

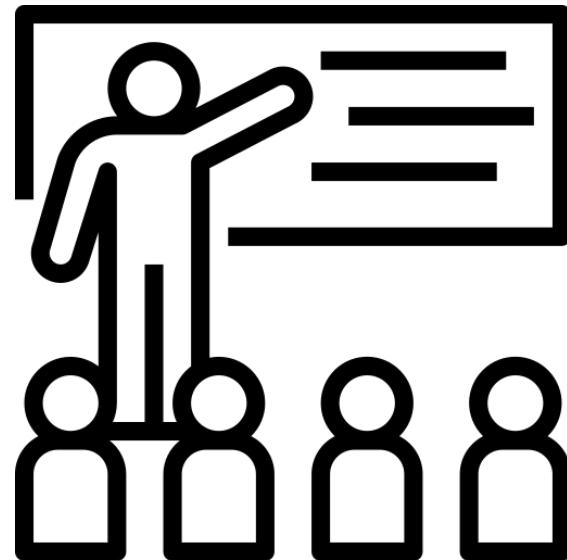
Data Science

Rafael Roberto Dias

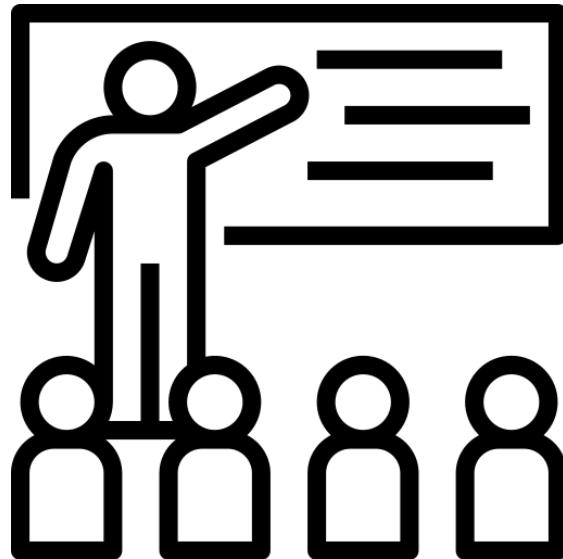
1ª Turma 2020



Quem é o Professor que vos fala?



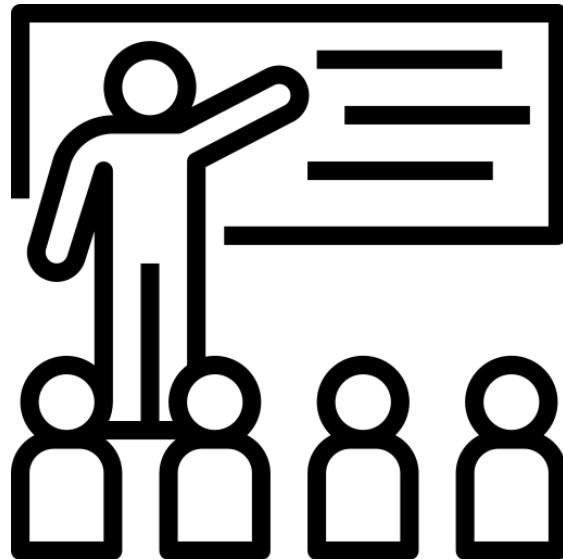
Quem é o Professor que vos fala?



Rafael Roberto Dias

- Bacharel em Estatística - UFPR
- Especialista em Data Science & Big Data - UFPR
- Manager Data Science & BI of Mkt - MadeiraMadeira

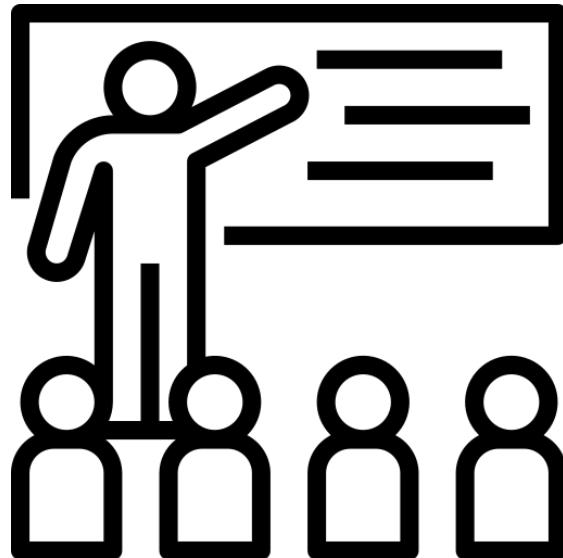
Quem é o Professor que vos fala?



Rafael Roberto Dias

- Bacharel em Estatística - UFPR
- Especialista em Data Science & Big Data - UFPR
- Manager Data Science & BI of Mkt - MadeiraMadeira

Quem é o Professor que vos fala?



Rafael Roberto Dias

- Bacharel em Estatística - UFPR
- Especialista em Data Science & Big Data - UFPR
- Manager Data Science & BI of Mkt - MadeiraMadeira

O que veremos hoje pela manhã?



Horário Assunto

09:30	Apresentação
10:00	Conceitos Estatísticos
11:30	Welcome to R



E hoje à tarde?



Horário Assunto

13:30	Aprendendo R & RStudio
14:30	Qualidade dos Dados
15:30	Variáveis Relevantes
16:30	Introdução Machine Learning

Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



A **Média** é um cálculo relativamente simples que possui muitas aplicações:

$$\text{Média} = \frac{\text{Soma dos Termos}}{\text{Quantidade dos Termos}}$$

Na estatística a média é chamada de **Esperança**, comumente utilizado em Inferências e Previsões

A tabela abaixo informa a cotação do dólar (moeda estrangeira) durante uma semana qualquer:

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

A tabela abaixo informa a cotação do dólar (moeda estrangeira) durante uma semana qualquer:

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

BORA CALCULAR ???

A tabela abaixo informa a cotação do dólar (moeda estrangeira) durante uma semana qualquer:

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

$$\text{Média} = \frac{\text{R\$3,79} + \text{R\$3,78} + \text{R\$3,76} + \text{R\$3,80} + \text{R\$3,90} + \text{R\$3,95} + \text{R\$3,86} + \text{R\$3,95}}{8}$$

$$\text{Média} = \frac{\text{R\$30,79}}{8} \quad \rightarrow \quad \text{Média} = \text{R\$3,85}$$

E se os valores da cotação fossem os abaixo, a média continua sendo um bom estimador?

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$1,79	R\$3,78	R\$3,76	R\$3,80	R\$15,90	R\$3,95	R\$3,86	R\$3,95

E se os valores da cotação fossem os abaixo, a média continua sendo um bom estimador?

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$1,79	R\$3,78	R\$3,76	R\$3,80	R\$15,90	R\$3,95	R\$3,86	R\$3,95

1 - Qual é a Média?

2 - Qual a Variação % entre os resultados?

3 - O que pode-se concluir?

Média

E se os valores da cotação fossem os abaixo, a média continua sendo um bom estimador?

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$1,79	R\$3,78	R\$3,76	R\$3,80	R\$15,90	R\$3,95	R\$3,86	R\$3,95

$$\text{Média} = \frac{\text{R\$1,79} + \text{R\$3,78} + \text{R\$3,76} + \text{R\$3,80} + \text{R\$15,90} + \text{R\$3,95} + \text{R\$3,86} + \text{R\$3,95}}{8}$$

$$\text{Média} = \frac{\text{R\$40,79}}{8}$$



- 1 - Média = R\$5,10
- 2 - Var % = 32,46%
- 3 - Não Recomendado

Mediana

Para o segundo exemplo é indicado utilizar a **Mediana**, como é possível verificar abaixo:

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$1,79	R\$3,78	R\$3,76	R\$3,80	R\$15,90	R\$3,95	R\$3,86	R\$3,95

1º Passo: Organizar os números em ordem crescente:

R\$ 1,79	R\$ 3,76	R\$ 3,78	R\$ 3,80	R\$ 3,86	R\$ 3,95	R\$ 3,95	R\$ 15,90
----------	----------	----------	----------	----------	----------	----------	-----------

Mediana

2º Passo: Calcular a média dos dois números do meio:

R\$ 1,79 R\$ 3,76 R\$ 3,78 **R\$ 3,80** **R\$ 3,86** R\$ 3,95 R\$ 3,95 **R\$ 15,90**

$$\text{Mediana} = \frac{\text{R\$}3,80 + 3,86}{2} \rightarrow \text{Mediana} = \text{R\$}3,83$$

Mediana

Qual é a **Mediana** dos números abaixo?

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

Mediana

Qual é a **Mediana** dos números abaixo?

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

$$\text{Mediana} = \frac{\text{R\$}3,80 + 3,86}{2} \rightarrow \text{Mediana} = \text{R\$}3,83$$

A Moda é o valor que mais aparece num conjunto de dados

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

A Moda é o valor que mais aparece num conjunto de dados

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

Moda = R\$3,95

Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação

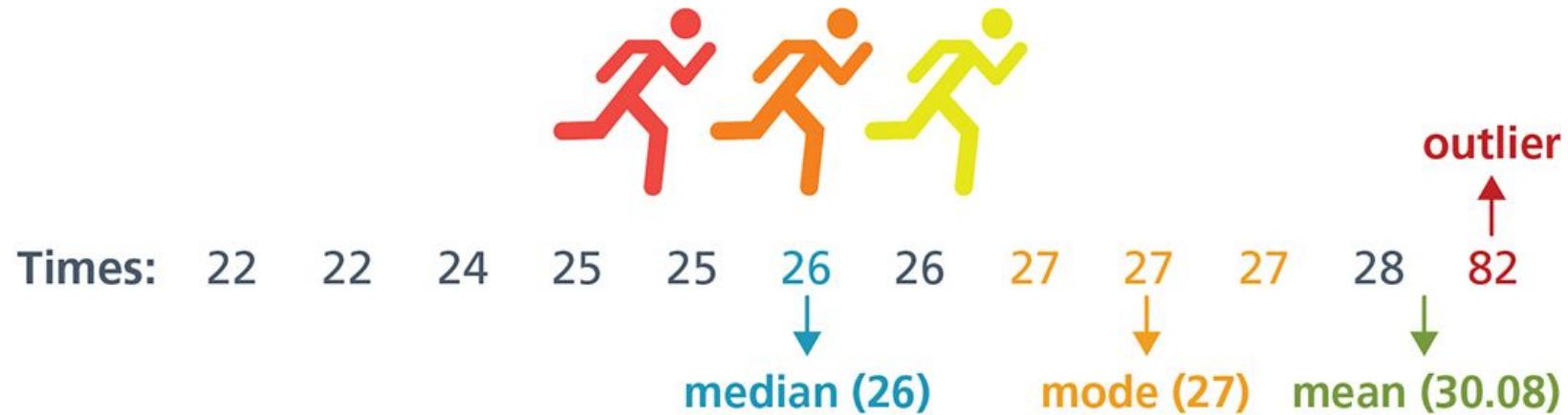


Outlier

Em estatística, **outlier**, valor aberrante ou **atípico**, é uma observação que apresenta um afastamento das demais da série, ou que é inconsistente

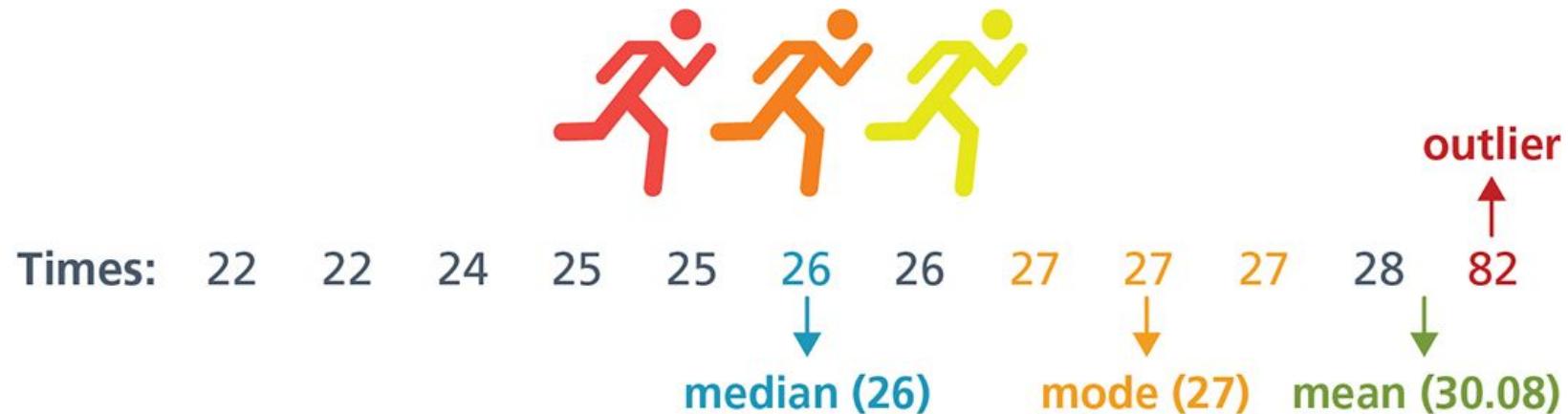
Outlier

Em estatística, **outlier**, valor aberrante ou **atípico**, é uma observação que apresenta um afastamento das demais da série, ou que é inconsistente



Outlier

Em estatística, **outlier**, valor aberrante ou **atípico**, é uma observação que apresenta um afastamento das demais da série, ou que é inconsistente

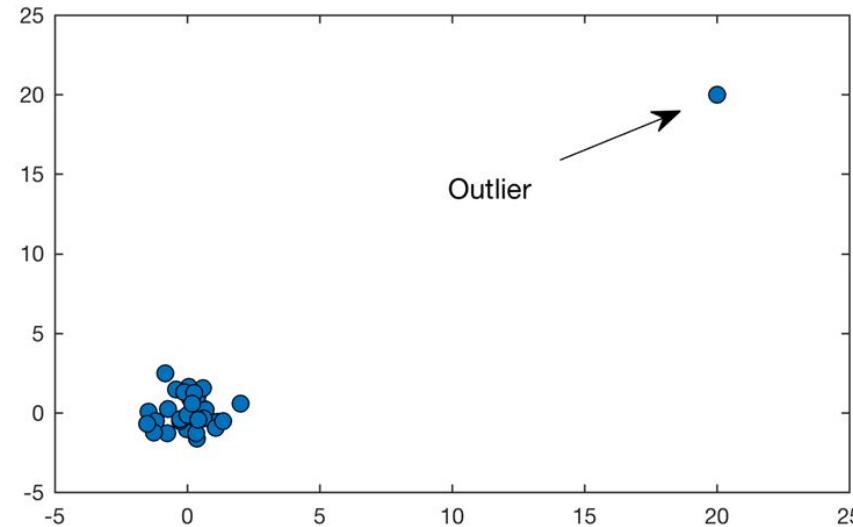


A existência de outliers implica em prejuízos a interpretação dos resultados dos testes estatísticos aplicados às amostras...

