

# Análise de Relação entre Variáveis



---

<https://advancedinstitute.ai>



# Análise de Relação entre Variáveis

---

Teste de Hipótese

## Referências e Fontes das Imagens

- [Estatística Básica](#) (Book)
- [Think Stats](#) (Book)
- [Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python](#) (Book)
- [Stats Normaltest](#)

## Grupos Pareados e Independentes

- ☐ Quando utilizar?

## Grupos Pareados e Independentes

- ☐ Quando utilizar?
- ☐ Sabe a diferença entre eles?

## Grupos Pareados e Independentes

- ☐ Grupos Independentes (ou Não Pareados)

## Grupos Pareados e Independentes

- Grupos Independentes (ou Não Pareados)
  - São aqueles que não guardam relação de dependência entre si

## Grupos Pareados e Independentes

- Grupos Independentes (ou Não Pareados)
  - São aqueles que não guardam relação de dependência entre si
  - Não são as mesmas unidades amostrais que são avaliadas



## Grupos Pareados e Independentes

- Grupos Independentes (ou Não Pareados)
  - São aqueles que não guardam relação de dependência entre si
  - Não são as mesmas unidades amostrais que são avaliadas
  - Cada grupo é composto por indivíduos ou unidades amostrais distintos

## Grupos Independentes (ou Não Pareados)

- **Exemplo:** Imagine que um psicólogo quer comparar o nível de ansiedade entre dois grupos diferentes de paciente.



Figure: Grupo A (Terapia A)



Figure: Grupo B (Terapia B)

## Grupos Independentes (ou Não Pareados)

- Os indivíduos no grupo A são diferentes dos indivíduos no grupo B.

## Grupos Independentes (ou Não Pareados)

- ☐ Os indivíduos no grupo A são diferentes dos indivíduos no grupo B.
- ☐ O nível de ansiedade seria medido usando uma escala numérica de 0 a 100 por exemplo, para cada paciente ao final do tratamento.

## Grupos Independentes (ou Não Pareados)

- ☐ Os indivíduos no grupo A são diferentes dos indivíduos no grupo B.
- ☐ O nível de ansiedade seria medido usando uma escala numérica de 0 a 100 por exemplo, para cada paciente ao final do tratamento.
- ☐ Não há obrigatoriedade estatística para que os grupos tenham o mesmo tamanho

## Grupos Dependentes (ou Pareados)

- ☐ São grupos que guardam relação de dependência entre si

## Grupos Dependentes (ou Pareados)

- ☐ São grupos que guardam relação de dependência entre si
- ☐ São as mesmas unidades amostrais avaliadas mais de uma vez

## Grupos Dependentes (ou Pareados)

- ❑ **Exemplo:** Imagine que um psicólogo quer avaliar o efeito de uma terapia no nível de ansiedade de um único grupo de pacientes.



Figure: Grupo A (Antes)



Figure: Grupo B (Depois)



## Grupos Dependentes (ou Pareados)

- O nível de ansiedade numa escala de 0 a 100 por exemplo, seria medido em cada paciente **antes** de iniciar a terapia e **depois** ao final da terapia

## Grupos Dependentes (ou Pareados)

- ☐ O nível de ansiedade numa escala de 0 a 100 por exemplo, seria medido em cada paciente **antes** de iniciar a terapia e **depois** ao final da terapia
- ☐ São as mesmas pessoas avaliadas em dois momentos distintos

## Grupos Dependentes (ou Pareados)

- ☐ O nível de ansiedade numa escala de 0 a 100 por exemplo, seria medido em cada paciente **antes** de iniciar a terapia e **depois** ao final da terapia
- ☐ São as mesmas pessoas avaliadas em dois momentos distintos
- ☐ **Obrigatoriamente** os grupos devem ter o mesmo tamanho

## Grupos Dependentes (ou Pareados)

- ☐ O nível de ansiedade numa escala de 0 a 100 por exemplo, seria medido em cada paciente **antes** de iniciar a terapia e **depois** ao final da terapia
- ☐ São as mesmas pessoas avaliadas em dois momentos distintos
- ☐ **Obrigatoriamente** os grupos devem ter o mesmo tamanho
- ☐ Caso isso não seja possível, o valor da quantidade da última análise é o que será válido

## Grupos Dependentes (ou Pareados)

- O conceito **pareado**, não se refere a ser "par", posso ter varias etapas de análise com o mesmo grupo



Figure: Mês 1



Figure: Mês 3



Figure: Mês 9



Figure: Mês 12

## Pareamento Estatístico X Pareamento Metodológico

- **Pareamento Estatístico:**

## Pareamento Estatístico X Pareamento Metodológico

### ☐ Pareamento Estatístico:

- Os grupos a serem comparados guardam relação de dependência entre si

## Pareamento Estatístico X Pareamento Metodológico

### ☐ Pareamento Estatístico:

- Os grupos a serem comparados guardam relação de dependência entre si
- Isso ocorre porque são as mesmas unidades amostrais (indivíduos, animais, etc.) que são avaliadas mais de uma vez



## Pareamento Estatístico X Pareamento Metodológico

- **Pareamento Metodológico:**

## Pareamento Estatístico X Pareamento Metodológico

### ☐ Pareamento Metodológico:

- Técnica utilizada durante o delineamento da pesquisa, antes mesmo da coleta de dados ou análise estatística

## Pareamento Estatístico X Pareamento Metodológico

### ☐ Pareamento Metodológico:

- Técnica utilizada durante o delineamento da pesquisa, antes mesmo da coleta de dados ou análise estatística
- Seu objetivo é garantir que grupos independentes tenham características semelhantes no início do estudo

## Associação entre Variáveis

- Tem como objetivo, descrever simultaneamente a variabilidade de **duas ou mais** variáveis de forma que cada conjunto seja observado para uma mesma unidade observacional (pessoas, animais, plantas, peças, etc.)

