

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

Probabilidade e Estatística (EAD)

 $\rm MAT02219$ - Probabilidade e Estatística - 2022/1

Plano Aula 05 e 06

(... cont.) Estatística Descritiva

Medidas Descritivas (capítulo 3, Livro Bussab e Morettin)

- Tabelas e gráficos × resumos numéricos (medidas descritivas).
- Como descrever o comportamento dos dados usando apenas números?
 - Para que tipos de variáveis podemos calcular?

Medidas de posição (ou localização) (Seção 3.1, Livro Bussab e Morettin)

Onde está o centro da distribuição de frequências? Ou qual o valor central dos dados? Ou qual o valor de maior frequência?

- Média aritmética: ponderada \times simples; em estatística, $\overline{x} = m\acute{e}dia \ aritmética \ simples.$
 - outros tipos de médias: geométrica, harmônica, ...;
- Mediana (md)
- Moda (*mo*)
 - Interpretações???

Exemplo 1: Variável em estudo - tempos (em segundos) de reação a um alarme de incêndio (Slides 1-4 página 17). Qual o tipo de variável?

No software R podemos calcular média e mediana usando as funções mean() e median(). (E para moda?)

```
x <- c(12, 9, 11, 7, 9, 14, 6, 10)
mean(x); median(x); names(table(x))[which.max(table(x))]</pre>
```

```
## [1] 9.75
## [1] 9.5
## [1] "9"
```

Medidas separatrizes (Seção 3.3, Livro Bussab e Morettin)

- Quantis: quartis, decis, percentis;
 - dividem os dados em grupos de mesma frequência.
- Quartis: três valores que dividem as observações em quatro grupos, $q_1, q_2 = md, q_3$.

Medidas de dispersão (ou variabilidade) (Seção 3.2, Livro Bussab e Morettin)

Qual o grau de dispersão/variabilidade/espalhamento dos dados?

- Amplitude total (a_t) e amplitude interquartílica (a_q) ;
- Variância (s^2) e desvio padrão (s);
- Coeficiente de variação (cv).

No R os comandos var() e sd() calculam a variância e o desvio padrão (ambos usam denominador n-1).

Medidas de formato (slides 1-4, página 50)

- Assimetria e curtose:
 - medidas de posição e dispersão são as duas principais, quando ha simetria nos dados;



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



MAT02219 - Probabilidade e Estatística - 2022/1

mas sob assimetria outras medidas podem ser de interesse ao se comparar distribuições de dados.

Dados Não Agrupados × Dados Agrupados em classe (slides 1-4, página 57)

- medidas descritivas podem ser calculadas mesmo quando os dados são apresentados apenas em forma de tabelas de frequências;
 - com perda de informação \times sem perda de informação.

Análise Exploratória de Dados (AED)

Resumo de Cinco Números (Slides 1-5, página 3)

- O que os quartis, juntamente com o mínimo e o máximo, nos dizem sobre a simetria dos dados?
 - Cinco números: $(x_{(1)}, q_1, q_2, q_3 \in x_{(n)})$
 - Dispersão inferior e dispersão superior: $di = q_2 x_{(1)}$ e $ds = x_{(n)} q_2$.

Valores atípicos (Outliers) (Slides 1-5, página 10)

Como identificar valores que fogem ao comportamento da maioria da maioria das observações?

- Calcular a cerca (limite) inferior e a cerca (limite) superior:
 - $-ci = q_1 1, 5 \times a_q e cs = q_1 + 1, 5 \times a_q;$
- Valores compreendidos entre as cercas (limites) são chamados valores adjacentes;
- Valores abaixo de ci e acima de cs são chamados outliers ou valores atípicos.

Gráfico de Caixa (Boxplot) (Seção 3.2, Livro Bussab e Morettin)

• Reúne informações sobre posição, dispersão, simetria, valores atípicos, ...

Qual a melhor maneira de descrever conjuntos de observações?

- Tabelas. gráficos, medidas resumo, ... depende do tipo de variável;
 - o que vemos na disciplina é apenas uma introdução a essas ideias.

Ler slides e ver vídeos da semana 3.

Fazer lista de exercícios 1-4 e 1-5.

Fazer o Quiz da semana 3 - VALE NOTA!!!

Exemplo 4: Variável em estudo - gasto (em reais) em supermercado por pessoa (Slides 1-5, página 18). Qual o tipo de variável?

No R as funções summary() e boxplot() retornam o resumo de cinco números e o gráfico de caixas.

 $x \leftarrow c(9.26, 10.81, 3.11, 85.76, 70.32, 82.70, 18.43, 19.54, 23.04, 24.47, 26.24, 26.26, 24.58, 28.38, summary(x)$

Min. 1st Qu. Median Mean 3rd Qu. Max. ## 3.11 19.33 27.86 34.78 45.22 93.34

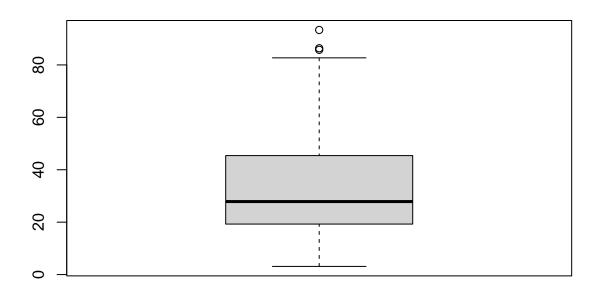
boxplot(x)



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



 $\operatorname{MAT}02219$ - Probabilidade e Estatística - 2022/1



Como incluir título e nomes nos eixos???