

1 Reverso conteúdos da última aula - RMarkdwon

2 Dados, Indicadores e escalas

3 Indicadores parlamentares e eleitorais

4 Medidas de posição

5 Medidas de dispersão

6 Surveys e planejamento

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Reverendo conteúdos da última aula - RMarkdwon

.rmd do zero e gh-pages

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

**Dados,
Indicadores
e escalas**

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Dados, Indicadores e escalas

Matriz de dados

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

```
# Atribuindo o dataframe de exercicio  
dfe <- read_rds("dados/dfe.rds")  
#
```

dataMaid

Frederico
Bertholini

Revendendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

```
# Codebook dataMaid -- explicar
attr(dfe$id, "label") <- "Identificação"
attr(dfe$id, "shortDescription") <- "Variável de identificação do aluno"

attr(dfe$media, "shortDescription") <- "Nota do aluno ao final da disciplina"
attr(dfe$faltas, "shortDescription") <- "Total de faltas ao longo do semestre"
attr(dfe$turma, "shortDescription") <- "Turma do aluno"
attr(dfe$idade, "shortDescription") <- "Variável de identificação do aluno"
attr(dfe$interess, "shortDescription") <- "Variável de identificação do aluno"
attr(dfe$tempocup, "shortDescription") <- "Variável de identificação do aluno"
attr(dfe$escola, "shortDescription") <- "Variável de identificação do aluno"
attr(dfe$estcivil, "shortDescription") <- "Estado civil do aluno"

# makeCodebook(dfe, file = "cb_dfe.pdf")
```

R Ladies - dataMaid

Índices e escalas

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Ordinais

■ Escala

- Atribuições ao padrão de resposta
- Pode conter vários itens

■ Índice

- Compostos por itens (multidimensional)
- cada item é uma variável

Cr terios

Frederico
Bertholini

Reverso
cont dos
da  ltima
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posi  o

Medidas de
dispers o

Surveys e
planeja-
mento

- Validade (interna e externa)
- Vari ncia
- Correla  o (bivariada e multivariada)
- Outras propriedades consagradas (Januzzi)

Frederico
Bertholini

Revendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Propriedades	Indicador 1	...	Indicador i	...	Indicador n
Relevância para a agenda política					
Validade de representação do conceito					
Confiabilidade da medida					
Cobertura populacional					
Sensibilidade às ações previstas					
Especificidade ao programa					
Transparência metodológica na sua construção					
Comunicabilidade ao público					
Factibilidade operacional para sua obtenção					
Periodicidade na sua atualização					
Desagregabilidade populacional e territorial					
Comparabilidade da série histórica					
Total de propriedades (+) :					
Obs.: Atribuir + quando a propriedade for verificada para o indicador					

tribuição de pontos

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Escalas de respostas

- amplitude
- Não resposta
- pesos

Tipos

- Likert
- Feeling thermometer

Menos conhecidas (e utilizadas)

- Bogartus (distância social)
- Thurstone
- Guttman

Children	Problems				
	Counting	Addition	Subtraction	Multiplication	Division
S_1	1	0	0	0	0
S_2	1	1	0	0	0
S_3	1	1	1	0	0
S_4	1	1	1	1	0
S_5	1	1	1	1	1

Testes de confiabilidade - Cronbach

Exemplo Cervi

Frederico
Bertholini

Revendendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

B.11. O papel do Estado é: Garantir educação universitária universal e gratuita

B.12. O papel do Estado é: Garantir acesso à saúde pública universal e gratuita

B.13. O papel do Estado é: Garantia de renda mínima para cidadãos em situação de pobreza

B.14. O papel do Estado é: Assegurar a oferta de cotas para minorias em universidades públicas

B.15. O papel do Estado é: Promover Reforma agrária com desapropriação de grandes propriedades

B.16. O papel do Estado é: Promoção de fontes alternativas de informação.

