

# Estatística descritiva

Conceitos importantes por trás da aplicação

FMVZ - USP

# Conteúdo

- Estatística descritiva:
  - Tipos de variáveis;
  - Medidas de posição e de dispersão;

# Tipos de Variáveis

Quantitativa

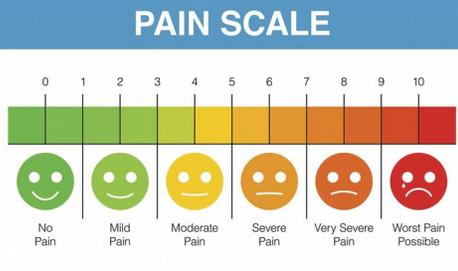
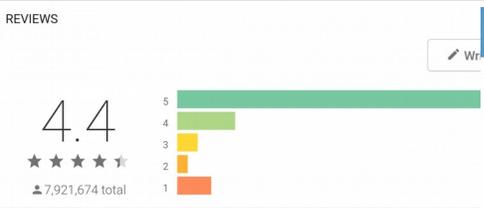
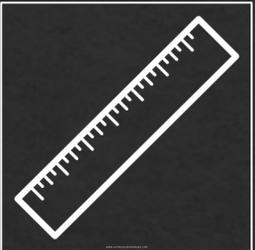
Contínua

