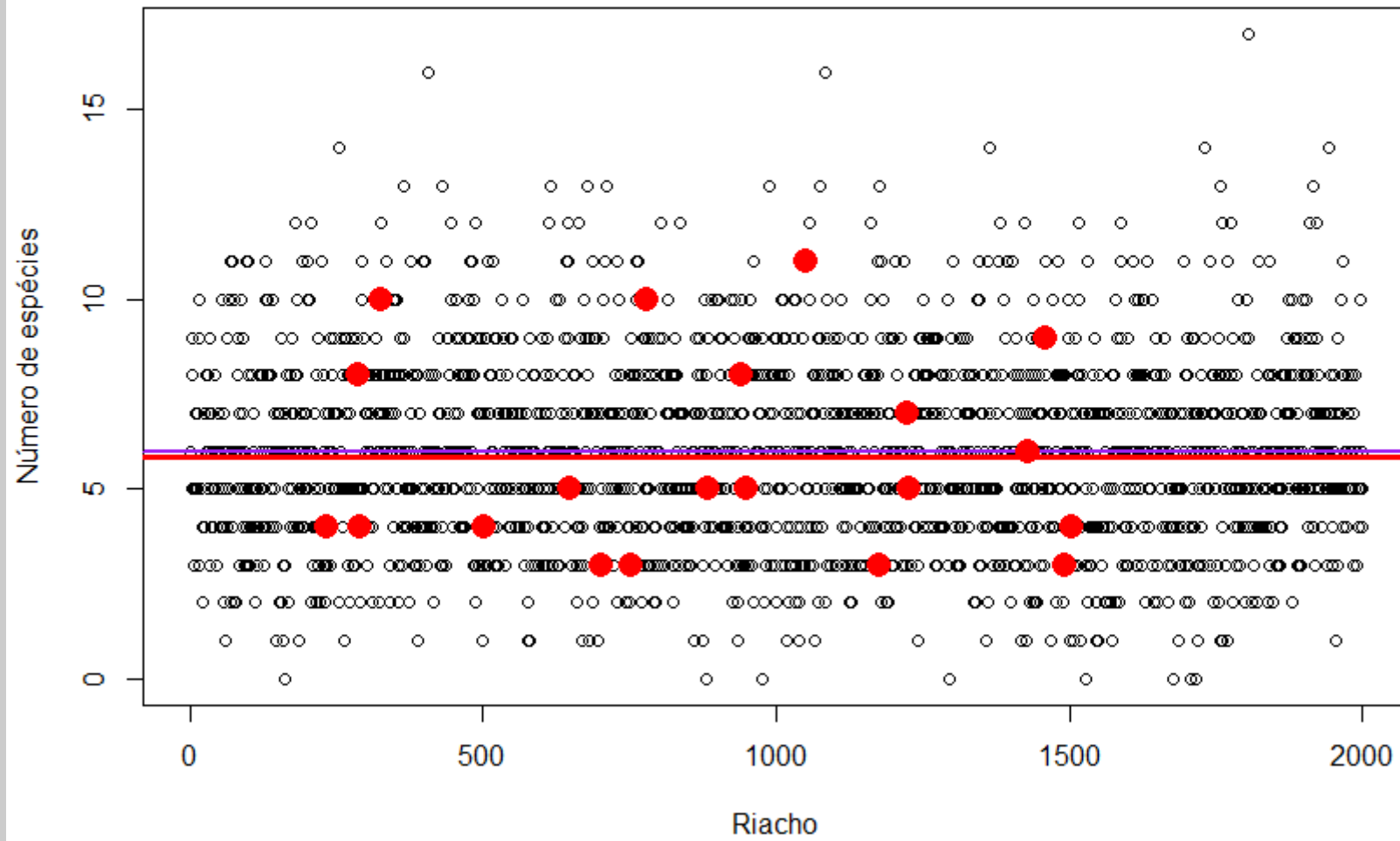
# Amostragem



Amostra 2

Média = 6.4

