

MAT02018 - Estatística Descritiva

Distribuição de Frequências (dados contínuos)

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2021



Distribuição de Frequências

Dados contínuos

- ▶ Dados contínuos podem assumir **diversos valores diferentes**¹, mesmo em amostras pequenas.
- ▶ Por essa razão, a menos que sejam em grande número, são apresentados na forma como foram coletados.

¹Aqui chamamos mais uma vez a atenção para a importância de distinguirmos os diferentes tipos de variáveis. Uma variável *quantitativa contínua* é uma **variável**! E portanto, **pode variar** de um indivíduo para outro! No entanto, a variável *quantitativa contínua* possui um conjunto de valores possíveis **infinito** (um intervalo da reta real), e assim, podemos observar um número de unidades com valores distintos para uma certa variável contínua maior que no caso de uma variável nominal. **Exercício:** compare os valores possíveis para as variáveis **altura** e **estado civil**.

Dados contínuos

- ▶ Considere, como exemplo, que o pesquisador resolveu organizar as idades dos empregados da seção de orçamentos da Companhia MB em uma tabela.
- ▶ Pode escrever os dados na ordem em que foram coletados, como segue:

26	20	41	23	37
32	40	43	33	44
36	28	34	27	30

Dados contínuos

- ▶ Quando em grande número, os dados contínuos podem ser organizados, para apresentação, em uma tabela de distribuição de frequências.
- ▶ Vamos entender como isso é feito por meio de novo exemplo.

Dados contínuos

- ▶ Foram propostas muitas maneiras de avaliar a capacidade de uma criança para o desempenho escolar.
- ▶ Algumas crianças estão “prontas” para aprender a escrever aos cinco anos, outras, aos oito anos.
- ▶ Imagine que um professor aplicou o *Teste de Desempenho Escolar* (TDE) a 27 alunos da 1ª série do Ensino Fundamental.
- ▶ Os dados obtidos pelo professor estão apresentados em seguida.

7	25	81	95	100	99	95	105	117
18	101	75	98	94	84	102	100	96
111	85	100	108	34	90	96	107	17

Dados contínuos

- ▶ Para **conhecer o comportamento** do desempenho escolar desses alunos, o professor deve organizar uma **distribuição de frequências**.
- ▶ No entanto, para isso, é preciso **agrupar os dados em faixas**, ou **classes**².
 - ▶ Em quantas faixas ou classes podem ser agrupados os dados?

²Note que se procedermos da mesma forma que procedemos para os casos anteriores, a nossa tabela de distribuição de frequências apresentaria um grande número de valores com baixas frequências. Isso nos daria tanta informação quanto a tabela de dados brutos, e portanto, não nos ajudaria a conhecer o comportamento da variável.

Dados contínuos

- ▶ Uma **regra prática** é a seguinte: **o número de classes deve ser aproximadamente igual à raiz quadrada do tamanho da amostra.**

$$\text{Número de classes} = \sqrt{n}.$$

- ▶ No exemplo, são 27 alunos.
 - ▶ O tamanho da amostra é, portanto, $n = 27$.
 - ▶ A raiz quadrada de 27 está entre $5(\sqrt{25})$ e $6(\sqrt{36})$. Portanto, podem ser organizadas **cinco classes**.
 - ▶ Mas como?

Dados contínuos

- ▶ Observe cuidadosamente o conjunto de dados.
 - ▶ Ache o **valor mínimo**, o **valor máximo** e a **amplitude**.
-
- ▶ **Valor mínimo** é o menor valor de um conjunto de dados.
 - ▶ **Valor máximo** é o maior valor de um conjunto de dados.
 - ▶ **Amplitude** é a diferença entre o valor máximo e o valor mínimo.

Dados contínuos

- ▶ Para os valores obtidos pelos 27 alunos no Teste de Desempenho Escolar, temos:
 - ▶ Valor mínimo = 7;
 - ▶ Valor máximo = 117;
 - ▶ Amplitude = $117 - 7 = 110$.
- ▶ Uma vez obtida a amplitude do conjunto de dados, é preciso calcular a **amplitude das classes**.

Dados contínuos

- ▶ **Amplitude de classe** é dada pela divisão da amplitude do conjunto de dados pelo número de classes.
- ▶ Para os dados do TDE, a amplitude (110) deve ser dividida pelo número de classes que já foi calculado (5):

$$110 \div 5 = 22.$$

Dados contínuos

- ▶ A **amplitude de classe** será, então, 22. Isso significa que:
 - ▶ a primeira classe vai do valor mínimo, 7 até $7 + 22 = 29$;
 - ▶ a segunda classe vai de 29 a $29 + 22 = 51$;
 - ▶ a terceira classe vai de 51 a $51 + 22 = 73$;
 - ▶ a quarta classe vai de 73 a $73 + 22 = 95$;
 - ▶ a quinta classe vai de 95 a $95 + 22 = 117$, inclusive.
- ▶ Os valores que delimitam as classes são denominados **extremos**.

Dados contínuos

- ▶ **Extremos de classe** são os valores que delimitam as classes.
- ▶ Uma questão importante é saber **como** as classes devem ser escritas. Alguém pode pensar em escrever as classes como segue:

$$\begin{array}{rcl} 7 & - & 28 \\ 29 & - & 51, \text{ etc.} \end{array}$$

- ▶ No entanto, essa notação traz dúvidas.

Dados contínuos

- ▶ Como saber, por exemplo, para qual classe vai o valor 28,5?
- ▶ Esse tipo de dúvida é evitado indicando as classes como segue:

$$\begin{array}{l} 7 \quad \vdash \quad 28 \\ 29 \quad \vdash \quad 51, \text{etc.} \end{array}$$

- ▶ Usando essa notação, fica claro que o intervalo é **fechado** à esquerda e **aberto** à direita.

