

Aula 01: Tipos de dados, Distribuição de frequência & Gráficos

Estatística e Probabilidades

André Victor Ribeiro Amaral (sala 3029)

avramaral@gmail.com

Revisão – Exercício

Exercício: Para as situações descritas a seguir, identifique a população e a amostra correspondente.

1. Para avaliar a eficácia de uma campanha de vacinação no estado de São Paulo, 200 mães de recém nascidos durante o primeiro semestre de um dado ano e em uma dada maternidade do estado foram entrevistadas a respeito da última vez que vacinaram seus filhos.
2. Para verificar a audiência de um programa de TV, 563 indivíduos foram entrevistados por telefone com relação ao canal em que estavam sintonizados.
3. A fim de avaliar a intenção de voto para presidente, 122 pessoas foram entrevistadas em Brasília.

Tipos de dados

Em relação às variáveis, podemos classificá-las como:

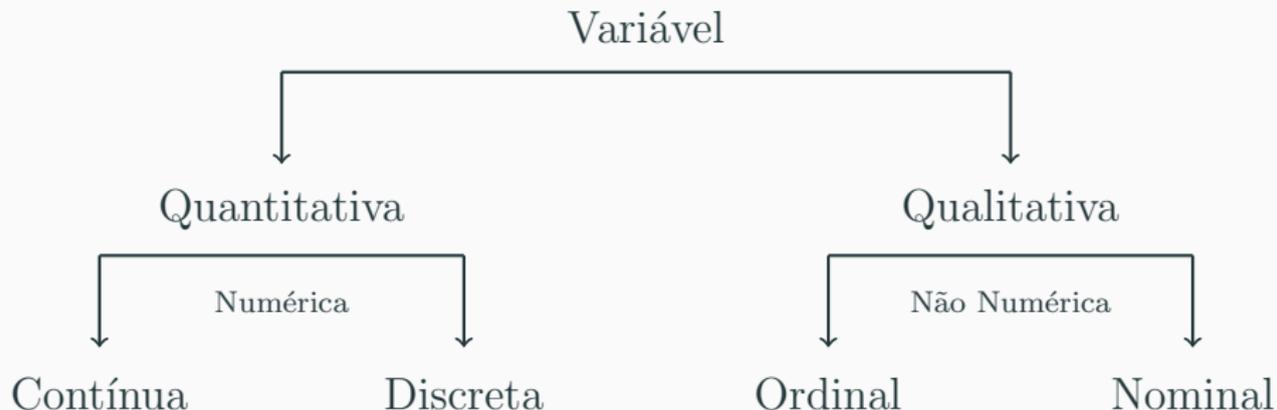


Figura 1: Tipos de dados.

Tipos de dados

1. Quantitativo

- Discreto: O conjunto resposta é enumerável. Em geral, é resultado de um processo de contagem. Exemplo: idade (em anos completos), número de irmãos/irmãs, etc.
- Contínuo: O conjunto resposta é não-enumerável. Potencialmente um subconjunto dos números reais. Exemplo: temperatura, tempo, peso, etc.

2. Qualitativo (Variáveis Categóricas)

- Nominal: O conjunto resposta não possui ordenação. Exemplo: raça, sexo, etc.
- Ordinal: O conjunto resposta possui uma ordenação natural. Exemplo: classe social, escolaridade, etc.

Tipos de dados – Exercícios

Exercício: Classifique cada variável abaixo em qualitativa (nominal ou ordinal) ou quantitativa (discreta ou contínua).

1. Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (“sim” ou “não” são as respostas possíveis).
2. Intenção de voto para presidente (“branco”, “nulo”, “Candidato X”, e “não sei” são as respostas possíveis).
3. Perda de peso (kg) de corredores na Corrida de São Silvestre.
4. Intensidade da perda de peso de corredores na Corrida de São Silvestre (“leve”, “moderada” e “forte” são as respostas possíveis).
5. Grau de satisfação de um visitante de determinado *website* (valores inteiros de 0 a 5 são as respostas possíveis).

Tabela de frequência

Assuma o seguinte conjunto de idades (em anos), com $n = 36$:

17, 18, 18, 18, 19, 20, 18, 21, 20, 22, 21, 19, 21, 22, 23, 23, 22, 22,
18, 19, 18, 20, 21, 18, 21, 22, 18, 21, 17, 20, 21, 21, 18, 21, 20, 19,

com frequência relativa $f_i = \frac{n_i}{n}$.

Aqui,

- n_i : é a contagem de cada classe de respostas; e
- f_i : é a frequência relativa para cada classe de respostas.

