

Conceitos Básicos

Rosineide da Paz

2022-05-03

Contents

1	Introdução	5
2	Análise Exploratória dos Dados	7
2.1	Variável	7
2.2	Escala de Likert para variáveis qualitativas	9
3	Estrutura de Dados e Notação	11
3.1	Notação	12

Chapter 1

Introdução

Neste texto são apresentados conceitos básicos para o entendimento das primeiras etapas da análise estatística. Inicialmente, veja alguns conceitos importantes.

- **População:** conjunto de elementos com pelo menos uma característica em comum que o delimita, podendo ser pensada como o conjunto que contém todos os indivíduos, fenômenos ou resultados que se pretende investigar.
- **Amostra:** um subconjunto da população, uma parte que pode ou não representar a população inteira.
- **Variável:** característica dos elementos da população que se deseja investigar.
- **Dados:** observações de uma ou mais variáveis a partir de elementos de uma população.
- **Censo:** as observações são realizadas a partir de todos os elementos da população.
- **Dados Amostrais:** as observações são realizadas a partir de uma amostra, ou seja, de parte da população.
- **Unidade Amostral:** entidade (ou elemento) da população sobre a qual a característica de interesse (a variável) é observada.

Exemplo 1.1. Considere que o interesse seja a investigação do número de focos do mosquito *Aedes aegypti* por bairro na cidade de Russas. Neste caso:

população: conjunto formado pelos bairros;

amostra: conjunto que não contém todos os bairros, mas parte deles;

variável: quantidade de focos (um número inteiro positivo);

dados: resultados de observações de focos por bairro;

censo: resultado de observações de focos de todos os bairros;

dados amostrais: resultados de observações de focos por bairro, considerando apenas os bairros que compõem a amostra;

unidades amostrais: bairros que serão efetivamente observados.

Chapter 2

Análise Exploratória dos Dados

Em qualquer análise estatística, deve-se, inicialmente, realizar uma análise exploratória dos dados, isso pode ser feito usando ferramentas da **Estatística Descritiva**. Nessa etapa da análise, procura-se obter a maior quantidade possível de informações a partir dos dados observados. Se a análise é realizada tendo como base uma amostra da população de interesse, é nessa fase que são obtidas as informações que podem dar ideia sobre qual modelo probabilístico pode descrever melhor o fenômeno investigado, assim pode-se ter uma ideia do modelo a ser utilizado na fase posterior da análise.

2.1 Variável

Os dados são observações de fenômeno de interesse, a esses fenômenos damos o nome de **variável**. Assim, uma variável é um atributo, cujo valor pode variar de uma unidade de investigação para outra. Por exemplo, as unidades podem ser pessoas contaminadas por um vírus e a variável o tempo de recuperação após um determinado tipo de tratamento ser iniciado. Nesse caso, espera-se que o tempo de recuperação varie entre os indivíduos, ou unidades amostrais.

Uma observação, ou resposta, é o valor assumido por uma variável em uma das unidades investigadas. A observação da variável em várias unidades dá origem aos dados observados.

Exemplo 2.1. Exemplos de variáveis.

- tempo de execução de um algoritmo em diferentes máquinas;
- rendimento das famílias de uma grande cidade;
- número de erros em pacotes de dados enviados por um servidor;

- número de clientes com a mesma dúvida em um site de suporte durante um período de tempo;
- opinião dos consumidores de um determinado produto (péssimo, regular, ótimo etc).

2.1.1 Tipos de Variáveis

Podemos classificar variáveis em qualitativas e quantitativas. As variáveis **qualitativas** têm seus valores expressos por atributos, exprimindo alguma qualidade.

As variáveis qualitativa podem ainda ser classificadas como **nominais** ou **ordinais**.

Uma variável é nominal se os atributos que esta representa não têm uma ordenação, por exemplo, cor de cabelo, sexo de indivíduos etc, enquanto que as variáveis ordinais exprimem alguma ordenação, como por exemplo, opinião sobre a qualidade de um produto (péssimo, regular, ótimo).

As variáveis *quantitativas* exprimem quantidades ou mensurações. Essas variáveis também podem ser classificadas em **contínuas** e **discretas**.

Variável discreta assume valores em um conjunto enumerável (ou contável), por exemplo, número de carros que passam por um posto de pedágio em um intervalo de tempo.

Uma variável é contínua quando assume valores em um conjunto não-enumerável, ou seja, em um intervalo da reta, por exemplo, alturas de pessoas em um determinado povoado. Veja um resumo dos tipos de variáveis no diagrama apresentado na Figura 2.1, criado com o uso do pacote [?].

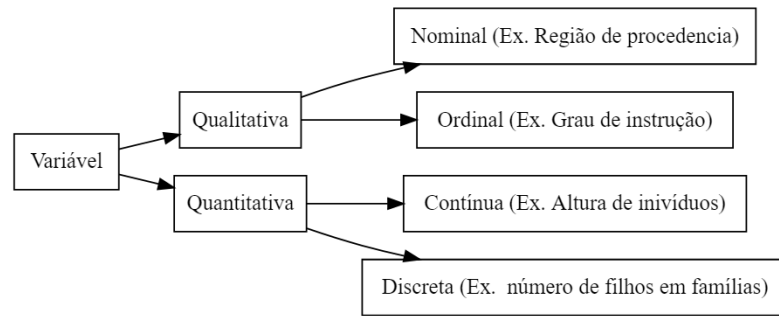


Figure 2.1: : Classificação de variáveis.