Depende!!!

Outlier

Em muitos casos encontrar o **outlier** é o objetivo do estudo, como por exemplo: detecção de fraudes, encontrar espécies mais resistentes, seleção de atletas de alto desempenho, etc



Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Amostragem Aleatória

A **População** é um conjunto de pessoas, itens ou eventos sobre os quais existe interesse em inferir

A **Amostra** é um subconjunto de pessoas, itens ou eventos de uma população que é coletada e analisada para fazer inferências



Amostragem Aleatória

Métodos de **Amostragem Probabilística** são os que selecionam os indivíduos da população de forma que todos tenham as mesmas chances de participar da amostra.



Amostragem Aleatória

Erro Amostral é a diferença entre um resultado amostral e o verdadeiro resultado populacional: tais erros resultam de flutuações amostrais aleatórias



Amostragem Aleatória

A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da **Média Populacional (μ)** é dada por:

$$n = \frac{N}{1+Ne^2}$$

n = Número de indivíduos na amostra

N = Tamanho da População

e = Margem de Erro



Amostragem Aleatória

Qual é o **tamanho de Amostra** para uma população de **859.347** indivíduos com uma margem de erro de **1,25%** ?



<https://pt.wikihow.com/Calcular-o-Tamanho-de-uma-Amostra>

Curso Data Science

Amostragem Aleatória

Qual é o **tamanho de Amostra** para uma população de **859.347** indivíduos com uma margem de erro de **1,25%** ?

$$n = 6.352$$



<https://pt.wikihow.com/Calcular-o-Tamanho-de-uma-Amostra>

Curso Data Science

Amostragem Aleatória

Qual é o **tamanho de Amostra** para uma população de **859.347** indivíduos com uma margem de erro de **1,25%** ?

$$n = 6.352$$

E para margem de erro de **5%**?



Amostragem Aleatória

Qual é o **tamanho de Amostra** para uma população de **859.347** indivíduos com uma margem de erro de **1,25%** ?

$$n = 6.352$$

E para margem de erro de **5%**?

$$n = 400$$



Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Cálculo de Probabilidades

A **probabilidade** é o estudo de **experimentos aleatórios**, ou seja, de eventos repetidos que não é possível prever o resultado.

A palavra **probabilidade** deriva do Latim probare (provar ou testar). Informalmente, provável é uma das muitas palavras utilizadas para eventos incertos ou desconhecidos.



Cálculo de Probabilidades

Condições de probabilidade

$\text{probabilidade} < 0 < 1$

Se $p = 1$ o evento é certo

Se $p = 0$ o evento é impossível



Cálculo de Probabilidades

Fórmulas para Cálculos de Probabilidade

Probabilidade

$$\text{Probabilidade de Um Evento} = \frac{\text{Nº de Casos Favoráveis (Evento)}}{\text{Nº de Casos Possíveis (Espaço Amostral)}}$$

Probabilidade Condicional

$$P(A \text{ acontecer}|B \text{ já aconteceu}) = \frac{\text{Elementos de } B \text{ que Pertencem a } A}{\text{Elementos de } B}$$

União de Eventos

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Cálculo de Probabilidades

No lançamento de um dado, observa-se os seguintes eventos e suas probabilidades:

A = Obter um número par:

$$A = \{2, 4, 6\} \text{ e } n(A) = 3$$

B = Sair um número primo:

$$B = \{2, 3, 5\} \text{ e } n(B) = 3$$

Cálculo de Probabilidades

No lançamento de um dado, observa-se os seguintes eventos e suas probabilidades:

A = Obter um número par:

$$A = \{2, 4, 6\} \text{ e } n(A) = 3 \quad P(A) = 3/6 = 0,5 \text{ ou } 50\%$$

B = Sair um número primo:

$$B = \{2, 3, 5\} \text{ e } n(B) = 3 \quad P(B) = 3/6 = 0,5 \text{ ou } 50\%$$

Cálculo de Probabilidades

No lançamento de um dado, observa-se os seguintes eventos e suas probabilidades:

C = Sair um número maior ou igual a 5:

$$C = \{5, 6\} \text{ e } n(C) = 2$$

D = Sair um número natural:

$$D = \{1, 2, 3, 4, 5, 6\} \text{ e } n(D) = 6$$

Cálculo de Probabilidades

No lançamento de um dado, observa-se os seguintes eventos e suas probabilidades:

C = Sair um número maior ou igual a 5:

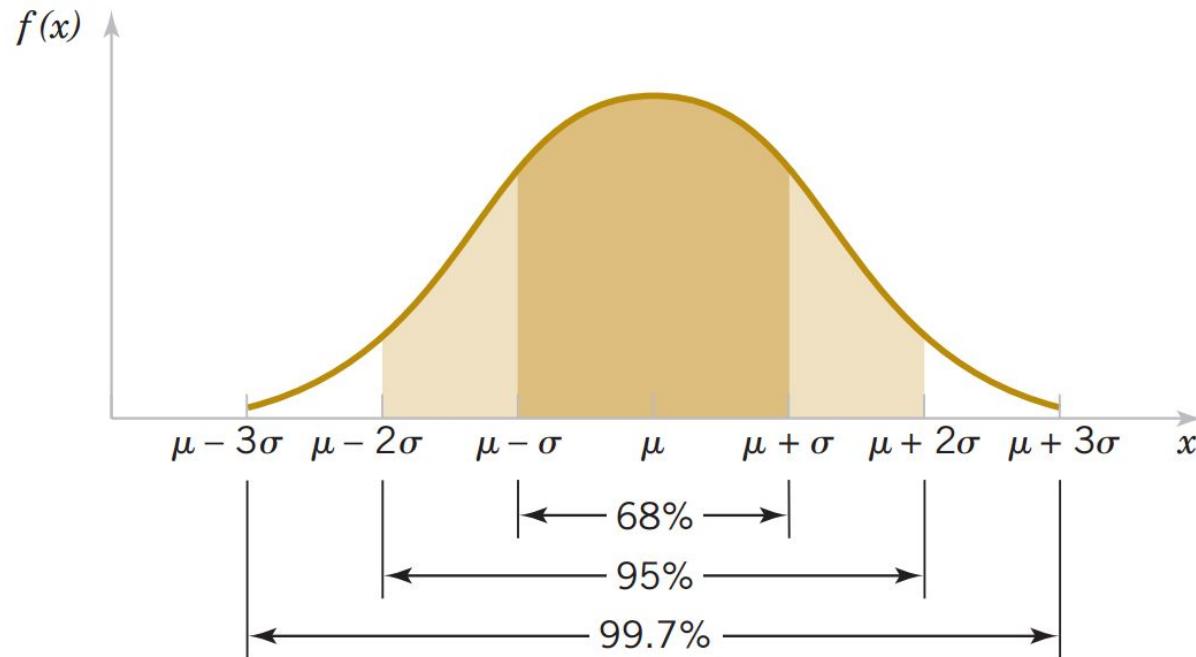
$$C = \{5, 6\} \text{ e } n(C) = 2 \quad P(C) = 2/6 = 0,333 \text{ ou } 33,3\%$$

D = Sair um número natural:

$$D = \{1, 2, 3, 4, 5, 6\} \text{ e } n(D) = 6 \quad P(D) = 6/6 = 1 \text{ ou } 100\%$$

Cálculo de Probabilidades

A **Distribuição Normal** é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais. Isso se deve ao fato de que um grande número de fenômenos naturais apresenta sua distribuição de probabilidade tão proximamente normal



Cálculo de Probabilidades

A **Distribuição Normal** é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais. Isso se deve ao fato de que um grande número de fenômenos naturais apresenta sua distribuição de probabilidade tão proximamente normal

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Variáveis Aleatórias

Discretas: características mensuráveis que podem assumir apenas um número inteiro finito ou infinito contável de valores e geralmente são o resultado de contagens. Exemplos: número de filhos, número de bactérias, quantidade de produtos vendidos



Variáveis Aleatórias

Contínuas: características mensuráveis que assumem valores em uma escala contínua, para as quais valores fracionais fazem sentido, usualmente medidas através de algum instrumento.

Exemplos: peso (balança), tempo (relógio), pressão arterial, idade.



Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Variância e Desvio Padrão

A **variância** e o **desvio padrão** são medidas que dão uma ideia da dispersão de uma distribuição de dados

A **variância** irá mostrar com eficiência a distância existente entre os valores em cada conjunto, mostrando a distância em que o conjunto se encontra com referência ao valor central

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

Variância e Desvio Padrão

A **variância** e o **desvio padrão** são medidas que dão uma ideia da dispersão de uma distribuição de dados

O **desvio padrão** serve para identificar onde existe um erro na amostragem de dados. Se existir discordância nos dados e assim pode-se substituí-los pela média aritmética do conjunto

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

Variância e Desvio Padrão

A variância e o desvio padrão são medidas que dão uma ideia da dispersão de uma distribuição de dados

	NOTA	DESVIOS EM RELAÇÃO À MÉDIA	DESVIOS AO QUADRADO
Julia	9,0	3,2	10,24
Marcos	7,0	1,2	1,44
Maria	5,0	-0,8	0,64
Andreza	4,0	-1,8	3,24
Yuri	4,0	-1,8	3,24

Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

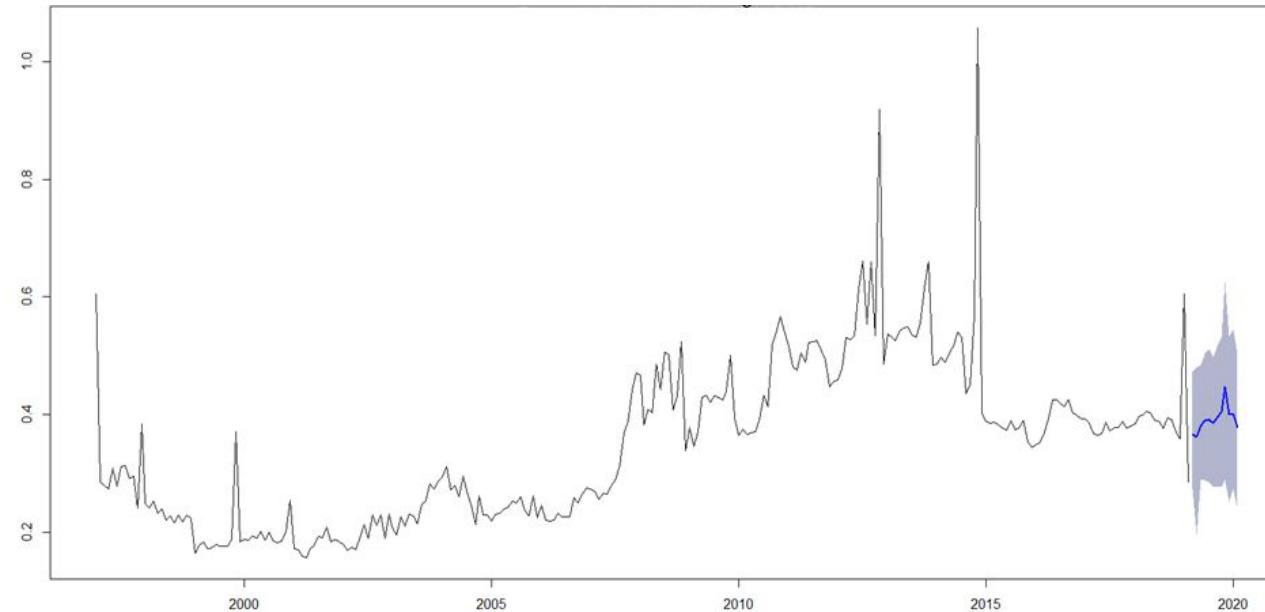
7 – Previsão

8 – Correlação

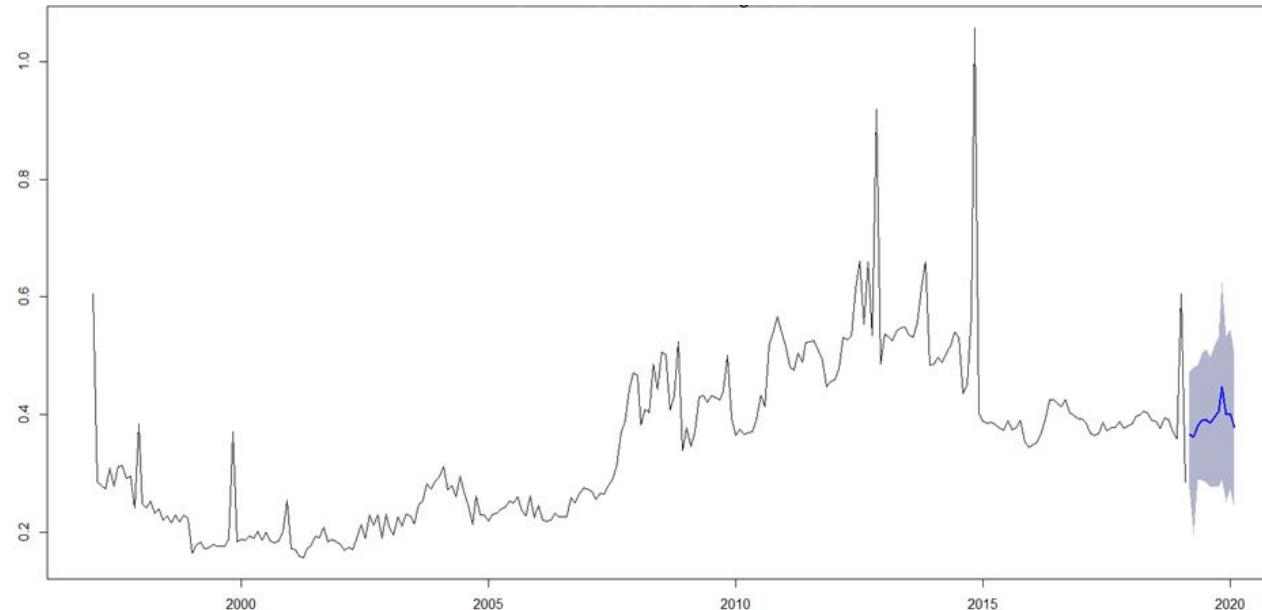


Previsão

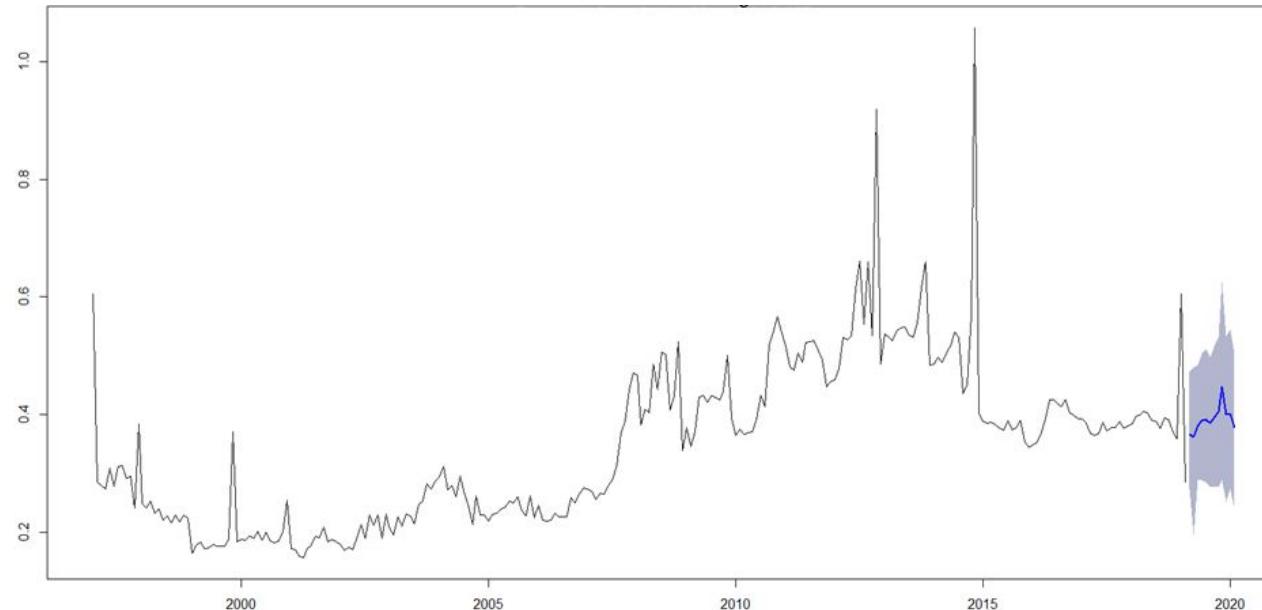
Previsão é o processo de estimativas em situações de incertezas e evoluiu para a prática do plano de demanda diária para tomada de decisões em negócios



O que podemos visualizar neste gráfico de uma Série Temporal?

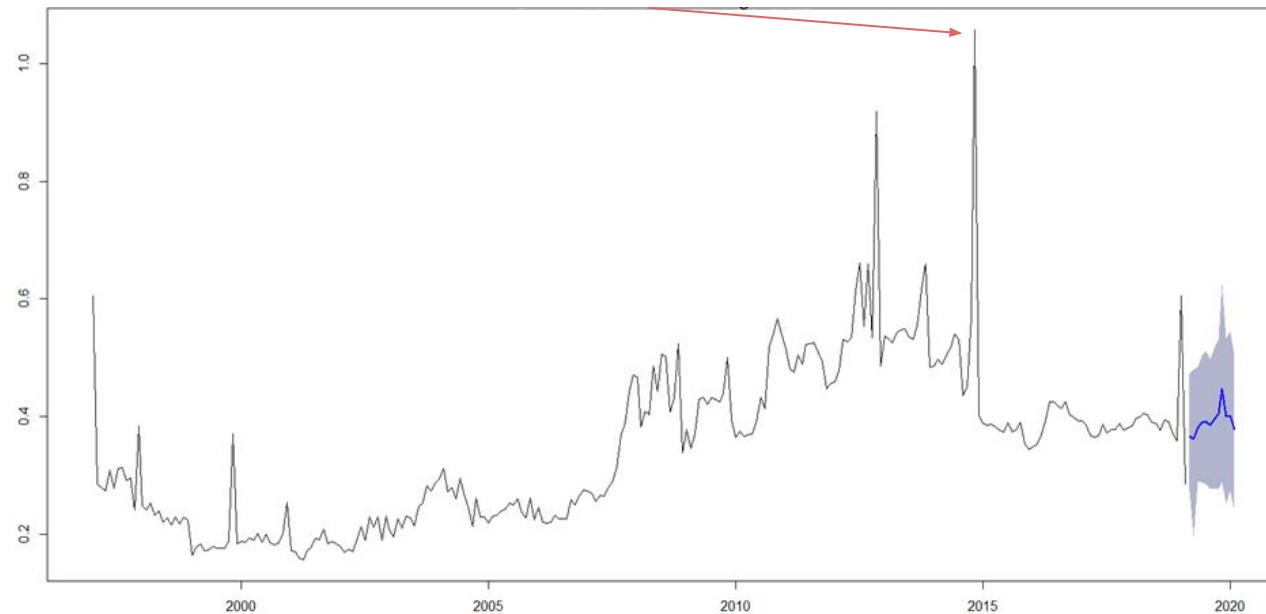


O que podemos visualizar neste gráfico de uma Série Temporal?

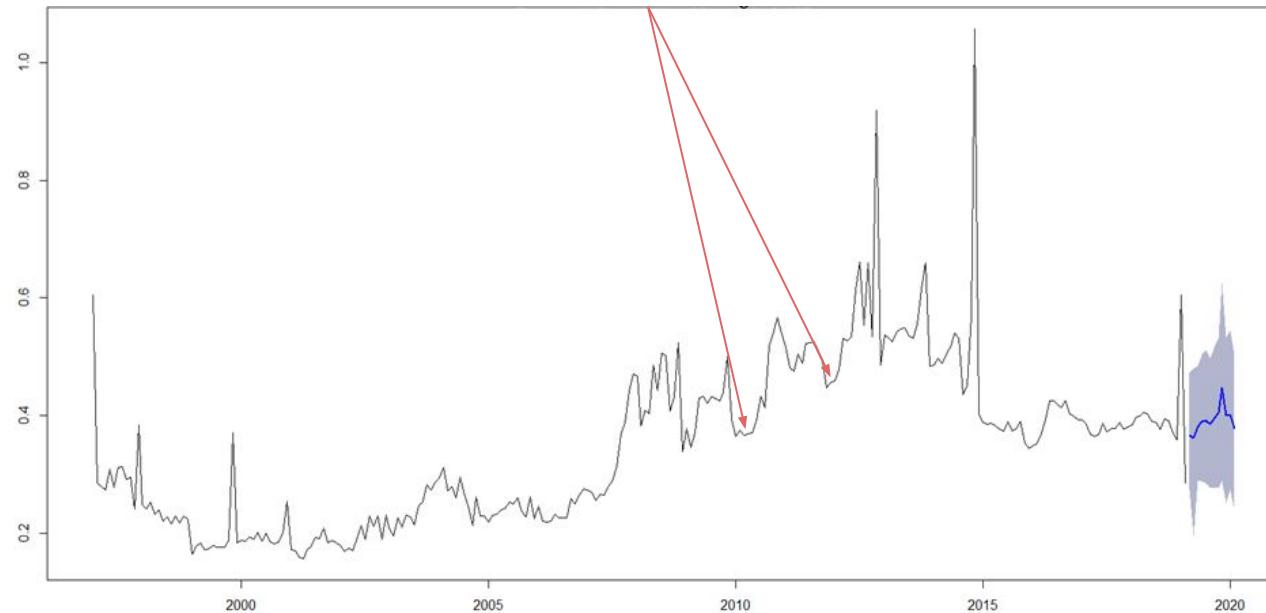


Previsão

Outlier

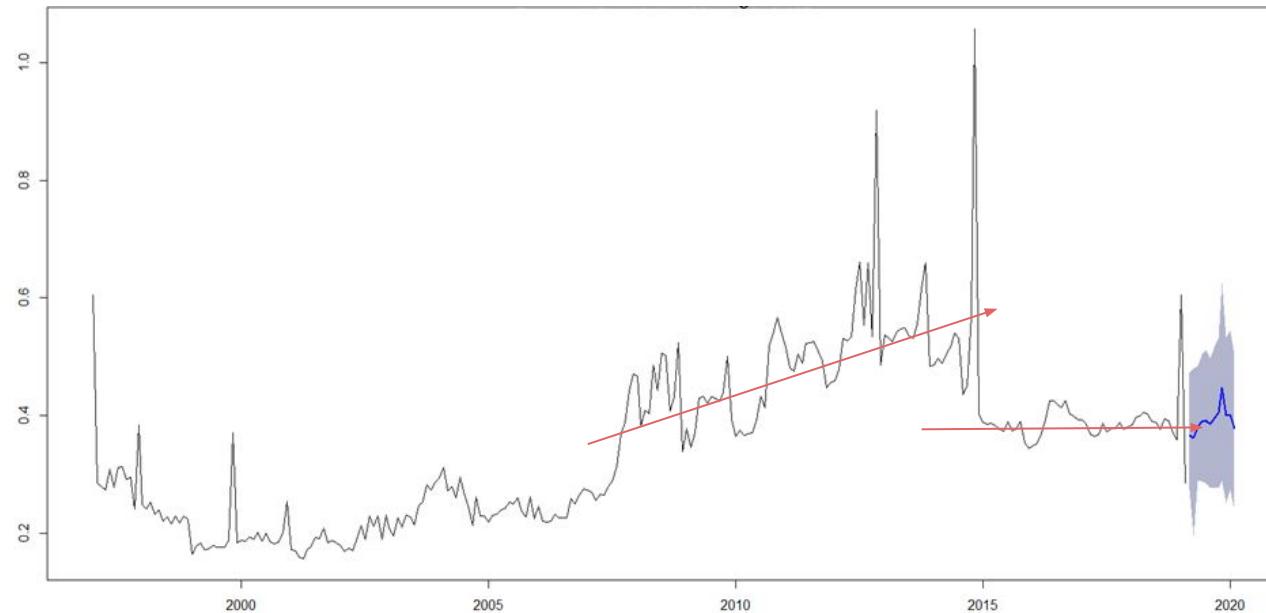


Sazonalidade



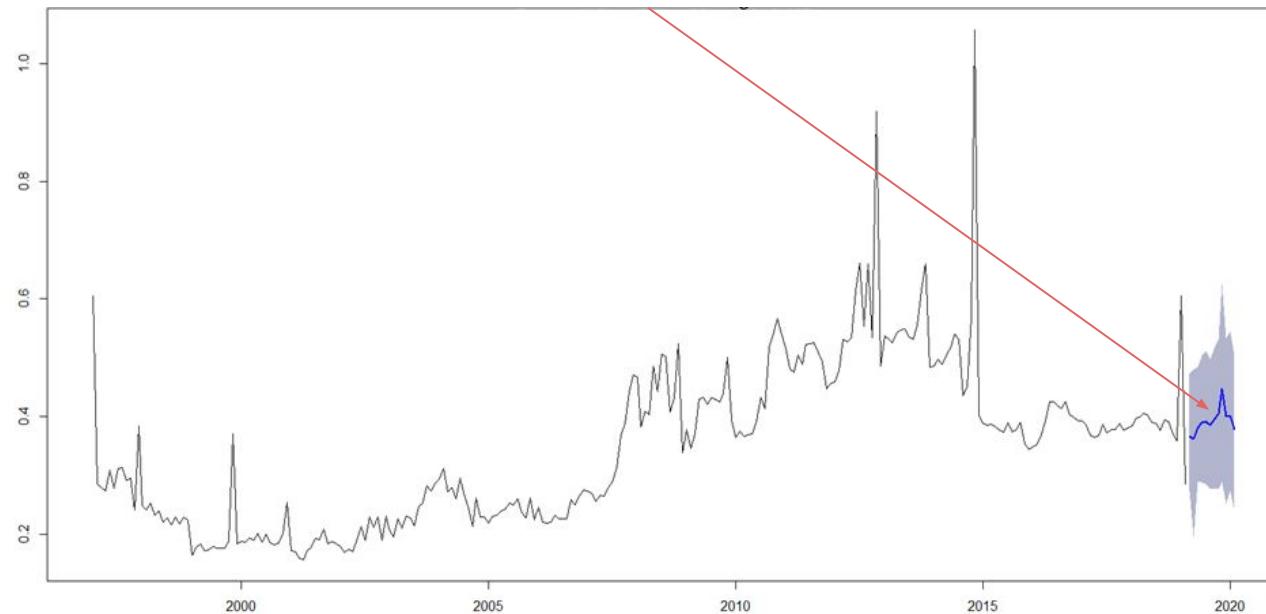
Previsão

Tendência



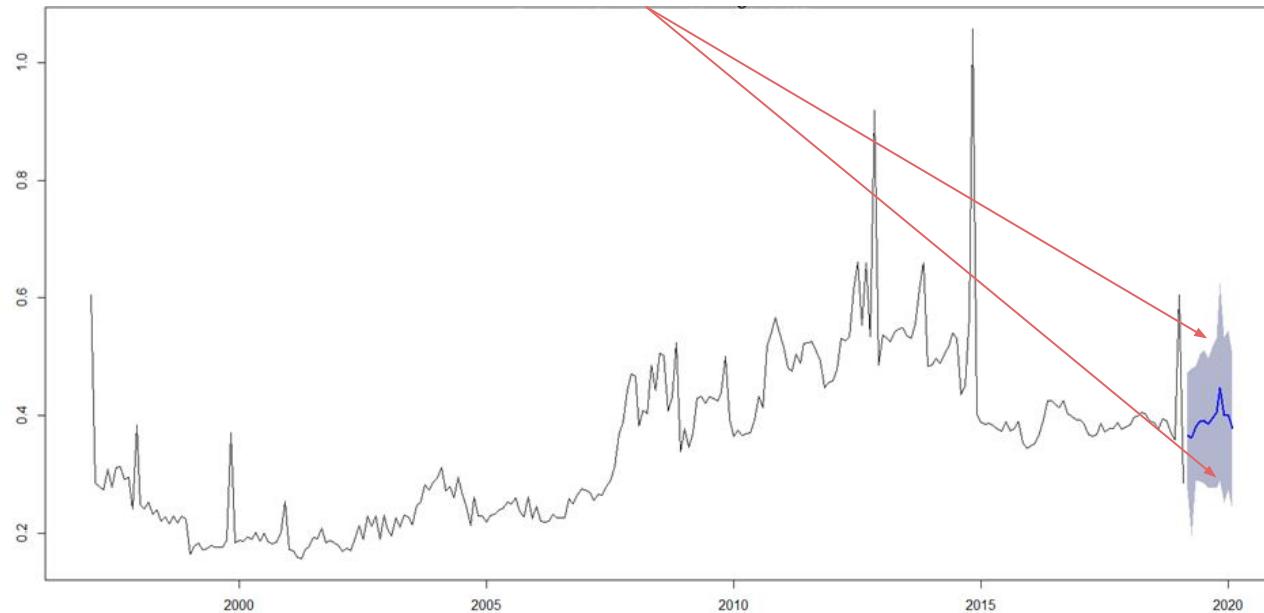
Previsão

Previsão



Previsão

Intervalos de Confiança



Previsão

Façam uma Previsão dos próximos 2 dias de valor de fechamento do dólar:

18/03	19/03	20/03	21/03	22/03	25/03	26/03	27/03
R\$3,79	R\$3,78	R\$3,76	R\$3,80	R\$3,90	R\$3,95	R\$3,86	R\$3,95

Justifique!

Conceitos Estatísticos

1 – Média, Mediana e Moda

2 – Outlier

3 – População e Amostra

4 – Probabilidade

5 – Variáveis: Discretas - Contínuas

6 – Variância e Desvio Padrão

7 – Previsão

8 – Correlação



Coeficiente de Correlação Linear

Correlação Linear

É o coeficiente que indica **a força da associação entre quaisquer duas variáveis métricas!**

Esta metodologia é amplamente utilizada para entender influências de fatores externos em relação a variável de estudo.

Coeficiente de Correlação Linear

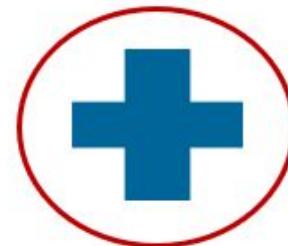
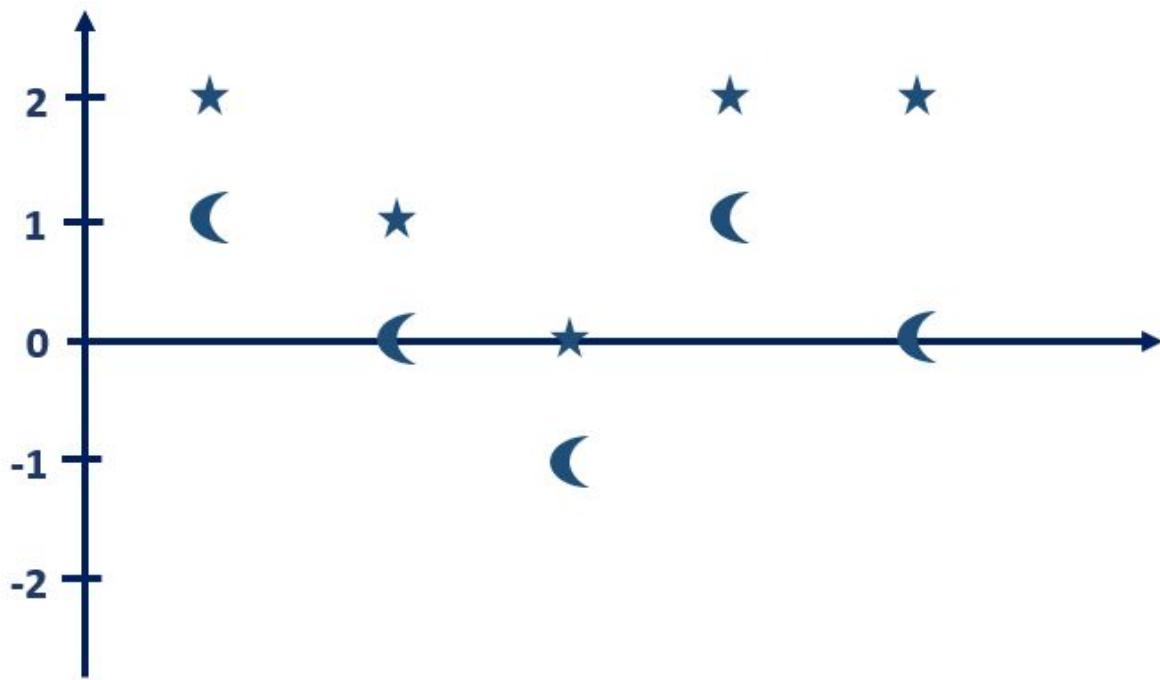
Correlação Linear

É o coeficiente que indica a **força da associação entre quaisquer duas variáveis métricas!**

Esta metodologia é amplamente utilizada para entender **influências de fatores externos em relação a variável de estudo.**

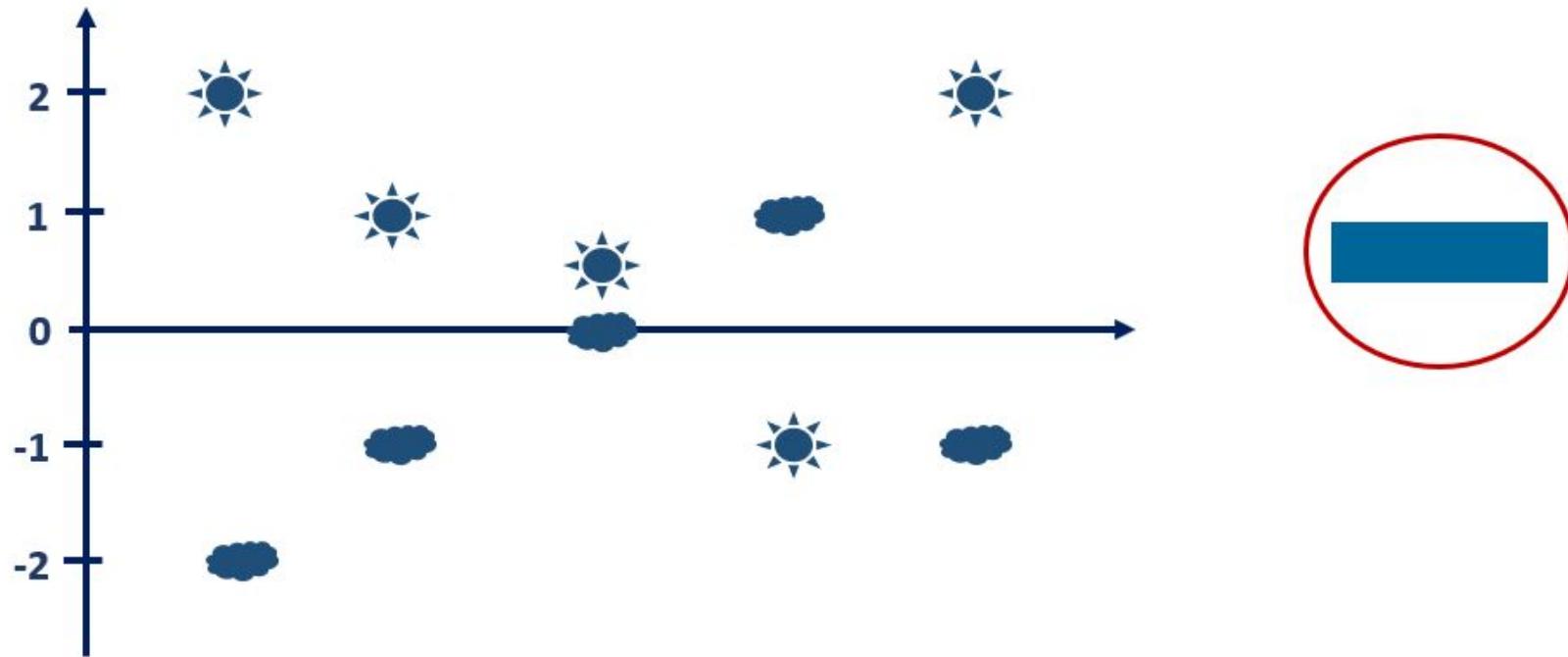
Coeficiente de Correlação Linear

Correlação Linear Positiva



Coeficiente de Correlação Linear

Correlação Linear Negativa



Coeficiente de Correlação Linear

O poder da Análise de Correlação está no método de cálculo que coloca todas as variáveis na mesma ordem de grandeza.

A notação para coeficiente de correlação:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Correlação é representada pela letra grega ρ !



Coeficiente de Correlação Linear

O poder da Análise de Correlação está no método de cálculo que coloca todas as variáveis na mesma ordem de grandeza.

A notação para coeficiente de correlação:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Correlação é representada pela letra grega ρ !



Coeficiente de Correlação Linear

Após calculada a correlação teremos um **único número**:

$$-1 \leq \rho_{X,Y} \leq 1$$

- Sendo que se $\rho_{X,Y} = 0$ não existe correlação entre as variáveis
- Se o valor for próximo de **+1** ou de **-1** indicará forte correlação entre as variáveis

Coeficiente de Correlação Linear

Após calculada a correlação teremos um **único número**:

$$-1 \leq \rho_{X,Y} \leq 1$$

- Sendo que se $\rho_{X,Y} = 0$ não existe correlação entre as variáveis
- Se o valor for próximo de **+1** ou de **-1** indicará forte correlação entre as variáveis

Coeficiente de Correlação Linear

Exercícios

- 1) Numa amostra de 8 funcionários de uma firma observou-se duas variáveis: anos de trabalho e número de promoções recebidas. Calcule a Correlação:

Funcionário	1	2	3	4	5	6	7	8
Anos Trabalho	5	6	6	7	7	8	8	8
Promoções	2	2	1	2	0	3	1	0

Coeficiente de Correlação Linear

Exercícios

- 1) Numa amostra de 8 funcionários de uma firma observou-se duas variáveis: anos de trabalho e número de promoções recebidas. Calcule a Correlação:
- 0,19**

Funcionário	1	2	3	4	5	6	7	8
Anos Trabalho	5	6	6	7	7	8	8	8
Promoções	2	2	1	2	0	3	1	0

Coeficiente de Correlação Linear

Exercícios

- 2) Pretende-se entender o efeito do teor de ferro na capacidade de carga de vigas de concreto. Calcule a Correlação:

Observações	1	2	3	4	5	6	7	8	9	10
Ferro (% peso)	5,4	6,8	6,9	7,3	7,7	8,1	8,2	8,5	8,6	8,9
Carga(ton./m ²)	2,1	2,2	2,9	2,9	3,0	3,1	3,1	3,1	3,4	3,5

Coeficiente de Correlação Linear

Exercícios

- 2) Pretende-se entender o efeito do teor de ferro na capacidade de carga de vigas de concreto. Calcule a Correlação:

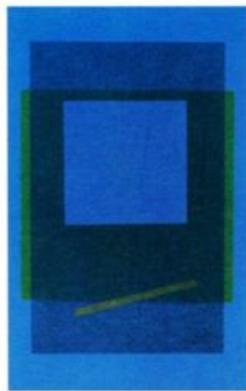
0,91

Observações	1	2	3	4	5	6	7	8	9	10
Ferro (% peso)	5,4	6,8	6,9	7,3	7,7	8,1	8,2	8,5	8,6	8,9
Carga(ton./m ²)	2,1	2,2	2,9	2,9	3,0	3,1	3,1	3,1	3,4	3,5

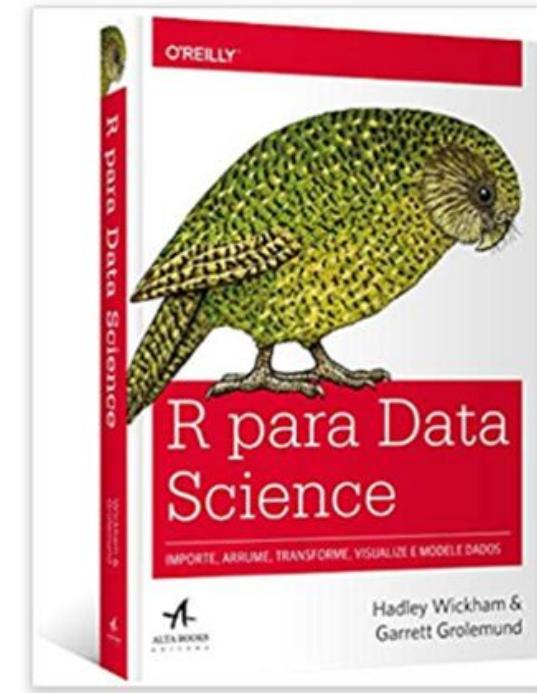
Referências

NOÇÕES DE PROBABILIDADE E ESTATÍSTICA

Marcos Nascimento Magalhães
Antonio Carlos Pedroso de Lima



**Noções de
Probabilidade e
Estatística** (Português)
Capa Comum – 5 jun
2013
por [Marcos
Nascimento
Magalhães](#) (Autor), [Ant
onio Carlos Pedroso de
Lima](#) (Autor)



**R Para Data
Science** (Português) Ca
pa Comum – 7 fev
2019
por [Hadley
Wickham](#) (Autor), [Garrett
Grolemund](#) (Autor), [Sa
mantha
Batista](#) (Tradutor)

Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estat

SAS

- Statistical Analysis System

SPSS

- Statistical Package for the S

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estat

SAS

- Statistical Analysis System

SPSS

- Statistical Package for the S

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estatístico

SAS

- Statistical Analysis System

SPSS

- Statistical Package for the Social Sciences

S-PLUS

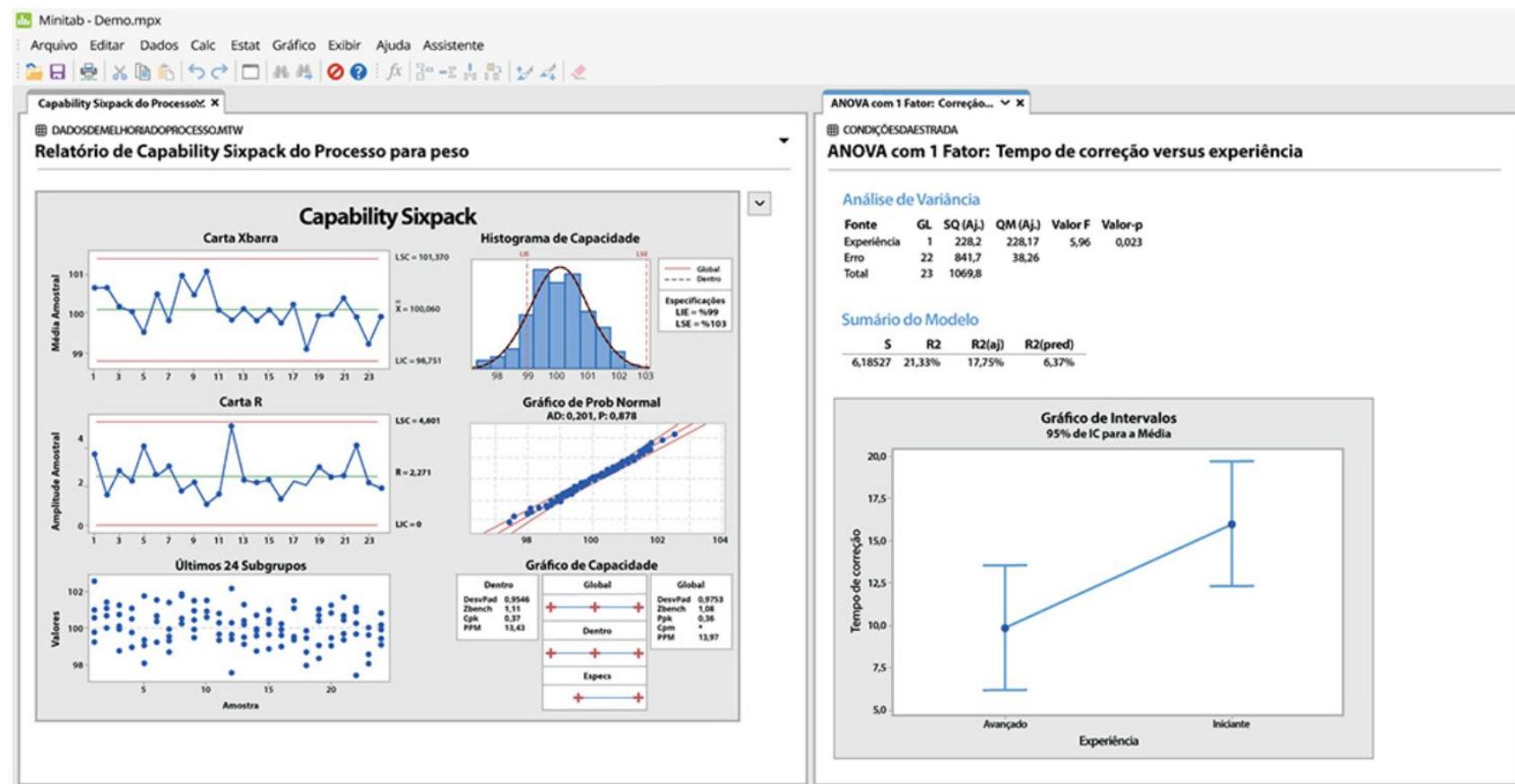
- Versão paga do R

Python

- Linguagem Interpretada



MiniTab



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estat

SAS

- **Statistical Analysis System**

SPSS

- Statistical Package for the S

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



SAS® Studio

Program 1 x Distribution Analysis 1 x

Sign Out

Search

Folders

Tasks

My Tasks

List Data

Data

Characterize Data

List Data

Rank Data

Random Sample

Sort Data

Table Attributes

Transpose Data

Graph

Statistics

Summary Statistics

Distribution Analysis

One-Way Frequencies

Correlations

Table Analysis

One-sample t Test

Paired-samplet Test

Two-sample t Test

One-Way ANOVA

Nonparametric One-Way ANOVA

Linear Regression

Snippets

Libraries

File Shortcuts

Settings Code/Results Split

DATA OPTIONS INFORMATION

CODE LOG RESULTS

Histogram

Add normal curve

Add kernel density estimate

Add inset statistics

Inset Statistics

Goodness-of-fit tests

Histogram with normal curve

Add inset statistics

Normal probability plot

Add inset statistics

Normal quantile-quantile plot

Add inset statistics

Inset Statistics

FITTING DISTRIBUTIONS

The SAS System

Distribution of Height

A histogram showing the distribution of height. The x-axis is labeled 'Height' and ranges from 52.5 to 72.5. The y-axis is labeled 'Percent' and ranges from 0 to 40. The histogram bars are light blue. A smooth blue curve represents a normal distribution fit to the data. Three vertical lines are drawn at approximately 55.5, 62.5, and 69.5.

Curve — Normal(Mu=62.337 Sigma=5.1271)

The SAS System

Fitted Normal Distribution for Height

Parameter	Symbol	Estimate
Mean	Mu	62.33684

ald
eia

Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estatístico

SAS

- Statistical Analysis System

SPSS

- **Statistical Package for the Social Sciences**

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



SPSS

Softpedia.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

4 : VAR00003 2

	VAR000...	
1	Number	Rating
2	1	4
3	2	5
4	3	6
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Visible: 3 of 3 Variables

Transform menu options:

- Compute Variable...
- Count Values within Cases...
- Shift Values...
- Recode into Same Variables...
- Recode into Different Variables...
- Automatic Recode...
- Visual Binning...
- Optimal Binning...
- Prepare Data for Modeling** (highlighted)
- Rank Cases...
- Date and Time Wizard...
- Create Time Series...
- Replace Missing Values...
- Random Number Generators...
- Run Pending Transforms Ctrl+G

Data View Variable View

Prepare Data for Modeling

IBM SPSS Statistics Processor is ready Unicode:ON



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estat

SAS

- Statistical Analysis System

SPSS

- Statistical Package for the S

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



S-Plus

S-PLUS - OptionPricing.ssc - S-PLUS Workbench

File Edit Source Navigate Search Project Run Window Help

Navigator publish... run.Opt... simopt... pptscr... FameGAR... OptionPricing.ssc

optionpricing
 data
 OptionPricing
 SP500.xls
 results
 Autocorrelation.jpg
 HistOfPrices.jpg
 HistOfReturns.jpg
 PriceHistory.jpg
 QQPlot.jpg
 Statistics.csv
 .project
 OptionPricing.ssc
 OptionPricing.wsda
 publishanalytic.ssc
 publishOptionPricing.ssc
 run.OptionPricing.ssc
 simulations.ssc

Outline History View

```

1 OptionPricing <- function(sources.dir, results.dir, properties.list, par...
2
3   library(winjava)
4   source(paste(sources.dir,"simoptions.ssc",sep=""))
5
6   ## parameters
7   nsims<-1000 ## number of simulations
8
9   optiondays <- switch(as.character(parameter.list$Expiration),
10     "3 Months"=92, "1 Year"=250, "2 Years"=500) ## length of option
11
12   strike <- as.numeric(as.character(parameter.list$Strike))
13
14   vol <- 40 ## returns volatility, annualized (40 = 40%) [Should be ca.
15   startprice <- 100 ## option start price
16
17   method <- switch(as.character(parameter.list$Model),
18     "Gaussian"="gbm", "Mixture"="normmix") ## method to estimate opt:
19
20   ## read in data
21   print(properties.list)
22   filename=paste(properties.list@inputs.root,"/OptionPricing/SP500.xls")

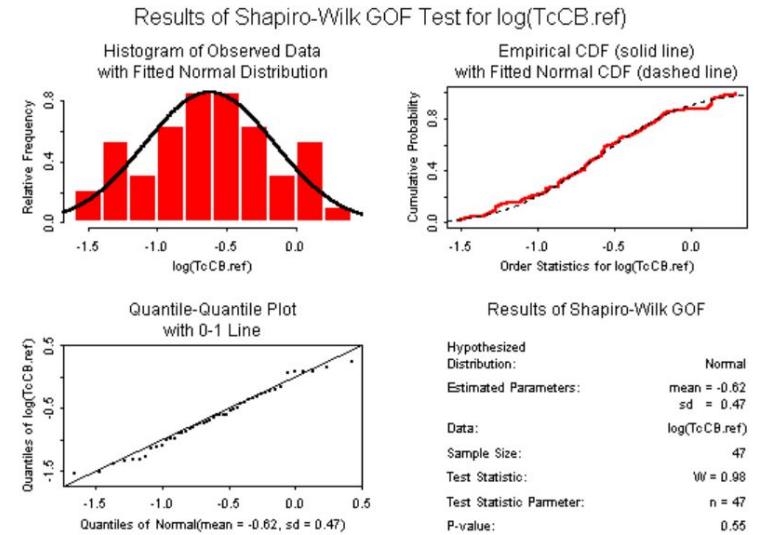
```

Console View Objects View Search Path View Output View Tasks Problems

```

> objects()
[1] ".Last.value"      ".Random.seed"    "OptionPricing"  "a.name"
[5] "attribute.vec"    "barraData"      "dat"          "dl.str"
[9] "dnormix"          "file_vec"       "g.mod"        "grp.co"
[13] "grp_nm"           "intra"         "keep"         "last.dump"
[17] "local.dir"        "lstdy"         "monthlyData" "monthlyData"
[21] "msft.ret"          "mu"            "parameter.list" "parameter.vec"
[25] "path.gbm"          "path.normix"  "pdata"        "pdfnormix"
[29] "pred"              "pred.len"       "price"        "properties.list"
[33] "q.99"              "results.dir"   "ret"          "rng"
[37] "sd"                "server.dir"   "sim.option"  "simnormix"
[41] "source_vec"        "sources.dir"  "stk"          "stockData"
[45] "sym"               "tody"          "var.95"      "varData"
[49] "weeklyData"        "x"             "xmax"        "xmin"
[53] "xx"               "ymax"         "ymin"        "ymin"

```



Quais são os principais softwares Estatísticos?

R

- (Ross e Robert)

MiniTab

- Software Matemático e Estat

SAS

- Statistical Analysis System

SPSS

- Statistical Package for the S

S-PLUS

- Versão paga do R

Python

- Linguagem Interpretada



Python & Jupyter

The screenshot shows a Jupyter Notebook interface with the following components:

- File Bar:** File, Edit, View, Run, Kernel, Tabs, Settings, Help.
- Left Sidebar:** Files, Running, Commands, Cell Tools, Tabs.
- File List:** Shows notebooks and files: Data.ipynb, Fasta.ipynb, Julia.ipynb, Lorenz.ipynb (selected), R.ipynb, iris.csv, lightning.json, lorenz.py.
- Code Cell:** In [4]:

```
from lorenz import solve_lorenz
t, x_t = solve_lorenz(N=10)
```
- Output View:** Displays sliders for parameters sigma (10.00), beta (2.67), and rho (28.00) and a 3D plot of the Lorenz attractor.
- Code Editor:** lorenz.py, containing the following code:

```
def solve_lorenz(N=10, max_time=4.0, sigma=10.0, beta=8./3., rho=28.0):
    """Plot a solution to the Lorenz differential equations."""
    fig = plt.figure()
    ax = fig.add_axes([0, 0, 1, 1], projection='3d')
    ax.axis('off')

    # prepare the axes limits
    ax.set_xlim((-25, 25))
    ax.set_ylim((-35, 35))
    ax.set_zlim((5, 55))

    def lorenz_deriv(x_y_z, t0, sigma=sigma, beta=beta, rho=rho):
        """Compute the time-derivative of a Lorenz system."""
        x, y, z = x_y_z
        return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]

    # Choose random starting points, uniformly distributed from -15 to 15
    np.random.seed(1)
    x0 = -15 + 30 * np.random((N, 3))
```

Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



Detalhes Linguagem R

Histórico

O nome R provém das iniciais dos criadores e a linguagem foi inspirada no software S (AT&T)

Ross Ihaka e Robert Gentleman trabalhavam e eram colegas no departamento de estatística na Universidade de Auckland, Nova Zelândia

O lançamento do R foi através lista de emails da S-news em 3 de Agosto de 1993



Detalhes Linguagem R

Histórico

O nome R provém em parte das iniciais dos criadores e a linguagem foi inspirada no software S

Ross Ihaka e Robert Gentleman trabalhavam e eram colegas no departamento de estatística na Universidade de Auckland, Nova Zelândia

O lançamento do R foi através lista de emails da S-news em 3 de Agosto de 1993



Detalhes Linguagem R

Histórico

O nome R provém em parte das iniciais dos criadores e a linguagem foi inspirada no software S

Ross Ihaka e Robert Gentleman trabalhavam e eram colegas no departamento de estatística na Universidade de Auckland, Nova Zelândia

O lançamento do R foi através lista de emails da S-news em 3 de Agosto de 1993



Detalhes Linguagem R

Vantagens

Linguagem de programação open source, com extensa comunidade contribuidora

Milhares de pacotes com diversas metodologias desenvolvidas

Utilizável para análise de pequenas quantidades de dados e também para Big Data (Mapreduce: Spark, Hadoop)



Detalhes Linguagem R

Vantagens

Linguagem de programação open source, com extensa comunidade contribuidora

Milhares de pacotes com diversas metodologias desenvolvidas

Utilizável para análise de pequenas quantidades de dados e também para Big Data (Mapreduce: Spark, Hadoop)



