MBA EM **DATA SCIENCE E ANALYTICS**

Fundamentos de Estatística

Prof. Dr. Wilson Tarantin Junior

Jonas Pereira Araujo 057.999.937-80

MBA USP ESALQ

A responsabilidade pela idoneidade, originalidade
e licitude dos conteúdos didáticos apresentados é do professor.

Proibida a reprodução, total ou parcial, sem autorização.

Lei nº 9610/98

Jonas Pereira Araujo 057.999.937-80

MBA USP ESALQ

Estrutura do banco de dados

- Estrutura tabular com as **observações** representadas nas linhas e as **variáveis** nas colunas

ID	Idade	Profissão	Renda Mensal	Estado (UF)	Escolaridade	...
Pessoa 1						
Pessoa 2						
Pessoa 3						
Pessoa 4						
Pessoa 5						
Pessoa 6						
Pessoa 7						
Pessoa n						

Jonas Pereira Araujo 057.999.037-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de variáveis

- As variáveis podem ser classificadas em:
  - **Qualitativas**: são variáveis não métricas, atribuem categorias ou classificações
    - Podem atribuir duas ou mais categorias
    - A análise descritiva de variáveis qualitativas é feita por meio de tabelas de frequência e gráficos, pois tais variáveis não permitem o cálculo de medidas de posição e dispersão
  - **Quantitativas**: são variáveis métricas, atribuem contagem ou mensuração
    - Podem ser discretas ou contínuas
    - A análise descritiva de variáveis quantitativas pode ser feita por diversas ferramentas estatísticas, incluindo as medidas de posição e dispersão

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de variáveis

- **Qualitativas**: exemplos
  - Faixa de renda
  - Nacionalidade
  - Estado civil
  - Escolaridade
  - Estação do ano
  - Cor do veículo
  - Crédito aprovado ou não
  - Escalas Likert

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de variáveis

- **Quantitativas**: exemplos

  - Idade (anos)
  - Renda (R$)
  - Quantidade de filhos
  - Altura da pessoa (cm)
  - Peso (kg)
  - Retorno de ações na bolsa (%)
  - Temperatura do ambiente (°C)
  - Lucro/prejuízo da empresa (R$)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Estatísticas Descritivas Univariadas

Jonas Pereira Arabio 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

1. Tabela de frequências

- Quantidade de ocorrências por categoria
  - **Qualitativas**
    - De forma direta, apresenta a quantidade de ocorrências para cada categoria
  - **Quantitativas**
    - Variável discreta: a análise assemelha-se ao caso da variável qualitativa, ou seja, mostra a quantidade de ocorrências para cada valor discreto da variável
    - Variável contínua: é necessária uma categorização inicial por classes ou faixas para, em seguida, apresentar a quantidade de ocorrências em cada categoria gerada

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

1. Tabela de frequências

- Elaborando uma tabela de frequências
  - Os tipos de frequências reportados podem ser:
    - **Frequência absoluta**: contagem de ocorrências em cada categoria
    - **Frequência relativa**: percentual de cada categoria em relação ao total de observações
    - **Frequência absoluta acumulada**: soma da frequência absoluta a cada nova categoria
    - **Frequência relativa acumulada**: soma da frequência relativa a cada nova categoria
  - **Exemplo:** Foram coletados dados sobre o país de origem de 300 pessoas que estavam em uma palestra. A tabela de frequências para a variável "país de origem" está disponível na planilha de suporte na **aba Tabela de Frequências**.

Jorge Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Média**
- **Mediana**
- **Moda**
- **Percentis**
- **Quartis**
- **Decis**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Média**
  - É média aritmética simples para a variável, ou seja, é a soma dos valores (X_i) contidos na variável dividido pela quantidade total de observações (n)
  - $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Mediana**
  - É o elemento central da distribuição da variável, considerando que a variável esteja com seus *n* valores organizados de forma crescente
  - Metade dos valores da variável são maiores ou iguais ao valor da mediana e metade dos valores são menores ou iguais ao valor da mediana

- $$Md(X) = \begin{cases} \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} \rightarrow \textbf{se n for par} \\ X_{\left(\frac{n+1}{2}\right)} \rightarrow \textbf{se n for impar} \end{cases}$$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Moda**
  - É o valor que aparece com maior frequência nas observações de uma variável
  - É possível que não exista a moda de uma variável (especialmente, se for uma variável contínua)
    - Ocorre quando nenhum valor se repete
  - A moda também pode ser interpretada em dados qualitativos, ou seja, é a categoria mais frequente

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Percentis**
  - São os elementos da distribuição da variável que dividem as observações em **cem partes iguais**, considerando que a variável esteja com seus valores organizados de forma crescente
    - 14º Percentil
    - 42º Percentil
    - 60º Percentil ...
  - $$\boldsymbol{Pos(P_i) = \left[(n-1).\left(\frac{P_i}{100}\right)\right] + 1}$$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Quartis**
  - São os elementos da distribuição da variável que dividem as observações em **quatro partes iguais**, considerando que a variável esteja com seus valores organizados de forma crescente
    - 1º Quartil: 25% das observações são menores do que o 1º quartil
    - 2º Quartil: trata-se da mediana
    - 3º Quartil: 25% das observações são maiores do que o 3º quartil
    - 1º Quartil = 25º Percentil
    - 2º Quartil = 50º Percentil
    - 3º Quartil = 75º Percentil

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

2. Medidas de posição

- **Decis**
  - São os elementos da distribuição da variável que dividem as observações em **dez partes iguais**, considerando que a variável esteja com seus valores organizados de forma crescente
    - 1º Decil
    - 3º Decil
    - 8º Decil ...
    - 1º Decil = 10º Percentil
    - 3º Decil = 30º Percentil
    - 8º Decil = 80º Percentil

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Amplitude**
- **Amplitude Interquartil**
- **Variância**
- **Desvio padrão**
- **Erro padrão**
- **Coeficiente de variação**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

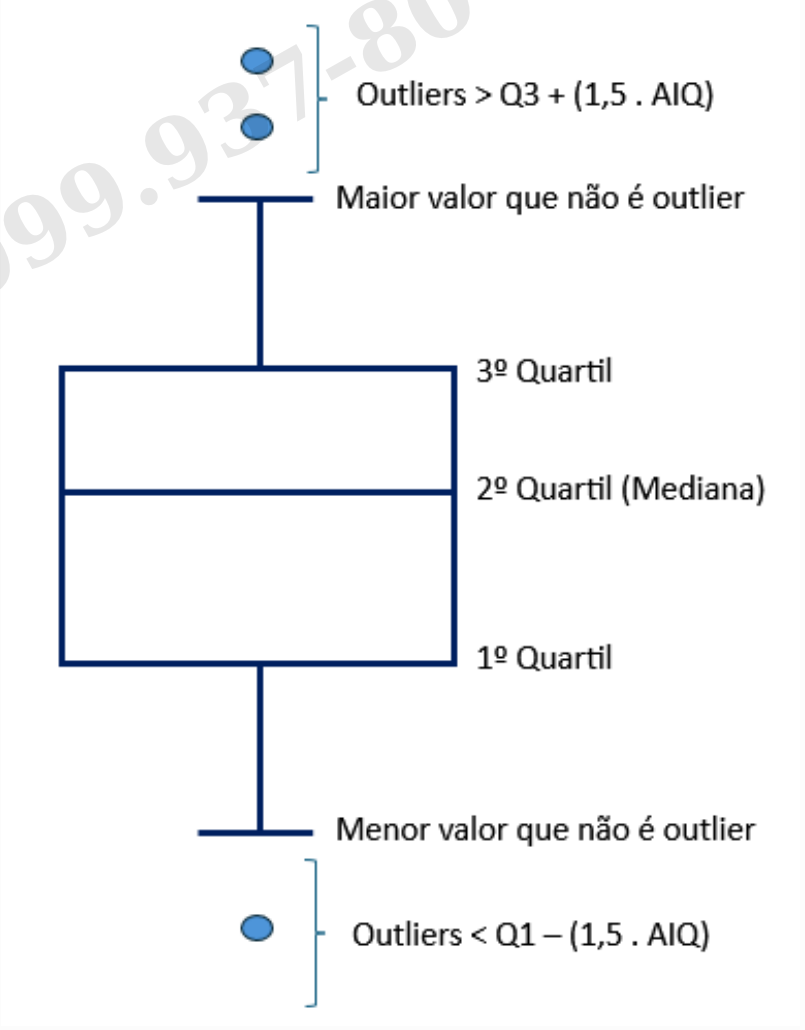
- **Amplitude**
  - Apresenta a diferença entre o valor máximo e o valor mínimo de uma variável
  - $A = X_{max} - X_{min}$
    - Valor máximo: maior valor da variável
    - Valor mínimo: menor valor da variável