As alternativas de respostas cada uma das perguntas são:

1. Concordo fortemente
2. Concordo
3. Discordo
4. Discordo fortemente

Frederico Bertholini

Revendo conteúdos da última aula - RMarkd-won

Dados, Indicadores e escalas

Indicadores parlamentares e eleitorais

Medidas de posição

Medidas de dispersão

Surveys e planejamento

Valor	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Categoria	Muito alta adesão				Alta adesão				Média adesão				Baixa adesão				Muito baixa adesão			

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

SOMA	Nº CASOS	% CASOS	CATEGORIA	% CASOS
6	77	17,3	Adesão muito alta	47%
7	33	7,42		
8	55	12,36		
9	44	9,89		
10	48	10,79	Adesão alta	41,54%
11	52	11,69		
12	49	11,01		
13	36	8,09		
14	18	4,04	Adesão média	9,21%
15	12	2,7		
16	11	2,47		
17	5	1,12		
18	2	0,45	Adesão baixa	2,25%
19	2	0,45		
20	1	0,22		
21	0	0		
22	0	0	Adesão muito baixa	0,00%
23	0	0		
24	0	0		

Princípios

Frederico
Bertholini

Reverso
contêidos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

- Criterion-related
- Content
- Construct

Lógica:

$$x = t + e$$

Onde:

x é a medida empírica a ser estudada

t é a parte explicada da variação

e é o erro randômico (aleatório)

Vamos olhar isso melhor em inferência.

Fórmula

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

A fórmula para o cálculo do α de *Cronbach* é a seguinte:

$$\alpha = \frac{k}{k-1} \times \frac{St^2 - \sum Si^2}{St^2}$$

Onde:

K = número de variáveis que compõem o índice;

St^2 = Variância Total (a partir da soma das variâncias dos respondentes);

$\sum Si^2$ = Soma das Variâncias das variáveis ou itens que compõem o índice.

Cronbach

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Usando o pacote psych

```
r4 <- psych::sim.congeneric()  
crb <- psych::alpha(r4)  
summary(crb)
```

FALSE

FALSE Reliability analysis

FALSE raw_alpha std.alpha G6(smc) average_r S/N median_r

FALSE 0.74 0.74 0.69 0.42 2.9 0.41

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Composição do índice	α de Cronbach
B11 + B12 + B13 + B14 + B15 + B16	0,762
B11 + B12 + B13 + B14 + B15	0,756
B11 + B12 + B13 + B14 + B16	0,676
B11 + B12 + B13 + B15 + B16	0,694
B11 + B12 + B14 + B15 + B16	0,702
B11 + B13 + B14 + B15 + B16	0,754
B12 + B13 + B14 + B15 + B16	0,736

Mais info

**Frederico
Bertholini**

Reverso
contêúdos
da última
aula -
RMarkd-
won

**Dados,
Indicadores
e escalas**

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Pacote psych

Consistência Interna

Indicadores estatísticos

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Razão: $Z = X/Y$

Proporção: $Z = X/(Y + X)$

Porcentagem: $\text{Proporção} * 100$

Taxa: $\text{eventos/exposição a risco}$

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Indicadores parlamentares e eleitorais

Parlamentares - Índice de fracionalização

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Parlamentares - Índice de fracionalização máxima

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Parlamentares - Índice de fragmentação

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Parlamentares - Número efetivo de partidos

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Parlamentares - Renovação

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Eleitorais

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Mais recursos

**Frederico
Bertholini**

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

**Indicadores
parla-
mentares e
eleitorais**

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

psData

rOpenGov

QoG

electoral

Representações descritivas de dados - Tabela

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

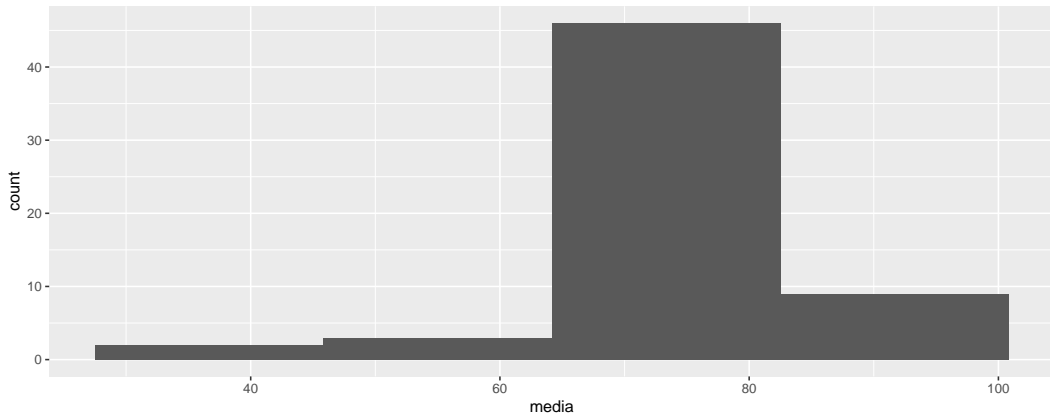
Surveys e
planeja-
mento

```
# Com tabyl
dfe %>%
  drop_na(interess,estcivil) %>% # retirando NAs
  janitor::tabyl(interess,estcivil) %>% # tabela cruzada
  janitor::adorn_percentages("col") %>%
  janitor::adorn_pct_formatting()
```

