

EDIT.

Módulo 7 – Sessão 1

QUANTITATIVE
STATISTICAL ANALYSIS
AND MODELING

TUTORA
Carla Cardoso
Freelancer AI Manager

18 de março 2025



1

ORGANIZAÇÃO DA SESSÃO

EDIT.



Alternar entre
conteúdos e
experimentação

Pausa para
jantar das
20:30 às 21:00





Quizz Final



2

AVALIAÇÃO DO MÓDULO

EDIT.



Assiduidade e Pontualidade
15%

Presença nas aulas



Motivação e Participação
5%

Participação nas aulas



Domínio dos Conteúdos das Aulas
30%

Questionários



Aplicação Técnica
20%

Exercícios Python



Trabalho Final
30%

Trabalhos

3

AGENDA

EDIT.



INTRODUÇÃO À ESTATÍSTICA



INTRODUÇÃO ÀS PROBABILIDADES



AMOSTRAGEM

4

AGENDA



INTRODUÇÃO À
ESTATÍSTICA



INTRODUÇÃO ÀS
PROBABILIDADES



AMOSTRAGEM

5

INTRODUÇÃO À ESTATÍSTICA - DEFINIÇÃO



Estatística é uma “**arte**” e uma **ciência** que permite tirar **conclusões** e fazer **inferências** a partir de um conjunto de **dados**.

Até 1900 a Estatística resumia-se ao que hoje chamamos **Estatística Descritiva**.

Nos anos seguintes foram-se desenvolvendo novos métodos e técnicas dedicados à inferência e, se olharmos para os trabalhos na área de estatística na **década de 60**, eles estão sobretudo dedicados ao tema da **Inferência Estatística**.

Nas últimas décadas, com o **crescimento da capacidade computacional**, entrámos numa era em que a atitude tem sido sobretudo “**Deixar os dados falar por si**”.



6

INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA



A **Estatística Descritiva** procura **evidenciar as principais características** e **propriedades dos dados** através da análise dos mesmos e recorrendo a técnicas especializados.

De modo gerar os dados são de elevada complexidade, e por isso requerem técnicas para mais facilmente podermos extrair informações uteis sobre os mesmos. A complexidade da população pode ser o resultado de:

- ❖ **Dimensão / características** da população (ex.: populações infinitas)
- ❖ **Número de variáveis** que descrevem a população (ex.: dados biométricos)
- ❖ **Diversidade da informação** (ex.: previsão de resultado eleições)



INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA



“Em média, cada cigarro que se fuma por dia, reduz o tempo de vida numa certa quantidade de tempo, a qual será estimada com base no estudo dos hábitos tabagísticos dos utentes fumadores registados no centro de saúde de Sete Rios.”

Que informação relevante podemos extrair da descrição deste estudo?



INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA



“Em média, cada cigarro que se fuma por dia, **reduz o tempo de vida** numa certa quantidade de tempo, a qual será estimada com base no estudo dos hábitos tabagísticos dos utentes fumadores registados no centro de saúde de Sete Rios.”

Variável de interesse: Esperança de vida



10

INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA



“**Em média**, cada cigarro que se fuma por dia, **reduz o tempo de vida** numa certa quantidade de tempo, a qual será estimada com base no estudo dos hábitos tabagísticos dos utentes fumadores registados no centro de saúde de Sete Rios.”

Variável de interesse: Esperança de vida

População: População geral



12

INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA

EDIT.

“Em média, cada cigarro que se fuma por dia, reduz o tempo de vida numa certa quantidade de tempo, a qual será estimada com base no estudo dos hábitos tabagísticos dos **utentes fumadores** registados no **centro de saúde de Sete Rios**.”

Variável de interesse: Esperança de vida

População: População geral

Amostra: Utentes fumadores registados no centro de saúde de Sete Rios



INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA

EDIT.