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Amplitude Interquartil**
  - Mostra a diferença entre o terceiro e o primeiro quartil
  - $AIQ = Q_3 - Q_1$
  - Representada no gráfico chamado de *boxplot*
  - Utilizada para identificar valores extremos univariados
  - $i_{outlier} < Q_1 - 1,5 \, . \, AIQ$
  - $i_{outlier} > Q_3 + 1,5 \, . \, AIQ$



Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Variância**
  - Mostra a dispersão das observações de uma variável em torno de sua média
  - $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
    - Neste caso, trata-se da variância amostral

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Desvio padrão**
  - É uma medida proveniente da variância, tornando mais simples a interpretação da dispersão
    - A variância é definida em termos quadrados, o que dificulta a interpretação
  - O desvio padrão é a raiz quadrada da variância
  - $S = \sqrt{S^2}$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Erro padrão**
  - É o desvio padrão da média da variável
  - $S_{\overline{X}} = \frac{S}{\sqrt{n}}$
    - Sendo que *S* é o desvio padrão da variável e *n* o tamanho da amostra
  - Quanto maior o tamanho da amostra, menor o erro padrão na estimativa da média da variável → mais precisa é a média estimada

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

3. Medidas de dispersão

- **Coeficiente de variação (CV)**
  - É uma medida de dispersão relativa, pois relaciona o desvio padrão e a média da variável
  - Pode ser utilizada para realizar comparações entre amostras, por exemplo
  - Quanto menor o CV, mais homogêneos são os valores da variável e mais concentrados estão os valores em torno da média
  - $CV = \frac{S}{\bar{X}} . 100$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

4. Medidas de forma

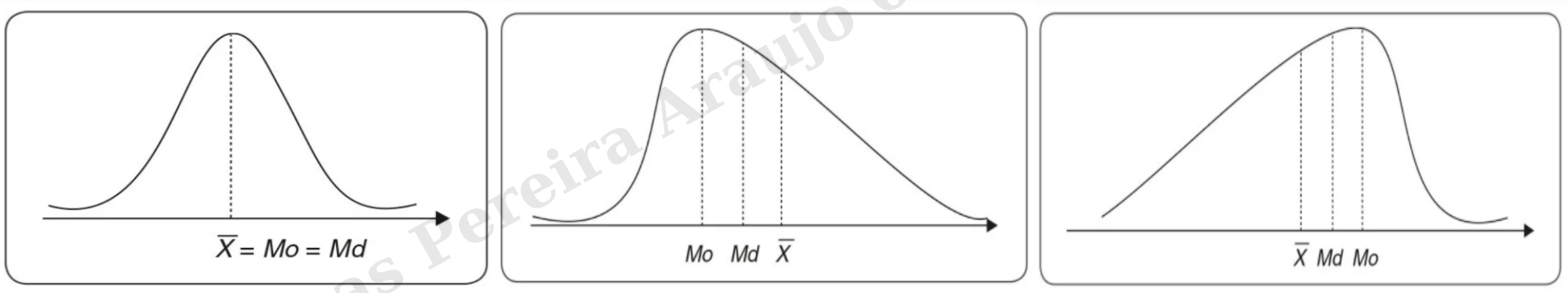
- **Assimetria e Curtose**
  - Assimetria: local de concentração da distribuição
    - Curva Simétrica: **Média = Mediana = Moda**
    - Curvas Assimétricas – Direta: tem cauda mais longa à direita → **Média > Mediana**
    - Curvas Assimétricas – Esquerda: tem cauda mais longa à esquerda → **Média < Mediana**
  - Coeficiente de Assimetria de Fisher:
  - $g_1 = \frac{n^2 \,.\, M_3}{(n-1).(n-2).S^3}$ em que $M_3 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n}$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

4. Medidas de forma

- **Assimetria**
  - Simétrica, Assimétrica à Direita e Assimétrica à Esquerda (respectivamente)



Fonte: Fávero e Belfiore (2024, Cap. 2)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

4. Medidas de forma

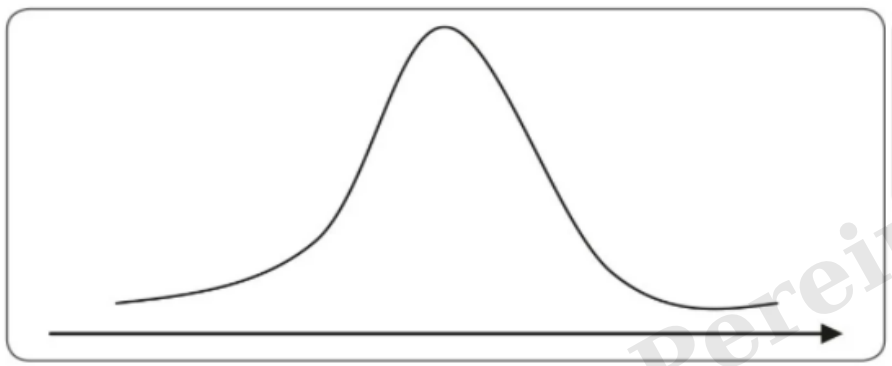
- **Assimetria e Curtose**
  - Curtose: achatamento da curva de distribuição
    - Com a curva normal como referência, é possível observar se as curvas têm menor dispersão em torno da média ou maior dispersão em torno da média
  - Coeficiente de curtose de Fisher:
  - $g_2 = \frac{n^2 \,.\, (n+1) \,.\, M_4}{(n-1).(n-2).(n-3).S^4} - 3 \,.\, \frac{(n-1)^2}{(n-2).(n-3)}$ em que $M_4 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n}$

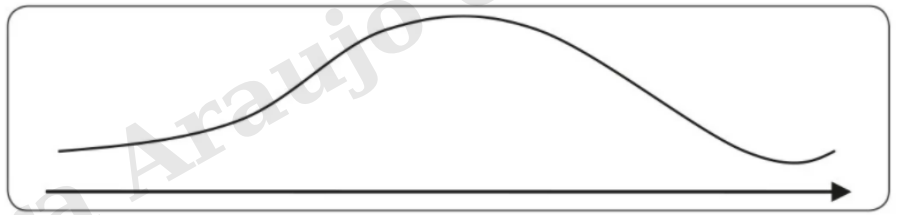
Jonas Pereira Aratijo 057.999.937-80

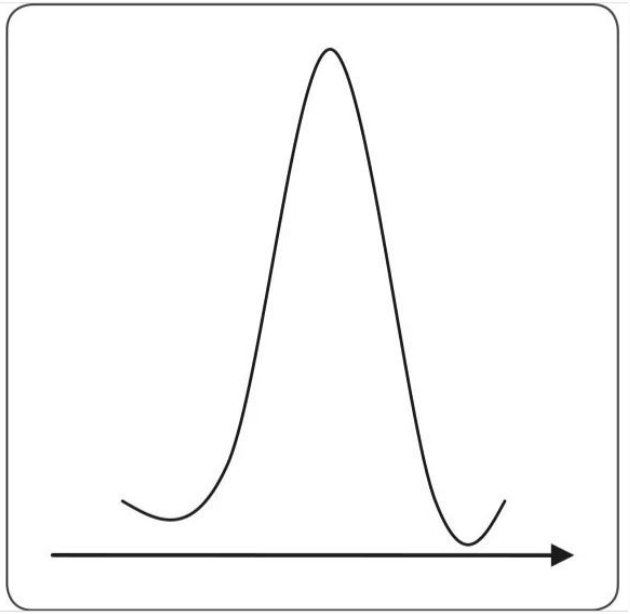
A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

4. Medidas de forma

- **Curtose**
  - Mesocúrtica, platicúrtica e leptocúrtica (respectivamente)







Fonte: Fávero e Belfiore (2024, Cap. 2)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

4. Medidas de forma

- **Assimetria e Curtose**
  - **Assimetria**
    - $g_1 = 0$ indica curva simétrica
    - $g_1 > 0$ indica curva assimétrica positiva (à direita)
    - $g_1 < 0$ indica curva assimétrica negativa (à esquerda)
  - **Curtose**
    - $g_2 = 0$ indica curva com distribuição normal
    - $g_2 > 0$ indica curva com distribuição alongada
    - $g_2 < 0$ indica curva com distribuição achatada

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Posição, dispersão e forma

- **Aplicação conjunta das medidas**

  - **Exemplo:** Um consumidor está analisando o preço de um produto que deseja comprar. Para gerar mais informações para sua tomada de decisão, ele coleta 100 preços praticados para o produto. Como o "preço" é variável quantitativa, serão realizadas as análises por meio das medidas de posição, dispersão e forma abordadas anteriormente.

    - O banco de dados com as 100 observações de preço está na planilha suporte **na aba Descritivas – Quantitativa**.