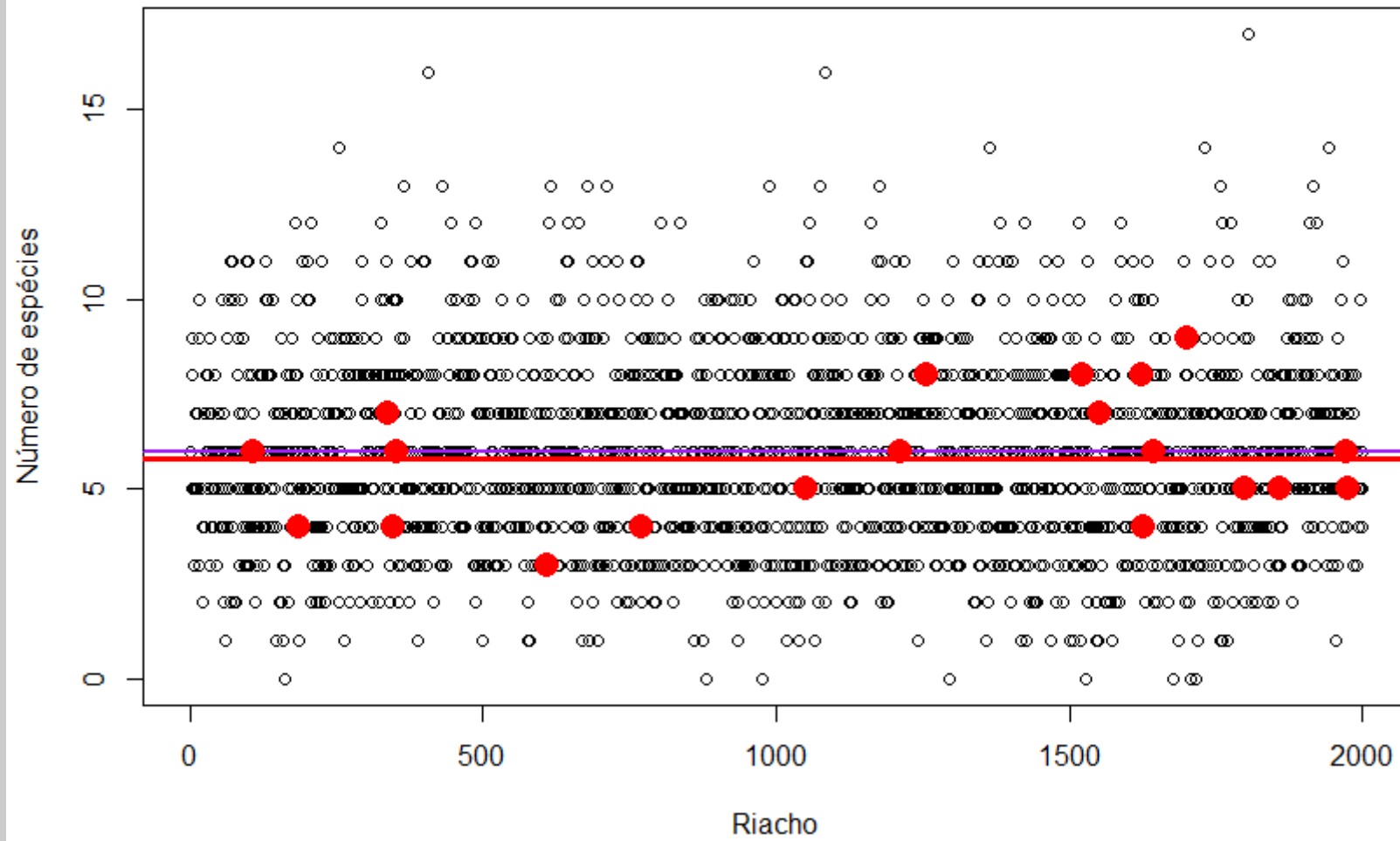
# Amostragem



Amostra 3

Média = 5.85