Note que, na prática, em geral variáveis quantitativas contínuas são resultados de mensurações, enquanto variáveis quantitativas discretas são resultados de

contagens. Portanto, para entender a natureza de uma variável, e classificá-la corretamente, deve-se entender o processo que a originou.

2.2 Escala de Likert para variáveis qualitativas

Variáveis qualitativas muitas vezes podem ser observadas usando a escala de Likert [?]. Essa escala pode ser uma boa alternativa aos dados categóricos, pois permite o uso de técnicas para análise que não seriam possíveis de serem utilizadas em dados categóricos.

Essa escala é principalmente utilizada em questionários para obter as preferências ou o grau de concordância do entrevistado com uma declaração ou conjunto de declarações. Essa é uma técnica de escala em que os entrevistados são solicitados a indicar seu nível de concordância com uma determinada declaração por meio de uma escala ordinal ou intervalar.

Exemplo 2.2. Como um exemplo de variável que pode ser observada em uma escala Likert, suponha que é desejado saber o quanto um cliente está satisfeito com um atendimento. Neste caso, pode-se perguntar:

em uma escala de 0 até 5, em que zero significa muito insatisfeito e 5 muito satisfeito, qual é o seu grau de satisfação com esse atendimento?

Para essa pergunta pode-se estabelecer valores intermediários (1, 2, 3 e 4), ou usar o intervalo de 0 até 5, em que a pessoa possa escolher qualquer valor no intervalo $[0,5]$.

Chapter 3

Estrutura de Dados e Notação

Os dados normalmente são dispostos em tabelas ou planilhas, de modo que em cada coluna podem ser observados os valores assumidos por uma dada variável. Assim, nas linhas da tabela são dispostos os valores observados para todas as variáveis em uma unidade amostral ou indivíduo. Essa estrutura permite que os dados sejam analisados por meio de diversos software, tais como R [R Core Team, 2021], Python [Van Rossum and Drake Jr, 1995], Library Office [Library Office Team, 2020], entre outros. No entanto, outras estruturas podem ser requeridas, dependendo das rotinas a serem utilizadas.

A Tabela 1 mostra as dez primeiras linhas e sete primeiras colunas do conjunto de dados coletados por meio de questionários respondidos por estudantes ingressantes nos cursos de engenharias do Campus da UFC de Russas no ano de 2020. Os discentes responderam as seguintes perguntas:

- Qual o local de residência no ato da matrícula no curso que está matriculado?
- Qual o seu sexo?
- Qual é sua idade, peso e altura?
- Qual curso está matriculado?
- Qual foi a sua nota no ENEM?
- Com qual idade decidiu fazer o curso que está cursando?
- Numa escala de 0 a 5, onde 0 significa nada e 5 significa muito, com que intensidade você gosta da área de exatas?

Table 3.1: Dados fornecidos por estudantes ingressantes do ano de 2020 nos cursos de engenharias do Campus da UFC de Russas.

	Cidade	UF	Sexo	Idade	Peso	Altura
1	Limoeiro do Norte	CE	Feminino	17	40.2	1.61
2	Aracati	CE	Feminino	17	100.0	1.60
3	Limoeiro do Norte	CE	Masculino	19	60.0	1.65
4	Quixadá	CE	Masculino	20	80.3	1.73
5	Russas	CE	Masculino	17	59.0	1.68
6	Limoeiro do Norte	CE	Masculino	19	75.0	1.80
7	Jaguaribe	CE	Masculino	18	63.0	1.75
8	Jaguaruana	CE	Feminino	18	57.0	1.53
9	Fortaleza	CE	Feminino	19	74.0	1.65
10	Russas	CE	Masculino	20	73.0	1.70

Note que a tabela mostra valores para as variáveis “local de residência dos estudantes”, “sexo”, idade, peso, curso, nota, idade de decisão pelo curso e gosto pelas área de exatas. Cada uma dessas variáveis foram observadas nos indivíduos entrevistados, os alunos. Assim, essas características são dispostas nas colunas da tabela e os indivíduos (que são anônimos) estão dispostos nas linhas da tabela.

3.1 Notação

Uma variável qualquer é geralmente denotada por uma letra maiúscula, exemplo, X , Y , etc. Um valor observado é denotado pela mesma letra, mas minúscula, exemplo, se a variável é denotada por X , então um valor observado dessa variável será denotado por x . Se for necessário representar uma sequência de n valores observados da variável X , essa sequência será denota por x_1, x_2, \dots, x_n . Uma sequência infinita de valores observados da variável X será denotada por x_1, x_2, \dots . Caso tenha sido observados dados a partir de todos os elementos da população, o tamanho da sequência de dados observados será denotado por N .

Resumidamente tem-se:

- X , Y , etc, representa variáveis;
- x e x_1, x_2, \dots, x_n representam um valor observado e uma sequência de tamanho n de valores observados de X , respectivamente;
- n representa a quantidade de elementos na amostra usada para obter os valores observados e
- N representa a quantidade de elementos na população.

Observações

1. Após a obtenção dos dados, é comum que seja feita uma crítica dos valores obtidos, pois erro de medições e equívocos na anotação de valores podem ocorrer. Assim, é comum que sejam passados “filtros” na tentativa de detectar esses valores.
2. Os dados após feita essa crítica é denominado **dados brutos**.

Bibliography

Library Office Team. *Library Office: The Document Foundation*. Library Office, DF, Alemanha, 2020. URL <https://pt-br.libreoffice.org/>.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.

Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.