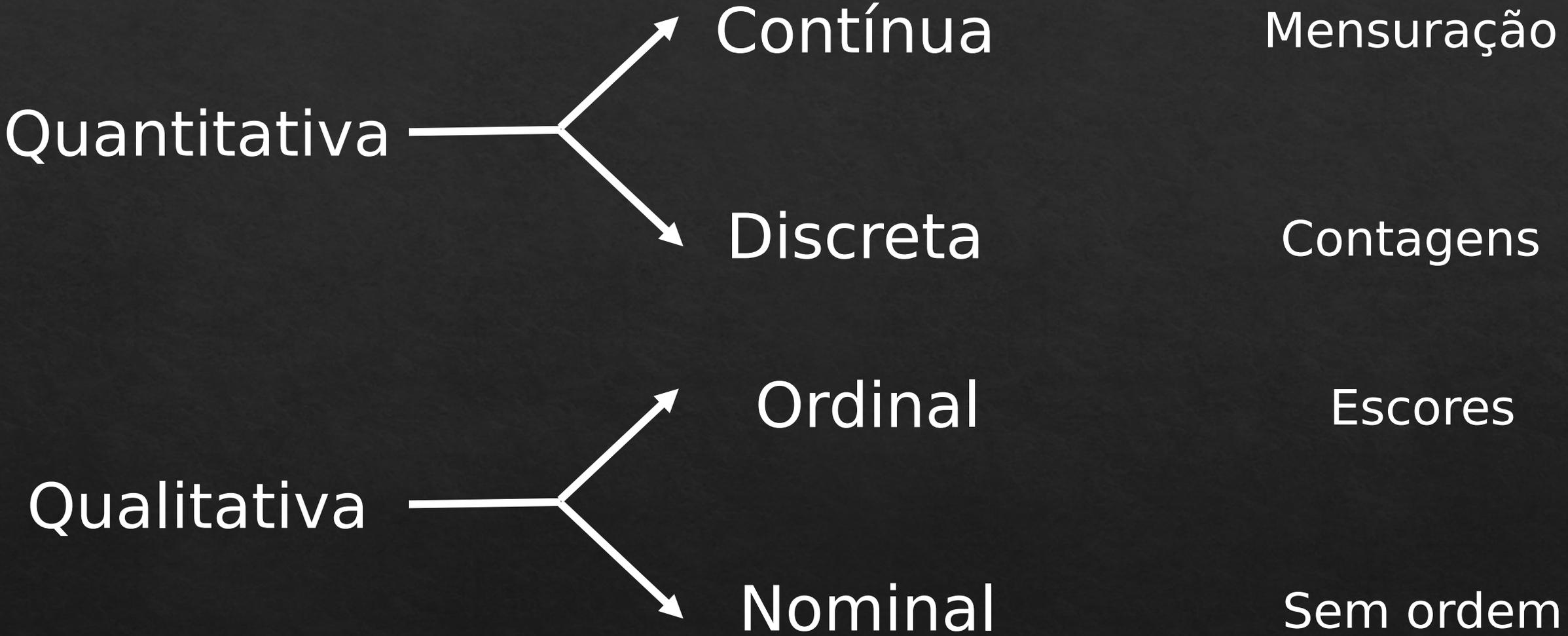
Discreta

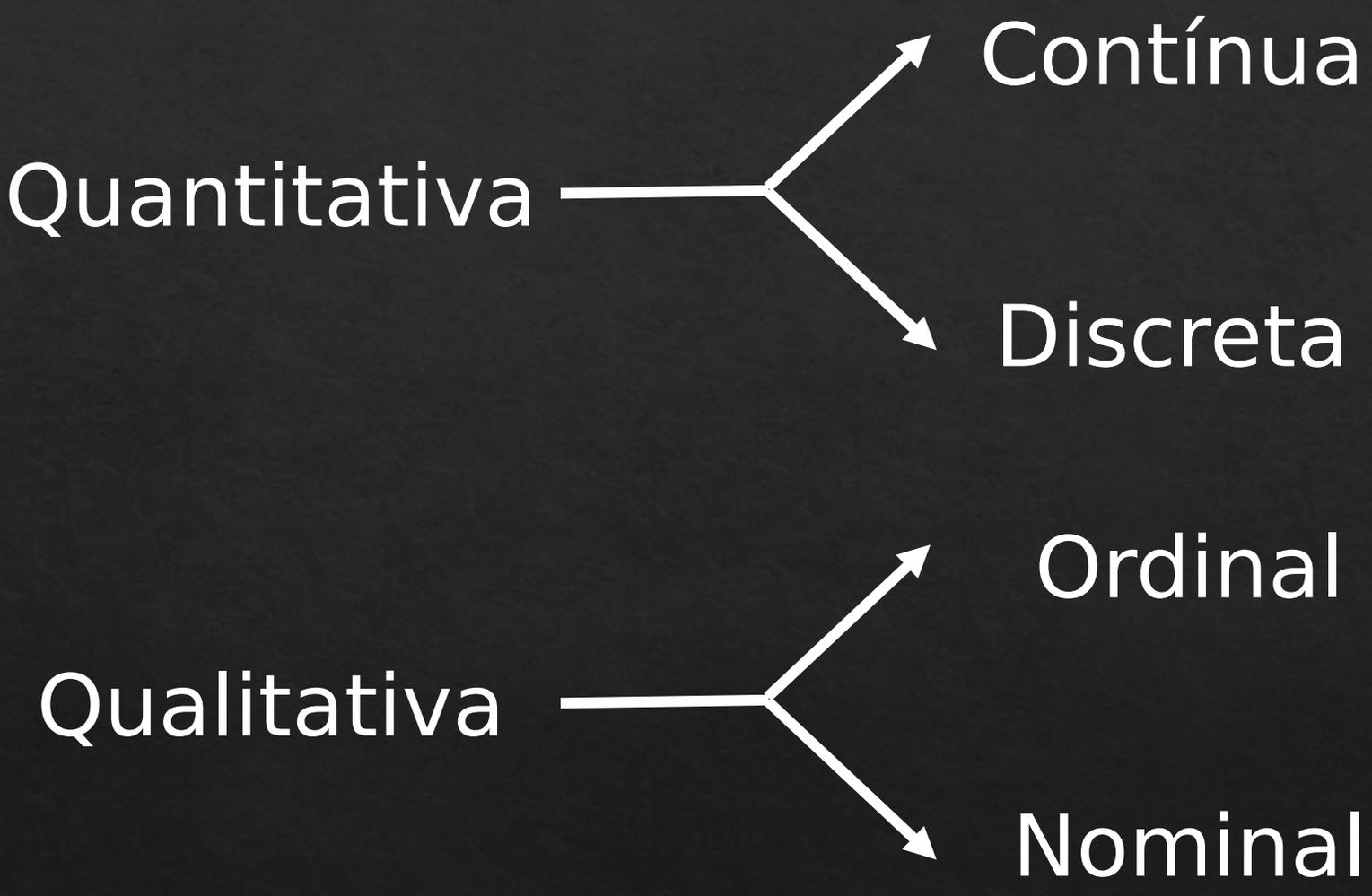
Qualitativa

Ordinal

Nominal



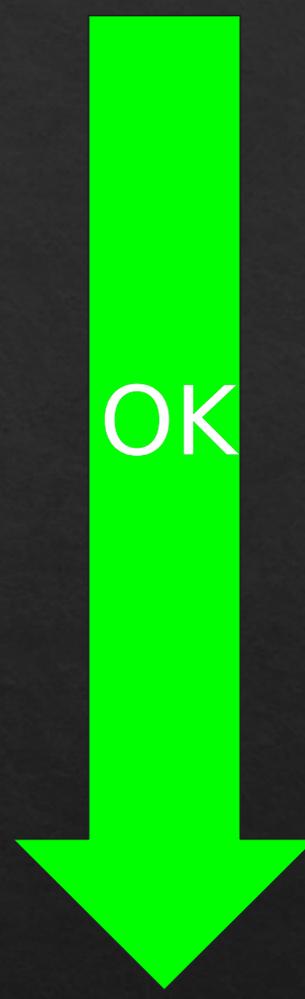
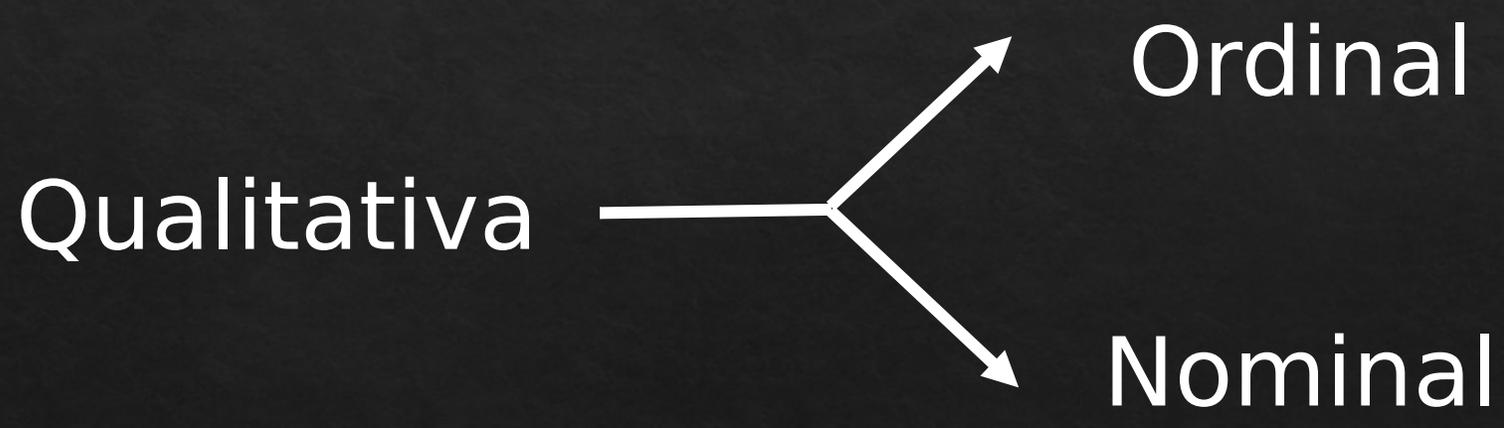
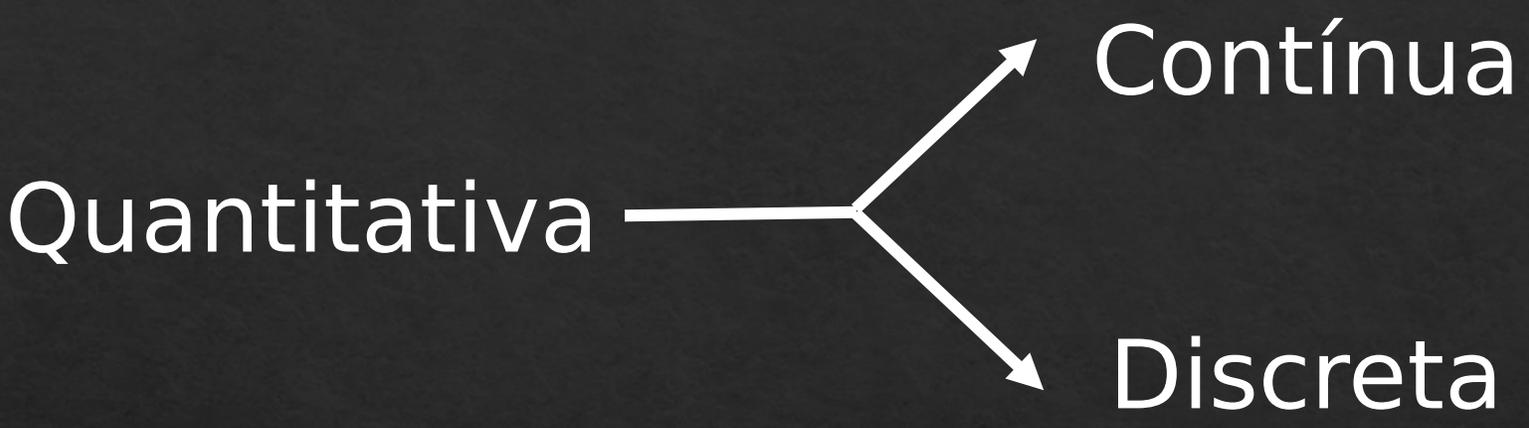