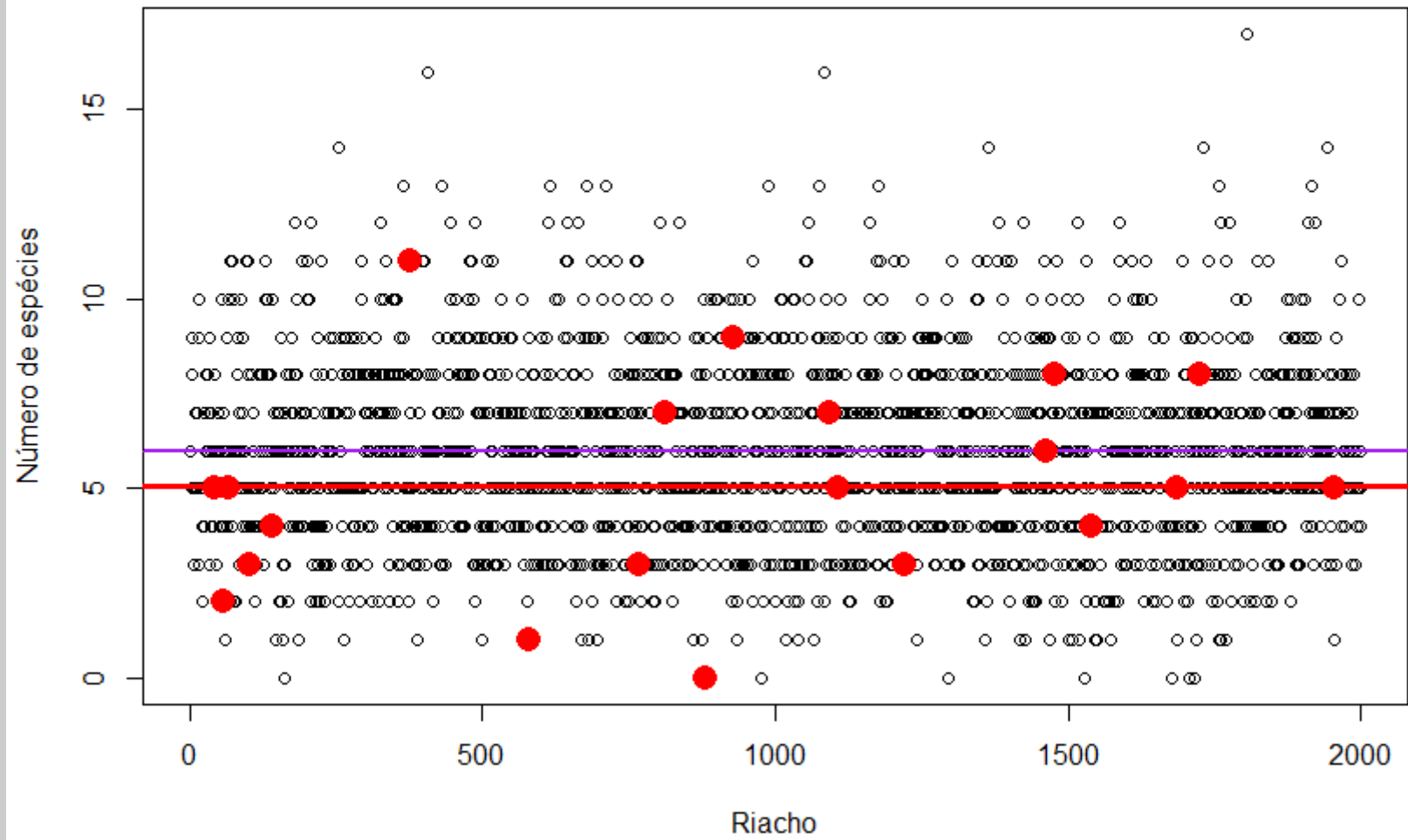
# Amostragem



Amostra 4

Média = 5.8

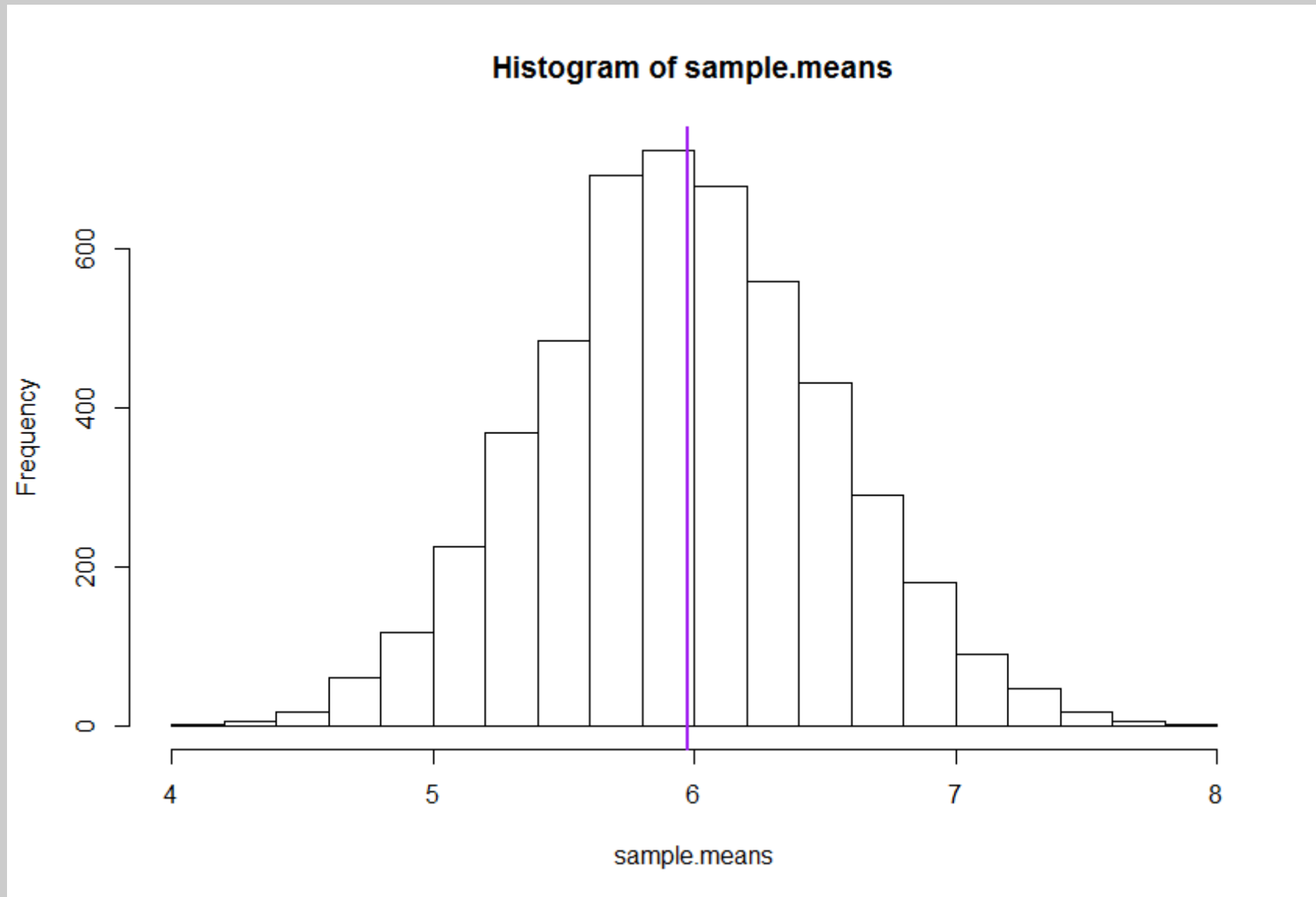