Jonas Peruchi Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Distribuições de Probabilidades

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis e as distribuições

- **Introdução**
  - Anteriormente, foram abordadas diversas estatísticas descritivas utilizadas para resumir as informações contidas nas variáveis amostradas
  - Adicionalmente, é fundamental conhecer as distribuições de probabilidades teóricas, porque, com base nelas, são determinadas as probabilidades de ocorrência dos possíveis resultados associados a um fenômeno caracterizado por determinada distribuição
    - Exemplo: considere que a altura, em metros, de mulheres brasileiras adultas seja uma variável que é caracterizada pela distribuição normal. Conhecendo os parâmetros específicos desta distribuição é possível identificar a probabilidade de encontrar uma mulher brasileira adulta com altura maior do que 1,70m.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis e as distribuições

- **Introdução**
  - Cada distribuição de probabilidades é representada por uma função matemática que mostra a probabilidade de ocorrência de um possível resultado
  - Como trata-se de probabilidade, a soma de todos os valores associados à distribuição deve ser igual a 1 (100% de probabilidade – todas as possibilidades de resultados do fenômeno)
  - A probabilidade de ocorrência de um resultado deve estar entre 0 e 1 (inclusive)
    - Quanto mais próximo de 1, maior é probabilidade de ocorrência do resultado
    - Quanto mais próximo de 0, menor é a probabilidade de ocorrência do resultado

Jones Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis e as distribuições

- **Distinção entre o tipo de variável**
  - **Variável aleatória discreta**: não assume valores decimais, isto é, é composta exclusivamente de valores inteiros e finitos
    - Exemplos: número de filhos; quantidade de pacientes por dia; quantidade de mudas por hectare; quantidade de casos da doença por dia, quantidade de produtos vendidos no mês...
  - **Variável aleatória contínua**: assume quaisquer valores contidos no intervalo dos números reais, portanto, inclui os valores decimais (medida em escala contínua)
    - Exemplos: distância ente cidades; salário mensal; altura da pessoa; retorno da ação na bolsa de valores, temperatura do ambiente...

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuições de probabilidade:**
  - **Uniforme**
  - **Bernoulli**
  - **Binomial**
  - **Binomial negativa**
  - **Poisson**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição uniforme discreta**
  - **Característica**: todos os possíveis resultados têm a mesma probabilidade de ocorrência
  - $\boldsymbol{P(X = x_i) = \frac{1}{n}}$
  - O parâmetro *n* representa a quantidade de possíveis resultados

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição uniforme discreta**
  - **Exemplo:** As probabilidades dos resultados possíveis ao lançar um dado são:
    - Como n = 6 (lados do dado)
    - **P(X = 1)** = 1/6
    - **P(X = 2)** = 1/6
    - **P(X = 3)** = 1/6
    - **P(X = 4)** = 1/6
    - **P(X = 5)** = 1/6
    - **P(X = 6)** = 1/6

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição de Bernoulli**
  - **Característica**: os valores da variável podem assumir apenas dois resultados possíveis, sendo que tais resultados são chamados de sucesso ($x = 1$) ou fracasso ($x = 0$)
  - A distribuição de Bernoulli apresenta a probabilidade de sucesso (p) ou de fracasso ($1 - p$) quando ocorre apenas um experimento
  - $\boldsymbol{P(X = x) = p^x \cdot (1 - p)^{1-x}}$

Regressão Logística Binária!

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição de Bernoulli**

  - **Exemplo:** A probabilidade (p) de que um candidato seja aprovado (x = 1) em um exame para um conselho de classes é de 48%. Se cada candidato só pode realizar o exame uma única vez, analise as probabilidades possíveis por meio da distribuição de Bernoulli.

    - Sendo que x = 1 é aprovação e x = 0 é a reprovação no exame:

    - **P(X = 1)** = $(0,48)^1 . (1 – 0,48)^0 = 0,48$ ou 48%
    - **P(X = 0)** = $(0,48)^0 . (1 – 0,48)^1 = 0,52$ ou 52%

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição binomial**
  - **Característica**: a distribuição binomial ocorre quando há (*n*) repetições independentes do experimento de Bernoulli e a probabilidade de sucesso (*p*) é constante em todas repetições
  - A variável no modelo binomial indica a quantidade de sucessos (*k*) nas (*n*) repetições do experimento
  - $\boldsymbol{P(X = k) = \binom{n}{k} . p^k . (1-p)^{n-k}}$ em que $\boldsymbol{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$

Regressão Logística Multinomial!

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição binomial**

  - **Exemplo:** Em uma indústria, sabe-se que a probabilidade (*p*) de encontrar peças defeituosas em cada lote produzido é 6,50%. São produzidos 12 lotes (*n*) da peça por mês. Analise as seguintes probabilidades (*k*):

    a) Qual a probabilidade de encontrar peças defeituosas em 2 lotes no mês?
    b) Qual a probabilidade de encontrar peças defeituosas em 4 lotes no mês?
    c) Qual a probabilidade de encontrar peças defeituosas em no máximo 2 lotes?

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição Binomial**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição binomial negativa**

  - **Característica**: na distribuição binomial negativa, são realizados (*x*) ensaios independentes de Bernoulli até que sejam obtidos (*k*) sucessos. A probabilidade de sucesso (*p*) é constante em todos os ensaios realizados

  - A variável no modelo binomial negativa indica a quantidade de ensaios (*x*) necessários para que sejam obtidas uma quantidade (*k*) fixa e pré-determinada de sucessos

  - $\boldsymbol{P(X = x) = \binom{x-1}{k-1} . p^k . (1-p)^{x-k}}$ em que $\boldsymbol{\binom{x-1}{k-1} = \frac{(x-1)!}{(k-1)![(x-1)-(k-1)]!}}$

Regressão Dados de Contagem!

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição binomial negativa**

  - **<u>Exemplo</u>:** Em um parque de diversões, existe uma máquina em que o jogador deve capturar algum item utilizando os comandos de um braço mecânico. Considere que a probabilidade (p) de que o jogador consiga capturar algum item em cada jogada é 11%. Identifique as seguintes probabilidades:

    a) De que o jogador necessite de 10 jogadas para capturar 3 itens.
    b) De que o jogador necessite de 20 jogadas para capturar 3 itens.
    c) De que o jogador necessite de 5 jogadas para capturar 1 item.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição Binomial Negativa**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição Poisson**
  - **Característica**: a distribuição Poisson indica a probabilidade do número de sucessos (*k*) em uma determinada exposição contínua
  - Exemplos de exposição: tempo e área
  - $P(X = k) = \frac{e^{-\lambda} \cdot \lambda^{k}}{k!}$
    - Em que λ é a taxa média estimada de ocorrência do evento (sucesso) em dada exposição

Regressão Dados de Contagem!

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis discretas

- **Distribuição Poisson**

  - **<u>Exemplo</u>:** Um médico analisou que a taxa média de ocorrência (λ) de pacientes com certa doença rara em seu consultório é de 2 por ano. Aceitando que a variável siga a distribuição Poisson, determine:

    a) A probabilidade de que o médico receba 1 paciente com a doença em um ano.
    b) A probabilidade de que o médico receba 3 pacientes com a doença em um ano.
    c) A probabilidade de que o médico receba 4 pacientes com a doença em um ano.
    d) A probabilidade de que o médico receba 6 pacientes com a doença em um ano.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição Poisson**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuições de probabilidade:**
  - **Normal (Normal Padrão)**
  - **Qui-quadrado**
  - ***t de Student***
  - **F de Snedecor**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

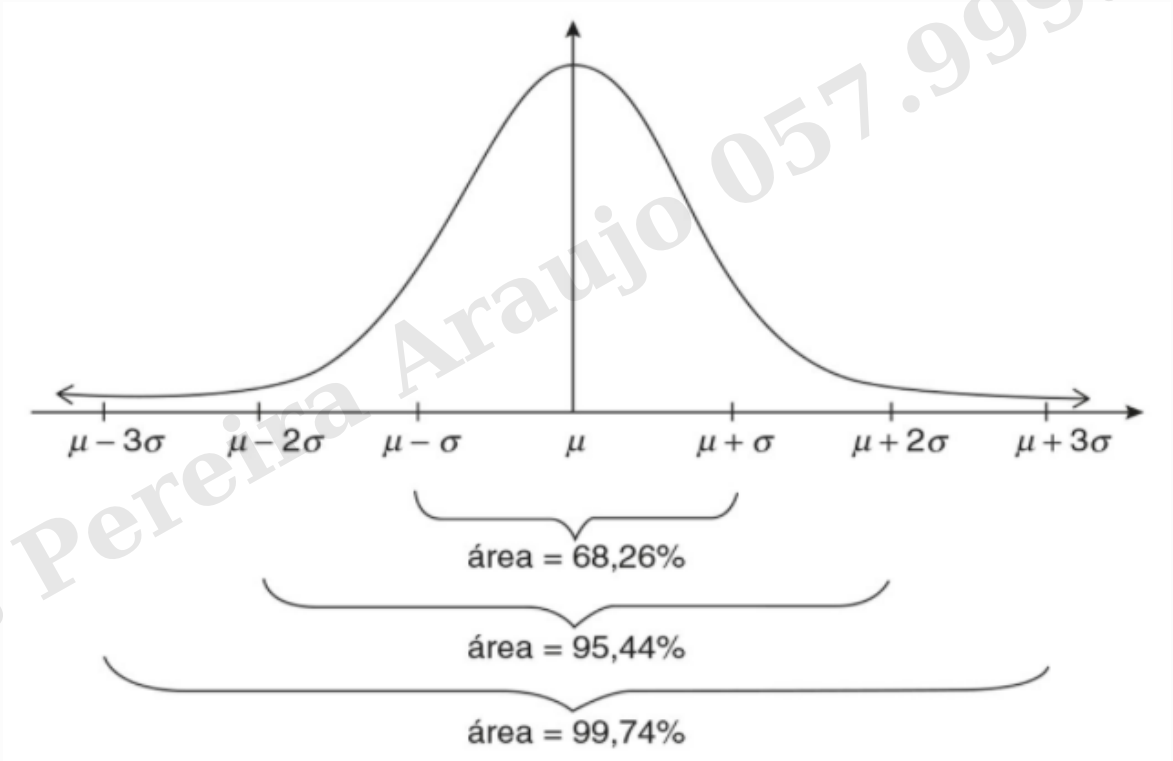
- **Distribuição normal**
  - **Característica**: é a distribuição Gaussiana, tem curva em formato de sino e é simétrica em torno da média. Os parâmetros relevantes da normal são a média (μ) e o desvio padrão (σ)
  - $$f(x) = \frac{1}{\sigma . \sqrt{2\pi}} . e^{-\frac{(x-\mu)^2}{2 . \sigma^2}}$$
    - Quanto menor for o desvio padrão, mais concentrados estão os valores em torno da média