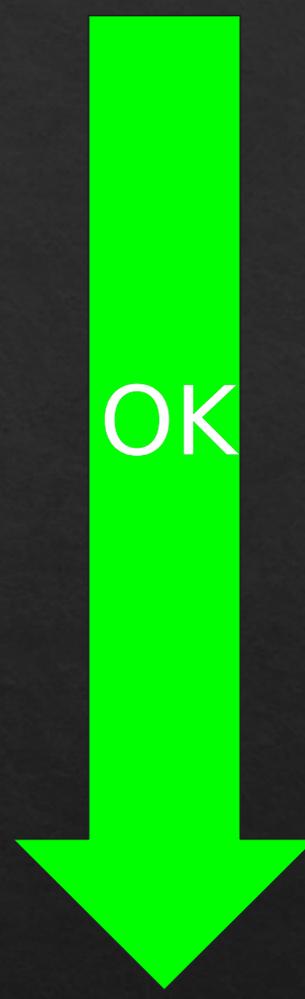
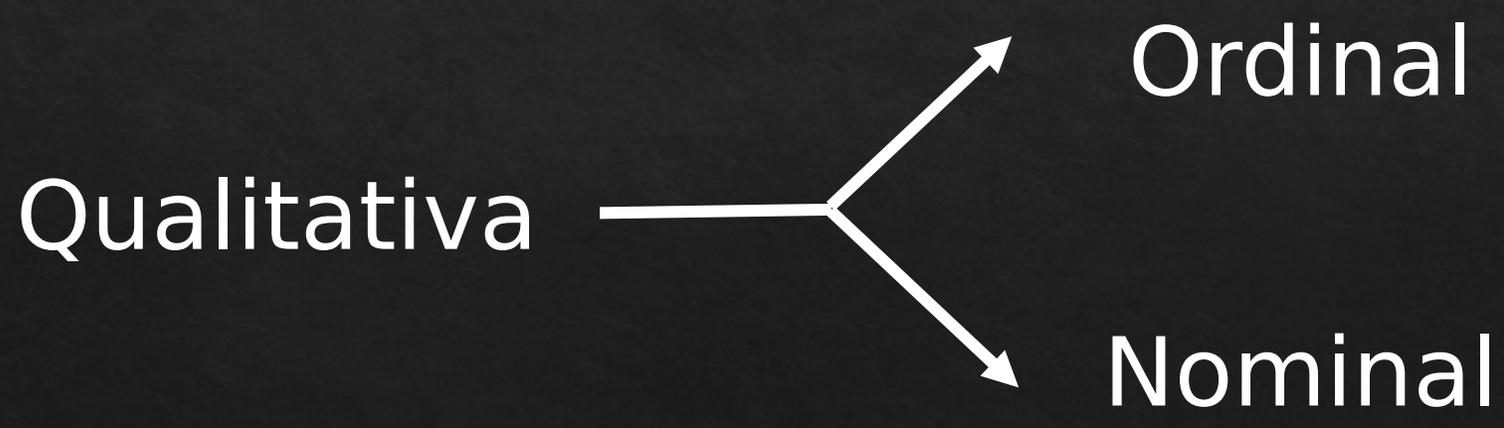
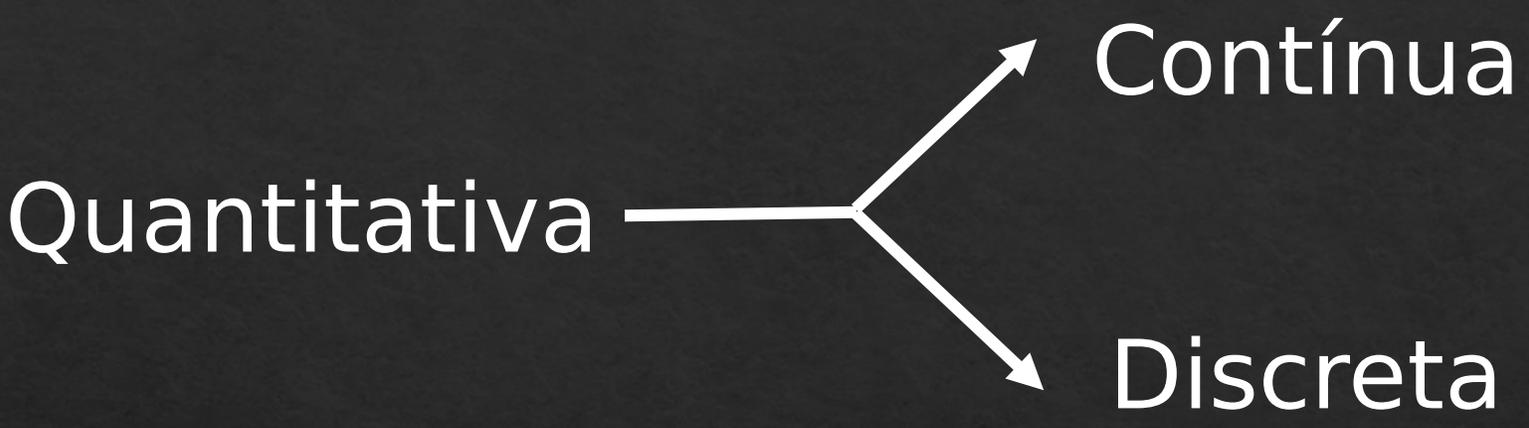


Mais informação



Menos informação





# Medidas Resumo

# Medidas

## Posição

- ◆ Média
- ◆ Mediana
- ◆ Moda

## Dispersão

- ◆ Amplitude
- ◆ Variância
- ◆ Desvio Padrão
- ◆ Intervalo interquartil

# Medidas

## Posição

- ◆ Média
- ◆ Mediana
- ◆ Moda

## Dispersão

- ◆ Amplitude
- ◆ Variância
- ◆ Desvio Padrão
- ◆ Intervalo interquartil

# Média aritmética

- ◆ Mais comum
- ◆ somam-se todos os valores e divide-se pelo número total de observações
- ◆ Desvantagem: é influenciada por “outliers” (dados discrepantes)
- ◆ É apropriada se a distribuição dos dados é simétrica

$$\bar{x} = \frac{\sum x}{n}$$

$$\mu = \frac{\sum x}{N}$$

# Mediana

- ◆ É o valor central em um conjunto de  $N$  observações colocadas em ordem de magnitude crescente (ou decrescente)
- ◆ Mesmo número de observações acima e abaixo da mediana
- ◆ Vantagem da mediana: não é afetada por “outliers” ou se a distribuição de dados é assimétrica

