

MBA
USP
ESALQ

*Estruturas de Bancos de Dados,
Tipos de Variáveis e Escalas de
Mensuração*

Rafael de Freitas Souza



Data Science

Conjunto de técnicas de programação voltadas à coleta, ao tratamento, à manipulação, à organização, à análise, à extração de informação e à apresentação de dados, na forma de relatórios ou gráficos, visando subsidiar o processo de tomada de decisão.

$$\beta_{yx} = r_{yx} * \frac{S_y}{S_x}, (4)$$

$$\frac{\overline{X-u}}{\sigma} = \sqrt{n-1} \frac{\overline{X-u}}{X^2}$$

$$\sqrt{\frac{S^2}{\sigma^2}}$$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

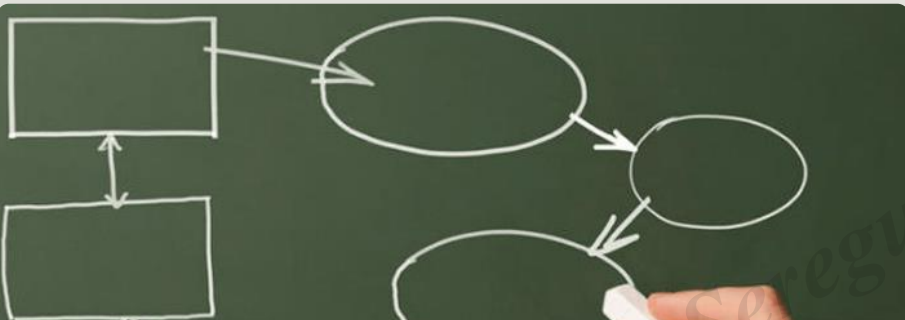
$$\tilde{G}^2(\epsilon) = \tilde{S}^2(\epsilon) = \frac{\sum_{i=1}^n e_i^2}{n-2}, (1)$$

$$\tilde{G}^2(\epsilon) = \tilde{S}^2(\epsilon) = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

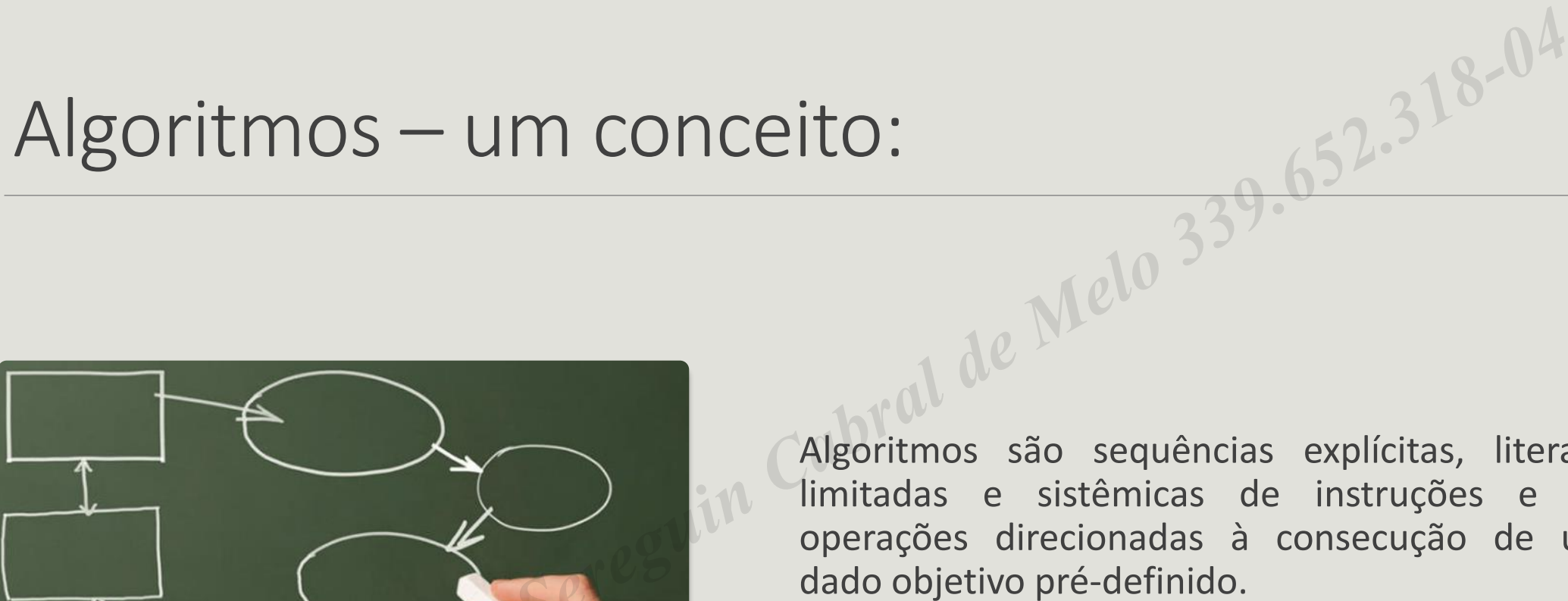
$$\tilde{G}^2(\epsilon) = \tilde{S}^2(\epsilon) = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

O que são algoritmos?