FALSE	interess	Casado	Solteiro
FALSE	Secundário	29.4%	45.0%
FALSE	Principal	70.6%	55.0%

Representações descritivas de dados - Histograma

```
dfe %>%  
  ggplot() +  
  geom_histogram(aes(media), bins=4)
```



Frederico
Bertholini

Revendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

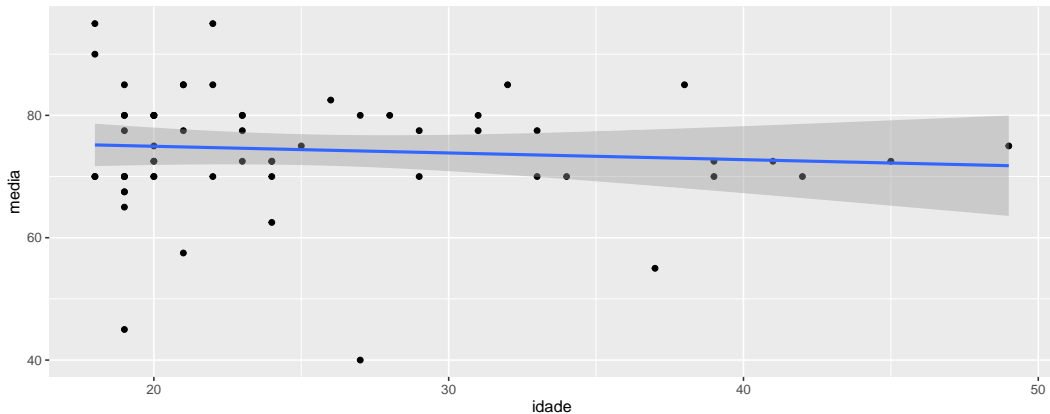
Medidas de
dispersão

Surveys e
planeja-
mento

Representações descritivas de dados - Gráfico de dispersão

dfe %>%

```
ggplot(aes(y=media,x=idade)) +  
geom_point() +  
geom_smooth(method="lm")
```



Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

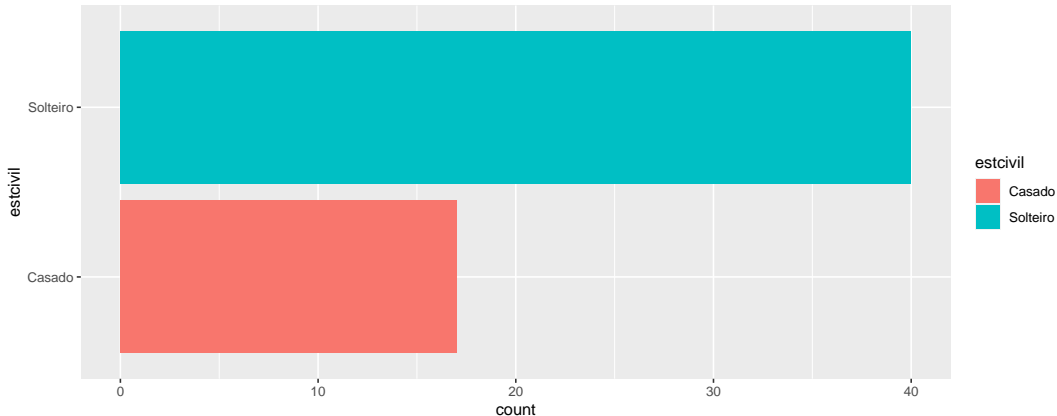
Representações descritivas de dados - Binário

```
df %>%
```

```
  drop_na() %>%
```

```
  ggplot() +
```

```
  geom_bar(aes(fill=estcivil,y=estcivil),position = position_stack(reverse = TRUE),stat="count")
```