## Associação entre Variáveis

- Tem como objetivo, descrever simultaneamente a variabilidade de **duas ou mais** variáveis de forma que cada conjunto seja observado para uma mesma unidade observacional (pessoas, animais, plantas, peças, etc.)
  - Vamos iniciar com um par de variáveis  $(x,y)$ , sendo  $(x_i, y_i)$ ,  $i = 1, \dots, n$  pares de observações de duas variáveis:

## Associação entre Variáveis

- Vamos iniciar com um par de variáveis  $(x,y)$ , sendo  $(x_i, y_i)$ ,  $i = 1, \dots, n$  pares de observações de duas variáveis:

## Associação entre Variáveis

- Vamos iniciar com um par de variáveis  $(x,y)$ , sendo  $(x_i, y_i)$ ,  $i = 1, \dots, n$  pares de observações de duas variáveis:

<b>x</b>	<b>y</b>
Qualitativa	Qualitativa
Qualitativa	Quantitativa
Quantitativa	Qualitativa

## Associação entre Variáveis

- Vamos iniciar com um par de variáveis  $(x,y)$ , sendo  $(x_i, y_i)$ ,  $i = 1, \dots, n$  pares de observações de duas variáveis:

<b>x</b>	<b>y</b>
Qualitativa	Qualitativa
Qualitativa	Quantitativa
Quantitativa	Qualitativa

Para cada combinação de pares de variáveis, um tipo de análise será realizada.



## Associação entre Variáveis

- Para entendermos a **associação** entre  $x$  e  $y$ , precisamos de uma medida associação, que deve avaliar se essa associação é **forte** ou **fraca**, **positiva** ou **negativa**.

## Associação entre Variáveis

- Para entendermos a **associação** entre  $x$  e  $y$ , precisamos de uma medida associação, que deve avaliar se essa associação é **forte** ou **fraca**, **positiva** ou **negativa**.
- Outra possibilidade é através da **representação gráfica**, podendo para esse tipo de análise ser:
  - Sentido da associação: Positiva ou Negativa;
  - Intensidade da Associação: Forte, Moderada ou Fraca

Entender os sentidos e sua intensidade é necessário e muito utilizado em processos de análise de **predições** nos dados.

## Associação entre Variáveis

- $x$ : Altura da planta
- $y$ : Largura da folha

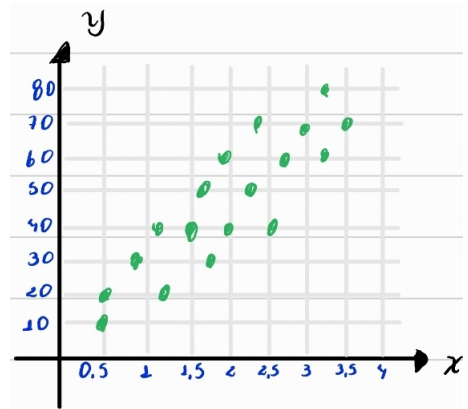


Figure: Correlação Positiva

## Associação entre Variáveis

- ☐  $x$ : Idade
- ☐  $y$ : N<sup>o</sup> de acidentes

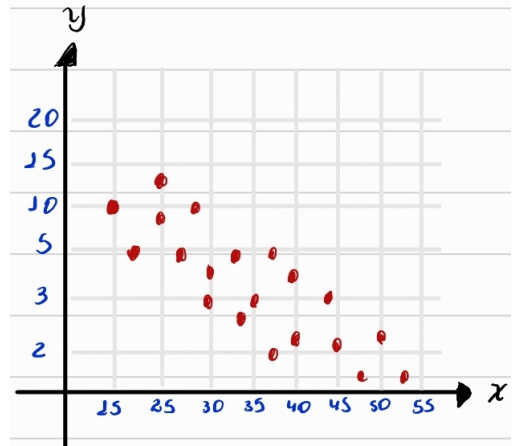


Figure: Correlação Negativa

## Associação entre Variáveis

- $x$ : N° do sapato
- $y$ : Nota final do semestre

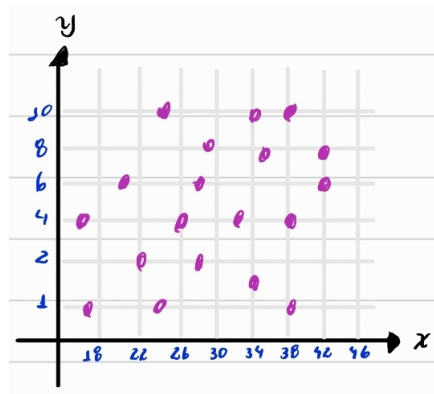


Figure: Sem Associação

## Tipos de relação entre Variáveis

- Causal unilateral:
  - $y$  depende de  $x$  ou  $x$  depende de  $y$ .

## Tipos de relação entre Variáveis

- Causal unilateral:

- $y$  depende de  $x$  ou  $x$  depende de  $y$ .

- Exemplo:

- Preço da venda de um produto (**y**) depende do local da venda (**x**).

$$x \Rightarrow y$$

## Tipos de relação entre Variáveis

### ☐ Causal unilateral (Preço X Local)

- É possível perceber então que o preço do produto **depende** do local da venda e também pode depender de outros fatores mas, o **local** não depende do preço do produto.



## Tipos de relação entre Variáveis

- Causal bilateral:
  - $y$  depende de  $x$  e  $x$  depende de  $y$ .

## Tipos de relação entre Variáveis

- Causal bilateral:

- $y$  depende de  $x$  e  $x$  depende de  $y$ .

- Exemplo:

- Peso ( $x$ ) e circunferência abdominal ( $y$ ) de uma pessoa.

$$x \Leftrightarrow y$$

## Variáveis Dependentes e Independentes

- **Variável Dependente (resposta):**

## Variáveis Dependentes e Independentes

### ☐ Variável Dependente (resposta):

- Representa uma variável cujo valor depende de outra variável (independente)

## Variáveis Dependentes e Independentes

### ☐ Variável Dependente (resposta):

- Representa uma variável cujo valor depende de outra variável (independente)
- É o que queremos encontrar ou medir

## Variáveis Dependentes e Independentes

### ☐ Variável Dependente (resposta):

- Representa uma variável cujo valor depende de outra variável (independente)
- É o que queremos encontrar ou medir
- Ela depende da independente. Ex: glicose, pressão, nota, etc.

## Variáveis Dependentes e Independentes

- ☐ **Variável Independente (grupamento):**

## Variáveis Dependentes e Independentes

- **Variável Independente (grupamento):**
  - É aquela que não depende de outras variáveis



## Variáveis Dependentes e Independentes

### ☐ Variável Independente (grupamento):

- É aquela que não depende de outras variáveis
- Geralmente é a que usamos para explicar ou prever algo. Ex: sexo, idade, tratamento, etc.