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição normal**



Fonte: Fávero e Belfiore (2024, Cap. 5)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

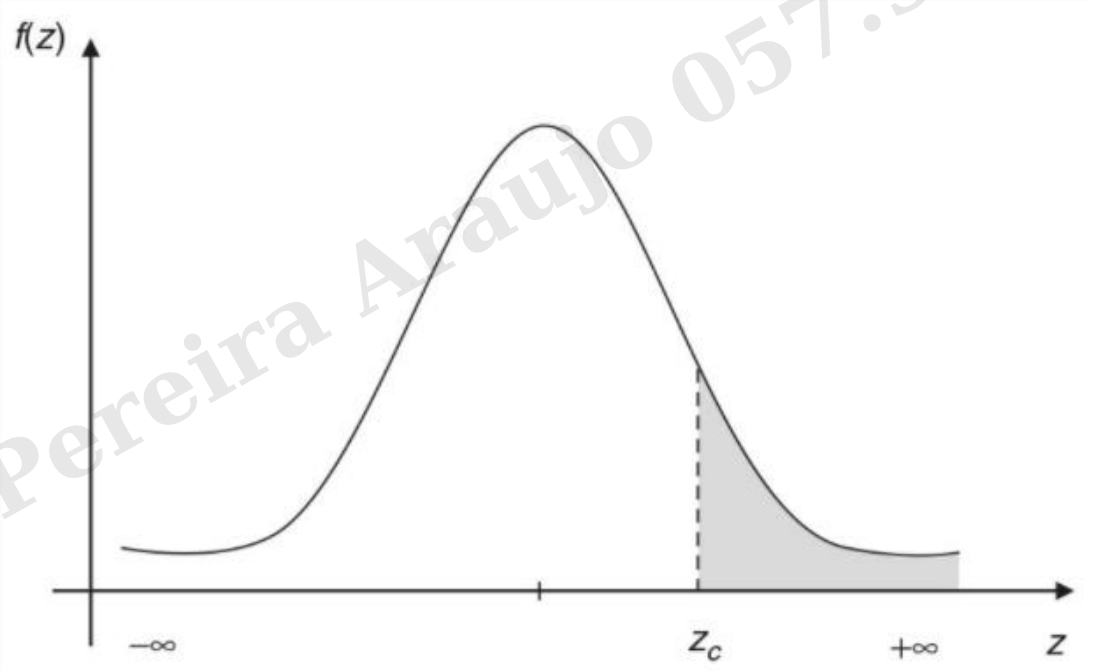
- **Distribuição normal padrão**
  - Para obter a normal padrão, transforma-se a variável por meio do Z-score
  - A transformação por meio do Z-score faz com que a variável passe a ter média = 0 e desvio padrão = 1 e não altera a distribuição original
  - $$Z = \frac{(X - \mu)}{\sigma}$$

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição normal padrão**



Fonte: Fávero e Belfiore (2024, Cap. 5)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição normal padrão**

  - **Exemplo:** Um investidor calculou que o retorno médio mensal de uma ação na bolsa de valores foi 2,80%. No mesmo período, o desvio padrão dos retornos da ação foi de 1,20%. Com base na distribuição normal, calcule:

    a) A probabilidade de que o retorno da ação seja maior do que 4% ao mês.
    b) A probabilidade de que o retorno da ação seja menor do que 3% ao mês.
    c) A probabilidade de que o retorno da ação seja negativo.
    d) A probabilidade de que o retorno da ação seja maior que 1% e menor que 5% ao mês.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição Normal**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

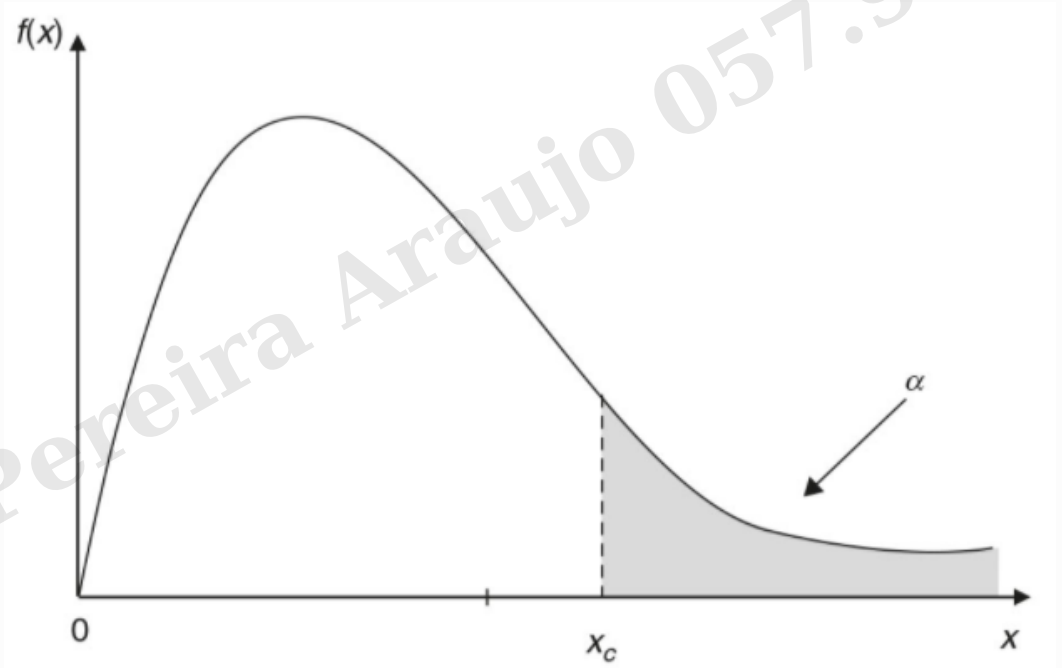
- **Distribuição qui-quadrado (χ^2)**
  - **Característica**: a distribuição χ^2 apresenta curva com forma influenciada pelos graus de liberdade
    - Para valores mais baixos nos graus de liberdade ela é assimétrica e positiva
    - Conforme os graus de liberdade aumentam, a qui-quadrado vai se tornando mais simétrica e, portanto, mais semelhante à curva normal
  - Uma aplicação da qui-quadrado é o teste de associação entre pares de variáveis categóricas

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição qui-quadrado (χ^2)**



Fonte: Fávero e Belfiore (2024, Cap. 5)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição qui-quadrado (χ^2)**

  - **Exemplo:** Um pesquisador em botânica identificou que um teste estatístico de seu estudo deve ser avaliado com base na distribuição qui-quadrado, sendo 5 graus de liberdade em seu teste. Neste contexto, avalie o seguinte:

    a) A probabilidade de que encontre um valor X>6.
    b) A probabilidade de que encontre um valor X<8.
    c) O valor de x que faz com que P(X>x) = 5%.
    d) O valor de x que faz com que P(X<x) = 90%.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição Qui-Quadrado**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição *t de Student***

  - **Característica**: a distribuição *t de Student* é parecida com a distribuição normal padrão, isto é, tem formato de sino e é simétrica em torno da média. Porém, a distribuição *t de Student* tem caudas mais longas e isso permite valores mais extremos

  - Tem sua forma influenciada pelos graus de liberdade; conforme os graus de liberdade vão aumentando, a distribuição t de *Student* vai se aproximando da normal