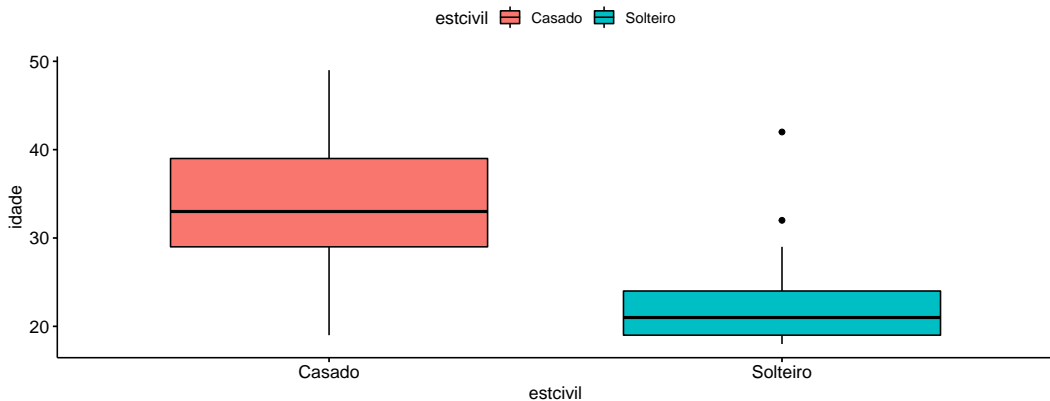


Representações descritivas de dados - Box plot

```
dfe %>%
```

```
drop_na() %>%
```

```
ggpubr::ggboxplot(x="estcivil",y="idade",fill = "estcivil")
```



**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

**Medidas de
posição**

Medidas de
dispersão

Surveys e
planeja-
mento

Medidas de posição

Medidas de posição

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

**Medidas de
posição**

Medidas de
dispersão

Surveys e
planeja-
mento

As medidas de posição trazem informação sobre a localização dos dados no seu conjunto de possíveis valores.

Média aritmética

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

A média aritmética é definida pela soma das observações dividida pelo número total de observações.

Exemplo: sejam x_1, \dots, x_n n observações de uma variável X qualquer. A média é dada pela expressão

$$\frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

No R:

```
dfe %>%  
  pull(media) %>%  
  mean(na.rm=T)
```

```
FALSE [1] 74.375
```

Média ponderada

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

**Medidas de
posição**

Medidas de
dispersão

Surveys e
planeja-
mento

Mediana

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

A mediana é a observação que ocupa a posição central dos dados.

Exemplo 1 (número ímpar de observações): vamos considerar o seguinte conjunto de observações: 3, 7, 10, 5, 2, 1, 1.

Para calcular a mediana, primeiramente ordenamos os dados: 1, 1, 2, 3, 5, 7, 10.

É fácil verificar que a observação que ocupa a posição central é o valor 3. Portanto, 3 é a mediana desse conjunto de valores.

Exemplo 2 (número par de observações): vamos considerar agora o seguinte conjunto de observações: 15, 3, 2, 0, 9, 17.

Ordenando os dados, temos: 0, 2, 3, 9, 15, 17.

Neste caso, a mediana será dada pela média entre as duas observações centrais, isto é, $(3 + 9)/2 = 6$.

Moda

Frederico
Bertholini

Reverso
contêidos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

A moda é a observação mais frequente do conjunto de valores observados.

Exemplo: no conjunto de observações $\{3, 5, 10, 11, 11, 20\}$, a observação 11 aparece 2 vezes enquanto as demais apenas 1 vez. Portanto, 11 é a moda desse conjunto.

**Frederico
Bertholini**

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

**Medidas de
dispersão**

Surveys e
planeja-
mento

Medidas de dispersão

Medidas de dispersão

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

**Medidas de
dispersão**

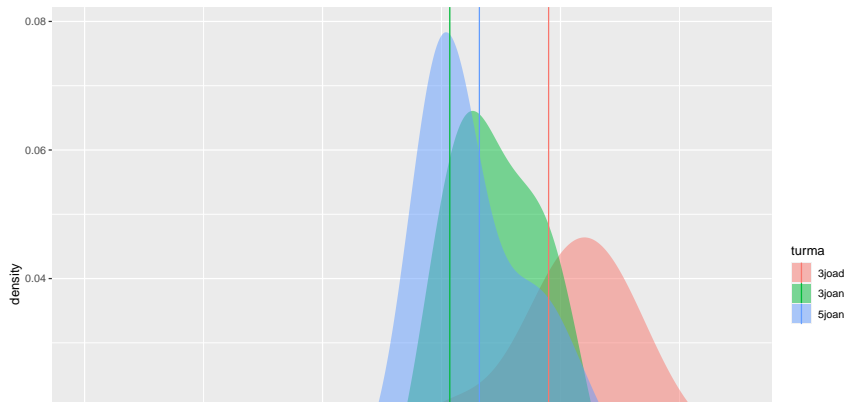
Surveys e
planeja-
mento

O resumo de um conjunto de dados por uma única medida de posição central não traz informação sobre a variabilidade das observações.