Além disso, podemos definir frequência acumulada como

$$f_{ac}^{(i)} = \sum_{j=1}^i f_j.$$

Tabela de frequência

Classe (Idade)	n_i	f_i	$f_{ac}^{(i)}$
17	2	0.06	0.06
18	9	0.25	0.31
19	4	0.11	0.42
20	5	0.14	0.56
21	9	0.25	0.81
22	5	0.14	0.95
23	2	0.05	1.00
total	$n = 36$	1	—

Tabela 1: Tabela de frequência (discreto) para conjunto de idades.

Tabela de frequência – Exercício

Exercício: Suponha que o número de alunos matriculados na turma de Estatística e Probabilidades, da UFMG, no 1º semestre de 2020, tenha sido de 42 alunos. Ao final do semestre, 5 indivíduos ficaram com conceito A, 6 tiraram B, 8 ficaram com C, 8 com D, 5 com E e os demais com F. Nesse caso, construa a tabela de frequência apropriada.

Tabela de frequência – Exercício

Resposta:

Classe (Conceito)	n_i	f_i	$f_{ac}^{(i)}$
A	5	0.119	0.119
B	6	0.143	0.262
C	8	0.190	0.452
D	8	0.190	0.642
E	5	0.119	0.761
F	10	0.239	1.000
total	$n = 42$	1	–

Tabela 2: Tabela de frequência para o exercício proposto.

Tabela de frequência

Assuma o seguinte conjunto de pesos (em kg), com $n = 24$:

60.8, 80.1, 70.3, 50.6, 52.0, 71.9, **99.9**, 90.0, 79.8, 96.1, 59.4, **50.1**,
58.8, 61.9, 70.2, 90.9, 59.9, 65.8, 99.2, 80.6, 70.8, 75.2, 55.1, 73.2.

Classe (Peso)	n_i	f_i	$f_{ac}^{(i)}$
[50, 60)	7	0.29	0.29
[60, 70)	3	0.12	0.42
[70, 80)	7	0.29	0.71
[80, 90)	2	0.08	0.79
[90, 100]	5	0.21	1.00
total	$n = 24$	1	—

Tabela 3: Tabela de frequência (contínuo) para conjunto de pesos.

Tabela de frequência

Para o caso anterior (dos pesos), perceba que acabamos por definir a amplitude de cada uma das classes de maneira arbitrária; nesse caso, 10 kg.

Via de regra, queremos definir um tamanho de amplitude tal que existam algo entre 5 e 8 classes. Assim:

$$\text{ampl.} = \frac{\text{max} - \text{min}}{\text{n}^\circ \text{ de classes}},$$

onde n° de classes $\in \{5, 6, 7, 8\}$.

No nosso exemplo, $\frac{99.9-50.1}{5} = 9.96 \approx 10$.

Tabela de frequência – Exercício

Exercício: Considerando o mesmo conjunto de dados com o qual estamos trabalhando (mín.: 50.1 kg e máx.: 99.9 kg, para $n = 24$), determine a amplitude e os novos limites de classe tal que “nº de classes” = 8.

Tabela de frequência – Exercício

Nesse caso, temos que

$$\text{ampl.} = \frac{99.9 - 50.1}{8} = 6.225 \approx 6.5.$$

Dessa forma, podemos definir as seguintes classes: $[50, 56.5)$, $[56.5, 63)$, $[63, 69.5)$, $[69.5, 76)$, $[76, 82.5)$, $[82.5, 89)$, $[89, 95.5)$, $[95.5, 102]$.

Tabela de frequência

Algumas observações:

1. Mesmo para variáveis discretas, no caso de existirem muitas classes (> 8 , como regra geral), também podemos utilizar a estratégia de dividir, por exemplos, as “idades” em intervalos apropriados.
2. Perceba que a utilização de tabelas de frequência oferece a vantagem de sumarizar os dados. Nesse sentido, vamos estudar mais a frente outras formas de resumir um determinado conjunto de informações.

Tabela de frequência

Também podemos construir tabelas bivariadas. Suponha que, para o conjunto de idades que listamos no Slide 6, os 18 primeiros indivíduos são do sexo feminino e os 18 indivíduos seguintes são do sexo masculino; então:

	Faixa Etária			
Sexo	[17, 19)	[19, 21)	[21, 23]	Total
F	5 (27.8%)	4 (22.2%)	9 (50.0%)	18 (100%)
M	6 (33.3%)	5 (27.8%)	7 (38.9%)	18 (100%)
Total	11 (30.5%)	9 (25.0%)	16 (44.5%)	36 (100%)

Tabela 4: Tabela de frequência bivariada (com soma por linha).