Detalhes Linguagem R

Vantagens

Linguagem de programação open source, com extensa comunidade contribuidora

Milhares de pacotes com diversas metodologias desenvolvidas

Utilizável para análise de pequenas quantidades de dados e também para Big Data (Mapreduce: Spark, Hadoop)



Detalhes Linguagem R

Detalhes Técnicos

Linguagem Alto Nível - Longe do código de máquina e mais próximo à linguagem humana

```
1 print("Hello World!")
```

VS

```
1 lea si, string
2 call printf
3 hlt
4 string db "Olá mundo!
5 printf PROC
6 mov AL, [SI]
7 cmp AL, 0
8 je pfend
9 mov AH, 0Eh
10 int 10h
11 inc SI
12 jmp printf
13 pfend:
14 ret
15 printf ENDP
```

Detalhes Linguagem R

Detalhes Técnicos

Interpretada - O programa resultante não é executado diretamente pelo sistema operacional ou processador



Detalhes Linguagem R

Detalhes Técnicos

Script - Programas escritos para um sistema de tempo que automatiza a execução de tarefas

```
import <- data.frame((export_graos %>%
  group_by(CO_PAIS) %>%
  dplyr::summarise(Dolares=sum(VL_FOB, NA, na.rm = TRUE)
  )),
  row.names = NULL)

import

summary(import)

## Agrupamento gerou varios valores nulos, serao substituidos

import$Dolares[which(is.na(import$Dolares))] <- 0
```

Detalhes Linguagem R

Detalhes Técnicos

Orientada a objetos - Abstração, Encapsulamento, Herança e Polimorfismo



Detalhes Linguagem R

Utilizações

Técnicas estatísticas

Análises de Séries Temporais

Gráficos

Classificação

Modelos Lineares

Agrupamento

Modelos não Lineares

Machine Learning

Testes estatísticos clássicos

Artificial Intelligence



Detalhes Linguagem R

Utilizações

Técnicas estatísticas

Gráficos

Modelos Lineares

Modelos não Lineares

Testes estatísticos clássicos

Análises de Séries Temporais

Classificação

Agrupamento

Machine Learning

Artificial Intelligence



Detalhes Linguagem R

Ao escrever $4 + 4$ na linha de comando, obtém-se o seguinte resultado:

```
> 4 + 4  
[1] 8  
> |
```

Detalhes Linguagem R

Ao escrever $4 + 4$ na linha de comando, obtém-se o seguinte resultado:

A linguagem R suporta matrizes aritméticas, escalares, vetores, matrizes, quadros de dados (tabelas numa base de dados relacional) e listas

Repositório de Pacotes - CRAN

Comprehensive R Archive Network

The screenshot shows the homepage of the Comprehensive R Archive Network (CRAN). The page features a large R logo on the left, followed by a sidebar with links to CRAN mirrors, what's new, task views, search, and about R. The main content area is titled "The Comprehensive R Archive Network" and contains two sections: "Download and Install R" and "Source Code for all Platforms". The "Download and Install R" section provides links for Linux, Mac OS X, and Windows users. It also notes that R is part of many Linux distributions. The "Source Code for all Platforms" section explains that Windows and Mac users should download precompiled binaries, while Linux users should use package management systems.

← → ⌛ 🔒 cran.r-project.org ☆ ABP ⚡ 📁

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

Detalhes RStudio

**RStudio é um software livre de
ambiente de desenvolvimento,
e que possui uma interface
gráfica amigável**



Detalhes RStudio

O R Studio é uma interface
para o R, com diversas
utilidades diferentes que a
tornam uma ferramenta mais
simples em comparação ao R



Detalhes RStudio

Ele possui duas versões:

RStudio Desktop, que roda

localmente em desktop

RStudio Server, que permite

acessá-lo usando um

navegador web enquanto ele

roda em um servidor

GNU/Linux remoto



Detalhes RStudio

Ele possui duas versões:

RStudio Desktop, que roda

localmente em desktop e

RStudio Server, que permite

acessá-lo usando um

navegador web enquanto ele

roda em um servidor

GNU/Linux remoto



Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



**Muita Conversa e
Pouca Prática!!!!**





Vamos
Praticar!



Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



Analisando Qualidade dos Dados

- Objetivo nesta etapa do estudo é verificar a qualidade dos dados para entender quais têm potencial de fazer parte do estudo
- Foco maior em verificar se existem dados faltantes ou nulos que podem interferir no estudo
- Também aqui começa o entendimento de como cada variável ajuda a explicar o evento em estudo
- Aqui começam as descobertas do Cientista de Dados

Analisando Qualidade dos Dados

- Objetivo nesta etapa do estudo é verificar a qualidade dos dados para entender quais têm potencial de fazer parte do estudo
- Foco maior em verificar se existem dados faltantes ou nulos que podem interferir no estudo
- Também aqui começa o entendimento de como cada variável ajuda a explicar o evento em estudo
- Aqui começam as descobertas do Cientista de Dados

Analisando Qualidade dos Dados

- Objetivo nesta etapa do estudo é verificar a qualidade dos dados para entender quais têm potencial de fazer parte do estudo
- Foco maior em verificar se existem dados faltantes ou nulos que podem interferir no estudo
- Também aqui começa o entendimento de como cada variável ajuda a explicar o evento em estudo
- Aqui começam as descobertas do Cientista de Dados

Analisando Qualidade dos Dados

- Objetivo nesta etapa do estudo é verificar a qualidade dos dados para entender quais têm potencial de fazer parte do estudo
- Foco maior em verificar se existem dados faltantes ou nulos que podem interferir no estudo
- Também aqui começa o entendimento de como cada variável ajuda a explicar o evento em estudo
- **Aqui começam as descobertas do Cientista de Dados**

Welcome to R

1 – Softwares Estatísticos

2 – Detalhes Linguagem R

3 – Aprendendo R & RStudio

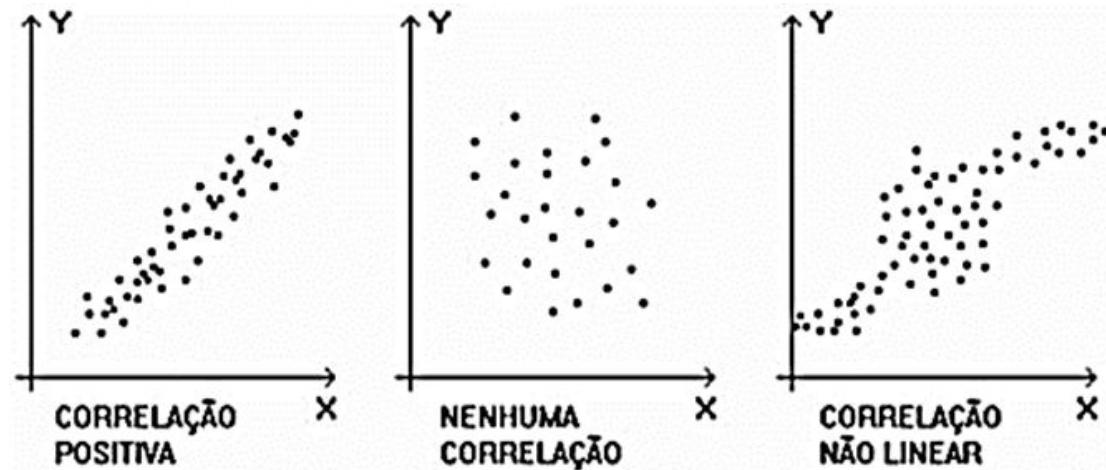
4 – Analisando Qualidade dos Dados

5 – Variáveis Relevantes



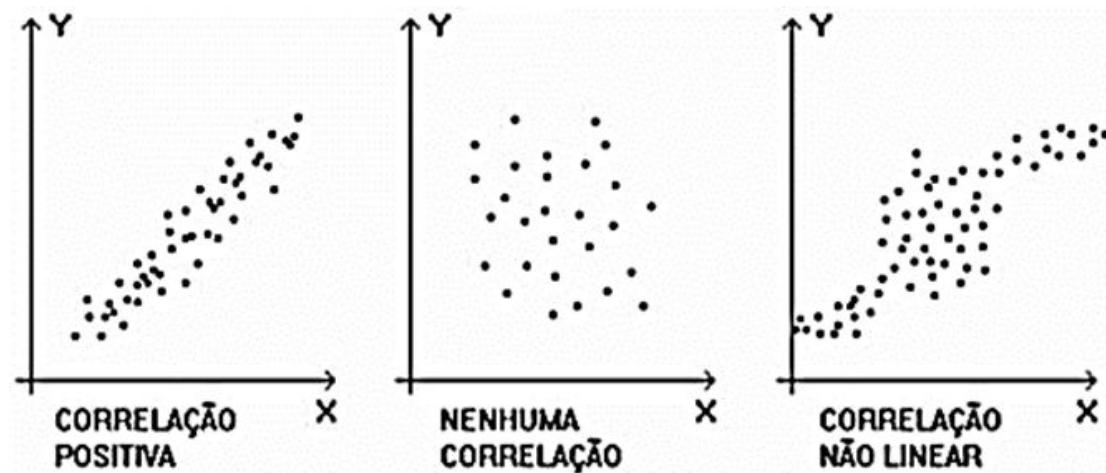
Variáveis Relevantes

- Objetivo nesta etapa do estudo é verificar a como as variáveis se relacionam entre si
 - Foco maior aqui é entender a correlação entre as variáveis



Variáveis Relevantes

- O modelo ou a metodologia que será utilizada para responder às perguntas do estudo dependem dos achados desta etapa



Machine Learning

1 – Machine Learning

2 – Associação

3 – Case MM

Machine Learning

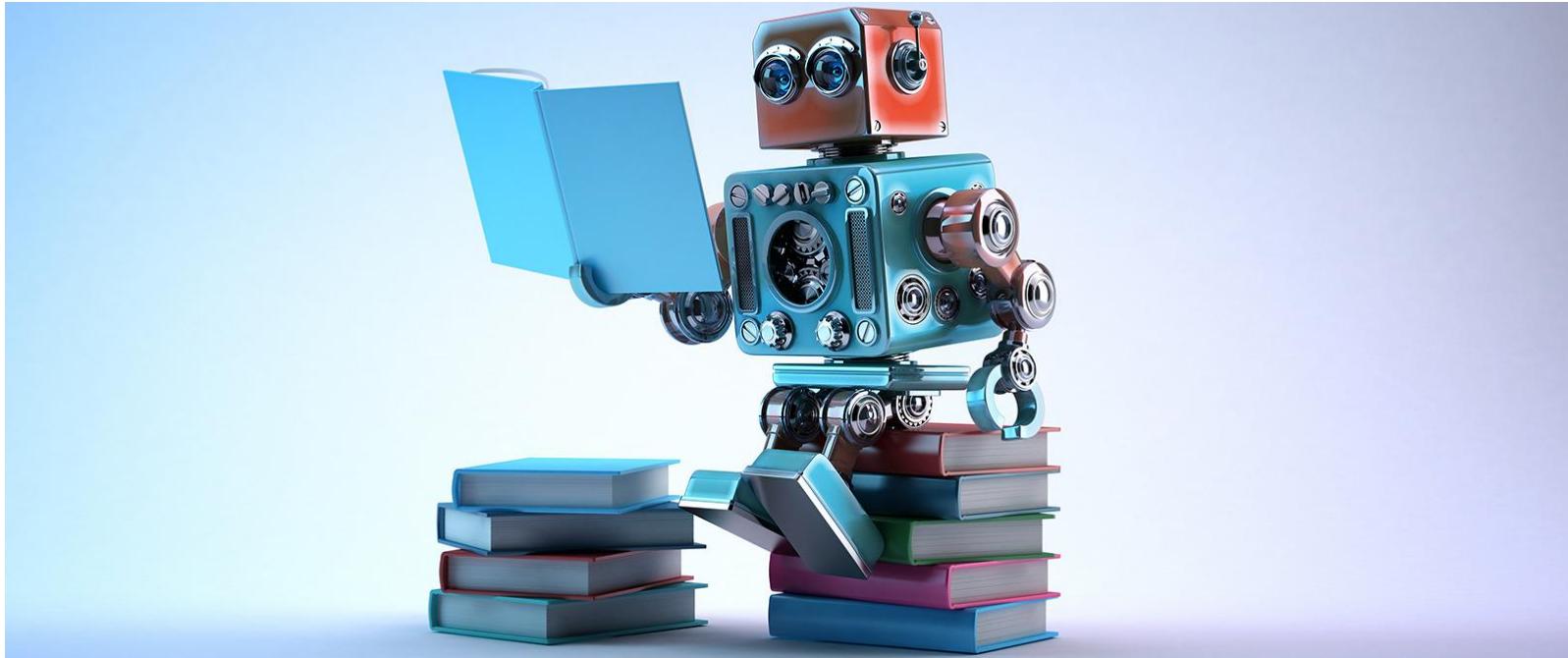
1 – Machine Learning

2 – Associação

3 – Case MM

Machine Learning

O que é Machine Learning?



O que é Machine Learning?

1959 - ML: "Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel

Ou seja, aprender com a experiência que existe intrínseca aos dados.

Como: Algoritmos de aprendizado de máquina analisam as correlações entre os atributos (variáveis) de um sistema (base de dados) a partir de dados amostrais (base de treinamento)



Machine Learning

O que é Machine Learning?

1959 - ML: "Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel

Ou seja, aprender com a experiência que existe intrínseca aos dados.

Como: Algoritmos de aprendizado de máquina analisam as correlações entre os atributos (variáveis) de um sistema (base de dados) a partir de dados amostrais (base de treinamento)



Machine Learning

O que é Machine Learning?