## Variáveis Dependentes e Independentes

- **Exemplo:** Diferença nos níveis de glicose entre homens e mulheres.

## Variáveis Dependentes e Independentes

- **Exemplo:** Diferença nos níveis de glicose entre homens e mulheres.
  - É aquela que não depende de outras variáveis

## Variáveis Dependentes e Independentes

- **Exemplo:** Diferença nos níveis de glicose entre homens e mulheres.
  - É aquela que não depende de outras variáveis
  - Geralmente é a que usamos para explicar ou prever algo. Ex: sexo, idade, tratamento, etc.

## Variáveis Dependentes e Independentes

- **Exemplo:** Diferença nos níveis de glicose entre homens e mulheres.

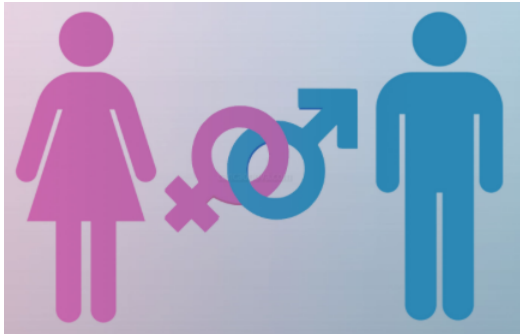


Figure: M—H (Independente)

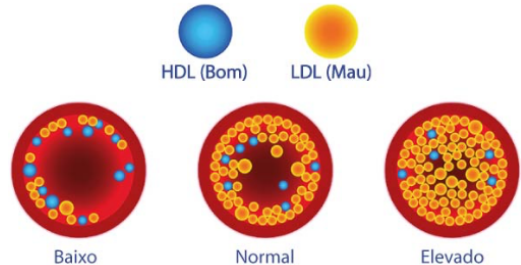


Figure: Células (Dependentes)

## Tipos de relação entre Variáveis

- Causal bilateral (Peso X Circunferência abdominal)
  - Neste caso, se o peso cresce, a circunferência também aumenta porém, se a circunferência aumenta, o peso também será maior, comprovando que uma variável depende da outra.

## Tipos de relação entre Variáveis

- Dependência Indireta:
  - Considerado uma condição que gera discussão em estatística devido ao fato de existir uma correlação mas não existir causa.

## Tipos de relação entre Variáveis

### □ Dependência Indireta:

- Considerado uma condição que gera discussão em estatística devido ao fato de existir uma correlação mas não existir causa.

### □ Exemplo:

- Vendas de sorvete na praia ( $x$ ), causas de afogamento ( $y$ ) e temperatura ( $w$ ).



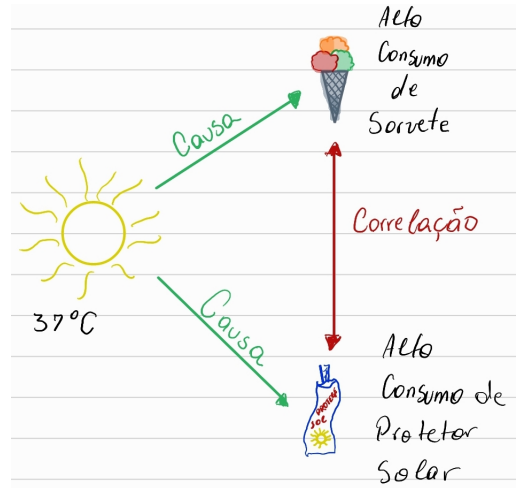


## Tipos de relação entre Variáveis

- Aqui,  $(w)$  possui uma **relação causal unilateral** com  $(x)$  e  $(w)$  também tem **relação causal unilateral** com  $(y)$ , ou seja, aumentar  $(w)$  irá causar aumento tanto em  $(x)$  como em  $(y)$ .

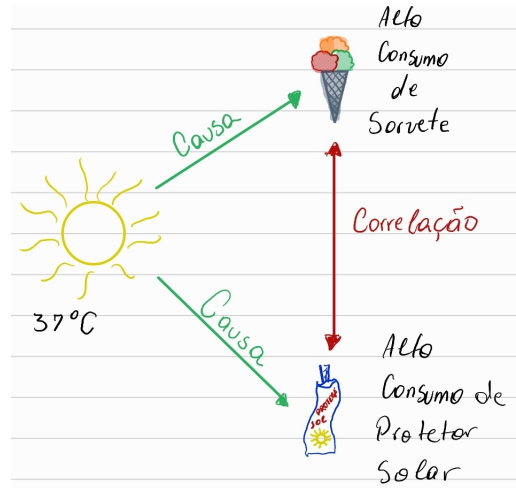
## Correlação não é Causalidade

- ☐ **Altas temperaturas:** Variável independente (x)
- ☐ **Aumento das vendas:** Variável dependente (y)



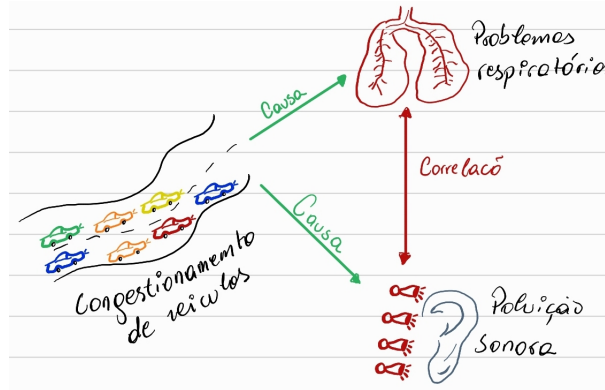
## Correlação não é Causalidade

- Existe uma correlação entre o consumo de sorvetes e o consumo de protetor solar porém, um não causa o outro (**Associação não causal**)