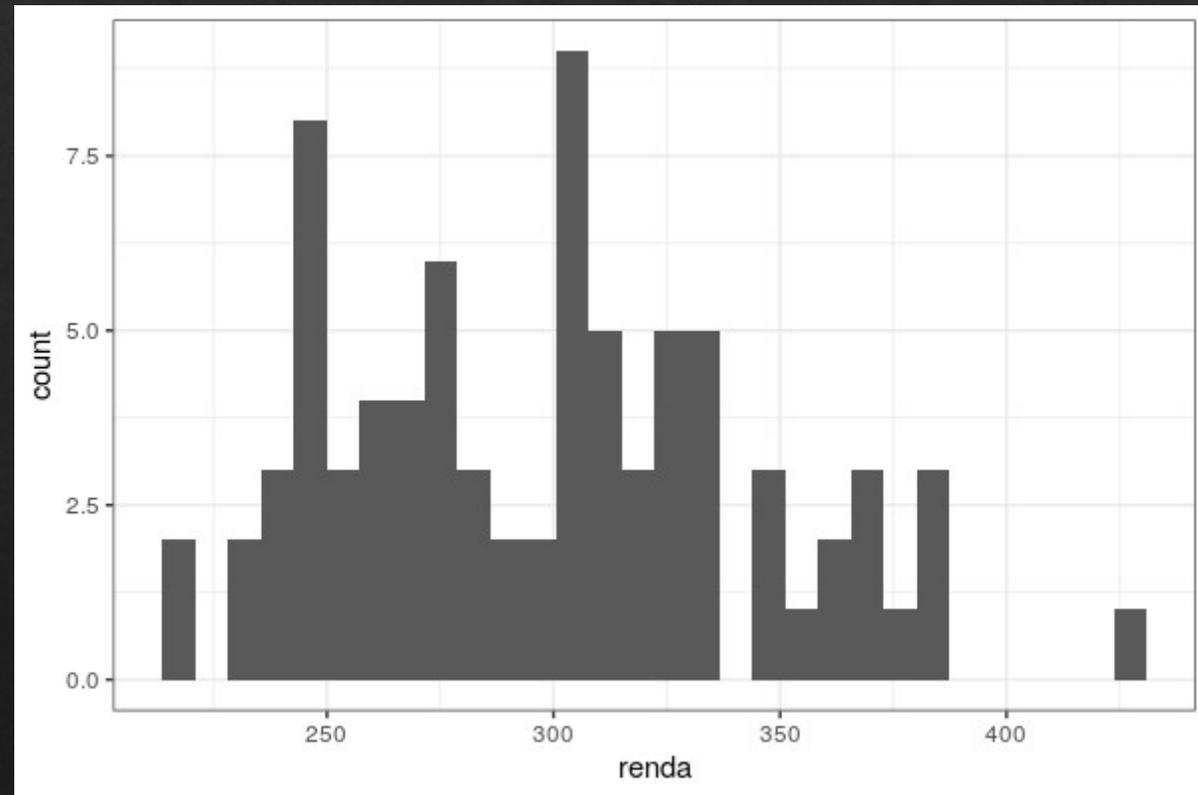
$$\textit{Mediana} = \frac{(n + 1)}{2}$$

# Moda

- ◆ É a medida mais frequente
- ◆ Raramente utilizada

# Média vs Mediana

# Renda de alunos

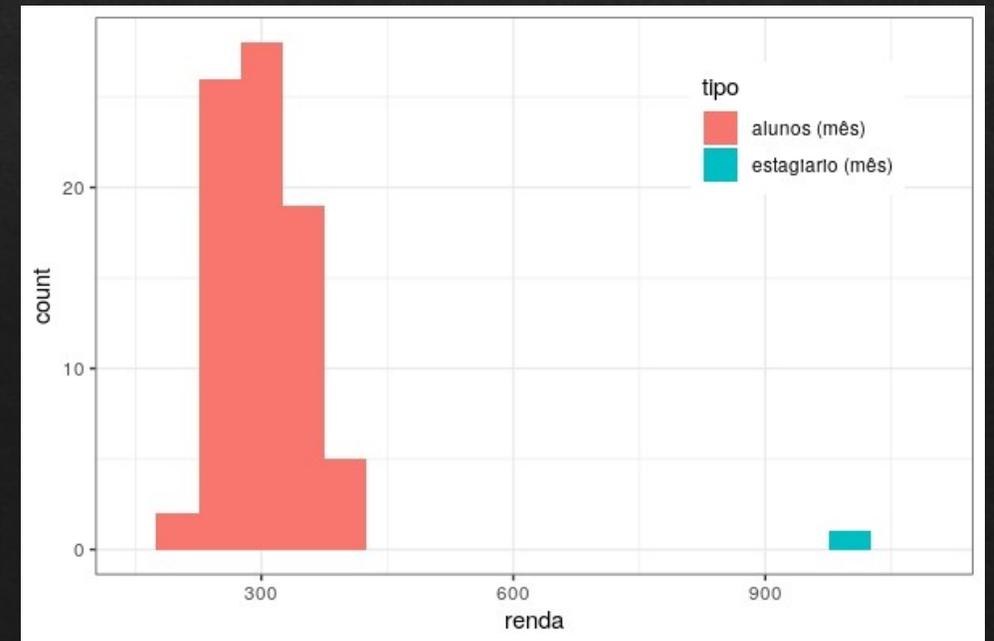
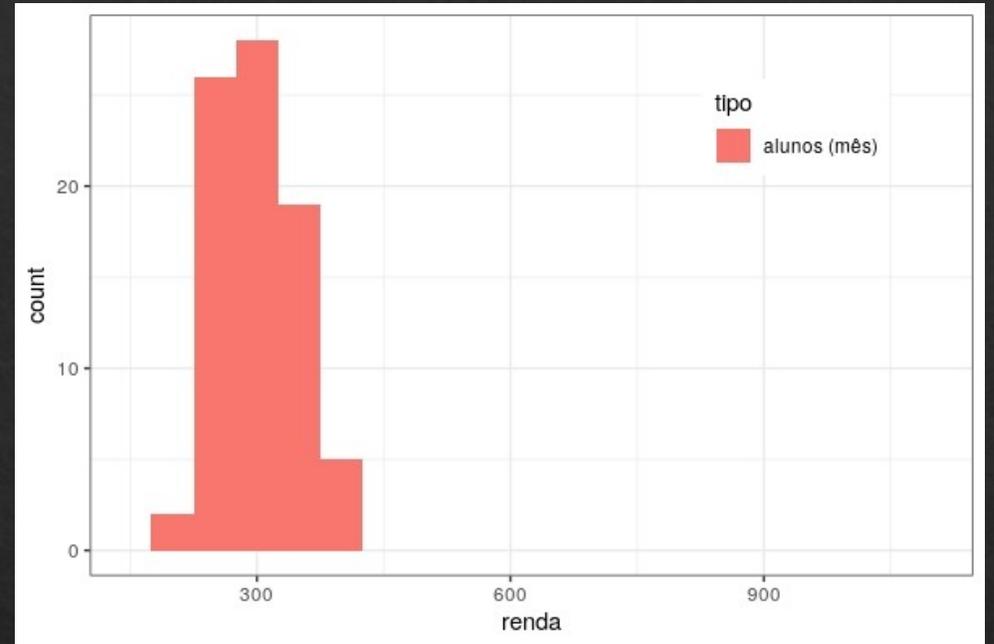


Distribuição mais ou menos simétrica  
Média ~ Mediana

Um aluno consegue estágio remunerado.