# Amostragem



Amostra 5

Média = 5.05

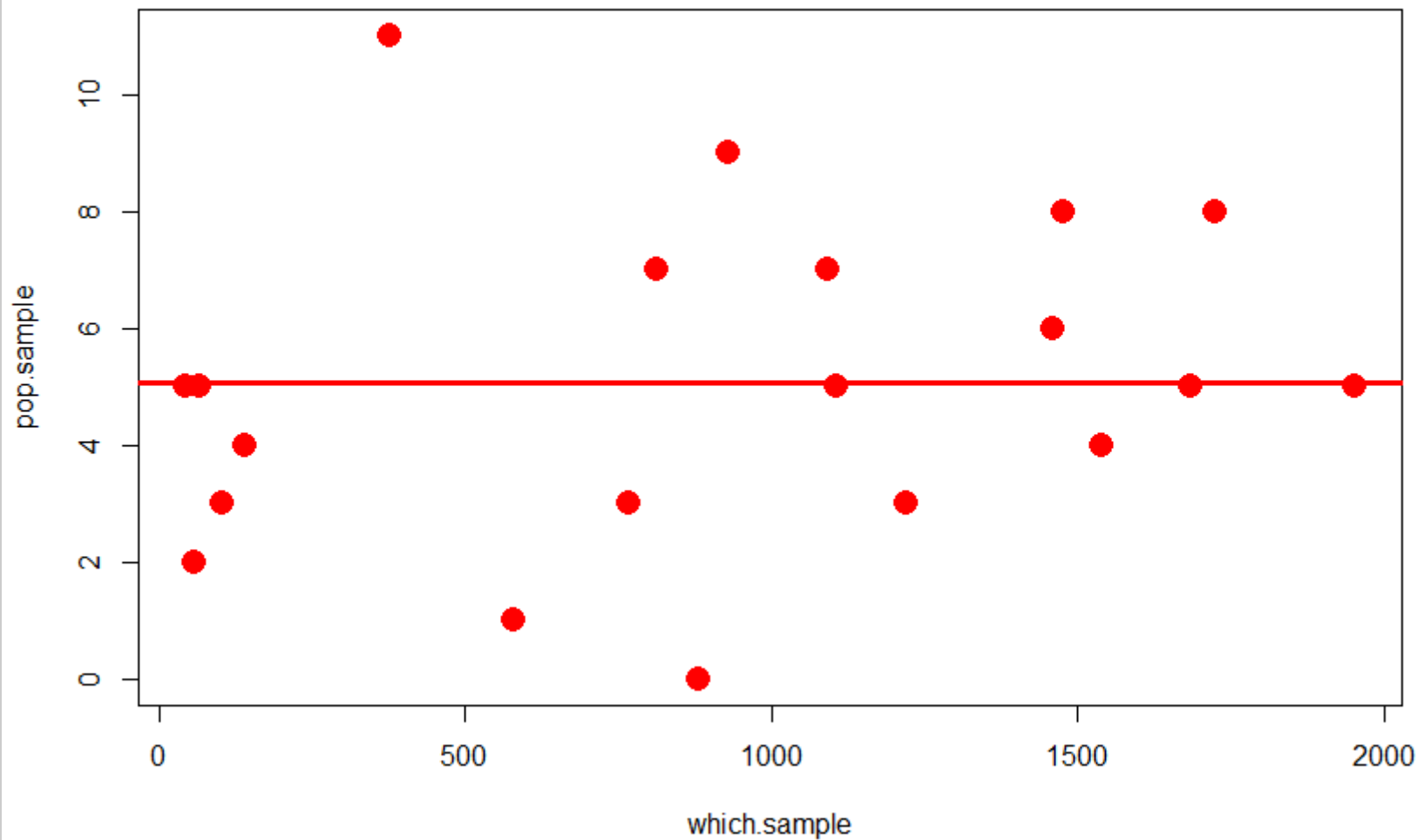
# Amostragem





# Amostra 5

Média = 5.05



Onde estará a média da população?