“Em média, cada cigarro que se fuma por dia, reduz o tempo de vida numa certa quantidade de tempo, a qual será estimada com base no estudo dos **hábitos tabagísticos** dos utentes fumadores registados no centro de saúde de Sete Rios.”

Variável de interesse: Esperança de vida

População: População geral

Amostra: Utentes fumadores registados no centro de saúde de Sete Rios

Características em estudo: Hábitos tabagísticos



INTRODUÇÃO À ESTATÍSTICA – ESTATÍSTICA DESCRITIVA

EDIT.

“Em média, cada cigarro que se fuma por dia, reduz o tempo de vida numa certa quantidade de tempo, a qual será estimada com base no estudo dos hábitos tabagísticos dos utentes fumadores registados no centro de saúde de Sete Rios.”

Que **problemas** apresenta este estudo?

Amostra não representativa

Características em estudo insuficientes



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

18

18

INTRODUÇÃO À ESTATÍSTICA – INFERÊNCIA ESTATÍSTICA

EDIT.

Ao contrário da **Estatística Descritiva** que procura salientar e descrever dados particulares, aos quais teve total acesso (ex.: amostra), a **Inferência Estatística** procura **generalizar as conclusões** obtidas na fase anterior e **testar certas hipóteses** levantadas.

Ainda que seja possível inferir determinadas **conclusões** (ex.: cada cigarro diariamente fumado reduz esperança de vida em 30 dias) essas conclusões **têm sempre um grau de incerteza**, visto que são baseadas em amostras.

Esse grau de incerteza é medido através de **probabilidades**.

Observamos X na amostra



Estimamos X na população, com determinada probabilidade



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

21

21


INTRODUÇÃO À ESTATÍSTICA – EXEMPLO

EDIT.

EXEMPLO

Universe / Population





Média da população: μ

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

22


22

INTRODUÇÃO À ESTATÍSTICA – EXEMPLO

EDIT.

EXEMPLO


Universe / Population



Média da população: μ

Amostragem

Amostra



Estatística Descritiva

Média da amostra: \bar{x}

Inferência

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1


25

25

8

INTRODUÇÃO À ESTATÍSTICA

EDIT.



All models are wrong, but some are useful.

— George E. P. Box —

George Edward Pelham Box (1919 – 2013) was a British statistician, who worked in the several areas of statistics and that has been called "**one of the great statistical minds of the 20th century**".

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

26

26

INTRODUÇÃO À ESTATÍSTICA

EDIT.

CONCEITOS BASE



125,058	154,568	95,054	124,500
125,487	56,845	97,511	125,000
124,000	110,000	99,011	154,000
		99,216	95,000
		101,090	154,200
		101,500	110,000
		101,962	89,000
			50,000
			700

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

27

27

CONCEITOS BASE: POPULAÇÃO

EDIT.

População ou **Universo**, é o conjunto de **TODOS** os elementos de interesse e que possuem uma, ou mais características, em comum.

Exemplos:

Finita ❖ Atuais adeptos de um clube de futebol

Infinita ❖ Temperaturas registadas no ponto mais alto Serra da Estrela

Finita ❖ Alunos 12º ano inscritos para provas de acesso ao ensino superior 2025

Infinita ❖ Parafusos produzidos numa fabrica nos Açores que produz 24h/dia

A população pode **finita** ou **infinita**.



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

32

32

CONCEITOS BASE: AMOSTRA

EDIT.

Nem sempre é possível, ou vantajoso, estudar uma População e muitas vezes opta-se por analisar apenas um **subconjunto finito da população**, a que chamamos **Amostra**.

Cenários em que é vantajoso recorrer a amostras:

- ❖ População de **elevada dimensão**, podendo a recolha de dados ser muito demorada
- ❖ **Custo** unitário para análise é muito elevada, tornando a recolha de dados dispendiosa
- ❖ Estudo implica **destruição** das observações (ex.: estudo do tempo que demora um fósforo a queimar)
- ❖ Populações **infinitas**



DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

33

33

CONCEITOS BASE: AMOSTRA

EDIT.