1959 - ML: "Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel

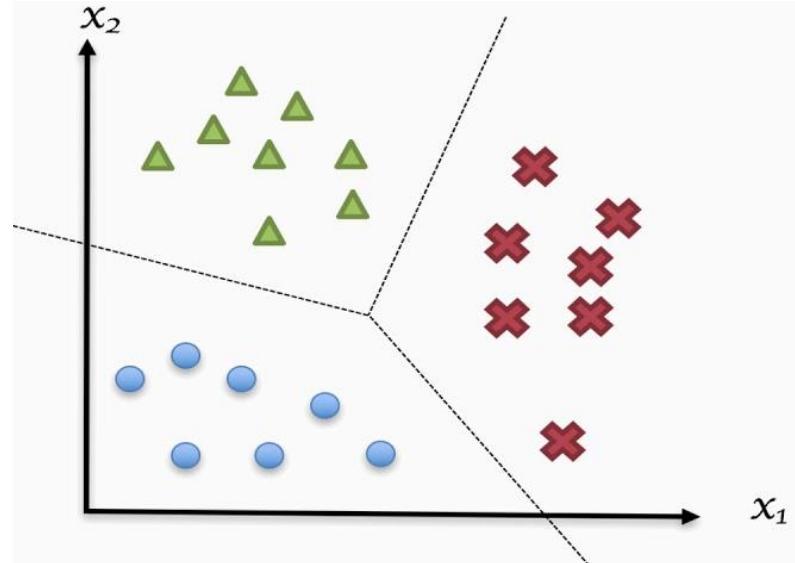
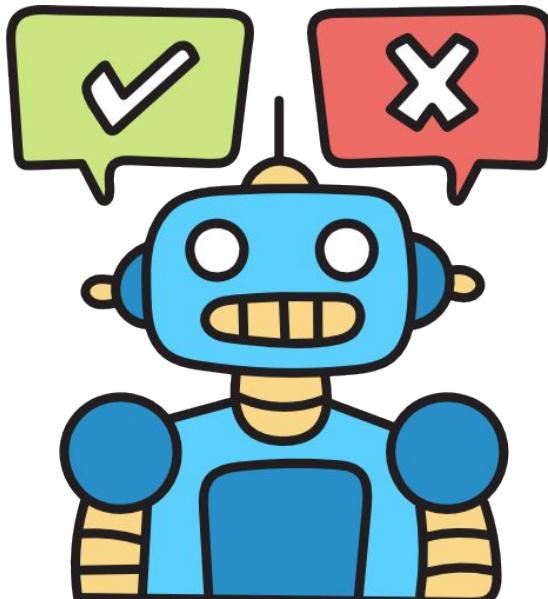
Ou seja, aprender com a experiência que existe intrínseca aos dados.

Como: Algoritmos de aprendizado de máquina analisam as correlações entre os atributos (variáveis) de um sistema (base de dados) a partir de dados amostrais (base de treinamento)



Machine Learning

Ocorre quando um algoritmo ou um programa de computador é utilizado para melhorar as tarefas desempenhadas por humanos com base na experiência



Machine Learning

Tipos de Machine Learning

Supervisionado

Rotulado com saídas esperadas. Modelo gera ao entrar com conjunto de características uma saída rotulada (Classificação) ou um valor futuro (Predição)

Não Supervisionado

Não existe rótulo prévio. Analisa a rede de relacionamento entre os dados para agrupá-los por características similares

Reforço

Maximizar o resultado. Baseado em recompensa e/ou punição. Com isso algoritmo encontra a “política” que mapeia os dados



Machine Learning

Tipos de Machine Learning

Supervisionado

Rotulado com saídas esperadas. Modelo gera ao entrar com conjunto de características uma saída rotulada (Classificação) ou um valor futuro (Predição)

Não Supervisionado

Não existe rótulo prévio. Analisa a rede de relacionamento entre os dados para agrupá-los por características similares

Reforço

Maximizar o resultado. Baseado em recompensa e/ou punição. Com isso algoritmo encontra a “política” que mapeia os dados



Machine Learning

Tipos de Machine Learning

Supervisionado

Rotulado com saídas esperadas. Modelo gera ao entrar com conjunto de características uma saída rotulada (Classificação) ou um valor futuro (Predição)

Não Supervisionado

Não existe rótulo prévio. Analisa a rede de relacionamento entre os dados para agrupá-los por características similares

Reforço

Maximizar o resultado. Baseado em recompensa e/ou punição. Com isso algoritmo encontra a “política” que mapeia os dados



Machine Learning

Tipos de Machine Learning

Supervisionado

Rotulado com saídas esperadas. Modelo gera ao entrar com conjunto de características uma saída rotulada (Classificação) ou um valor futuro (Predição)

Não Supervisionado

Não existe rótulo prévio. Analisa a rede de relacionamento entre os dados para agrupá-los por características similares

Reforço

Maximizar o resultado. Baseado em recompensa e/ou punição. Com isso algoritmo encontra a “política” que mapeia os dados



Machine Learning

Aplicações

Machine Learning

Aplicações

Predição de
Eventos

Machine Learning

Aplicações

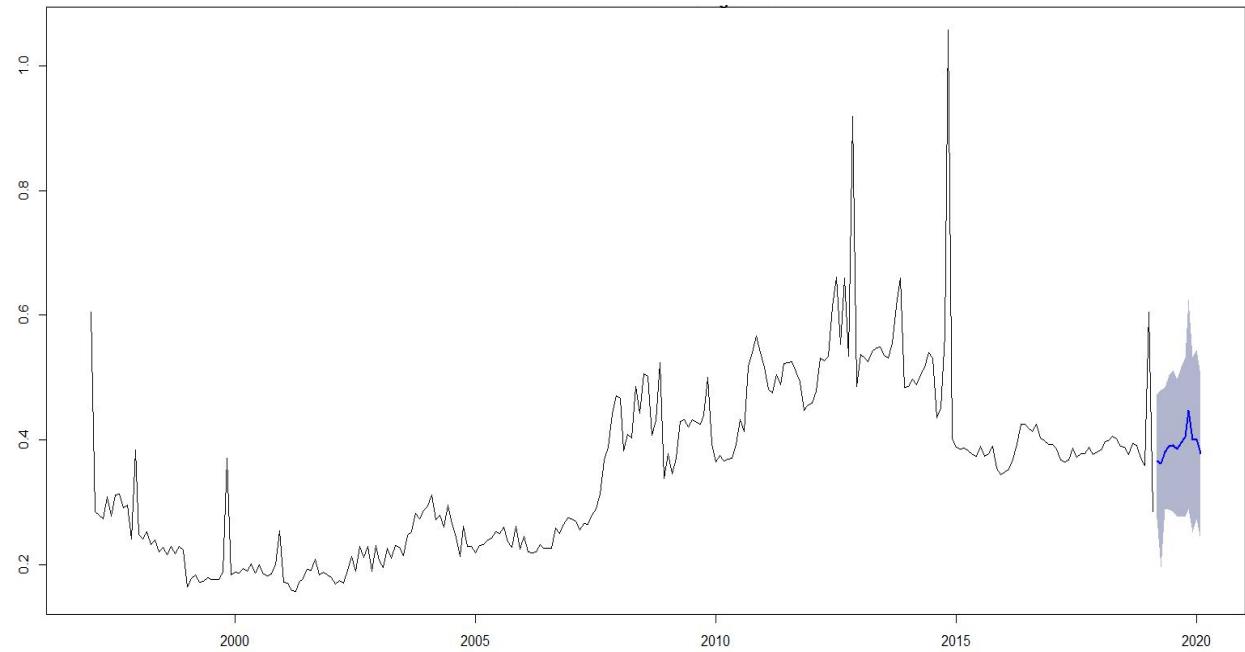
Predição de
Eventos



Machine Learning

Aplicações

Predição de
Eventos



Machine Learning

Aplicações

Predição de
Eventos

Eventos

Temporais:

Vendas, Estoque,
Visitas, Safras de
Grãos, Clima,
Ações Financeiras



Machine Learning

Aplicações

Predição de
Eventos

Anomalias:
Detecção de
Falhas, Fraudes,
Oportunidades,
Invasões (Hackers)



Machine Learning

Aplicações

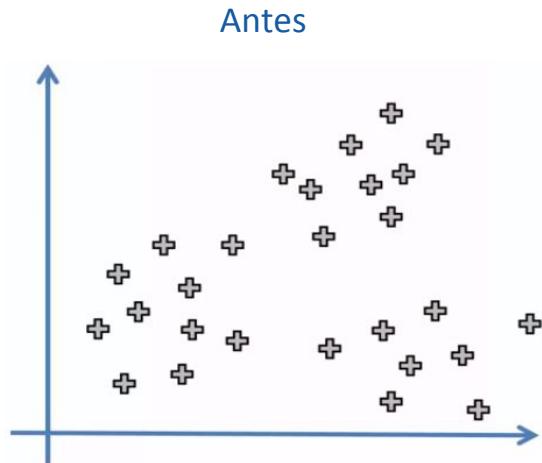
Predição de
Eventos

Sobrevivência:
**Duração de
Equipamentos,
Mortalidade,
Reincidência, Cura**

Machine Learning

Aplicações

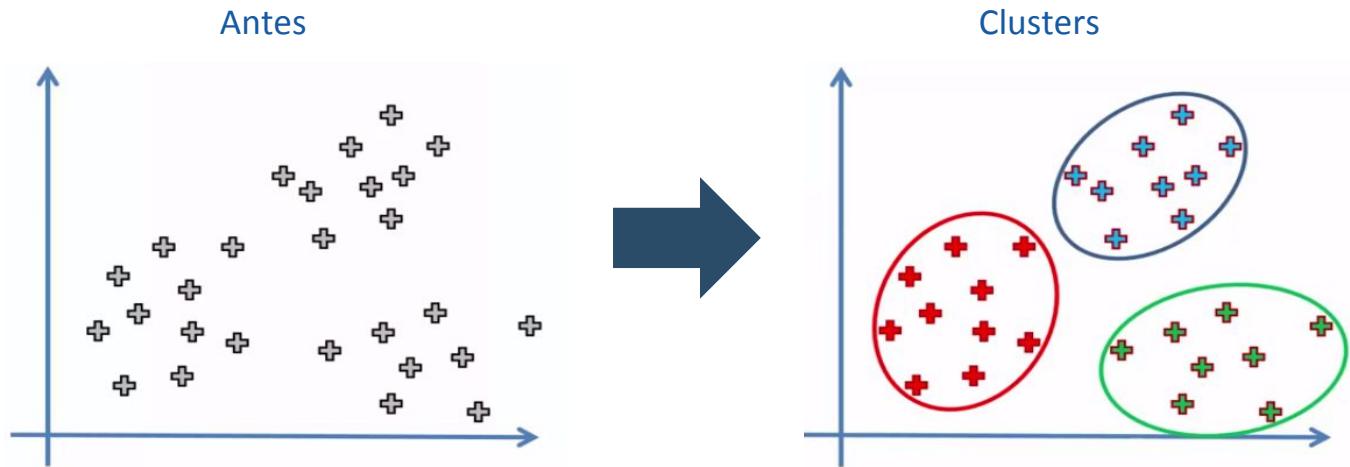
Classificação



Machine Learning

Aplicações

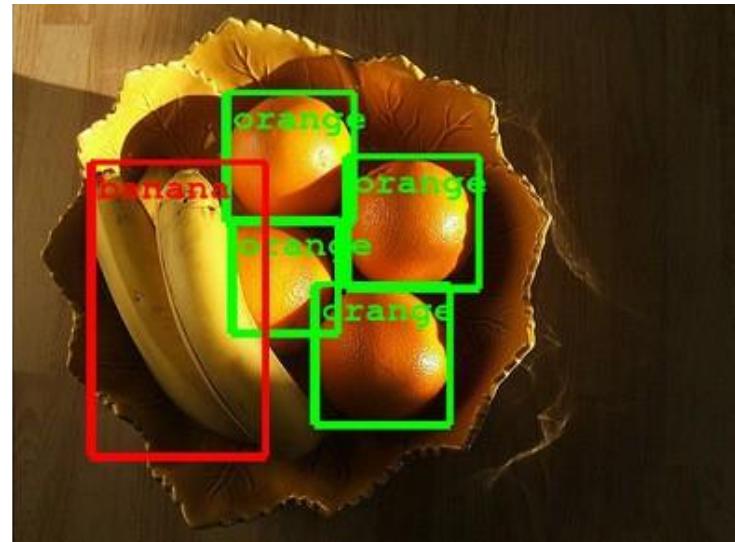
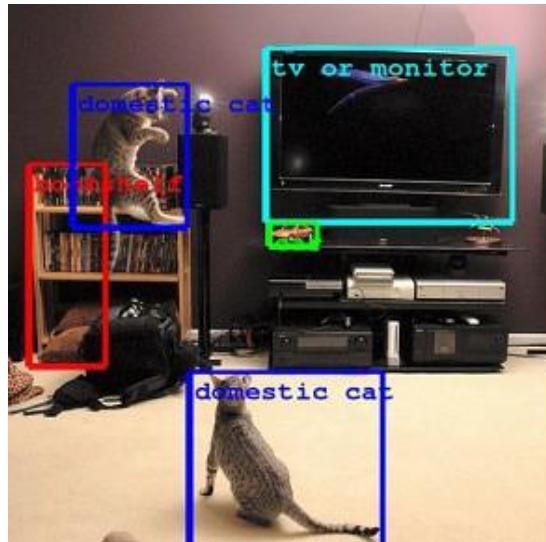
Classificação



Machine Learning

Aplicações

Classificação



Machine Learning

Aplicações

Associação



Machine Learning

Aplicações

Associação

Científica:

Bioinformática,
Diagnóstico,
Mineração de
Dados, Dados
Científicos

Machine Learning

Aplicações

Associação

Entidades:

Cartéis,

Concorrência,

Oportunidades,

Fornecedores,

Quadrilhas



Machine Learning

Aplicações

Associação

E-Commerce:
Transações,
Navegação, Wish
List, Pesquisas,
Clientes



Machine Learning

Fluxo para Construção de Machine Learning

Machine Learning

Fluxo para Construção de Machine Learning



Associação

1 – Machine Learning

2 – Associação

3 – Case MM

Associação

Utilizado para descobrir relacionamentos interessantes em grandes conjuntos de dados

- A metodologia Apriori é a utilizada pela MM
- Motivo: simplicidade e resultados comprovados

Comumente implementado para análise de transações de compras (market basket analysis)

Os resultados são regras com:

- Suporte
- Confiança
- Lift

Associação

Utilizado para descobrir relacionamentos interessantes em grandes conjuntos de dados

- A metodologia Apriori é a utilizada pela MM
- Motivo: simplicidade e resultados comprovados

Comumente implementado para análise de transações de compras (market basket analysis)