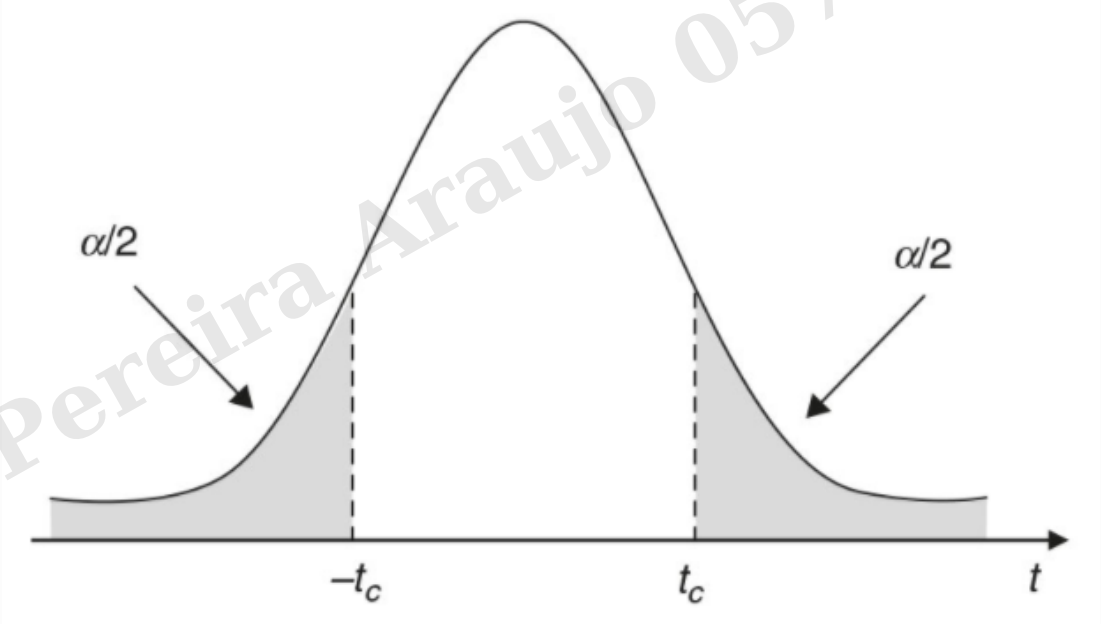
  - Uma aplicação da distribuição *t de Student* é o teste de médias, comumente utilizada para trabalhar com pequenas amostras (pois tem caudas mais longas e evita que se subestime a variabilidade dos dados provenientes de pequenas amostras)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição *t de Student***



Fonte: Fávero e Belfiore (2024, Cap. 5)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição *t de Student***

  - **<u>Exemplo</u>:** O gestor do controle de qualidade de uma empresa está realizando uma análise com base na distribuição t de Student. Em seu teste, há 7 graus de liberdade. Quais são os resultados nas seguintes situações:

    a) A probabilidade de que encontre T > 2,5.
    b) A probabilidade de que encontre T < - 2,5.
    c) A probabilidade de que encontre T>-1 e T<2.
    d) O valor de t para que P(T>t) = 5%.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição t Student**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

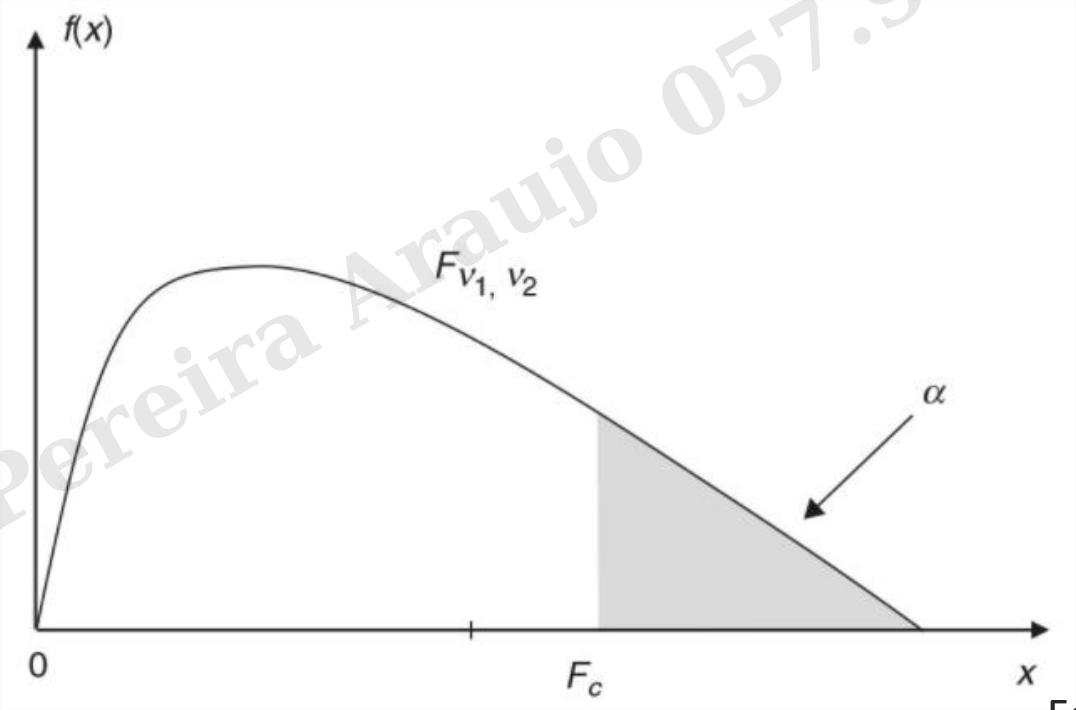
- **Distribuição F de Snedecor (distribuição de Fischer)**
  - **Característica**: distribuição de probabilidades que trabalha com razões entre valores, assim tem sua forma influenciada pelos graus de liberdade no numerador e no denominador
  - É assimétrica e positiva quando os graus de liberdade são pequenos
  - Uma aplicação comum da distribuição F é a análise de razão entre variâncias

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição F de Snedecor**



Fonte: Fávero e Belfiore (2024, Cap. 5)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Variáveis contínuas

- **Distribuição F de Snedecor**

  - **Exemplo:** Um cientista de dados está realizando um teste estatístico que tem como base a distribuição F de Snedecor. No teste, há 17 graus de liberdade no numerador e 28 graus de liberdade no denominador. Determine:

    a) A probabilidade de que encontre $X > 1,5$.
    b) A probabilidade de que encontre $X < 1,0$.
    c) A probabilidade de que encontre $2<X<3$.
    d) O valor de x para que $P(X>x) = 5\%$.

  - A planilha para auxílio está na planilha de suporte na **aba Distribuição F de Snedecor**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Graus de liberdade

- **O que são os graus de liberdade?**
  - As distribuições qui-quadrado, t de Student e F de Snedecor têm sua forma e, portanto, seus valores alterados em função dos graus de liberdade
  - "Graus de liberdade para variar": é a quantidade de observações da amostra que pode variar de forma independente e aleatória e ainda assim obter o valor em análise
    - Cada teste estatístico tem um cálculo específico dos graus de liberdade, ou seja, não é padrão para todos os testes. Normalmente, leva-se em consideração o tamanho da amostra e a quantidade de parâmetros estimados

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Graus de liberdade

- **O que são os graus de liberdade?**
  - Exemplo
    - Considere uma amostra de 5 números cuja soma deve ser igual a 18. Os 4 primeiros elementos da amostra podem ser escolhidos aleatoriamente, mas o 5º elemento deve ser um valor específico → o valor que tornará a soma = 18
    - {1; 3; 5; 7; _}: dado que os valores 1, 3, 5 e 7 foram selecionados, o 5º elemento deve ser obrigatoriamente igual a **2** para atender à restrição da soma
    - Portanto, há **4 graus de liberdade (poderia ser escrito como n-1 graus de liberdade, onde n é o tamanho da amostra)**

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Graus de liberdade

- **Qual a relação dos graus de liberdade com os testes estatísticos?**
  - Analisando as distribuições de probabilidade notaremos que:
    - Para uma dada probabilidade (área sob a curva), os valores críticos da distribuição se alteram em função das alterações nos graus de liberdade
    - Isto significa que os graus de liberdade do teste estatístico são fundamentais, pois irão impactar o teste de hipótese (conforme veremos a seguir)
    - Influencia o valor crítico que será utilizado como base de comparação para a rejeição ou não da hipótese nula do teste

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Introdução aos Testes de Hipóteses

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Finalidade**
  - Em certos casos, podemos estar interessados em testar características sobre parâmetros populacionais como a média e o desvio padrão (variância)
  - Dada a impossibilidade de obter os dados sobre a população, utilizamos as amostras da população
  - Para testar o parâmetro de interesse por meio de amostras, utilizamos os testes de hipóteses estatísticas