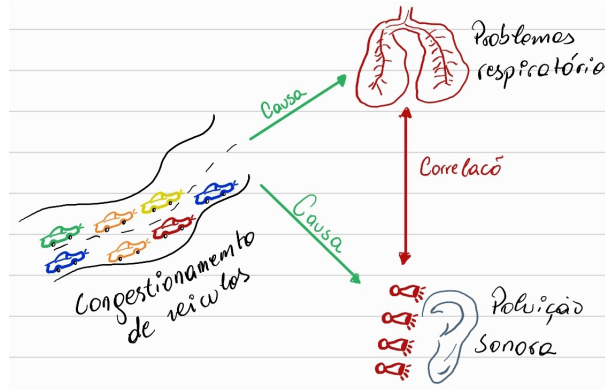
## Correlação não é Causalidade

- **Congestionamento de veículos:**  
Variável independente (x)
- **Nível de Poluição:** Variável dependente (y)



## Correlação não é Causalidade

- Existe uma correlação entre os problemas respiratórios e a poluição sonora porém, um não causa o outro (**Associação não causal**)



## Testes para Comparação

- ☐ A primeira coisa a ser feita é encontrar qual é a **variável dependente**

## Testes para Comparação

- ☐ A primeira coisa a ser feita é encontrar qual é a **variável dependente**
- ☐ Vamos dividir nosso exemplo em 3 partes:

## Testes para Comparação

- A primeira coisa a ser feita é encontrar qual é a **variável dependente**
- Vamos dividir nosso exemplo em 3 partes:
  - Qualitativa Nominal;
  - Qualitativa Ordinal;
  - Quantitativa



## Testes para Comparação

- ☐ Vamos iniciar com as variáveis **Qualitativas Nominais**

## Testes para Comparação

- Vamos iniciar com as variáveis **Qualitativas Nominais**
  - Qualitativa Nominal: Como já sabemos, para quando possuímos uma variável que tem uma característica que é uma qualidade e não uma ordem.

## Testes para Comparação (Qualitativas Nominais)

### ☐ Teste Hipotético I

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para isso, aplicou um questionário contendo 20 questões.

## Testes para Comparação (Qualitativas Nominais)

### □ Teste Hipotético I

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para isso, aplicou um questionário contendo 20 questões.
- De acordo com os valores das questões no questionário será dado o diagnóstico inicial sendo, caso obtenha 10 pontos ou mais, será considerado como **estressado**, caso obtenha menos de 10 pontos, seu diagnóstico terá um quadro **normal**.

## Testes para Comparação (Qualitativas Nominais)

### □ Teste Hipotético I

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para isso, aplicou um questionário contendo 20 questões.
- De acordo com os valores das questões no questionário será dado o diagnóstico inicial sendo, caso obtenha 10 pontos ou mais, será considerado como **estressado**, caso obtenha menos de 10 pontos, seu diagnóstico terá um quadro **normal**.
- Podemos verificar então que ao final, a resposta da nossa pesquisa será se o indivíduo está estressado ou não.

## Testes para Comparação (Qualitativas Nominais)

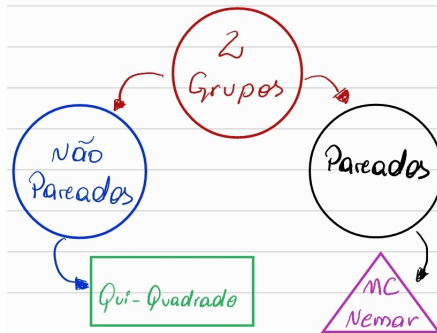
- Mas... qual seria o melhor teste comparativo?



## Testes para Comparação (Qualitativas Nominais)

### Não Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos)



## Testes para Comparação (Qualitativas Nominais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos)
  - **Qui-quadrado:** trata-se de um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis **categóricas nominais** e avaliar a associação existente entre essas variáveis.



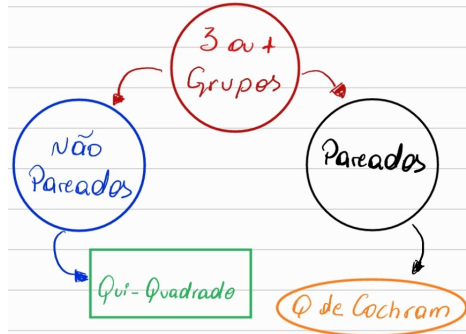
## Testes para Comparação (Qualitativas Nominais) Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos)
  - **Qui-quadrado**: trata-se de um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis **categóricas nominais** e avaliar a associação existente entre essas variáveis.
  - **MC Nemar**: é um teste não paramétrico que se baseia em dados ordinais e nominais e não requerem os pressupostos dos testes paramétricos.

## Testes para Comparação (Qualitativas Nominais)

### Pareados

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)



## Testes para Comparação (Qualitativas Nominais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Qui-quadrado:** trata-se de um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis **categóricas nominais** e avaliar a associação existente entre essas variáveis.

## Testes para Comparação (Qualitativas Nominais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Qui-quadrado**: trata-se de um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis **categóricas nominais** e avaliar a associação existente entre essas variáveis.
  - **Q de Cochran**: é um teste estatístico não paramétrico para verificar se  $k$  tratamentos têm efeitos similares ao realizado nas análises de delineamentos em blocos aleatorizados, onde a variável de resposta pode assumir apenas dois valores possíveis (**0** ou **1**).

## Testes para Comparação

- ☐ Vamos agora com as variáveis **Qualitativas Ordinal**

## Testes para Comparação

- Vamos agora com as variáveis **Qualitativas Ordinal**
  - Qualitativa Ordinal: Como próprio nome diz, existe uma ordem (hierarquia) na classificação dos dados na qual precisas ser respeitada

## Testes para Comparação (Qualitativas Ordinais)

### ☐ Teste Hipotético II

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para esse novo teste, o psicólogo aplicou agora um questionário contendo 50 questões.

## Testes para Comparação (Qualitativas Ordinais)

### □ Teste Hipotético II

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para esse novo teste, o psicólogo aplicou agora um questionário contendo 50 questões.
- Nesse novo experimento, a análise realizada ao final será aplicada dependendo das respostas do indivíduo, classificando como: **ansiedade**, **ansiedade moderada**, **ansiedade leve**, **muito ansioso** e **sem ansiedade**.