Os resultados são regras com:

- Suporte
- Confiança
- Lift

Associação

Utilizado para descobrir relacionamentos interessantes em grandes conjuntos de dados

- A metodologia Apriori é a utilizada pela MM
- Motivo: simplicidade e resultados comprovados

Comumente implementado para análise de transações de compras (market basket analysis)

Os resultados são regras com:

- Suporte
- Confiança
- Lift

Associação

Utilizado para descobrir relacionamentos interessantes em grandes conjuntos de dados

- A metodologia Apriori é a utilizada pela MM
- Motivo: simplicidade e resultados comprovados

Comumente implementado para análise de transações de compras (market basket analysis)

Os resultados são regras com:

- Suporte
- Confiança
- Lift

Associação

Análise de Transações de Compras => Apriori

ID_Transação	Itens_Comprados
1	biscoito, cerveja, chá, salaminho
2	cerveja, couve, lingüiça, pão, queijo
3	café, brócolis, couve , pão
4	brócolis, café, cerveja, couve, pão, salaminho
5	brócolis, café, couve, pão, refrigerante
6	couve, lingüiça

Associação

Funcionamento Apriori

ID_Transação	Itens_Comprados
1	biscoito, <u>cerveja</u> , chá, salaminho
2	<u>cerveja</u> , couve, lingüiça, pão, queijo
3	café, brócolis, couve , pão
4	brócolis, café, <u>cerveja</u> , couve, pão, <u>salaminho</u>
5	brócolis, café, couve, pão, refrigerante
6	couve, lingüiça



{ cerveja => salaminho }

Suporte ou Frequência

$$Suporte = \frac{Freq\{Cerveja \Rightarrow Salaminho\}}{Quantidade\ Transacoes}$$

$$Suporte = \frac{2}{6}$$

$$Suporte = 0,334 \text{ ou } 33,4\%$$

Associação

Funcionamento Apriori

ID_Transação	Itens_Comprados
1	biscoito, <u>cerveja</u> , chá, salaminho
2	<u>cerveja</u> , couve, lingüiça, pão, queijo
3	café, brócolis, couve , pão
4	brócolis, café, <u>cerveja</u> , couve, pão, <u>salaminho</u>
5	brócolis, café, couve, pão, refrigerante
6	couve, lingüiça



{ cerveja => salaminho }

Confiança ou Probabilidade

$$\text{Confiança} = \frac{\text{Freq } \{Cerveja \Rightarrow Salaminho\}}{\text{Qt Tr Cerveja ou Salaminho}}$$

$$\text{Confiança} = \frac{2}{3}$$

$$\text{Confiança} = 0,667 \text{ ou } 66,7\%$$

Associação

Funcionamento Apriori

ID_Transação	Itens_Comprados
1	biscoito, <u>cerveja</u> , chá, salaminho
2	<u>cerveja</u> , couve, lingüiça, pão, queijo
3	café, brócolis, couve , pão
4	brócolis, café, <u>cerveja</u> , couve, pão, <u>salaminho</u>
5	brócolis, café, couve, pão, refrigerante
6	couve, lingüiça



{ cerveja => salaminho }

Lift ou Chances de Ocorrer

$$Lift = \frac{Confiança}{Suporte}$$

$$Lift = \frac{0,667}{0,334}$$

$$Lift = 2$$

Associação

1 – Machine Learning

2 – Associação

3 – Case MM

Case MadeiraMadeira

A compra de
Painel de TV é
precedida pela de
Guarda Roupas,
com:

- 6% de Suporte
- 45% de
Confiança
- 7,5 de Lift



Case MadeiraMadeira

A compra de Painel de TV é precedida pela de Guarda Roupa, com:

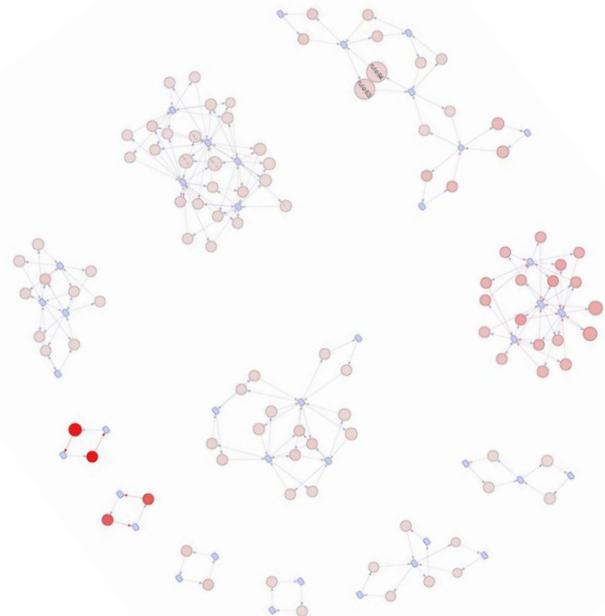
- 6% de Suporte
- 45% de Confiança
- 7,5 de Lift



- Os itens acima aparecem na compra de 6% dos clientes
- Destes, 45% das vezes estes são vendidos juntos
- Tem-se 7,5 vezes mais chances de vender Painel de TV para quem comprou Guarda Roupa

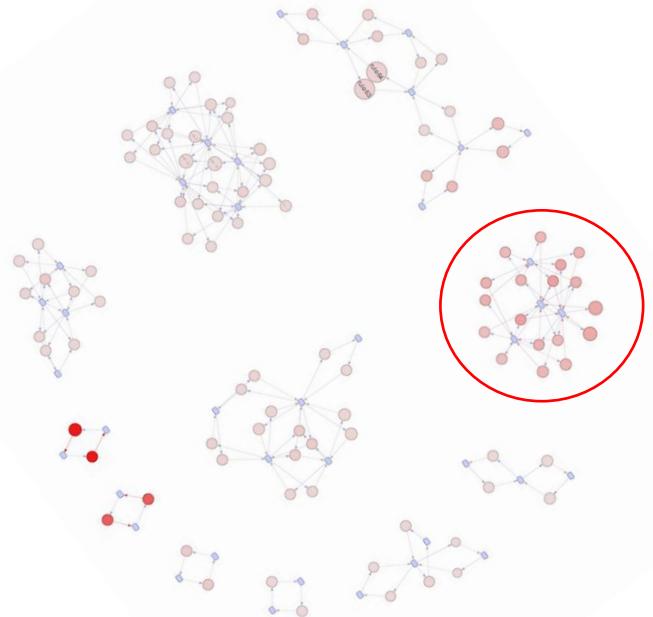
Case MadeiraMadeira

Visualização de regras utilizando o pacote 'arulesViz' do software R



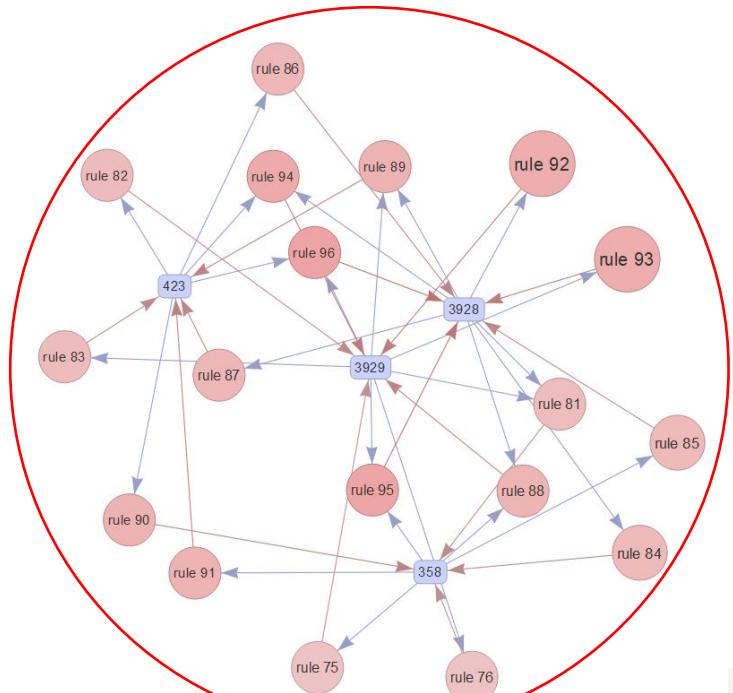
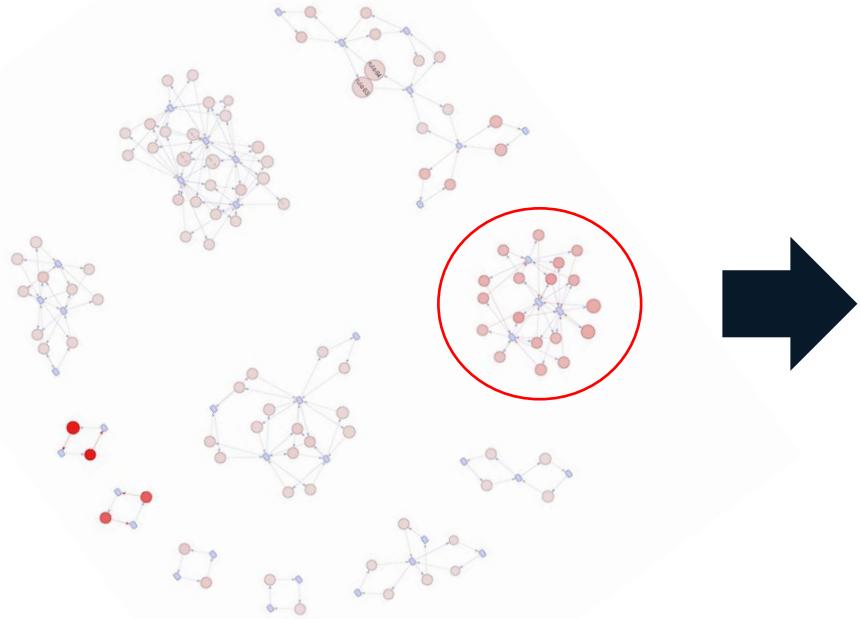
Case MadeiraMadeira

Visualização de regras utilizando o pacote 'arulesViz' do software R



Case MadeiraMadeira

Visualização de regras utilizando o pacote 'arulesViz' do software R



Case MadeiraMadeira

Utilização Prática

Prever a
próxima melhor
oferta (NBO)
dado o último
item que o
cliente comprou

Olá, Rafael

Um produto que pode ser do seu interesse baixou de preço.
Não perca esta chance!



Produto com desconto

Painel para TV até 47 Polegadas com Nicho Flash Artely

Os momentos de maior descontração em frente a tv ou em confraternizações com amigos em casa ocorrem na sala. O painel Flash surgiu para...

Comprar

Recomendamos para você



Painel para TV até 42 Polegadas com Nicho Flash Artely

Comprar



Painel para TV até 42 Polegadas 2 Nichos Navi Móveis...

Comprar



Case MadeiraMadeira

Utilização Prática

Prever a
próxima melhor
oferta (NBO)
dado o último
item que o
cliente comprou

madeiramadeira Busque aqui tudo para sua casa... Chat Online Olá, entre, ou cadastre-se

Departamentos Por Ambiente Móveis Kits & Conjuntos Frete Grátis SP +Design Veja mais

MadeiraMadeira / Móveis / Móveis para Sala de Estar / Painéis / Painéis / 146882



Vendido e entregue por MadeiraMadeira

Painel para TV até 60 Polegadas 4 Prateleiras Veneza Siena Móveis Chocolate Wood

Envio imediato de: 086,98 por **460,70** à vista no cartão (4% desconto) ou 479,90 em até 12x de 39,99 sem juros

Variações deste produto:
 Chocolate Wood
 Off
 White/Nogueira

[Ver detalhes deste produto](#) [Qtd 1](#) [Comprar](#)

Produtos visitados por quem procura este item



< >

Painel para TV até 42 Polegadas Capri Siena Móveis
261,86 **172,70** à vista 12x de 14,99 sem juros

Rack com Painel para TV até 55 Polegadas 1 Porta Miami Siena Móveis
545,86 **316,70** à vista 12x de 27,49 sem juros

Painel para TV até 65 Polegadas Itaparica Siena Móveis
558,92 **335,90** à vista 12x de 29,15 sem juros

Painel para TV até 80 Polegadas com LED 3 Gavetas 182cm Decore I...
1.007,86 **691,10** à vista 12x de 58,99 sem juros

Sala de Estar com Painel, Rack e Mesa de Centro Requisite Siena M...
430,42 **301,35** à vista 12x de 26,15 sem juros

MACHINE LEARNING PARA RECOMENDAÇÕES DE PRODUTOS
Data Sciences

aldeia

Case MadeiraMadeira

Utilização Prática

Juntar as categorias com regras com maiores lifts de vendas conjuntas



Busque aqui tudo para sua casa...



Quer ajuda?



Olá, entre ou cadastre-se



Item	Preço	Pagamento	Juros
Estante Home para TV	604,70	à vista no cartão	até 12x de 50,48 sem juros
Guarda Roupa Casal com Espelho	489,50	à vista no cartão	até 12x de 40,88 sem juros
Guarda Roupa Casal com Espelhos	777,50	à vista no cartão	até 12x de 67,48 sem juros
Painel para TV	691,10	à vista no cartão	até 12x de 58,98 sem juros
Armário Multiuso	364,70	à vista no cartão	até 12x de 31,86 sem juros

Produtos que baixaram de preço

Item	Preço	Pagamento	Juros
Guarda Roupa Casal	739,10	à vista no cartão	até 12x de 61,66 sem juros
Painel para TV	412,70	à vista no cartão	até 12x de 34,38 sem juros
Sala de Estar	301,35	à vista no cartão	até 12x de 25,16 sem juros
Painel para TV	259,10	à vista no cartão	até 12x de 22,48 sem juros
Painel para TV	326,30	à vista no cartão	até 12x de 27,22 sem juros

Case MadeiraMadeira

Utilização Prática

**Utilizar as
regras com
maiores lifts
para montar
conjuntos**



In Data We Trust !

-  **Rafael Roberto Dias**
-  rafael.dias@madeiramadeira.com.br
-  [rafael-roberto-dias-00b39123](https://www.linkedin.com/in/rafael-roberto-dias-00b39123)
-  **41 99672 7170**