



Desempenho de Negócios

Aula 04 – Estatística Descritiva

Renato Rodrigues Oliveira da Silva
renato.silva@impacta.edu.br

Sumário

- Classificação de variáveis
- Análise qualitativa (nominal e ordinal)
 - Moda
 - Frequência absoluta e relativa
- Análise quantitativa (discreta e contínua)
 - Medidas de centralidade
 - Média, Mediana
 - Medidas de dispersão
 - Amplitude, Variância, Desvio Padrão

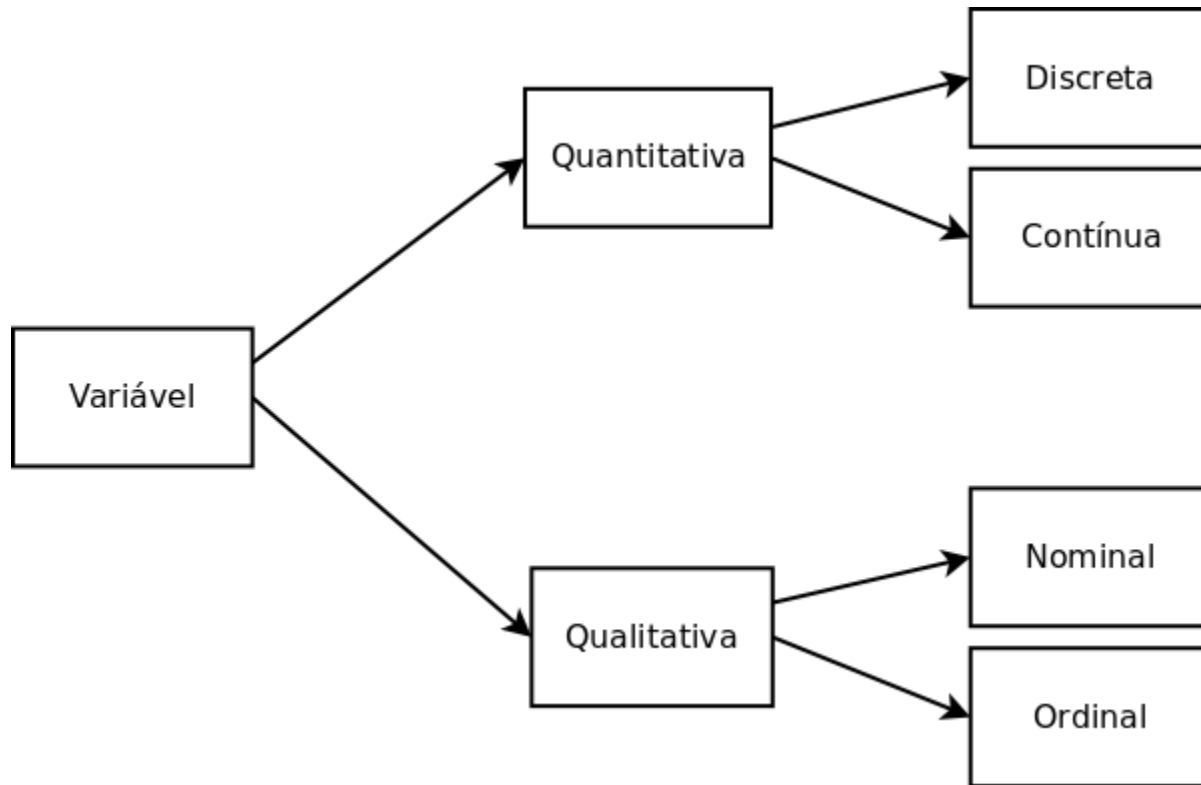
Classificação de variáveis

- Em um processo de análise, as variáveis são os atributos relacionados aos objetos observados
- Exemplo:
 - Idade, Sexo, Grau de instrução, etc.
- As variáveis podem assumir diversos valores, separados em qualitativos e quantitativos
 - Qualitativos: categorias
 - Quantitativos: numéricos

Classificação de variáveis

- As variáveis quantitativas podem ser:
 - **Discretas:** assumem apenas valores inteiros.
 - Número de carros, número de passageiros
 - **Contínuas:** assumem valores em um intervalo real.
 - Preço de um produto, peso de um objeto.
- As variáveis qualitativas podem ser:
 - **Nominais:** categorias sem ordem natural.
 - Sexo, Fumante/Não fumante, Doente/Sadio.
 - **Ordinais:** categorias podem ser ordenadas
 - Tamanho (alto, baixo), Instrução (fundamental, médio)

Classificação de variáveis



Fonte: http://fernandomayer.github.io/ce001n-2016-01/02_Analise_Exploratoria_de_Dados.html