Uma **amostra probabilística** é representativa da população e por isso permite fazer inferência sobre a mesma.

Dizemos que uma amostra é **representativa** quando possui as **mesmas características** da população nas variáveis em estudo.

Métodos mais comuns de amostragem probabilística:

- ❖ **Aleatória**: Cada unidade tem uma probabilidade (ou ‘peso’) associada de pertencer à amostra.
- ❖ **Amostragem Estratificada**: a população é dividida em subconjuntos (estratos) e é feita uma amostragem aleatória dentro de cada substrato. Utilizada para garantir a representatividade desses subconjuntos na amostra.

Falaremos de técnicas de amostragem mais à frente no módulo.



INTRODUÇÃO À ESTATÍSTICA

EDIT.

BORA LÁ POR A CABEÇA A BOMBAR



EXERCICIO 1



Considerem o ficheiro **laptop_pricing** que vos foi disponibilizado.
Explore os dados e **definam uma hipótese** sobre a informação apresentada, como por exemplo:

“A única característica relevante para a definição do preço dos laptops é Storage_GB_SSD”

Descrevam o vosso estudo, começando por definir:

- **Variável de interesse** (variável dependente)
- Qual a **população** do estudo e suas características
- Que **amostra** vão usar e que características têm
- Que **informação** (variáveis independentes) vão usar para testar a nossa hipótese

Expliquem a **abordagem** que vão seguir para, a partir dos dados da amostra, chegarem a alguma conclusão sobre a vossa hipótese. **Podem chegar a uma conclusão 100% certa?**

EXERCICIO 1



Equipa 1

Andreia
Gonçalo



Equipa 2

Joana
João Balinha



Equipa 3

José PM
Nuno



Equipa 4

Sara
Filipa



Equipa 5

Carolina L
Stefane



Equipa 6

Ana
Alexandre



Equipa 7

Yhoanna
Susana



Equipa 8

José F
Rui



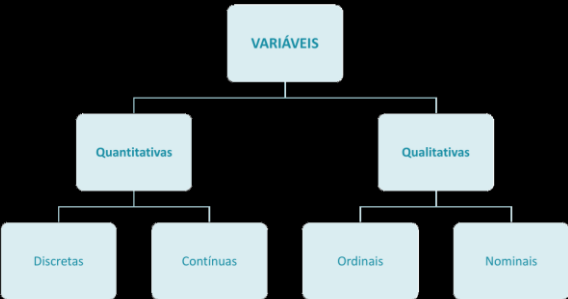
Equipa 9

Tamara
Carolina M

CONCEITOS BASE: TIPOS DE VARIÁVEIS



Já tínhamos visto em módulos anteriores que as **variáveis** de estudo podem ser de **diferentes tipos**:



CONCEITOS BASE: TIPOS DE VARIÁVEIS



Já tínhamos visto em módulos anteriores que as **variáveis** de estudo podem ser de **diferentes tipos**:



As variáveis **quantitativas** representam uma característica **suscetível de ser medida**, apresentando-se com diferentes níveis de intensidade.

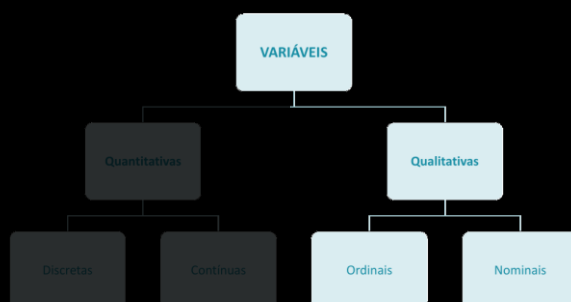
Estas variáveis podem ter uma natureza **Discreta**, quando representam medidas descontínuas, como por exemplo **número de filhos**.

Ou podem ser **Contínua**, quando representam medidas contínuas, como por exemplo **distância entre 2 cidades**.