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

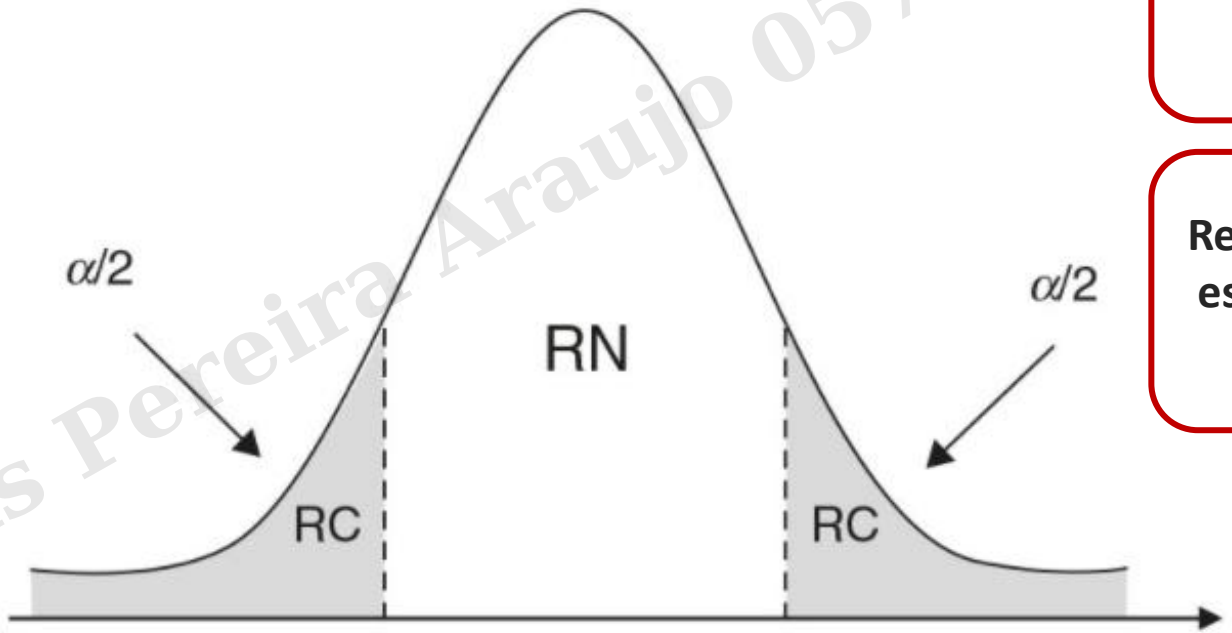
- **Teste bilateral (bicaudal)**
  - No teste bilateral, para um parâmetro θ, o interesse é testar:
    - H_0: $\theta = \theta_0$ (hipótese nula)
    - **H_1: $\theta \neq \theta_0$ (hipótese alternativa)**
  - O objetivo é verificar se o parâmetro é **estatisticamente diferente** do valor de interesse (θ_0)
  - É necessário definir o **nível de significância (α)** desejado para a análise

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

- **Teste bilateral (bicaudal)**



Região Crítica (RC): se a estatística do teste cair em RC, a hipótese nula é rejeitada

Região de Não Rejeição (RN): se a estatística do teste cair em RN, a hipótese nula não é rejeitada

Fonte: Fávero e Belfiore (2024, Cap. 7)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

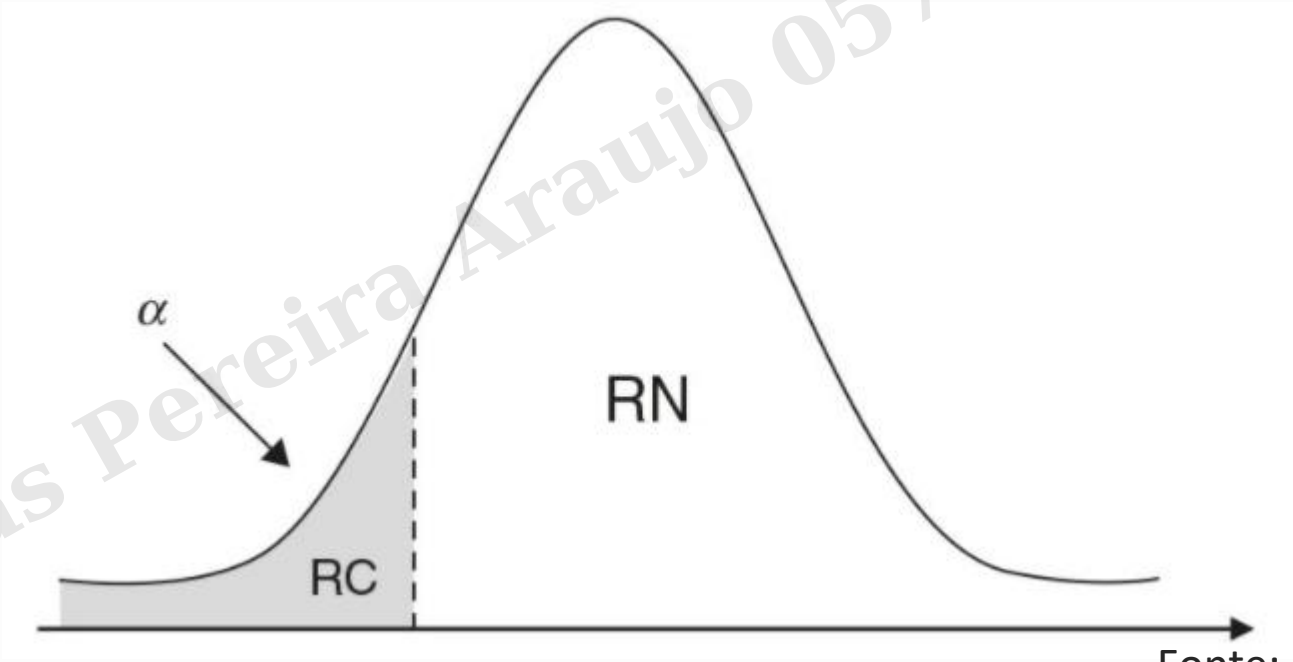
- **Teste unilateral à esquerda**
  - No teste unilateral à esquerda, para um parâmetro θ, o interesse é testar:
    - H_0: $\theta = \theta_0$ (hipótese nula)
    - **H_1: $\theta < \theta_0$ (hipótese alternativa)**
  - O objetivo é analisar se o parâmetro é **estatisticamente menor** do valor de interesse (θ_0)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

- **Teste unilateral à esquerda**



Fonte: Fávero e Belfiore (2024, Cap. 7)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

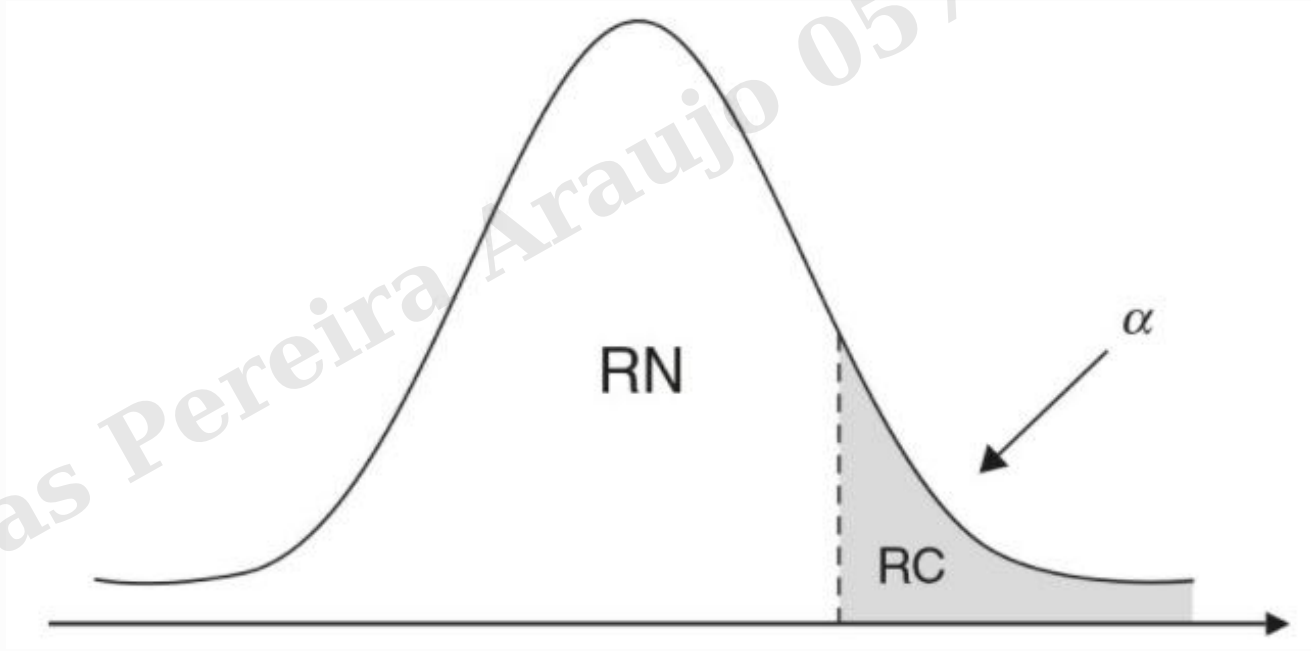
- **Teste unilateral à direita**
  - No teste unilateral à direita, para um parâmetro θ, o interesse é testar:
    - H_0: $\theta = \theta_0$ (hipótese nula)
    - **H_1: $\theta > \theta_0$ (hipótese alternativa)**
  - O objetivo é verificar se o parâmetro é **estatisticamente maior** do valor de interesse (θ_0)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de testes