## Testes para Comparação (Qualitativas Ordinais)

### □ Teste Hipotético II

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, para esse novo teste, o psicólogo aplicou agora um questionário contendo 50 questões.
- Nesse novo experimento, a análise realizada ao final será aplicada dependendo das respostas do indivíduo, classificando como: **ansiedade**, **ansiedade moderada**, **ansiedade leve**, **muito ansioso** e **sem ansiedade**.
- Podemos verificar então que ao final, a resposta da nossa pesquisa irá apresentar em que nível de estresse se encontra um determinado indivíduo.

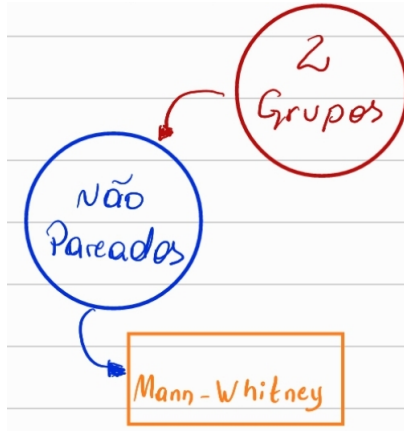
## Testes para Comparação (Qualitativas Ordinais)

- Mas... qual seria o melhor teste comparativo?



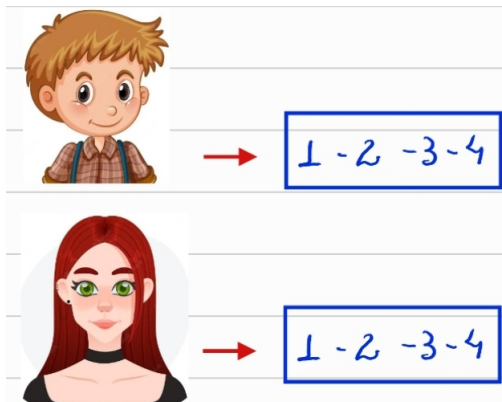
## Testes para Comparação (Qualitativas Ordinais)

- Diagrama para variáveis qualitativas nominais (2 grupos)



## Testes para Comparação (Qualitativas Ordinais) Não Pareados

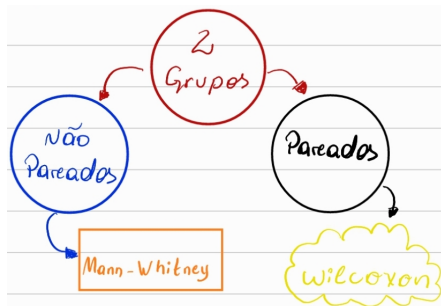
- Diagrama para variáveis qualitativas nominais (2 grupos)



## Testes para Comparação (Qualitativas Ordinais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos)



## Testes para Comparação (Qualitativas Ordinais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos ou mais)
  - **Mann-Whitney**: indicado para comparação de dois grupos não pareados para se verificar se pertencem ou não à mesma população e cujos requisitos para aplicação do **teste t de Student** não foram cumpridos

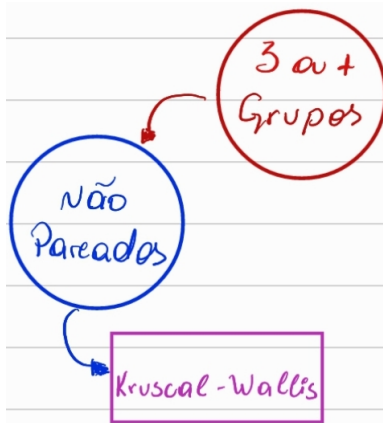
## Testes para Comparação (Qualitativas Ordinais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (2 grupos ou mais)
  - **Mann-Whitney**: indicado para comparação de dois grupos não pareados para se verificar se pertencem ou não à mesma população e cujos requisitos para aplicação do **teste t de Student** não foram cumpridos
  - **Wilcoxon**: é um método não-paramétrico para comparação de duas amostras pareadas ou não pareadas.

## Testes para Comparação (Qualitativas Ordinais)

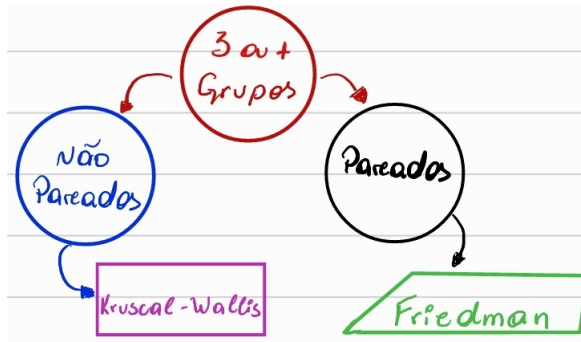
- Diagrama para variáveis qualitativas nominais (3 grupos)





## Testes para Comparação (Qualitativas Ordinais)

- Diagrama para variáveis qualitativas nominais (3 grupos)



## Testes para Comparação (Qualitativas Ordinais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Kruskal-Wallis:** é um método não paramétrico para testar se amostras se originam ou não da mesma distribuição. Muito utilizado para comparar duas ou mais amostras independentes de tamanhos iguais ou diferentes.

## Testes para Comparação (Qualitativas Ordinais)

### Não Pareados e Pareados

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Kruskal-Wallis:** é um método não paramétrico para testar se amostras se originam ou não da mesma distribuição. Muito utilizado para comparar duas ou mais amostras independentes de tamanhos iguais ou diferentes.
  - **Friedman:** trata-se de um teste não-paramétrico utilizado para comparar dados amostrais vinculados, ou seja, quando o mesmo indivíduo é avaliado mais de uma vez. Esse teste não utiliza os dados numéricos diretamente, mas sim os postos ocupados por eles após a ordenação feita para cada grupo separadamente.

## Testes para Comparação (Quantitativas)

### □ Teste Hipotético III

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, só que agora, o psicólogo aplicou um teste psicológico que classificava o nível de ansiedade do indivíduo através de uma escala numérica.

## Testes para Comparação (Quantitativas)

### □ Teste Hipotético III

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, só que agora, o psicólogo aplicou um teste psicológico que classificava o nível de ansiedade do indivíduo através de uma escala numérica.
- Nessa escala que vai de 0 à 100, quanto maior o *score* alcançado, maior as chances do indivíduo estar em um quadro de estresse.