CONCEITOS BASE: TIPOS DE VARIÁVEIS

EDIT.

Já tínhamos visto em módulos anteriores que as **variáveis** de estudo podem ser de **diferentes tipos**:



As variáveis **qualitativas** representam alguma qualidade, categoria ou característica, **não suscetível de medida**, mas de classificação.

Podem ser **Ordinais**, quando a categoria permite uma ordenação, como por exemplo **nível de escolaridade**.

Ou podem ser **Nominais**, quando as categorias não permitem ordenação, como por exemplo **marca de um equipamento**.

40

CONCEITOS BASE: DISTRIBUIÇÃO

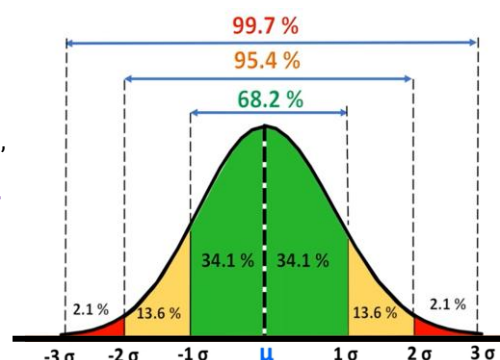
EDIT.

Quando nos referimos à **distribuição** de uma variável procuramos saber **que valores ela toma**, e com que **frequência**. A distribuição ajuda-nos a compreender, por exemplo, que valores são **mais prováveis** e que valores são **atípicos**.

Existem **distribuições teóricas** que permitem recorrer a determinados **teoremas** e **axiomas** teóricos **facilitando assim o cálculo** dos indicadores de interesse.

Por exemplo, se a distribuição da variável X for uma Normal(μ , σ), como no gráfico à direita, sabemos que **95,4%** dos dados estão entre $\mu - 2\sigma$ e $\mu + 2\sigma$ e que **99,7%** estão entre $\mu - 3\sigma$ e $\mu + 3\sigma$.

É com base na distribuição Normal e nas suas características teóricas, que se classificam como **outliers** valores fora do intervalo: $\mu - 3\sigma$ e $\mu + 3\sigma$. Tal classificação só é válida se a variável seguir uma distribuição Normal.



41

INTRODUÇÃO À ESTATÍSTICA

EDIT.

CARACTERÍSTICAS AMOSTRAIS

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

42

42

CARACTERÍSTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.

Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

43

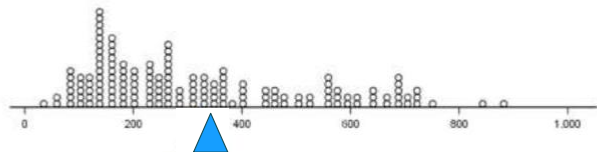
43

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS



Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



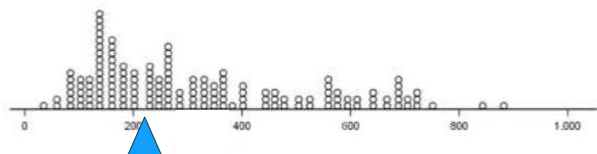
Média

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS



Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



Média

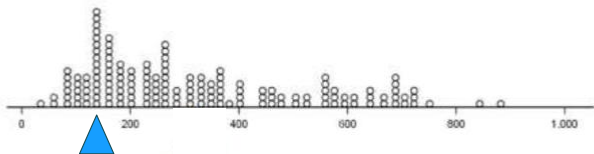
Mediana

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.

Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



Média

Moda

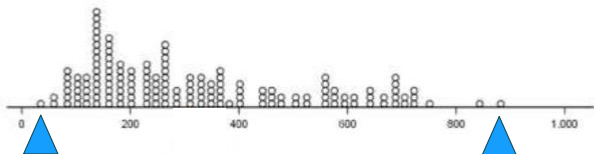
Mediana

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.

Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



Média

Moda

Mínimo

Mediana

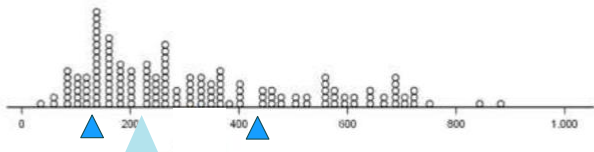
Máximo

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.

Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



Média

Moda

Mínimo

Mediana

Máximo

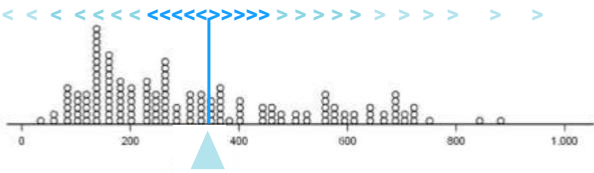
Quartis

CARACTERISTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.

Imaginem que temos acesso à distribuição de uma **variável**, que representa uma característica dos nossos dados, e que a sua **representação gráfica** é a que está espelhada abaixo.

Que **informação** seria interessante **extrair** sobre esta variável?



Média

Moda

Mínimo

Variância /
Desvio Padrão

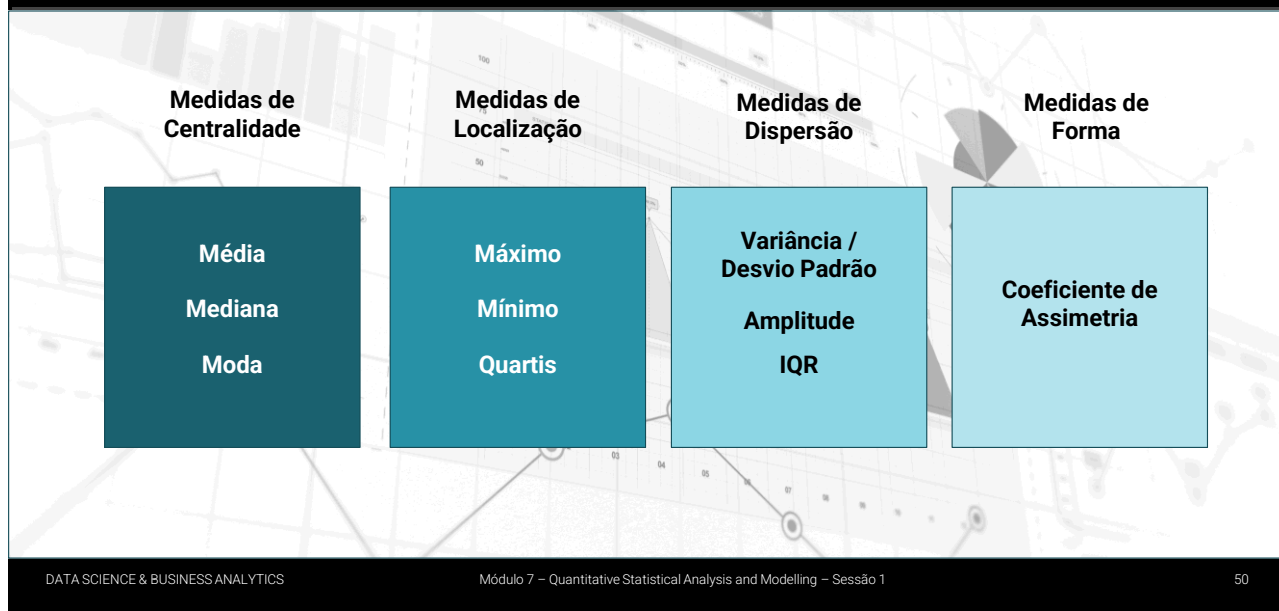
Mediana

Máximo

Quartis

CARACTERÍSTICAS AMOSTRAIS: TIPOS DE MEDIDAS

EDIT.



50

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE LOCALIZAÇÃO

EDIT.