# Amostra 5

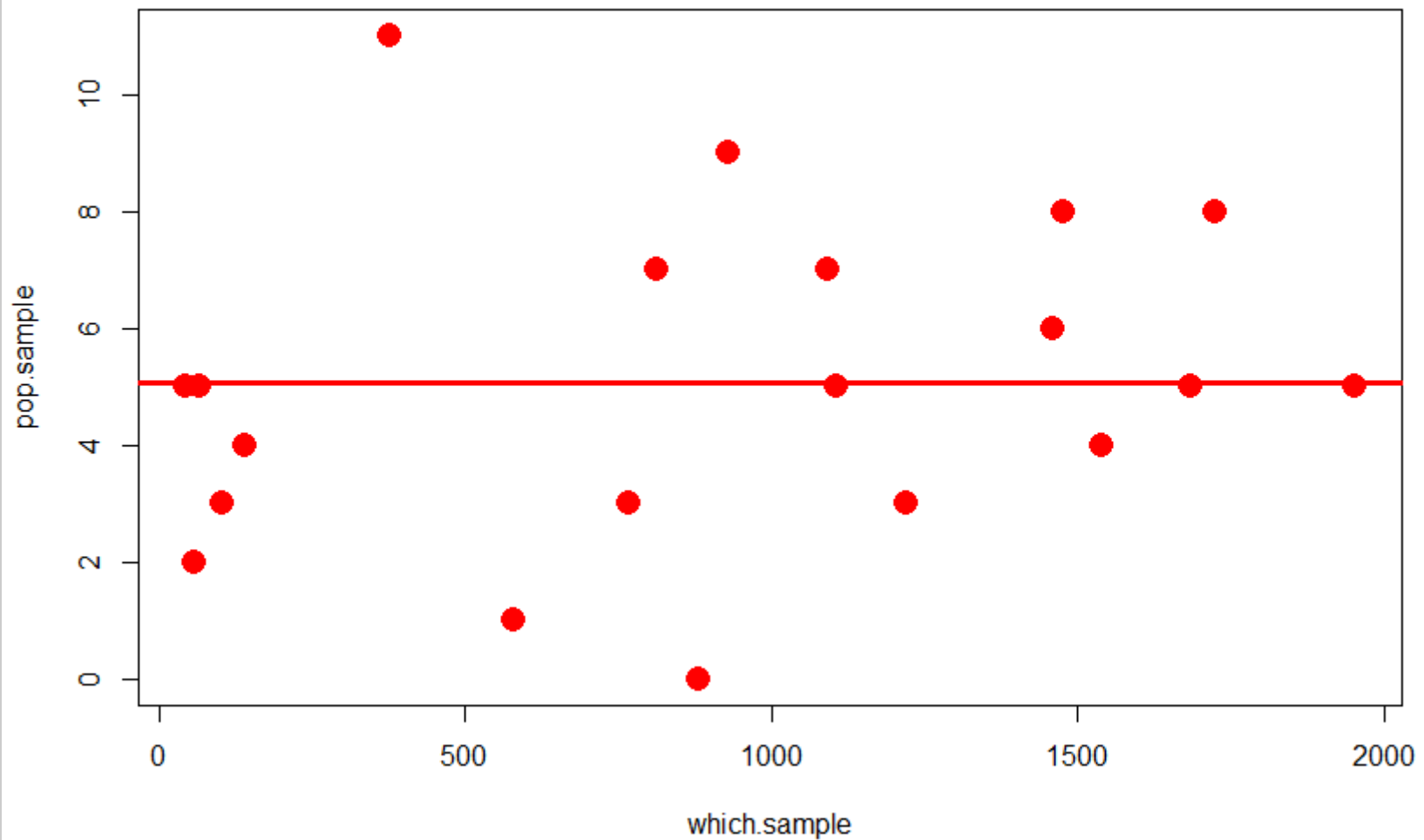
Média = 5.05



Onde estará a média da população?