Um critério frequentemente usado para avaliar a dispersão de um conjunto de observações é aquele que mede a dispersão dos dados em torno da sua média.

```

dfe %>%
  ggplot() +
  geom_density(aes(fill=turma,x=media),alpha=.5,color=NA) +
  geom_vline(data=.%>%group_by(turma)%>%
              summarise(media=mean(media,na.rm=T)),
              aes(color=turma,xintercept=media))
  
```



Amplitude e Quantis

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

Amplitude = valor máximo - valor mínimo

A mediana é uma medida que deixa metade dos dados abaixo dela e metade acima. De modo geral, podemos definir a medida q_p , $0 < p < 100$, tal que p das observações sejam menores que q_p . Esta medida é chamada de quantil de ordem p .

Alguns quantis são bastante utilizados na prática. São eles

$Q1 = q_{25\%}$ (primeiro quartil)

$Q2 = q_{50\%}$ (mediana)

$Q3 = q_{75\%}$ (terceiro quartil)

Observe que os $q_{0\%}$ e $q_{100\%}$ denotam, respectivamente, o mínimo e o máximo de conjunto de dados.

Variância

Sejam x_1, x_2, \dots, x_n n observações de uma variável quantitativa X . Seja \bar{x} a média dessas observações.

Vamos observar, então, os desvios das observações em relação à média \bar{x} , isto é, $x_i - \bar{x}$, para $i = 1, \dots, n$.

Ideia 1: usar como medida de dispersão a soma desses desvios, isto é,

$$\sum_{i=1}^n (x_i - \bar{x}).$$

Problema 1: para qualquer conjunto de observações

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Ideia 2: considerar a soma do quadrado dos desvios, dada por

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

Desvio padrão

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento

A utilização da variância como medida de dispersão pode ainda causar um problema de interpretação, pois a sua dimensão é igual ao quadrado da dimensão dos dados. Sendo assim, costuma-se usar o desvio padrão, definido por

$$DP(X) = \sqrt{VAR(X)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Coeficiente de variação

Frederico
Bertholini

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

**Medidas de
dispersão**

Surveys e
planeja-
mento

Medida relativa -> que proporção da média é o desvio-padrão

$$\frac{dp}{m\acute{e}dia}$$

Medidas de assimetria e curtose

Frederico
Bertholini

Reverendo
conteúdos
da última
aula -
RMarkd-
won

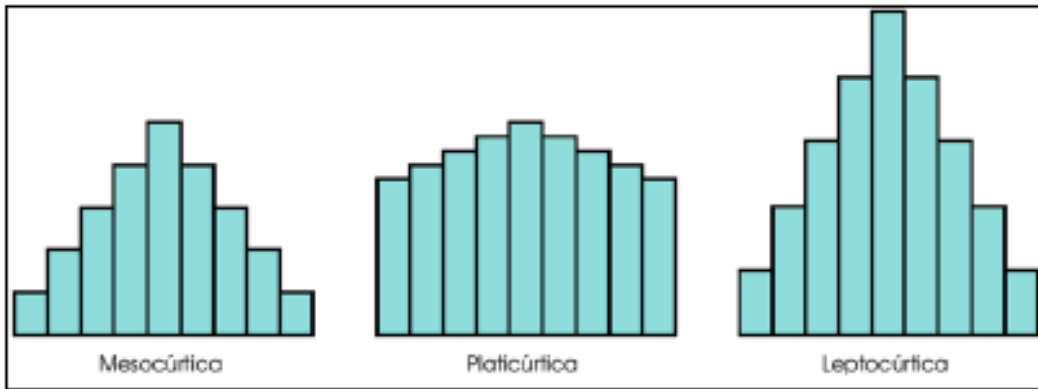
Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

Surveys e
planeja-
mento



Coeficiente de Assimetria de Pearson = $média - moda$

Assimetria adimensional = $\frac{média - moda}{dp}$

Coeficiente quartílico de assimetria = $3 * \frac{média - mediana}{dp}$

**Frederico
Bertholini**

Reverso
conteúdos
da última
aula -
RMarkd-
won

Dados,
Indicadores
e escalas

Indicadores
parla-
mentares e
eleitorais

Medidas de
posição

Medidas de
dispersão

**Surveys e
planeja-
mento**

Surveys e planejamento