As **medidas de localização** procuram identificar o **centro da distribuição**.

❖ Média

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Medida mais utilizada, como **indicador resumo** das características dos dados, como parâmetro para modelos analíticos, etc.

Bastante **sensíveis a valores atípicos**, por isso muitas vezes é calculada com a exclusão dos extremos, passado a ser uma **média aparada**.

51

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE LOCALIZAÇÃO

EDIT.

As **medidas de localização** procuram identificar o **centro da distribuição**.

❖ Mediana

1. Ordenação ascendentes dos valores de X
2. Se a dimensão dos dados é **ímpar**, então identificar a **observação central**, se for **par**, então identificar as **duas observações centrais** e calcular a **média**

2, 2, 3, **7**, 8, 9, 9
Mediana = **7**

1, 4, 4, **5**, **6**, 7, 7, 7
Mediana = $(5+6) \div 2 = 5.5$

É a observação centra da amostra **ordenada**.

Quando dizemos que m_x é a mediana da variável X isso significa que **50% dos valores** de X são **inferiores a m_x** e 50% são **superiores a m_x** .

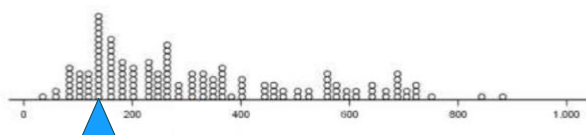
CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE LOCALIZAÇÃO

EDIT.

As **medidas de localização** procuram identificar o **centro da distribuição**.

❖ Moda

É o valor **mais frequente** da amostra.



Ao contrário das anteriores medidas, no caso da moda ela **pode tomar mais de um valor**.

A moda é mais **frequentemente** usada para **dados qualitativos**, visto ser a única medida possível de calcular, mas também pode ser utilizada para dados quantitativos agrupados em intervalos, passando a representar o **intervalo mais frequente**.

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE LOCALIZAÇÃO RELATIVA

EDIT.

Estas medidas permitem descrever os dados com respeito à alguns **indicadores uteis**, mas **não de centralidade**.

- ❖ **Mínimo**: valor mais reduzido tomado pela variável na amostra
- ❖ **Máximo**: valor mais elevado tomado pela variável na amostra
- ❖ **Quantil** de ordem p (onde p é um valor entre 0 e 1): x_p é o quantil de ordem p sempre e quando $p\%$ dos valores de X são **inferiores a x_p** e $(1-p)\%$ dos valores de X são **superiores a x_p** .

Casos particulares dos Quantis:

- Mediana, que corresponde ao Quantil 50, ou x_{50}
- 1º Quartil, que corresponde ao Quantil 25, ou x_{25}
- 3º Quartil, que corresponde ao Quantil 75, ou x_{75}

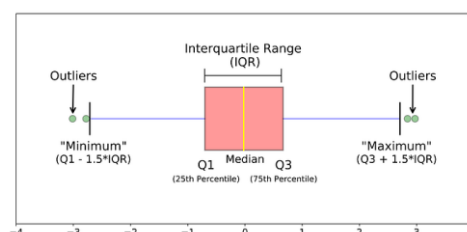
55

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE DISPERSÃO

EDIT.

As **medidas de dispersão** procuram medir a **diversidade** dos dados com respeito a uma característica.

- ❖ **Amplitude**: Valor Máximo – Valor Mínimo, é a medida mais simples de calcular, mas muito sensível a valores extremos
- ❖ **IQR (InterQuartile Range)**: Quantil 75% (ou Q3)– Quantil 25% (ou Q1)



56

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE DISPERSÃO



As **medidas de dispersão** procuram medir a **diversidade** dos dados com respeito a uma característica.

- ❖ **Variância:** Trata-se da média do quadrado dos desvios das observações com respeito à média da amostra, e a sua formula é:

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

É muitas vezes apresentada na forma corrigida, onde a divisão é feita por n-1, em vez de n.