# Amostra 5

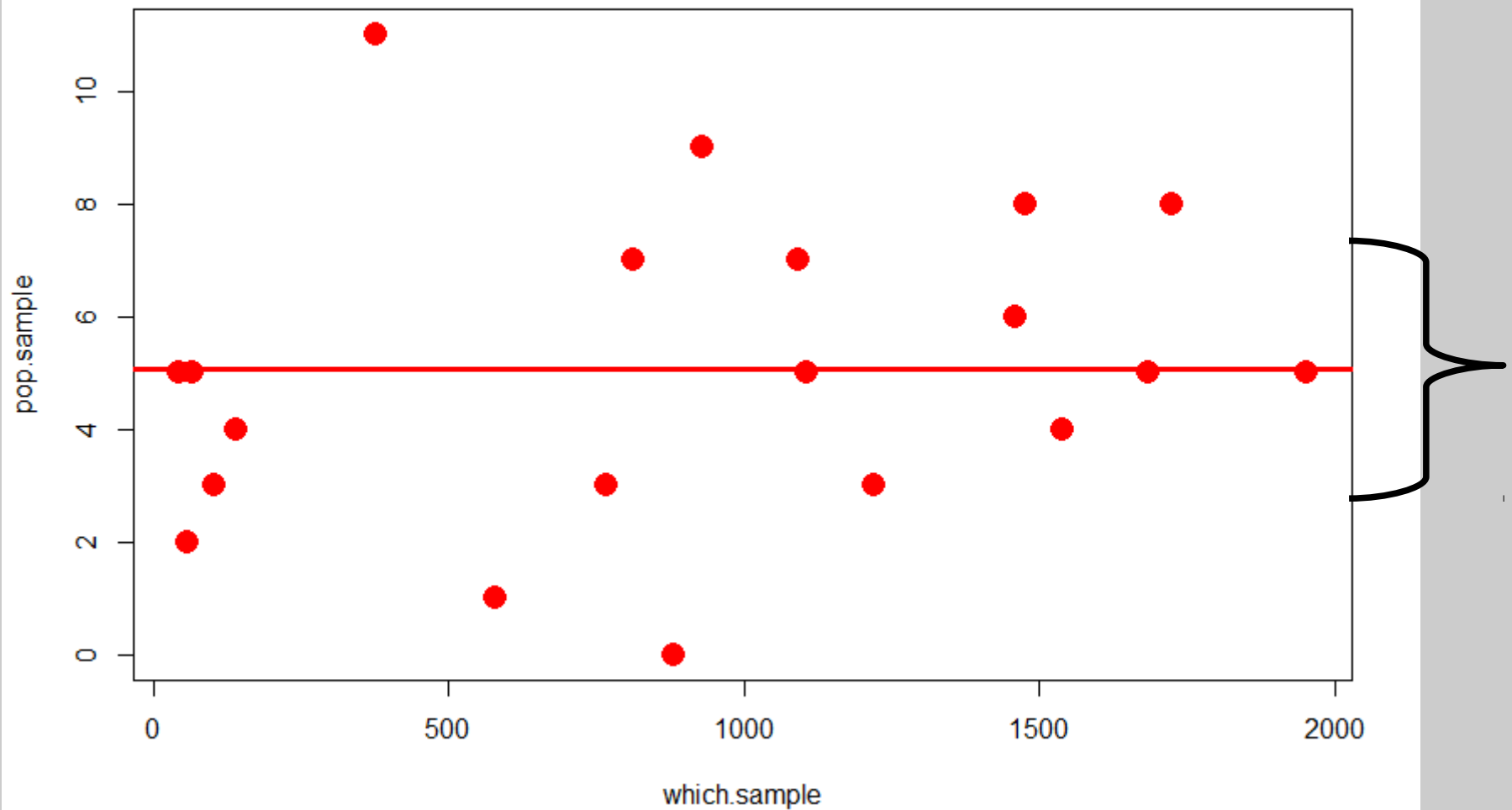
Média = 5.05



Onde poderia estar a média da população?

# Amostra 5

Média = 5.05



Onde *poderia* estar a média da população?

# Erro amostral

Resulta de termos  
**amostras da população, e**  
**não a população inteira**

# Erro amostral

Resulta de termos  
**amostras da população, e**  
**não a população inteira**

Diminui com o  
esforço amostral

# Erro amostral

Resulta de termos  
**amostras da população, e**  
**não a população inteira**

Diminui com o  
esforço amostral

$$\textit{Erro padrão} = \frac{\sigma}{\sqrt{n}}$$

# Intervalo de confiança

A **probabilidade** de que a  
**média real** está **dentro**  
deste **intervalo** = 95%  
(Gotelli & Ellison 2004)



# Intervalo de confiança

~~A probabilidade de que a  
média real está dentro  
deste intervalo = 95%  
(Gotelli & Ellison 2004)~~

Ou está ou não está; o  
experimento já foi feito.  
(Vários autores...)

# Intervalo de confiança

Se o procedimento fosse **repetido em múltiplas amostras**, o **IC** calculado **englobaria a média** da população **95%** das vezes  
(Cox & Hinkley 1974)