- **Teste unilateral à direita**



Fonte: Fávero e Belfiore (2024, Cap. 7)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de erros

- **Possíveis erros na tomada de decisão**
  - **Erro tipo I:** rejeitar a hipótese nula (H_0) quando ela é verdadeira
  - **Erro tipo II:** não rejeitar a hipótese nula (H_0) quando ela é falsa
- As decisões corretas são:
  - Rejeitar H_0 quando ela é falsa
  - Não rejeitar H_0 quando ela é verdadeira

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Tipos de erros

- **Possíveis erros na tomada de decisão**

		H_0 é Verdadeira	H_0 é Falsa
Decisões	Não Rejeitar H_0	Correto	Erro Tipo II
	Rejeitar H_0	Erro Tipo I (α)	Correto

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Significância do teste

- **Nível de significância do teste**
  - O nível de significância (α) indica a probabilidade de rejeitar H_0 quando ela é verdadeira, ou seja, a probabilidade de cometer o erro tipo I
  - Alguns níveis de significância recorrentemente utilizados:
    - α = 1%
    - α = 5%
    - α = 10%
  - O nível de confiança do teste é definido como 1 – α

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

P-valor e teste de hipótese

- **P-valor e nível de significância**
  - Uma forma de testar estatisticamente a hipótese é comparando o valor da estatística calculada nos dados com o valor crítico para o nível de significância
  - Também é possível obter o p-valor para a estatística calculada e, em seguida, compará-lo ao nível de significância escolhido
    - Se p-valor < nível de significância (α) → **rejeita-se H_0**
    - Se p-valor > nível de significância (α) → **não rejeita H_0**
  - **O p-valor é a probabilidade associada ao valor da estatística de teste calculada**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Aplicações dos Testes de Hipóteses

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Estatísticas descritivas bivariadas

- **Medidas para análises bivariadas**
  - Muitas vezes, o interesse pode estar na relação entre duas variáveis. Nestes casos, antes, é importante conhecer o tipo de variável:
    - **Variáveis qualitativas**: análise da **associação** pelo teste qui-quadrado (χ^2)
    - **Variáveis quantitativas**: análise da **correlação** por meio da covariância e coeficiente de correlação de Pearson

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**
  - Inicia-se com uma tabela de distribuição conjunta de frequências, a tabela de contingência (tabela de classificação cruzada) que apresenta as **frequências absolutas observadas** para cada par de categorias das variáveis

		Variável B					
		Categoria 1	**Categoria 2**	**Categoria 3**	**...**	**Categoria J**	**Total**
Variável A	**Categoria 1**	n_{11}	n_{12}	n_{13}	...	n_{1J}	Σ_{L1}
	Categoria 2	n_{21}	n_{22}	n_{23}	...	n_{2J}	Σ_{L2}
	Categoria 3	n_{31}	n_{32}	n_{33}	...	n_{3J}	Σ_{L3}
	...	...	...	...	...	...	...
	Categoria I	n_{I1}	n_{I2}	n_{I3}	...	n_{IJ}	Σ_{LI}
	Total	Σ_{C1}	Σ_{C2}	Σ_{C3}	...	Σ_{CJ}	N

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**
  - Em seguida, são identificadas as **frequências absolutas esperadas** para cada par de categorias das variáveis
  - $$\boldsymbol{freq.absoluta\ esperada_{11}} = \frac{(\sum \boldsymbol{L1} \,.\, \sum \boldsymbol{C1})}{\boldsymbol{N}}$$
    - O mesmo cálculo é realizado para cada par de categorias da tabela de contingência, mantendo-se as respectivas correspondências de linha e coluna no numerador

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**
  - Posteriormente, são identificados os **resíduos** para cada par de categorias das variáveis
  - $\boldsymbol{resíduo_{11} = freq.absoluta\ observada_{11} - freq.absoluta\ esperada_{11}}$
    - Os resíduos são identificados para cada par de categorias da tabela de contingência

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**
  - Por fim, são calculados os valores χ^2 individuais de cada par de categorias
  - $$\chi^2_{11} = \frac{(resíduo_{11})^2}{(freq.\ absoluta\ esperada_{11})}$$
  - E são somados valores χ^2 individuais para obter o valor χ^2 total da análise
  - **O χ^2 total é a estatística de teste utilizada para avaliar a significância da associação entre as variáveis qualitativas em análise**

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**
  - Com base no valor χ^2 total, realiza-se o seguinte teste:
    - H_0: as variáveis se associam de forma aleatória.
    - H_1: a associação entre as variáveis não se dá de forma aleatória.
  - Dados o nível de significância e os graus de liberdade, se o valor da estatística χ^2 for maior do que seu valor crítico, há associação significante entre as duas variáveis (H_1)
    - Valor crítico da distribuição χ^2 com (I – 1) . (J – 1) graus de liberdade

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis qualitativas

- **Teste qui-quadrado: variáveis qualitativas**

  - **<u>Exemplo</u>:** Um estudo foi realizado com 200 pessoas para analisar o comportamento conjunto das variáveis “operadora de plano de saúde” e o “nível de satisfação” do consumidor. O objetivo é analisar se existe a associação estatisticamente significativa entre tais variáveis. (Fonte: Fávero e Belfiore, 2024, Cap. 8)

    - Os dados da tabela de contingência obtida a partir da amostra está na planilha suporte na **aba Associação – Qui²**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis métricas

- **Coeficiente de correlação de Pearson**
  - É utilizado para identificar a correlação entre duas varáveis quantitativas
  - Inicia-se pelo cálculo da covariância entre as duas variáveis e, posteriormente, obtém-se o coeficiente de correlação de Pearson
  - $$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X}) \,.\, (Y_i - \bar{Y})}{n-1}$$
  - $r_{XY} = \frac{cov(X,Y)}{S_X \,.\, S_Y}$, sendo que r_{XY} varia entre -1 e 1

r_{XY} = -1 (perfeita negativa)
r_{XY} = 0 (sem correlação)
r_{XY} = 1 (perfeita positiva)

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis métricas

- **Teste *t* para correlações de Pearson**
  - Após estimado o coeficiente de correlação (*r*) entre duas variáveis quantitativas é possível testar a significância do parâmetro estimado
  - Estatística do teste:
  - $$t = \frac{r}{\sqrt{\frac{(1-r^2)}{(n-2)}}}$$
  - A distribuição relevante é a *t de student* com *n-2* graus de liberdade

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Relação entre variáveis métricas

- **Coeficiente de correlação de Pearson**

  - **<u>Exemplo</u>:** O coordenador de um curso deseja analisar se existe correlação entre as notas dos alunos em diferentes disciplinas. Para tanto, montou um banco de dados com as notas de 30 alunos para as disciplinas de matemática, física e literatura. Em seguida, deseja calcular os pares de correlações entre as notas de matemática – física, matemática – literatura e física – literatura. Quais são as correlações de Pearson? Os coeficientes de correlações obtidos para as amostras de notas são significantes ao nível de significância de 5%?

    - Os dados estão na planilha suporte na **aba Correlação de Pearson**.

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de Hipóteses

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste Z para médias de uma amostra**
  - Aplicado quando o desvio padrão populacional é conhecido e a distribuição da variável é normal (ou utilizando grandes amostras)
  - A estatística do teste é:
  - $$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
  - A distribuição relevante para os valores críticos é a normal padrão

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste Z para médias de uma amostra**

  - **<u>Exemplo</u>:** Um fabricante de caixas de papelão deseja verificar se a quantidade de papelão que está sendo utilizada em cada caixa do tipo 1 está de acordo com seu padrão histórico, pois existem indícios de que o consumo aumentou. Historicamente, são utilizados, em média, 100 g de papelão em cada caixa e o desvio padrão é de 12g. Coletou-se uma amostra para verificar se a média atual é maior do que a média histórica.

    - A amostra coletada está na planilha suporte na **aba Teste Z médias**.

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste *t* para médias de uma amostra**
  - Aplicado quando o desvio padrão populacional não é conhecido e, assim, é utilizado o desvio padrão amostral
  - Estatística do teste é semelhante ao Z, porém com desvio padrão amostral:
  - $$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$
  - A distribuição relevante é a *t de student* com *n-1* graus de liberdade

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste *t* para médias de uma amostra**

  - **Exemplo:** O tempo médio de processamento de determinada tarefa em uma máquina tem sido de 18 minutos. Foram introduzidos novos conceitos para reduzir o tempo médio de processamento. Desta forma, após certo período, coletou-se uma amostra de 25 elementos, obtendo-se o tempo médio de 16,808 minutos com desvio-padrão de 2,733 minutos. Verifique se esse resultado evidencia uma melhora no tempo médio de processamento. Considere α = 1%. (Fonte: Fávero e Belfiore, 2024, Cap. 7)

    - Os dados estão na planilha suporte na **aba Teste t médias**.