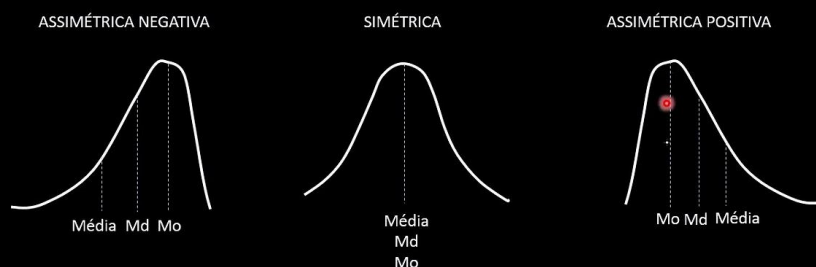
- ❖ **Desvio Padrão:** Não é mais que a raiz quadrada da variância, e tem a vantagem de estar medida na mesma "escala" que o valor original dos dados, e a sua formula é:

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

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE FORMA



Quando representamos a **distribuição dos dados**, podemos verificar diferentes **formas** com base na sua **simetria** ou **assimetria**:



Temos uma **assimetria negativa** quando os valores se concentram do **lado direito** (valores mais altos) da distribuição, e uma **assimetria positiva** quando existe uma maior concentração do **lado esquerdo** (valores mais reduzidos) da distribuição.

CARACTERÍSTICAS AMOSTRAIS: MEDIDAS DE FORMA

EDIT.

- ❖ **Coefficiente de assimetria (skewness)**: medida amostral que devolve o valor 0 para distribuições totalmente simétricas e valores diferentes de 0, positivos ou negativos, para medir nível de assimetria.

Uma das formulas (existem diversas) para o seu cálculo é:

$$skew = \frac{1}{n-2} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Um valor de assimetria entre **-1** e **+1** é indicador de um nível de simetria **excelente**, enquanto **-2** a **+2** é geralmente **aceitável**.

Valores além de -2 e +2 sugerem não normalidade substancial.

CARACTERÍSTICAS AMOSTRAIS: COEFICIENTE DE CORRELAÇÃO

EDIT.

Existe vários coeficientes de correlação, sendo o coeficiente de **correlação de Pearson**, o mais usado, e este mede o grau da correlação (e a direção dessa correlação - se positiva ou negativa) entre duas variáveis quantitativas.

Este coeficiente, normalmente representado por ρ assume apenas valores entre -1 e 1.

- ❖ Se $\rho = 1$ então existe uma **correlação perfeita positiva** entre as duas variáveis.
- ❖ Se $\rho = -1$ então existe uma **correlação perfeita negativa** entre as duas variáveis.
- ❖ Se $\rho = 0$ então as duas variáveis **não dependem linearmente uma da outra**. No entanto, pode existir uma dependência não linear.

O coeficiente é calculado da seguinte forma:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

INTRODUÇÃO À ESTATÍSTICA

EDIT.

REPRESENTAÇÃO DE DADOS

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

62

62

REPRESENTAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

As tabelas de frequência são a maneira mais **simples de organizar os dados**, sendo que no caso de dados contínuos podem ser um pouco mais trabalhosas.

Exemplo: **Tipo de Loja**

Tipo de Loja	Freq. Total	Freq. Relativa
e-Shop	9 311	40%
Flagship store	4 577	20%
MBR	4 661	20%
TeleShop	4 504	20%
Total	23053	100%

Frequência Total

Total das Observações

Total das Observações

Sempre igual a 100%

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

63

63

REPRESENTAÇÃO DOS DADOS: TABELAS DE FREQUÊNCIA

EDIT.

Novo exemplo: Rate

Rate Class	Freq. Total	Freq. Relativa
[-1499 : -899 [942	4%
[-899 : -299 [875	4%
[-299 : 301 [3 773	16%
[301 : 901 [8 671	38%
[901 : 1501 [8 792	38%
Total	23 053	100%

Mas como definir o k?