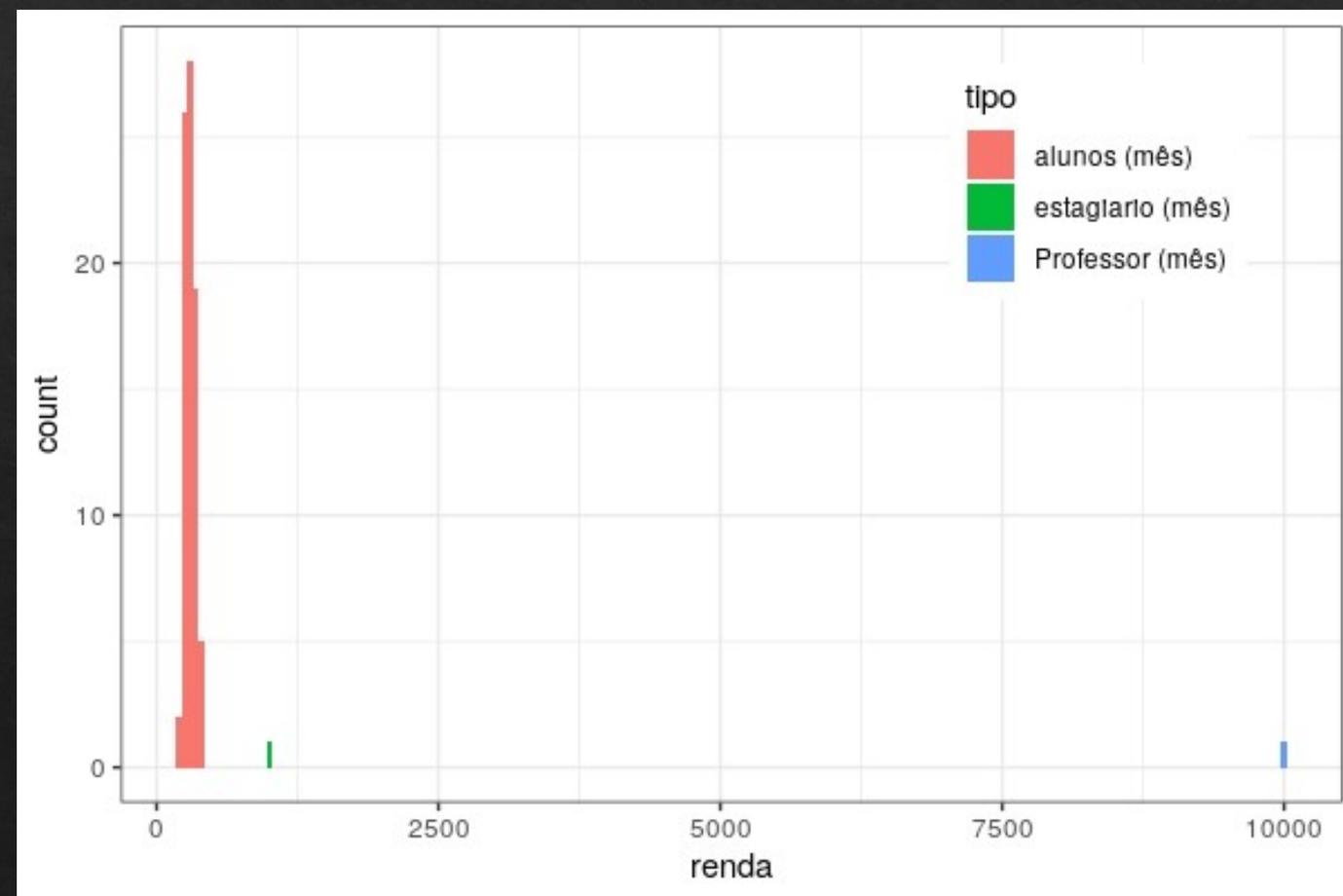
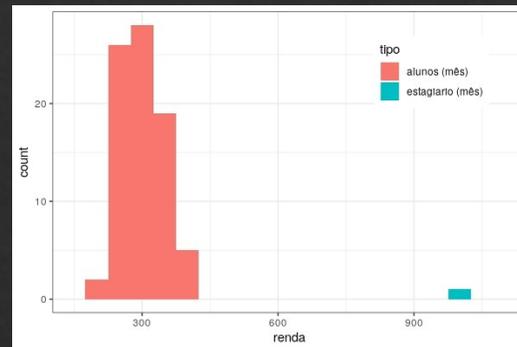
O que acontece com a média?

O que acontece com a mediana?



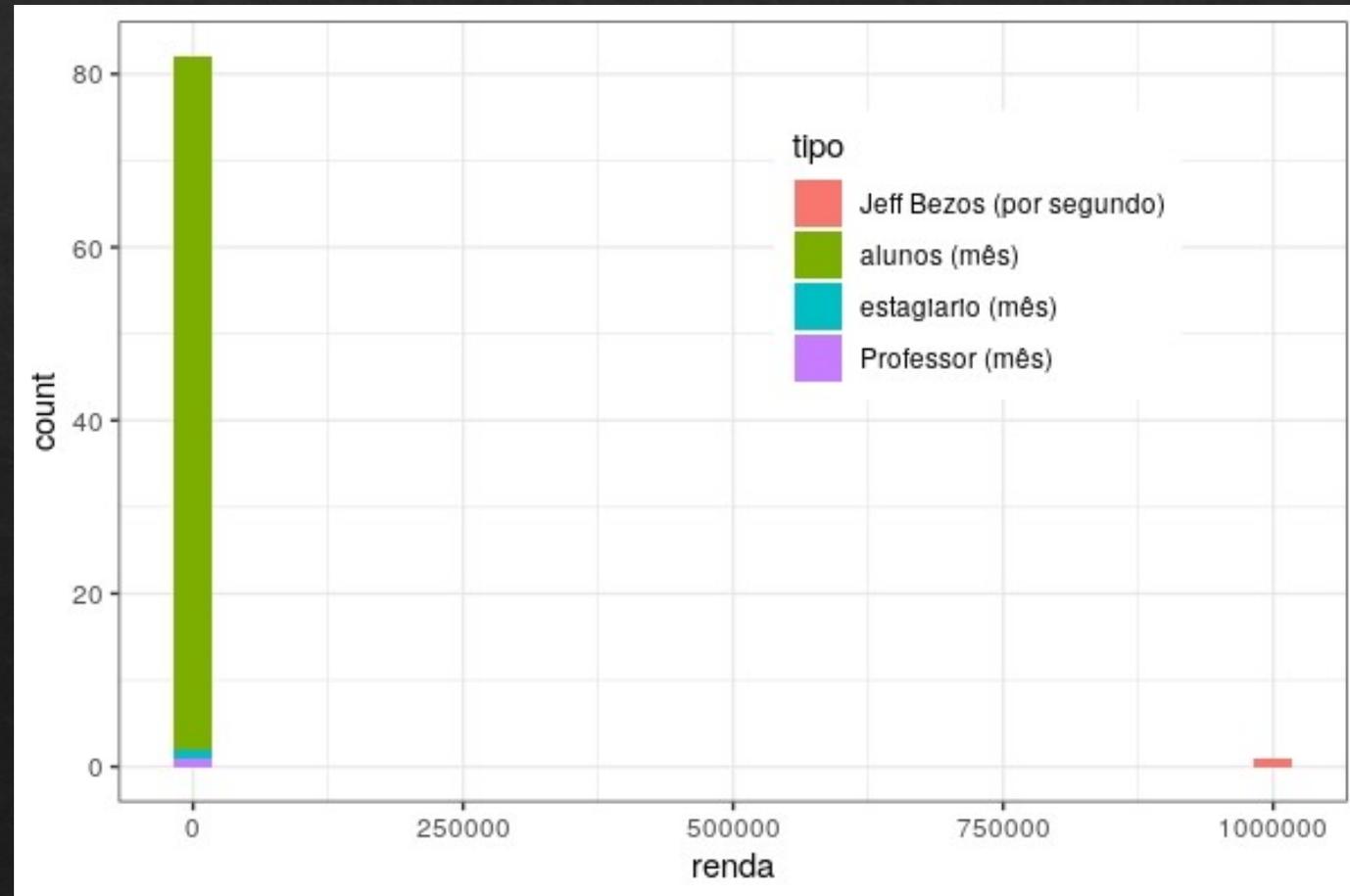
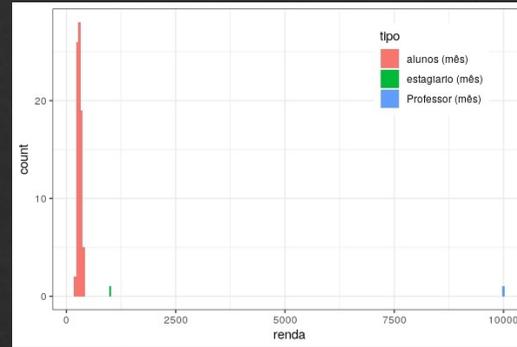
Professor entra na sala