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste qui-quadrado para uma amostra**
  - Aplicado quando a variável assume duas ou mais categorias (K) e o objetivo é verificar se há diferenças entre as frequências observada (O) e esperada (E)
  - A estatística do teste é:
  - $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
  - A distribuição relevante é a qui-quadrado com *k-1* graus de liberdade

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste qui-quadrado para uma amostra**

  - **Exemplo:** Uma loja deseja verificar se a quantidade vendida em cada dia da semana varia em função do dia da semana. Os dados para as vendas em cada dia de uma semana escolhida aleatoriamente foram tabulados. Neste caso, o objetivo é testar se a frequência observada e esperada são iguais ou se são diferentes. (Fonte: Fávero e Belfiore, 2024, Cap. 8)

    - Os dados tabulados estão na planilha de suporte na **aba Teste Qui² Uma Amostra**.

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste F para comparação de variâncias**
  - Aplicado para comparar as variâncias de duas amostras independentes
  - A estatística do teste é:
  - $$F = \frac{S^2_{maior}}{S^2_{menor}}$$
  - A distribuição relevante é a F de Snedecor, com *n-1* graus de liberdade no numerador e *n-1* graus de liberdade no denominador

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste F para comparação de variâncias**

  - **<u>Exemplo</u>:** Uma empresa de logística está analisando qual entre duas rotas oferece a melhor previsibilidade para o horário de entrega ao seu maior cliente. Foram coletados dados sobre o tempo de entrega durante 35 dias para cada rota. O diretor de logística deseja testar a hipótese que a rota B tem maior variabilidade no tempo de entrega, comparativamente à rota A.

    - Os dados estão na planilha suporte na **aba Teste F Variâncias**.

Jonas Pereira Araújo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Intervalo de confiança

- **Intervalo de confiança para a média**
  - Quando obtemos a estimativa para a média populacional a partir de uma amostra, também podemos construir seu intervalo de confiança, isto é, um intervalo de valores possíveis para o parâmetro populacional
  - É necessário estabelecer o nível de confiança da análise (por exemplo, 95%)
  - $IC = (\overline{x} - Z \cdot \frac{\sigma}{\sqrt{n}} \ , \ \overline{x} + Z \cdot \frac{\sigma}{\sqrt{n}})$ **ou** $IC = \left(\overline{x} - t \cdot \frac{s}{\sqrt{n}} \ , \ \overline{x} + t \cdot \frac{s}{\sqrt{n}}\right)$

Grandes amostras / Variância conhecida **Pequenas amostras / Variância desconhecida**

  - **Z** e **t** são os valores bicaudais; na distribuição **t** utiliza-se *n-1* graus de liberdade

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Intervalo de confiança

- **Intervalo de confiança para a média**
  - **Exemplo:** Um engenheiro coletou uma amostra de 25 peças saídas da linha de montagem e encontrou que o tamanho médio foi de 47cm e o desvio padrão foi 1cm. Qual é o intervalo de confiança com 95% para esta média estimada?
    - Os dados estão na planilha de suporte na **aba Intervalo de Confiança – Média**.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste t para comparação de médias em duas amostras independentes**
  - Para comparar as médias de duas amostras independentes de uma mesma população por meio do teste t, antes é necessário comparar as variâncias populacionais dos dois grupos
    - Por exemplo, antes pode ser feito um teste F para comparação das variâncias
  - O cálculo da estatística t e graus de liberdade depende: se as variâncias populacionais forem diferentes ou se são homogêneas

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste t para comparação de médias em duas amostras independentes**
  - Estatística T para variâncias populacionais diferentes
  - $T = \frac{(\overline{X_1} - \overline{X_2})}{\sqrt{\frac{S^2 1}{n_1} + \frac{S^2 2}{n_2}}}$
  - Os graus de liberdade são $\nu = \frac{\left(\frac{S^2 1}{n_1} + \frac{S^2 2}{n_2}\right)^2}{\frac{\left(S^2 1 / n_1\right)^2}{(n_1 - 1)} + \frac{\left(S^2 2 / n_2\right)^2}{(n_2 - 1)}}$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste t para comparação de médias em duas amostras independentes**
  - Estatística T para variâncias populacionais homogêneas
  - $$T = \frac{(\overline{X_1} - \overline{X_2})}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
  - Em que $$S_p = \sqrt{\frac{(n_1 - 1).S^2{}_1 + (n_2 - 1).S^2{}_2}{n_1 + n_2 - 2}}$$ e os graus de liberdade são $n_1 + n_2 - 2$

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Testes de hipóteses

- **Teste t para comparação de médias em duas amostras independentes**

  - **Exemplo:** Em uma indústria, o gerente de produção fez um levantamento com 30 medições da temperatura (em °C) dos dois principais fornos da linha de produção que produzem os produtos do mesmo tipo. Destas, 15 medições foram do forno A e 15 medições foram para o forno B. O objetivo é verificar se a temperatura média está consideravelmente diferente entre os fornos.

    - Os dados estão na planilha suporte na **aba Teste t Duas Amostras Indep**.

Jonas Pereira Araújo 057.999.937-80

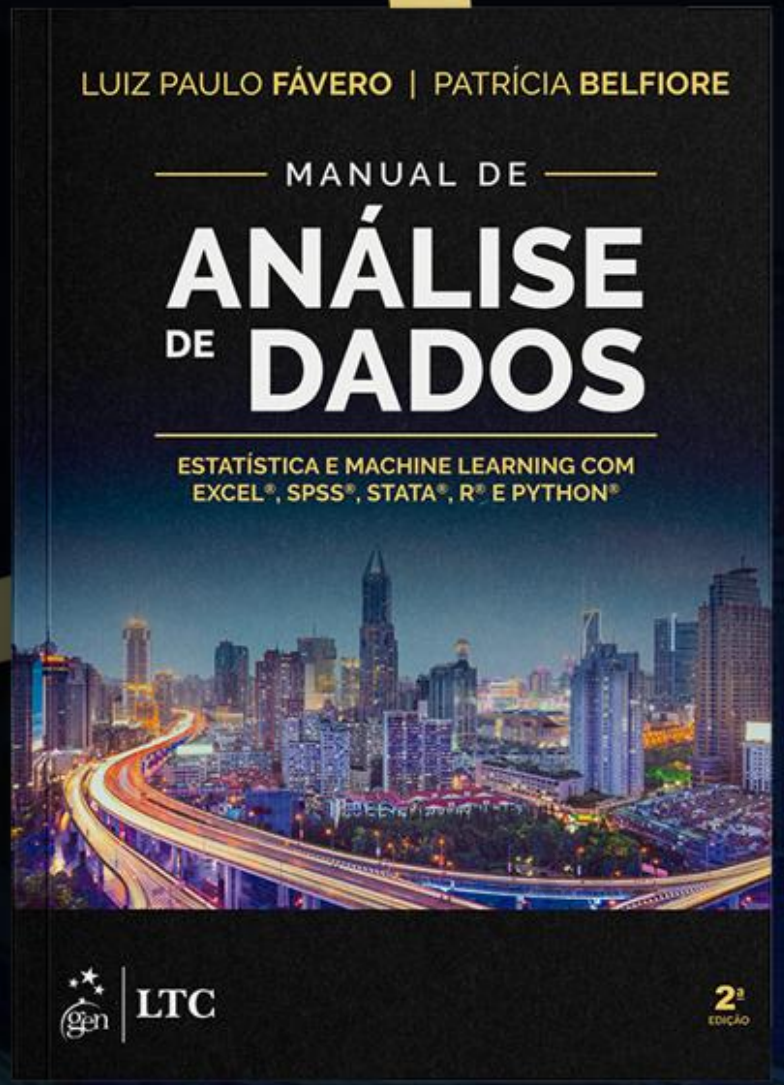
A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Sugestões de Leituras/Referências

- Fávero, Luiz Paulo; Belfiore, Patrícia. (2024). Manual de análise de dados: estatística e machine learning com Excel®, SPSS®, Stata®, R® e Python®. 2 ed. Rio de Janeiro: LTC.

Jonas Pereira Araujo 057.999.937-80

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98



35%* OFF

na compra do **livro impresso** ou **e-book** apenas no site do Grupo GEN

Use o cupom

FAVERO35

*Cupom válido sobre preço de capa até 31/10/2025.
Não cumulativo com outras promoções do site.

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.
Proibida a reprodução, total ou parcial, sem autorização. Lei nº 9610/98

Obrigado!

Wilson Tarantin Jr. | linkedin.com/in/wilson-tarantin-junior-359476190

Jonas Pereira Araujo 057.999.937-80

MBA USP ESALQ