## Testes para Comparação (Quantitativas)

### □ Teste Hipotético III

- Vamos imaginar que um psicólogo quer **comparar** o nível de ansiedade de seu grupo de pacientes, só que agora, o psicólogo aplicou um teste psicológico que classificava o nível de ansiedade do indivíduo através de uma escala numérica.
- Nessa escala que vai de 0 à 100, quanto maior o *score* alcançado, maior as chances do indivíduo estar em um quadro de estresse.
- Agora, nesse tipo de análise nossa variável resposta independente se trata de um número, sendo esse responsável por impactar o resultado.

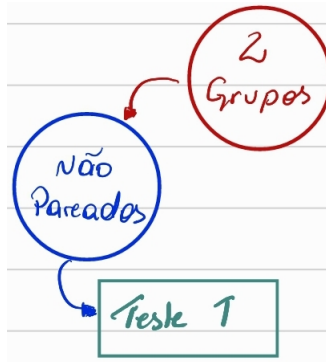
## Testes para Comparação (Quantitativas)

- Mas... qual seria o melhor teste comparativo?



## Testes para Comparação (Quantitativas Não Pareados

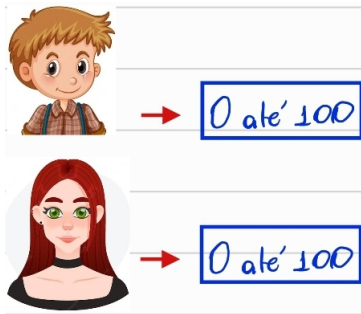
- Diagrama para variáveis quantitativas (2 grupos)





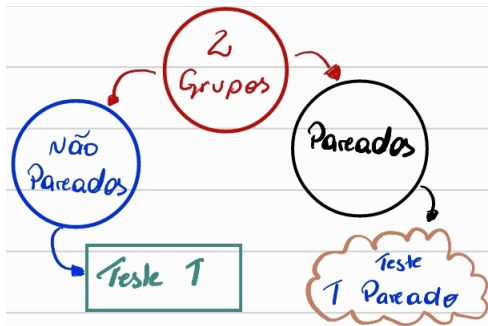
## Testes para Comparação (Quantitativas Não Pareados)

- Diagrama para variáveis quantitativas (2 grupos)



## Testes para Comparação (Quantitativas) Não Pareados e Pareados

- Diagrama para variáveis quantitativas (2 grupos)



## Testes para Comparação (Quantitativas Não Pareados e Pareados)

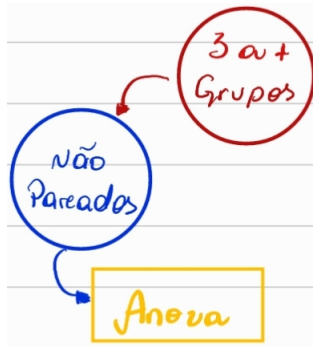
- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Teste T:** é um teste de hipóteses utilizado quando queremos tirar conclusões de um grupo inteiro de indivíduos com base em apenas uma pequena amostra coletada.

## Testes para Comparação (Quantitativas Não Pareados e Pareados)

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Teste T:** é um teste de hipóteses utilizado quando queremos tirar conclusões de um grupo inteiro de indivíduos com base em apenas uma pequena amostra coletada.
  - **Teste-t pareado:** simplesmente calcula a diferença entre observações emparelhadas (por exemplo, antes e depois) e, em seguida, realiza um teste-t para 1 amostra sobre as diferenças.

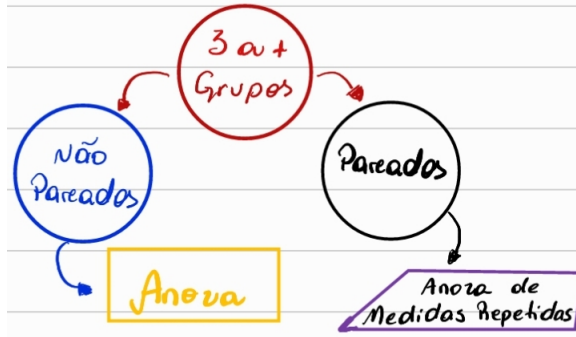
## Testes para Comparação (Quantitativas Não Pareados

- Diagrama para variáveis quantitativas (3 grupos)



## Testes para Comparação (Quantitativas) Não Pareados e Pareados

- Diagrama para variáveis quantitativas (3 grupos)



## Testes para Comparação (Quantitativas Não Pareados e Pareados)

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Anova:** é uma fórmula estatística usada para comparar as variâncias entre as medianas (ou médias) de grupos diferentes.

## Testes para Comparação (Quantitativas Não Pareados e Pareados)

- Diagrama para variáveis qualitativas nominais (3 grupos ou mais)
  - **Anova:** é uma fórmula estatística usada para comparar as variâncias entre as medianas (ou médias) de grupos diferentes.
  - **Anova de Medidas Repetidas:** compara as médias de uma ou mais variáveis que se baseiam em observações repetidas. Um modelo ANOVA de medidas repetidas também pode incluir zero ou mais variáveis independentes.



## Iniciando os Testes

- Basicamente, sempre que iremos utilizar uma amostra de dados, calculamos algumas métricas estatísticas conhecidas (como média, mediana, desvio padrão e outras) e generalizamos esse valor para toda uma população, esse processo é conhecido como **inferência estatística**.
  - Mas essa generalização pode realmente ser feita?

## Iniciando os Testes

- Basicamente, sempre que iremos utilizar uma amostra de dados, calculamos algumas métricas estatísticas conhecidas (como média, mediana, desvio padrão e outras) e generalizamos esse valor para toda uma população, esse processo é conhecido como **inferência estatística**.
  - Mas essa generalização pode realmente ser feita?
  - Será que nossa amostra é realmente uma boa representação da população?

## Iniciando os Testes

- Basicamente, sempre que iremos utilizar uma amostra de dados, calculamos algumas métricas estatísticas conhecidas (como média, mediana, desvio padrão e outras) e generalizamos esse valor para toda uma população, esse processo é conhecido como **inferência estatística**.
  - Mas essa generalização pode realmente ser feita?
  - Será que nossa amostra é realmente uma boa representação da população?
  - Como provar isso estatisticamente?