# Intervalo de confiança

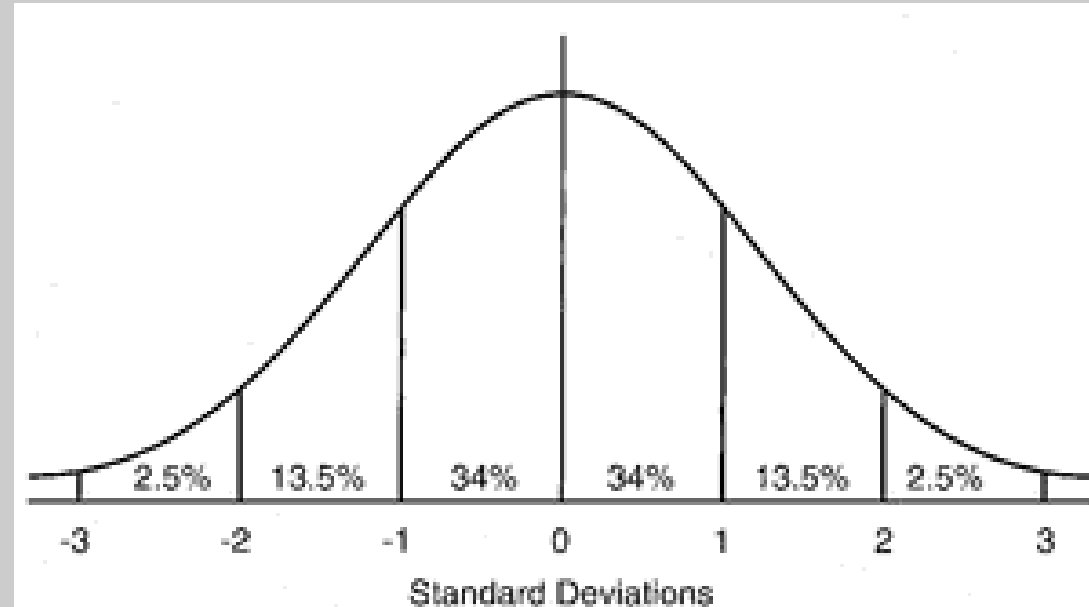
Calculado a partir da  
**variação nas médias**

**Erro-padrão da média:**  
**variação nas médias**  
com **repetições do**  
**experimento**

# Intervalo de confiança

- Pode ser calculado analiticamente

$$CI = \bar{X} \pm Z_{\alpha(2)} \sigma_{\bar{X}}$$
$$Z_{\alpha(2)} = 1.96$$

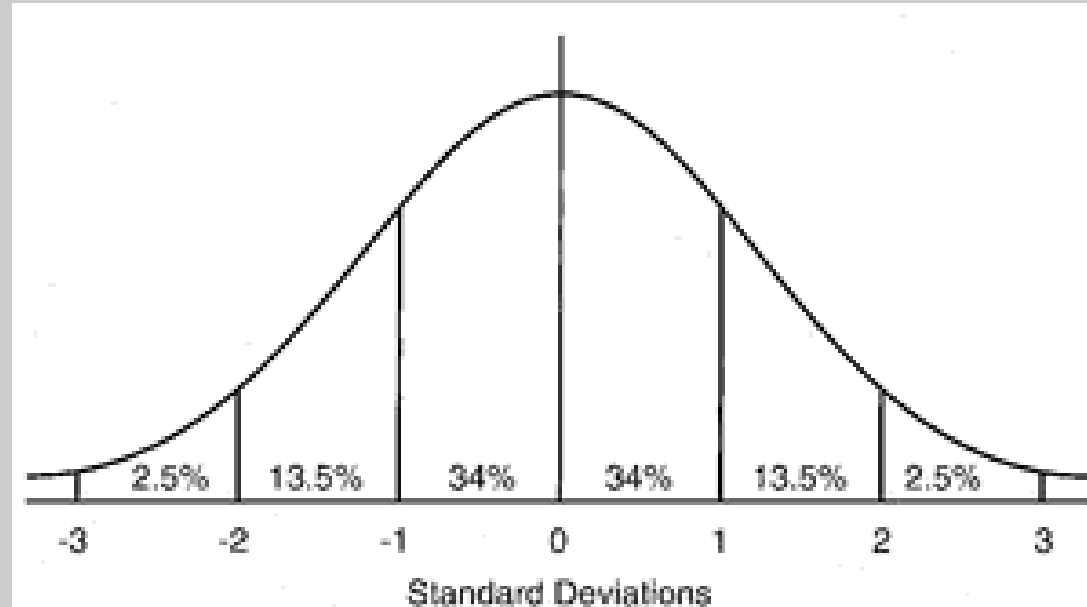


# Intervalo de confiança

- Pode ser calculado analiticamente

$$CI = \bar{X} \pm Z_{\alpha(2)} \sigma_{\bar{X}}$$

$$Z_{\alpha(2)} = 1.96$$



$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

Assumindo normalidade dos dados