Quiz

- Responder ao quiz no *software* Socrative
 - <https://www.socrative.com/>
 - Login do estudante
 - Sala **RENATOIMPACTA**

Processo de análise dos dados

1. Obtenção de dados

- Leitura de arquivos, bases de dados, internet

2. Análise estatística

- Uso de funções estatísticas para entender o comportamento dos dados

3. Interpretação dos resultados

- Análise de significância estatística, visualizações

Dados Companhia “Milsa”

- Conjunto de dados hipotético elaborado no livro “Estatística Básica” de W. O. Bussab e P. A. Morettin

Funcionario	Est.civil	Inst	Filhos	Salario	Anos	Meses	Regiao
1	solteiro	1oGrau	NA	4.00	26	3	interior
2	casado	1oGrau	1	4.56	32	10	capital
3	casado	1oGrau	2	5.25	36	5	capital
4	solteiro	2oGrau	NA	5.73	20	10	outro
5	solteiro	1oGrau	NA	6.26	40	7	outro
6	casado	1oGrau	0	6.66	28	0	interior
7	solteiro	1oGrau	NA	6.86	41	0	interior
8	solteiro	1oGrau	NA	7.39	43	4	capital
9	casado	2oGrau	1	7.59	34	10	capital
10	solteiro	2oGrau	NA	7.44	23	6	outro

Dados Companhia “Milsa”

Variável

Funcionario

Est.civil

Inst

Filhos

Salario

Anos

Meses

Regiao

Classificação

Quantitativa discreta

Qualitativa nominal

Qualitativa ordinal

Quantitativa discreta

Quantitativa contínua

Quantitativa discreta

Quantitativa discreta

Qualitativa nominal

Análise Qualitativa Nominal

- Para variáveis qualitativas nominais, podemos obter:
 - Uma tabela de **frequências** (absolutas ou relativas)
 - Absoluta: Contagem de cada valor
 - Relativa: Proporção de cada valor, em relação ao total
 - Um gráfico de barras ou de “pizza”
 - A “moda” da variável: o valor que mais ocorre

Análise Qualitativa Ordinal

- A análise desse tipo de variável é feita de modo semelhante à análise do tipo anterior
 - Registro de frequências, moda.
- No entanto, agora existe uma ordenação natural das categorias
 - Afeta a exibição dos resultados, devendo respeitar a ordenação.
 - Necessário definir de antemão qual é essa ordenação

Análise Qualitativa

- **Exercício**
 - Computar a **frequência absoluta** da variável “estado civil” do conjunto Milsa.csv
 - Computar a **frequência relativa** da mesma variável
 - Computar a **moda** da variável

Análise Quantitativa

- Para variáveis **discretas**, pode ser feita a análise de frequências, do mesmo modo que as categorias anteriores.
 - Nesse caso assumimos que cada valor numérico é uma categoria
 - Caso existam muitos valores diferentes, a análise de frequências pode não ser adequada

Análise Quantitativa

- Tanto no tipo de variável quantitativa **discreta** quanto na **contínua**, podemos utilizar medidas-resumo para entender o seu comportamento
 - Média
 - Mediana

Medidas de Centralidade

- **Média**

- Indica o valor onde estão concentrados os dados

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \longrightarrow \frac{4 + 36 + 45 + 50 + 75}{5} = \frac{210}{5} = 42$$

- **Mediana**

- Valor que separa a metade maior e a menor da amostra
 - Menos propenso a “*outliers*” que a média

Medidas de Dispersão

- **Amplitude**

- A diferença entre o maior e o menor valor

- **Variância**

- Valor médio do quadrado dos desvios dos elementos em relação à média
- Mostra o quanto cada elemento está distante da média

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$$

Medidas de Dispersão

- **Desvio Padrão**

- Indica o “erro” em cada observação, caso quiséssemos substituir uma observação pela média

$$\sigma = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N}}$$

Exemplos

- Conjunto de Dados: [1,2,3,4,5]
- Média = $(1+2+3+4+5)/5 = 3$
- Mediana = 3
- Amplitude = $5 - 1 = 4$
- Variância = 2.5
- Desvio Padrão = 1,58

Análise Quantitativa

- **Exercício**
 - Calcular a **amplitude, média e mediana** da quantitativa “Salário”

Referências

- **Curso de Análise Exploratória de Dados**
Fernando Mayer
 - http://fernandomayer.github.io/ce001n-2016-01/02_Analise_Exploratoria_de_Dados.html

Referências

- Bussab, W.O., Morettin, P.A., **Estatística Básica**, Editora Saraiva.