## Iniciando os Testes

- Basicamente, sempre que iremos utilizar uma amostra de dados, calculamos algumas métricas estatísticas conhecidas (como média, mediana, desvio padrão e outras) e generalizamos esse valor para toda uma população, esse processo é conhecido como **inferência estatística**.
  - Mas essa generalização pode realmente ser feita?
  - Será que nossa amostra é realmente uma boa representação da população?
  - Como provar isso estatisticamente?
  - **Simples, usando teste de hipóteses!**

## Iniciando os Testes

- O teste de hipóteses é uma ótima ferramenta para validar as nossas inferências, mas mesmo sendo ótima, muitas vezes ela é utilizada de forma errada ou simplesmente é esquecido o que realmente o teste representa.

## Teorema Central do Limite

- Antes de apresentar as características de um teste de hipótese, é importante entendermos o que é o **Teorema Central do Limite - TCL**, pois trata-se de um dos principais conceitos por trás da inferência estatística.

## Teorema Central do Limite

- Antes de apresentar as características de um teste de hipótese, é importante entendermos o que é o **Teorema Central do Limite - TCL**, pois trata-se de um dos principais conceitos por trás da inferência estatística.
- Basicamente, esse teorema nos diz que conforme aumentamos o tamanho de uma amostra, a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal, independentemente da distribuição da população.

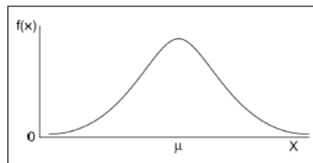


Figure: Distribuição normal: simétrica e com média

## Teorema Central do Limite

- Mas o que isso significa???



## Teorema Central do Limite

- Mas o que isso significa???
- Se utilizarmos a *feature* **Renda** do nosso conjunto de dados para analisar o salário médio dos brasileiros. Pela definição de média, teríamos que somar o salário de todos os brasileiros e dividir pelo tamanho da população do Brasil.

## Teorema Central do Limite

- Mas o que isso significa???
- Se utilizarmos a *feature* **Renda** do nosso conjunto de dados para analisar o salário médio dos brasileiros. Pela definição de média, teríamos que somar o salário de todos os brasileiros e dividir pelo tamanho da população do Brasil.
- Isso é um pouco inviável de ser feito não acham?

## Teorema Central do Limite

- Nesse caso, como não é possível ter informações da população toda, teríamos que utilizar amostras. Com isso, decidimos perguntar o salário de **10 pessoas** que encontramos aleatoriamente na rua (vamos ignorar o fato de que praticamente ninguém responderia a uma pergunta dessa rss)

## Teorema Central do Limite

- Nesse caso, como não é possível ter informações da população toda, teríamos que utilizar amostras. Com isso, decidimos perguntar o salário de **10 pessoas** que encontramos aleatoriamente na rua (vamos ignorar o fato de que praticamente ninguém responderia a uma pergunta dessa rss)
- Então, pegaríamos a média desses 10 salários, o que nos retornaria um valor  $X_1$ .

## Teorema Central do Limite

- Nesse caso, como não é possível ter informações da população toda, teríamos que utilizar amostras. Com isso, decidimos perguntar o salário de **10 pessoas** que encontramos aleatoriamente na rua (vamos ignorar o fato de que praticamente ninguém responderia a uma pergunta dessa rss)
- Então, pegaríamos a média desses 10 salários, o que nos retornaria um valor  $X_1$ .
- Em seguida, perguntamos o salário de mais 10 pessoas para obtermos mais uma média,  $X_2$ .

## Teorema Central do Limite

- Nesse caso, como não é possível ter informações da população toda, teríamos que utilizar amostras. Com isso, decidimos perguntar o salário de **10 pessoas** que encontramos aleatoriamente na rua (vamos ignorar o fato de que praticamente ninguém responderia a uma pergunta dessa rrr)
- Então, pegaríamos a média desses 10 salários, o que nos retornaria um valor  $X_1$ .
- Em seguida, perguntamos o salário de mais 10 pessoas para obtermos mais uma média,  $X_2$ .
- Repetimos esse processo com mais 10 pessoas, obtemos então outra média,  $X_3$ .

## Teorema Central do Limite

- Nesse caso, como não é possível ter informações da população toda, teríamos que utilizar amostras. Com isso, decidimos perguntar o salário de **10 pessoas** que encontramos aleatoriamente na rua (vamos ignorar o fato de que praticamente ninguém responderia a uma pergunta dessa rrr)
- Então, pegaríamos a média desses 10 salários, o que nos retornaria um valor  $X_1$ .
- Em seguida, perguntamos o salário de mais 10 pessoas para obtermos mais uma média,  $X_2$ .
- Repetimos esse processo com mais 10 pessoas, obtemos então outra média,  $X_3$ .
- Vamos supor agora que repetimos esse processo até obter mil médias amostrais,  $(X_1, X_2, \dots, X_{1000})$ .

## Teorema Central do Limite

- Pelo **Teorema Central do Limite**, a distribuição dessas **mil médias** tende a ser normal.



## Teorema Central do Limite

- Pelo **Teorema Central do Limite**, a distribuição dessas **mil médias** tende a ser normal.
- Se ao invés de perguntarmos a 10 pessoas, perguntássemos a 30, a distribuição dessas médias obtidas por meio de amostras de 30 pessoas, se aproximaria mais ainda de uma normal!

## Teorema Central do Limite

- Pelo **Teorema Central do Limite**, a distribuição dessas **mil médias** tende a ser normal.
- Se ao invés de perguntarmos a 10 pessoas, perguntássemos a 30, a distribuição dessas médias obtidas por meio de amostras de 30 pessoas, se aproximaria mais ainda de uma normal!
- Conforme aumentássemos o número de pessoas questionadas sobre o seu salário, mais a distribuição dessas médias se aproximaria de uma normal.

## Teorema Central do Limite

- E o melhor, a média dessa distribuição normal é uma ótima aproximação da média dos salários de toda a população.