- que acontece com a média?
- que acontece com a mediana?

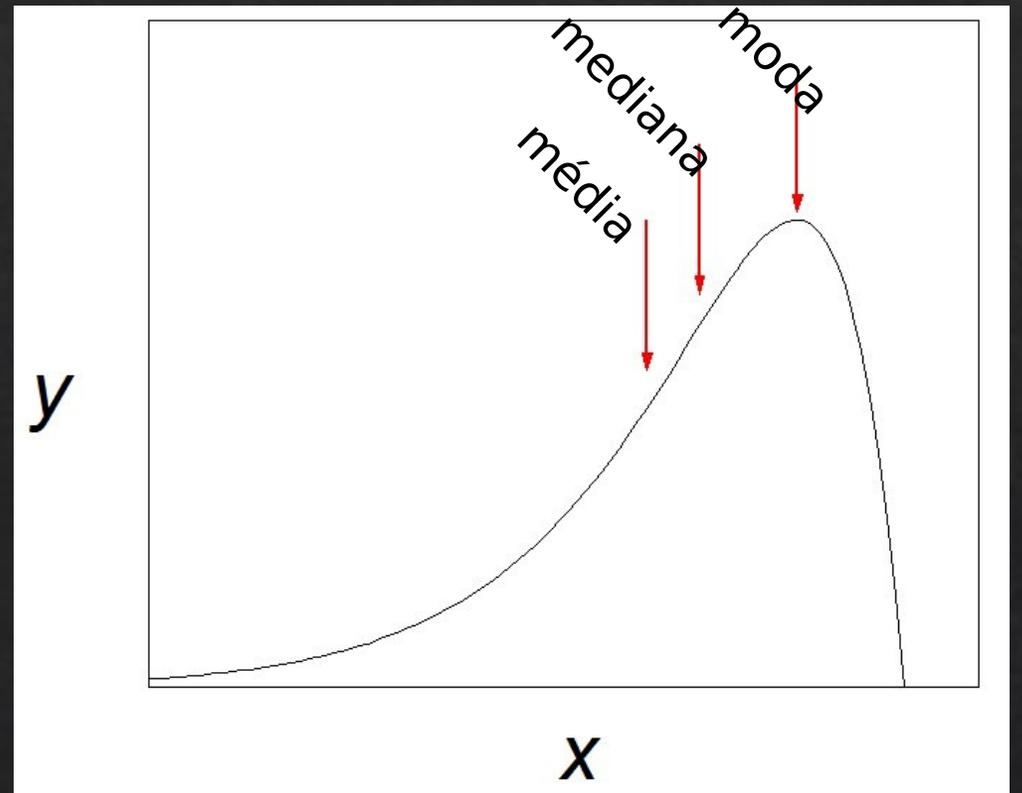
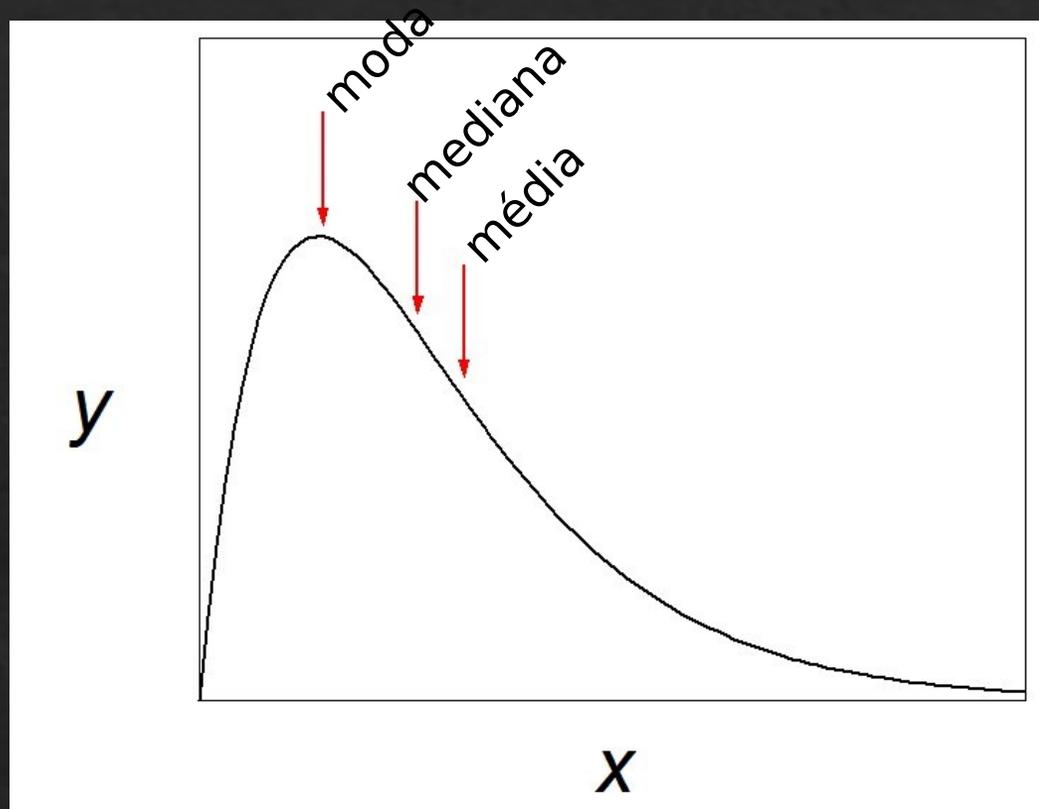


## Jeff Bezos entra na sala

- que acontece com a média?
- que acontece com a mediana?