Um regra da estatística é considerar como k o menor inteiro tal que $2^k \geq \text{dimensão}$ da dados.

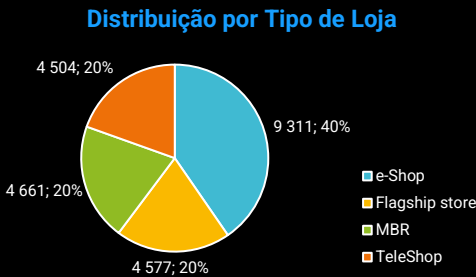
No nosso caso, temos que a dimensão dos dados é 23.053, logo $k = 15$, pois $2^{14} = 16.384$

REPRESENTAÇÃO DOS DADOS: PIE CHART

EDIT.

A partir dos dados organizados é simples criar representações gráficas para melhor visualização.

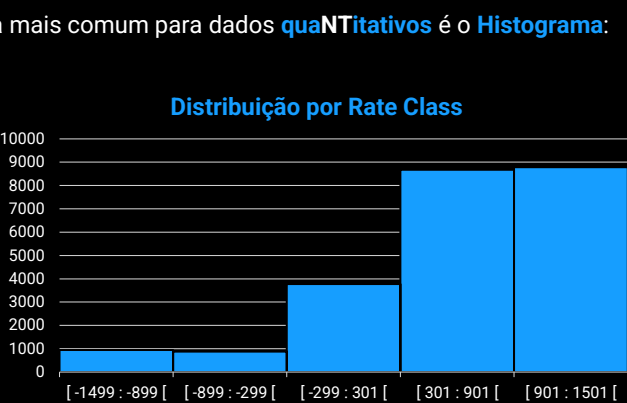
A representação gráfica mais comum para dados qualitativos é o Pie Chart:



REPRESENTAÇÃO DOS DADOS: HISTOGRAMA

E D I T.

A partir dos dados organizados é simples criar **representações gráficas** para melhor visualização.
A representação gráfica mais comum para dados **quanTitativos** é o **Histograma**:



REPRESENTAÇÃO DOS DADOS: SCATTER PLOT

E D I T.

Continuando nas variáveis quantitativas, podemos querer analisar relações entre 2 variáveis, e a forma mais simples de o fazer é através de um **Scatter Plot**:



INTRODUÇÃO À ESTATÍSTICA

EDIT.

BORA LÁ POR A MÃO NA MASSA

DATA SCIENCE & BUSINESS ANALYTICS


Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

68

68

EXERCICIO 2

EDIT.



Voltem a considerem o ficheiro **laptop_pricing** que vos foi disponibilizado no exercício 1.

Explore as variáveis:

- Manufacturer
- Screen_Size_cm
- Price

Calculem em Python as medidas de localização, localização relativa, dispersão e forma que estudamos neste módulo e **que façam sentido** para cada uma dessas variáveis.

Justifiquem se existe, ou não, alguma **relação entre a variável Price**, e alguma das outras variáveis.

Façam uma **representação gráfica** da **distribuição** de cada uma das variáveis, e da **correlação** entre Screen_Size_cm e o Price.


DATA SCIENCE & BUSINESS ANALYTICS


Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1


69


69


EXERCICIO 2






Equipa 1
Andreia
Gonçalo



Equipa 2
Joana
João Balinha



Equipa 3
José PM
Nuno



Equipa 4
Sara
Filipa


Equipa 5
Carolina L
Stefane


Equipa 6
Ana
Alexandre


Equipa 7
Yhoanna
Susana


Equipa 8
José F
Rui


Equipa 9
Tamara
Carolina M


DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

70

70

INTRODUÇÃO À ESTATÍSTICA



EDIT.

Questionário

DATA SCIENCE & BUSINESS ANALYTICS

Módulo 7 – Quantitative Statistical Analysis and Modelling – Sessão 1

71

71

28