## Teorema Central do Limite

- E o melhor, a média dessa distribuição normal é uma ótima aproximação da média dos salários de toda a população.
- Com isso teríamos uma estimativa, ou seja, teríamos uma inferência da média dos salários de toda a população brasileira sem precisar perguntar isso para cada pessoa que vive no país. :-)

## Teorema Central do Limite

- E o melhor, a média dessa distribuição normal é uma ótima aproximação da média dos salários de toda a população.
- Com isso teríamos uma estimativa, ou seja, teríamos uma inferência da média dos salários de toda a população brasileira sem precisar perguntar isso para cada pessoa que vive no país. :-)
- Podemos observar também que em nenhum momento há a definição de que os salários são normalmente distribuídos.

## Teorema Central do Limite

- E o melhor, a média dessa distribuição normal é uma ótima aproximação da média dos salários de toda a população.
- Com isso teríamos uma estimativa, ou seja, teríamos uma inferência da média dos salários de toda a população brasileira sem precisar perguntar isso para cada pessoa que vive no país. :-)
- Podemos observar também que em nenhum momento há a definição de que os salários são normalmente distribuídos.
- Pelo TCL essa característica não é necessária. A distribuição das médias amostrais dos salários seguirão uma normal, mesmo que os salários sigam qualquer outra distribuição.

## Teorema Central do Limite

- Vamos supor que os salários seguem uma distribuição normal.

## Teorema Central do Limite

- Vamos supor que os salários seguem uma distribuição normal.
- Muitas pessoas recebem um salário médio e o número de pessoas que recebem um salário menor que a média é o mesmo número de pessoas que recebem um salário maior que a média.



## Teorema Central do Limite

- Vamos supor que os salários seguem uma distribuição normal.
- Muitas pessoas recebem um salário médio e o número de pessoas que recebem um salário menor que a média é o mesmo número de pessoas que recebem um salário maior que a média.
- Por outro lado, acreditamos que o salário das pessoas seguem uma **distribuição exponencial**.

## Teorema Central do Limite

- Vamos supor que os salários seguem uma distribuição normal.
- Muitas pessoas recebem um salário médio e o número de pessoas que recebem um salário menor que a média é o mesmo número de pessoas que recebem um salário maior que a média.
- Por outro lado, acreditamos que o salário das pessoas seguem uma **distribuição exponencial**.
- Muitas pessoas recebem um salário mais baixo, enquanto poucas pessoas recebem um salário mais alto.

## Teorema Central do Limite

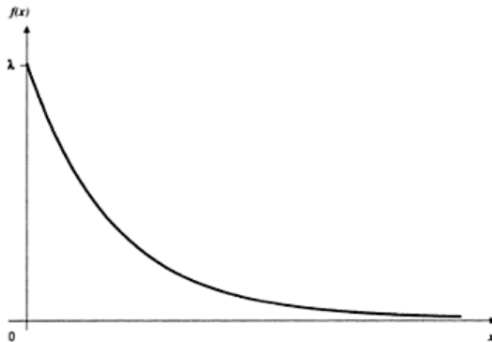


Figure: Distribuição exponencial.

## Teorema Central do Limite

- Uma pessoa sonhadora acredita que vivemos em uma sociedade muito igualitária, onde os salários seguem uma **distribuição uniforme**, ou seja, as pessoas recebem salários parecidos.

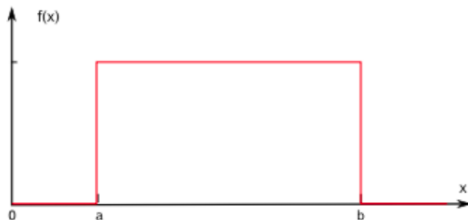


Figure: Distribuição uniforme.

## Teorema Central do Limite

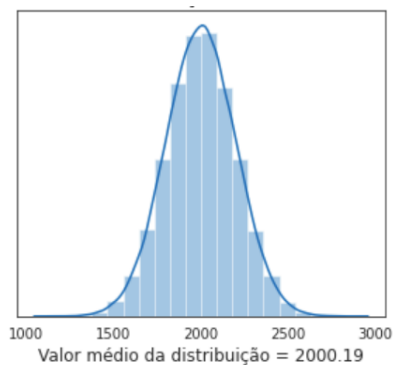
- Para obtermos uma melhor visualização, geramos uma distribuição contendo 100.000 valores aleatórios que seguem, respectivamente, uma distribuição **normal**, **exponencial** e **uniforme**.

## Teorema Central do Limite

- Para obtermos uma melhor visualização, geramos uma distribuição contendo 100.000 valores aleatórios que seguem, respectivamente, uma distribuição **normal**, **exponencial** e **uniforme**.
- O valor médio das três distribuições é aproximadamente **2.000**, ou seja, a suposição é que o salário médio da população é = R\$2.000.

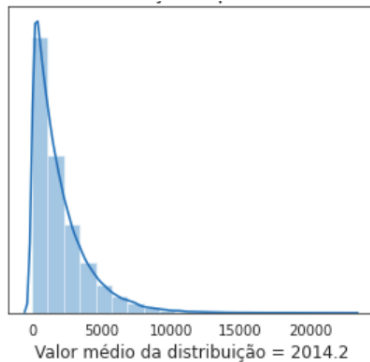
## Teorema Central do Limite

- Distribuição Normal



## Teorema Central do Limite

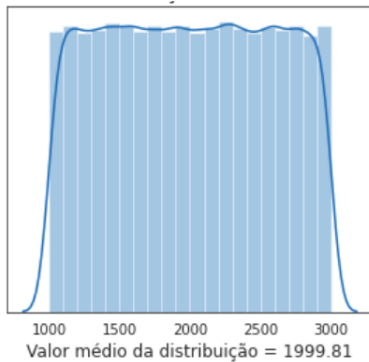
### □ Distribuição Exponencial





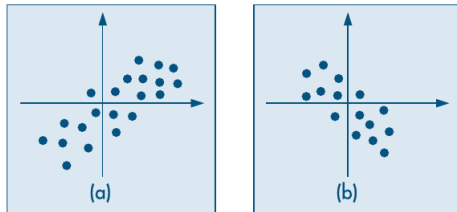
## Teorema Central do Limite

- Distribuição Uniforme



## Teorema Central do Limite

- Tipos de associações entre duas variáveis
  - (a) Associação linear direta (ou **positiva**)
    - Soma do produto das coordenadas será sempre positivo
  - (b) Dependência linear inversa (ou **negativa**)
    - Soma dos produtos das coordenadas será negativa



Dúvidas?