Dados contínuos

- ▶ Então, na classe $7 \vdash 29$ estão **incluídos** os valores iguais ao extremo inferior da classe, que é 7 (o intervalo é fechado à esquerda), mas **não estão incluídos** os valores iguais ao extremo superior da classe, que é 29 (o intervalo é aberto à direita).
 - ▶ A indicação de que o intervalo é fechado é dada pelo lado esquerdo do traço vertical do símbolo \vdash .
 - ▶ A indicação de intervalo aberto é dada pela ausência de traço vertical no lado direito do símbolo \vdash .
- ▶ Uma alternativa a esta notação é dada por **colchetes** e **parênteses**.

Dados contínuos

- ▶ Considere *ei* e *es* os **extremos inferior** e **superior** de uma classe qualquer, respectivamente.
 - ▶ " $(ei; es]$ ", ou " $-]$ " é um intervalo aberto à esquerda e fechado à direita;
 - ▶ " $[ei; es)$ ", ou " $[-$ " é um intervalo aberto à direita e fechado à esquerda;
 - ▶ " $(ei; es)$ ", ou " $]ei; es[$ ", ou " $-$ " é um intervalo aberto;
 - ▶ " $[ei; es]$ ", ou " $[-]$ " é um intervalo fechado.

Dados contínuos

- ▶ Estabelecidas as classes, é preciso obter as **frequências**.
- ▶ Para isso, contam-se quantos alunos estão na classe de 7 a 29 (**exclusive**)³, quantos estão na classe de 29 a 51 (**exclusive**), e assim por diante.

Apuração

- ▶ Aqui uma abordagem poderia ser a criação de uma “nova variável” (transformada) de idade em classes na planilha de dados brutos, e então proceder com a apuração desta “nova variável” como no caso de uma variável qualitativa.
- ▶ Afinal de contas, as classes de idade são categorias.
 - ▶ Neste caso, categorias de uma variável qualitativa ordinal.

³Ou, seja, sem incluir o extremo direito do intervalo de classe; neste caso, o valor 29.

Dados contínuos

- A distribuição de frequências pode então ser organizada como segue.

Classe TDE	Frequência (n_i)
7 ┤ 29	4
29 ┤ 51	1
51 ┤ 73	0
73 ┤ 95	6
95 ┤ 117	15
Total	27

Observações

- ▶ Embora a **regra prática** apresentada aqui para a determinação do número de classes seja útil, ela não é a única forma de determinar classes em uma tabela de frequências para dados contínuos.
- ▶ O pesquisador pode especificar as classes de acordo com “convenções”.
- ▶ É comum vermos as frequências da variável idade serem apresentadas em classes de amplitude 5 ou 10 anos.
- ▶ Ainda, podem ser especificadas classes com amplitudes distintas (Idade de 0 a 19 anos, 20 a 59 anos, 60 a 79 anos, 80 anos ou mais).

Observações

- ▶ Outro ponto importante é que nem sempre existe interesse em apresentar todas as classes possíveis.
- ▶ Em alguns casos, a primeira classe pode incluir todos os elementos menores que determinado valor.
- ▶ Diz-se, então, que o extremo inferior da primeira classe não está definido.
- ▶ Como exemplo, veja a distribuição de frequências das pessoas conforme a altura, com as seguintes classes:

Menos de 150 cm

150 ┤ 160cm

160 ┤ 170cm, etc.

Observações

- ▶ Do mesmo modo, todos os elementos iguais ou maiores que determinado valor podem ser agrupados na última classe.
- ▶ Diz-se, então, que o extremo superior da última classe não está definido.
- ▶ Muitos dados de idade publicados pelo **Instituto Brasileiro de Geografia e Estatística (IBGE)** estão em tabelas de distribuição de frequências com intervalos de classes diferentes (em relação a amplitude) e não possuem extremo superior definido.
- ▶ Veja o exemplo a seguir.

Observações

Grupo de idade	Frequência
0 a 4 anos	13796159
5 a 9 anos	14969375
10 a 14 anos	17166761
15 a 19 anos	16990870
20 a 24 anos	17245190
25 a 29 anos	17104413
30 a 34 anos	15744512
35 a 39 anos	13888581
40 a 44 anos	13009367
45 a 49 anos	11833351
50 a 54 anos	10140402
55 a 59 anos	8276219
60 a 64 anos	6509119
65 a 69 anos	4840810
70 a 74 anos	3741637
75 a 79 anos	2563448
80 a 84 anos	1666972
85 a 89 anos	819483
90 a 94 anos	326559
95 a 99 anos	98335
Mais de 100 anos	24236
Total	190755799

Figure 1: População residente, segundo grupos de idade no Brasil (Censo 2010; <https://censo2010.ibge.gov.br/sinopse/index.php?dados=12>).

Para casa

1. Resolver os exercícios 4, 5 e 6 do Capítulo 3.5 do livro **Fundamentos de Estatística**⁴ (disponível no Sabi+).
2. Para os dados contínuos do seu levantamento estatístico, construa tabelas de frequências e compartilhe no Fórum Geral do Moodle. Discuta como você definiu as classes e suas amplitudes.

⁴Vieira, S. **Fundamentos de Estatística**, Atlas, 2019, pg. 37-38.

Próxima aula

- ▶ Distribuição de frequências: **frequências relativa, acumulada, relativa acumulada e porcentagem.**

Por hoje é só!

Bons estudos!