Tabela de frequência

	Faixa Etária			
Sexo	[17, 19)	[19, 21)	[21, 23]	Total
F	5 (27.8%)	4 (22.2%)	9 (50.0%)	18 (100%)
M	6 (33.3%)	5 (27.8%)	7 (38.9%)	18 (100%)
Total	11 (30.5%)	9 (25.0%)	16 (44.5%)	36 (100%)

Comentários:

1. 27.8% das mulheres têm menos que 19 anos, ao passo que 33.3% dos homens pertencem à essa faixa etária;
2. Existe maior porcentagem de indivíduos (independente do sexo) entre 21 e 23 anos (inclusive), etc.

Gráfico

A fim de facilitar a visualização e interpretação de um determinado conjunto de dados, podemos utilizar diferentes tipos de **gráficos** para representá-lo.

Via de regra, é necessário que um gráfico contenha:

- Título (quando o gráfico está inserido em um documento que apresente contexto das informações, o “título” pode ser dispensado);
- Eixo horizontal;
- Eixo Vertical; e
- Legenda.

Gráfico de barra

Gráfico de barra: esse tipo de gráfico se adapta melhor às variáveis discretas ou qualitativas ordinais.

Sobre o Slide 7,

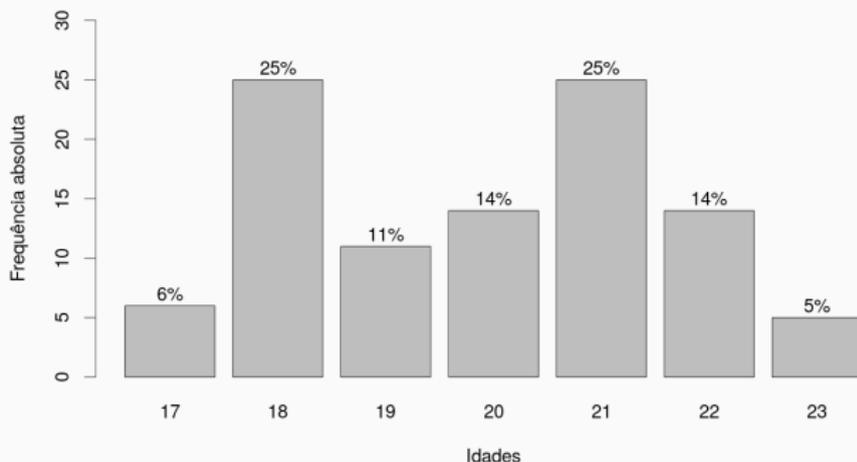


Figura 2: Gráfico de barras para frequência de idades.

Gráfico de linhas

Gráfico de linhas: aplicável, principalmente, para séries temporais. Exibe um a evolução de uma variável ao longo de, por exemplo, o tempo.

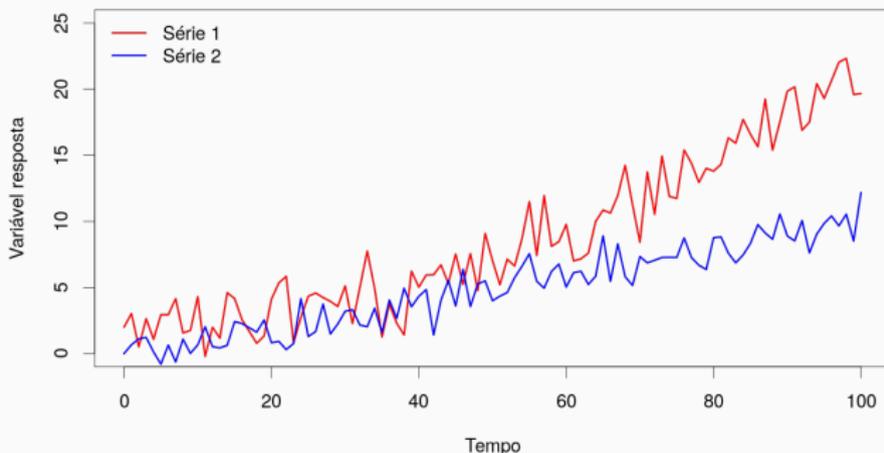


Figura 3: Gráfico de linhas para duas séries temporais arbitrárias.

Histograma

Histograma: utilizado para representar variáveis contínuas. Como exemplo, vamos fazer referência à tabela de frequência que construímos no Slide 10.