Algoritmos – um conceito:



Algoritmos são sequências explícitas, literais e limitadas e sistêmicas de instruções e operações direcionadas à consecução de um dado objetivo pré-definido.



Algoritmos são sequências explícitas, literais, limitadas e sistêmicas de instruções e de operações direcionadas à consecução de um dado objetivo pré-definido.

Basicamente, qualquer verbo conhecido, desde que denote uma ação destinada a humanos, pode ser considerado um algoritmo.



Link: <https://www.youtube.com/watch?v=pdhqwbUWf4U>

Taxonomia Básica de Espécies de Algoritmos de *Machine Learning*

NÃO SUPERVISIONADOS

Técnicas de *machine learning* assentadas com algoritmos **não supervisionados** (*unsupervised learning*) não possuem capacidade de inferência. Dedicam-se, portanto, a uma análise exploratória, diagnóstica, de dado fenômeno estudado. Exemplos comuns: Análise de *Clusters*, Análise Fatorial de Componentes Principais, Análises de Correspondências Simples e Múltiplas, etc.

SUPERVISIONADOS

Técnicas de *machine learning* assentadas com algoritmos **supervisionados** (*supervised learning*) possuem capacidade de inferência. Dedicam-se, portanto, a uma análise confirmatória, preditiva, de dado fenômeno estudado. Exemplos comuns: Regressões Lineares, Regressões Logísticas, Árvores de Decisão, *Random Forests*, Redes Neurais, etc.



Como escolher os algoritmos?

Passo 1: Algoritmos Não Supervisionados ou Algoritmos Supervisionados?



O primeiro passo é a definição do problema que se quer resolver, seja ele acadêmico ou não.

- Se há objetivos de **fazer inferências** para observações não presentes na amostra que foi utilizada para o treino do algoritmo, **o ideal é a utilização de algoritmos supervisionados**;
- Se há objetivos de **fazer diagnósticos**, sem a intenção de fazer inferências para observações não presentes na amostra que foi utilizada para o treino do algoritmo, **o ideal é a utilização de algoritmos não supervisionados**.

Passo 2: A Construção e a Estruturação de uma Base de Dados

Regra geral, as bases de dados são estruturadas da seguinte maneira: variáveis em colunas e observações em linhas.

id	palavras suspeitas	remetente desconhecido	presença de imagens	classificação
1	sim	não	sim	spam
2	sim	sim	não	spam
3	sim	sim	não	spam
4	não	sim	sim	genuíno
5	não	não	não	genuíno
6	não	não	não	genuíno

Passo 3: Quais as Escalas de Mensuração de suas Variáveis?

A definição incorreta das escalas de mensuração das variáveis da base de dados é um dos principais erros na aplicação das técnicas de *machine learning*. Tal erro é irreparável, implicando no reinício de todo o processo de modelagem, em razão dos vieses criados (e.g.: a ponderação arbitrária).

Em suma: suas variáveis são apenas quantitativas, apenas qualitativas ou há a presença dos dois tipos?



A black and white photograph of numerous wooden letter blocks scattered on a dark surface. In the center, the word "VARIÁVEL" is spelled out using ten blocks, arranged in a gentle upward curve. Other blocks with various letters like 'D', 'O', 'E', 'N', 'L', 'F', 'S', 'H', 'G', 'Y', 'I', 'M', 'X', 'A', 'C', 'T', 'U' are visible in the background and foreground, some partially obscured.

VARIÁVEL

O que é uma variável?

O que é uma variável?

Variáveis podem ser entendidas como uma característica de dada amostra ou população, que pode ser medida, ou contada ou categorizada.

São bons exemplos introdutórios, a altura e/ou o peso das pessoas, suas faixas de renda, a cor e/ou o modelo de carros que dirigem.

Os indivíduos de uma dada amostra ou população, não necessariamente, precisam ser pessoas em seu sentido físico. Podem ser objetos, distritos, municípios, organizações, grupos, células, moléculas, astros, etc. Dessa forma, as características dos indivíduos mencionados, seriam consideradas suas variáveis.

Para o curso, estabeleceremos a escala de mensuração das variáveis em duas: i) variáveis qualitativas; e ii) variáveis quantitativas.

Variáveis Qualitativas

- Também são conhecidas como variáveis latentes ou variáveis categóricas. São variáveis que não podem ser medidas; tão somente, categorizadas ou contadas.
- Por não poderem ser medidas, não permitem o cálculo de estatísticas descritivas de posição – e.g.: a média e a mediana.
- Por outro lado, podemos estabelecer tabelas de frequências para as suas categorias.
- Dividem-se em categóricas nominais e categóricas ordinais.

Variáveis Quantitativas

- Também conhecidas como variáveis métricas, ao contrário das variáveis qualitativas, as variáveis quantitativas podem ser mensuradas, possuindo, por óbvio, uma respectiva unidade de medida.
- Permitem o cálculo da média e da mediana, por exemplo.
- Dividem-se em variáveis contínuas e variáveis discretas.

Obrigado!



Rafael de Freitas Souza

[Linkedin](#)