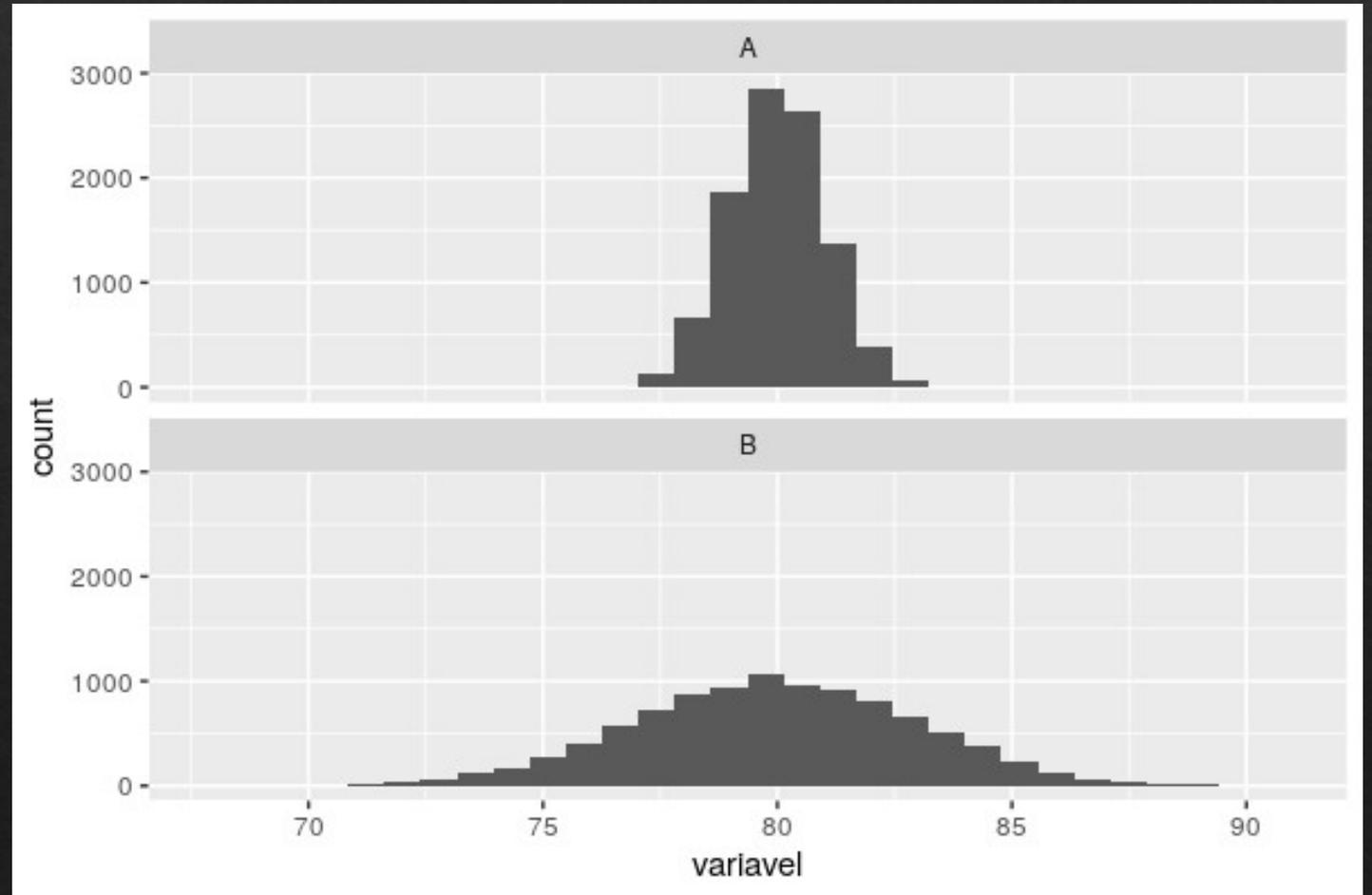


# Distribuições assimétricas



Médias iguais  
Medianas iguais  
Modas iguais

Distribuições são iguais?



# Medidas

## Posição

- ◆ Média
- ◆ Mediana
- ◆ Moda

## Dispersão

- ◆ Amplitude
- ◆ Variância
- ◆ Desvio Padrão
- ◆ Intervalo interquartil

# Amplitude

- ◆ Diferença entre o maior e o menor dado observado
- ◆ Simples de calcular
- ◆ Não é uma boa medida de dispersão, porque seu cálculo se baseia nos valores extremos da amostra e não em todos os dados

$$\text{Amplitude} = \textit{máximo} - \textit{mínimo}$$

# Variância

- ◆ Variância mede a dispersão dos dados em torno da média
- ◆ Utiliza todas as observações
- ◆ É uma medida sensível de dispersão

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

# Desvio padrão ( $s$ )

- ◆ Raiz quadrada da variância
- ◆ Apresenta a mesma dimensão que os dados originais
- ◆ Muito útil para distribuições simétricas (p.ex. gaussiana)