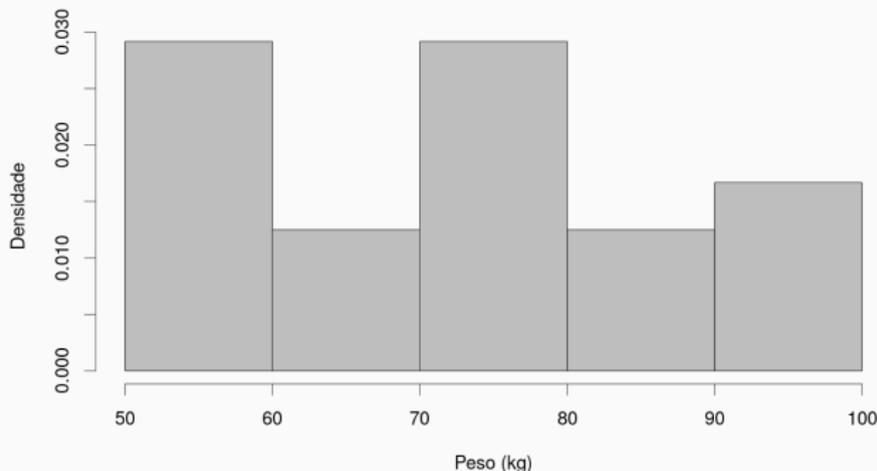
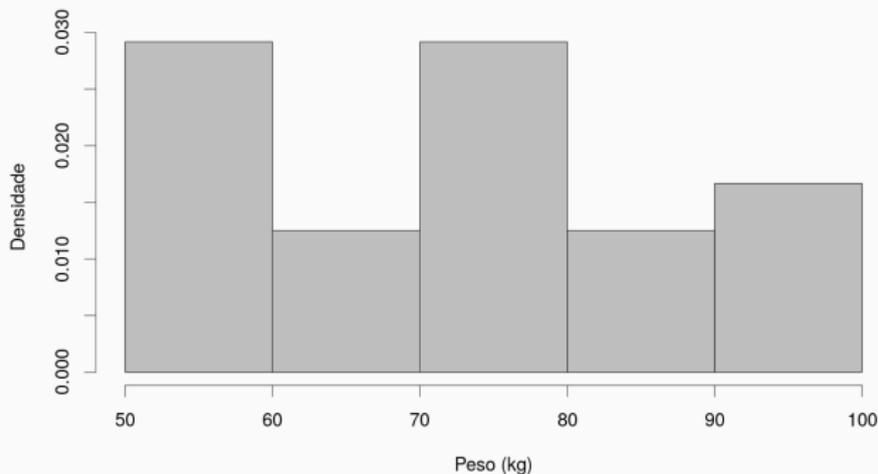


Figura 4: Histograma para os pesos (em kg).

Histograma

Como observação,



note que, no eixo vertical, estamos utilizando a medida de densidade d_i (em contraponto à frequência absoluta); tal que $d_i = \frac{f_i}{\text{ampl.}}$.

Gráfico de pizza

O problema com o gráfico de pizza:

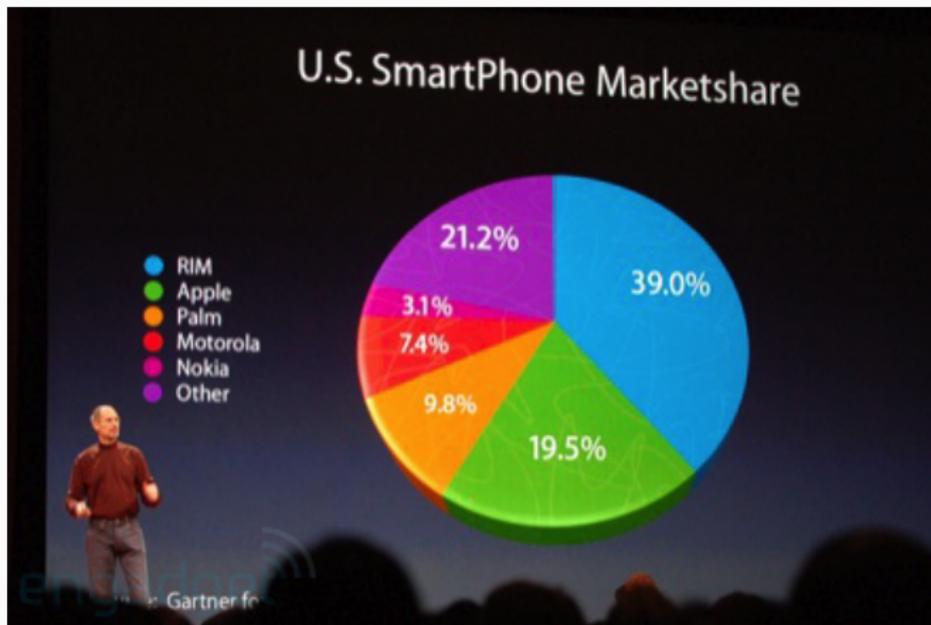


Figura 5: Fonte: <https://bit.ly/2T6UD1B>.