$$SD = \sqrt{S^2}$$

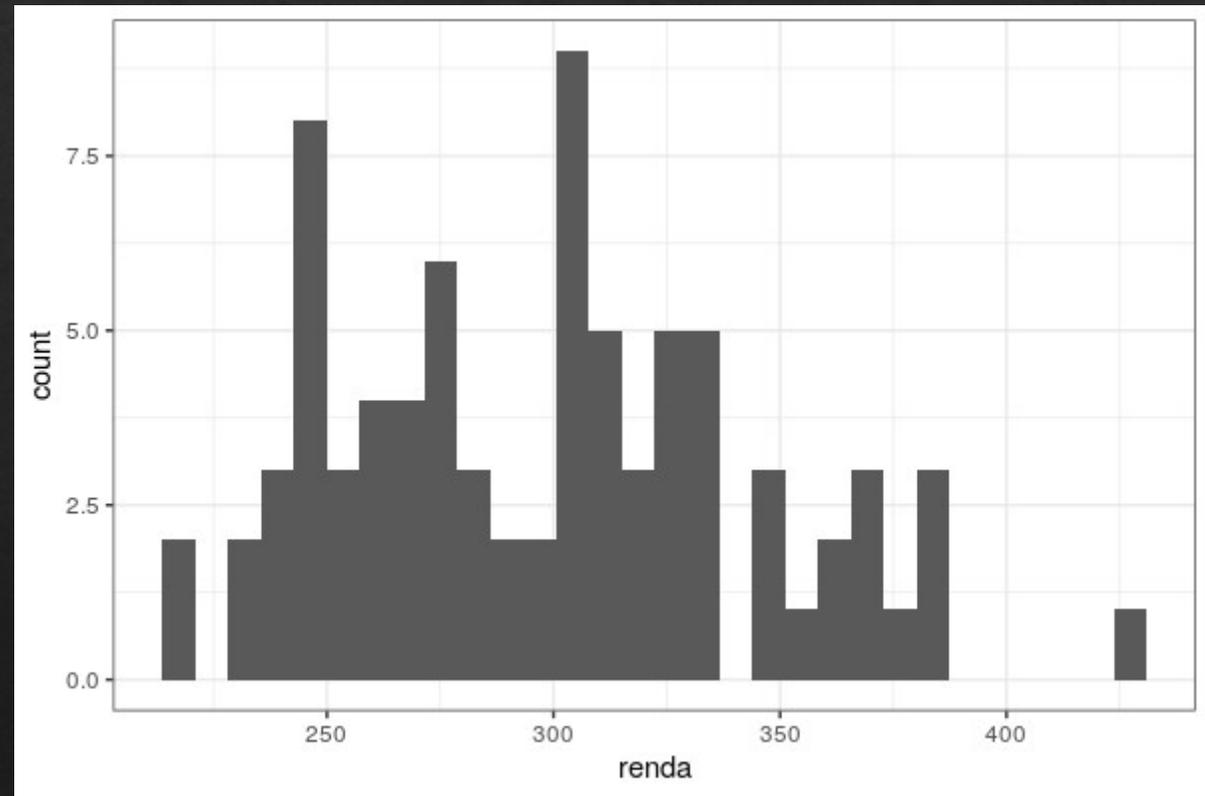
# Intervalo interquartil

- ◆ Ou distância interquartil, ou amplitude interquartil
- ◆ É a amplitude de valores que abrange os dados centrais (50%) das observações (diferença entre o terceiro e o primeiro quartil)
- ◆ não é influenciada pela presença de outliers
- ◆ ignora a maioria das observações

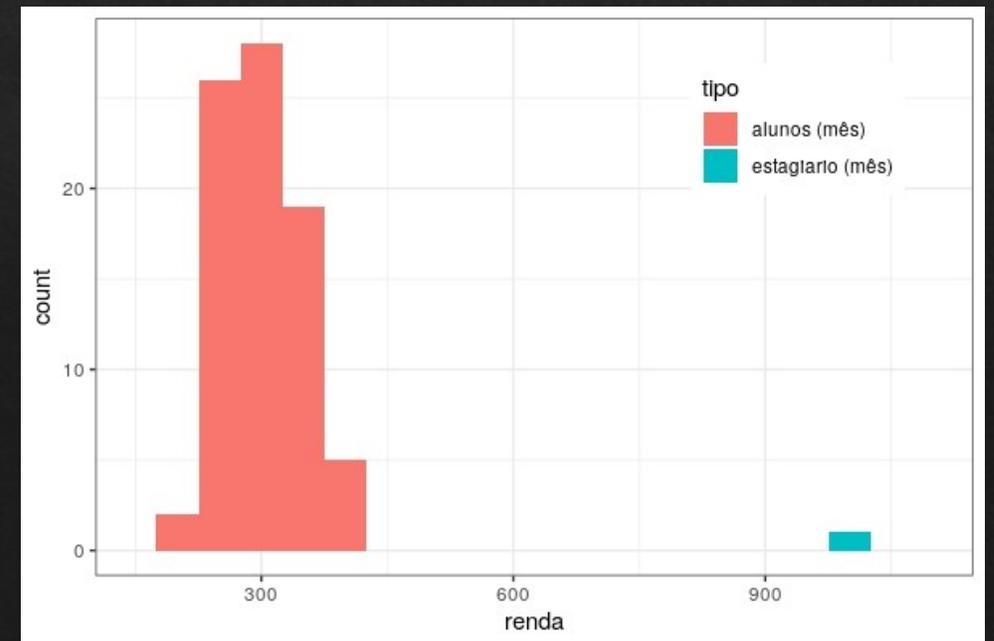
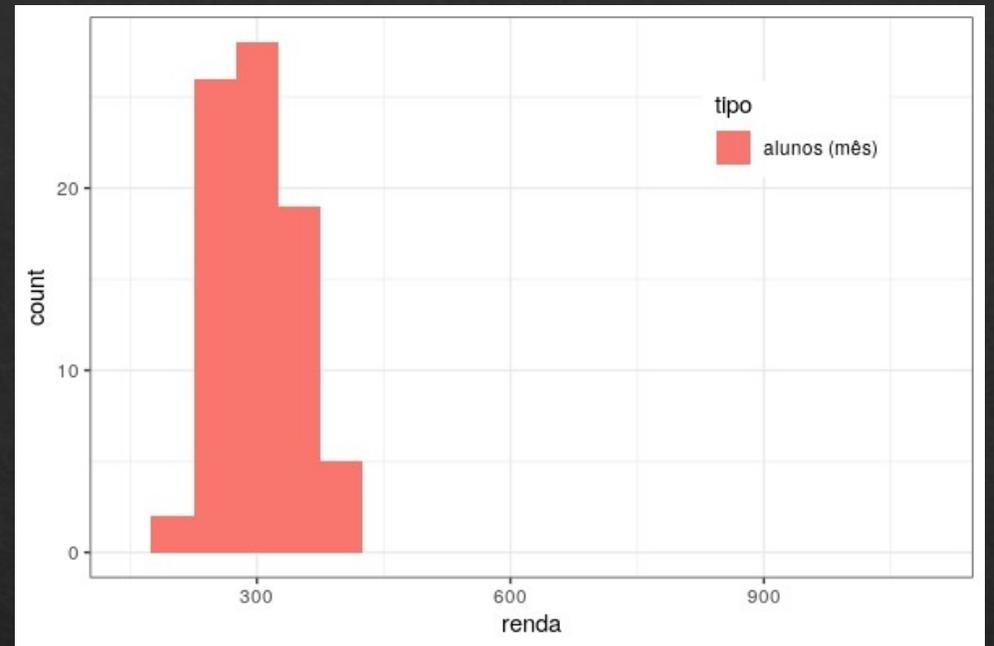
$$IIQ = Q3 - Q1$$



# Renda de alunos



Um aluno consegue estágio remunerado.  
O que acontece com a variância?  
O que acontece com o intervalo interquartil?



Desvio-padrão = 45,8

Desvio-padrão = 90,2

