

Análise de variância e correlação

Métodos Quantitativos Aplicados à Ciência Política

Frederico Bertholini

07.dez.2020

1 Formalizando Hipóteses

2 Testando diferentes hipóteses

3 Correlação

4 Modelo Linear Simples

Análise de
variância e
correlação

Frederico
Bertholini

**Formalizando
Hipóteses**

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Formalizando Hipóteses

Um caso comum - Esta análise está correta?

Análise de
variância e
correlação

Frederico
Bertholini

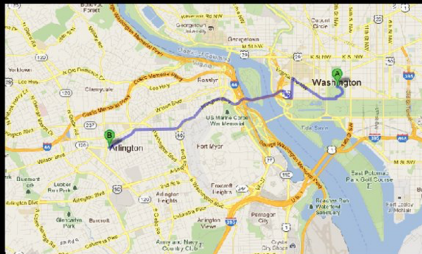
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

CIDADÃOS ARMADOS: MENOS CRIMES



- Para ir de **Washington** a **Arlington** você só precisa cruzar um rio.
- Em Washington, as armas são **proibidas**, em Arlington, as armas são **liberadas**.
- A taxa de homicídios de Washington é de 57/100.000, em Arlington é de 1,6/100.000, cerca de **35 vezes menor**.
- Mais armas, menos crimes.
- **DEFENDA SUA VIDA, COMPARTILHE**

<http://www.facebook.com/campodarmamento>
@campodarmamento
O direito a legítima defesa armada é um direito fundamental. Lutamos pelo seu.

Comparação de médias

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

A resposta é não. (por quê?)

Em todo caso, ela ilustra uma situação bem comum na prática, onde se deseja comparar médias. No caso, deseja-se comparar as taxas médias de mortalidade em cidades onde as armas são proibidas ou liberadas. Deseja-se testar se a média de homicídios em cidades onde armas são liberadas é menor que a média de homicídios em cidades onde armas são proibidas. (Como você colheria dados para esse estudo?)

As hipóteses, portanto, são:

$$H_0 : \mu_L = \mu_P$$

$$H_A : \mu_L < \mu_P$$

Lembrando a base de trabalho

```
summary(dfe %>% dplyr::select(-id))
```

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

media	faltas	turma	idade
Min. :40.00	Min. : 0.00	Length:60	Min. :18.00
1st Qu.:70.00	1st Qu.: 2.00	Class :character	1st Qu.:19.75
Median :73.75	Median : 4.00	Mode :character	Median :22.00
Mean :74.38	Mean : 4.25		Mean :25.23
3rd Qu.:80.00	3rd Qu.: 6.00		3rd Qu.:29.00
Max. :95.00	Max. :10.00		Max. :49.00

interess	tempocup	escola	estcivil
Secundário:24	não tem : 4	Tudo privada :20	Casado :17
Principal :34	até 2h : 3	Maior parte privada:15	Solteiro:42
NA's : 2	de 2h a 4h:11	Maior parte pública:18	NA's : 1
	de 4h a 6h:42	Tudo pública : 7	
	+ de 6h : 0		

nota1	nota2
Min. :40.00	Min. :40.00
1st Qu.:67.88	1st Qu.:71.75
Median :72.00	Median :77.00

Diferença entre médias (amostras não pareadas)

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

H_0 : A média de notas de casados e solteiros é igual ou $H_0 : \mu_c - \mu_s = 0$ ou $H_0 : \mu_c = \mu_s$

H_1 : A média de notas de casados e solteiros é diferente ou $H_1 : \mu_c - \mu_s \neq 0$ ou $H_1 : \mu_c \neq \mu_s$

Variável **dependente**: Notas

Variável **independente**: Situação conjugal

O que eu quero testar? Se a situação conjugal *faz diferença* na nota.

É **efeito**? Não! (Pearl, 2020) Inferência vs. Causalidade

Como testar na prática? Distribuições:

Análise de
variância e
correlação

Frederico
Bertholini

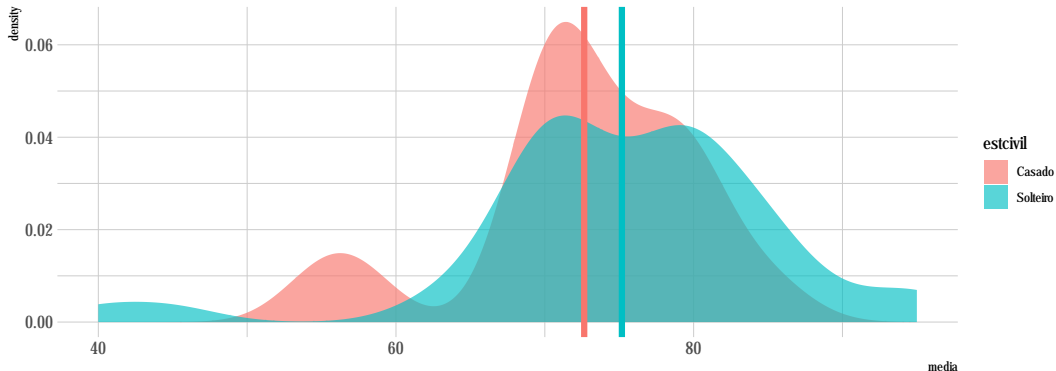
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% drop_na(estcivil) %>%  
  ggplot(aes(fill=estcivil,x=media,color=estcivil,group=estcivil)) +  
  geom_density(color=NA,alpha=.65) +  
  geom_vline(data=., %>% group_by(estcivil) %>% summarise(media=mean(media,na.rm = T)),  
             size=2,aes(xintercept=media,color=estcivil)) +  
  guides(color="none") + theme_ipsum_mod
```



Como testar na prática? Vamos construir intervalos

Análise de
variância e
correlação

Frederico
Bertholini

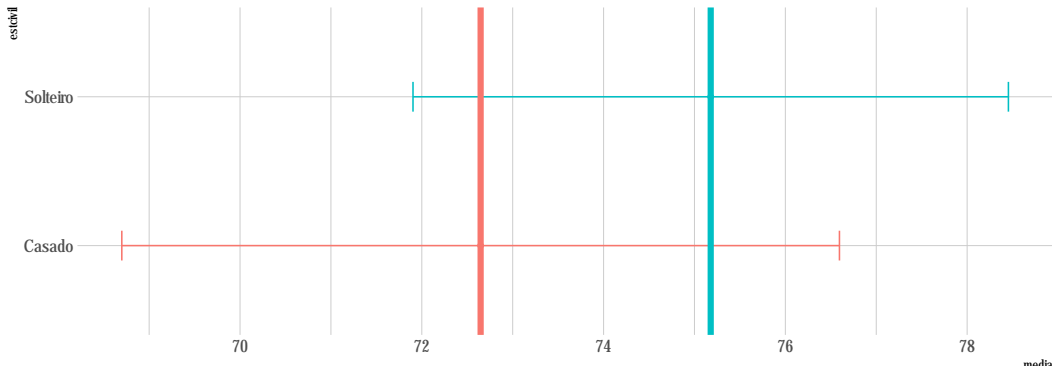
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% drop_na(estcivil) %>%  
  ggplot(aes(fill=estcivil,x=media,color=estcivil,y=estcivil)) +  
  stat_summary(fun=mean, geom="point") +  
  stat_summary(fun.data=mean_ci, geom="errorbar", width=0.2) +  
  geom_vline(data=. %>% group_by(estcivil) %>% summarise(media=mean(media,na.rm = T)),  
             size=2,aes(xintercept=media,color=estcivil)) +  
  theme_ipsum_mod +theme(legend.position = "none")
```



Como testar na prática? Teste-t

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
t_test_results <- dfe %>%  
  t_test(formula = media ~ estcivil,  
         order = c("Casado", "Solteiro"))  
t_test_results
```

```
# A tibble: 1 x 6
```

	statistic	t_df	p_value	alternative	lower_ci	upper_ci
	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>
1	-1.03	40.4	0.311	two.sided	-7.52	2.46

Outro exercício

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

$$H_0 : \mu_{3joan} = \mu_{3joad}$$

$$H_1 : \mu_{3joan} \neq \mu_{3joad}$$

Variável **dependente**: Notas

Variável **independente**: Turma (apenas 3joan e 3joad)

Intervalos

Análise de
variância e
correlação

Frederico
Bertholini

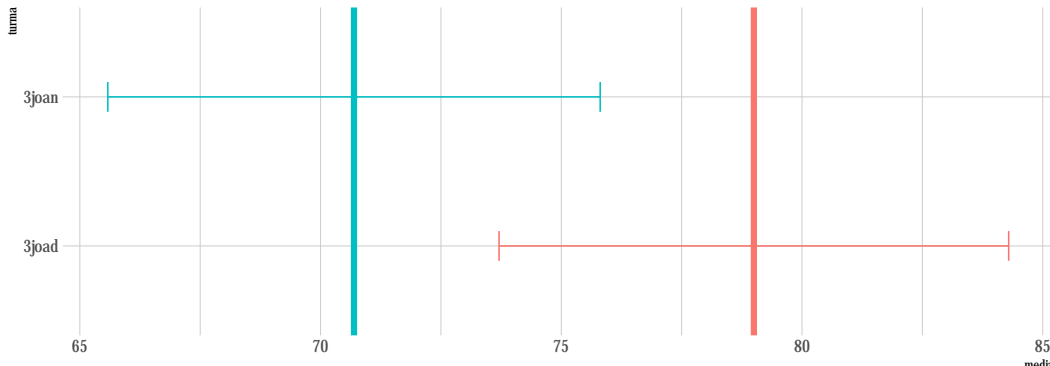
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% dplyr::filter(turma %in% c("3joan","3joad")) %>%  
  ggplot(aes(fill=turma,x=media,color=turma,y=turma)) +  
  stat_summary(fun=mean, geom="point") +  
  stat_summary(fun.data=mean_ci, geom="errorbar",width=0.2) +  
  geom_vline(data=. %>% group_by(turma) %>% summarise(media=mean(media,na.rm = T)),  
             size=2,aes(xintercept=media,color=turma)) +  
  theme_ipsum_mod +theme(legend.position = "none")
```



Como testar na prática? Teste-t

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
t_test_results <- dfe %>%  
  dplyr::filter(turma %in% c("3joan", "3joad")) %>%  
  t_test(formula = media ~ turma)  
t_test_results
```

```
# A tibble: 1 x 6
```

	statistic	t_df	p_value	alternative	lower_ci	upper_ci
	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>
1	2.37	36.0	0.0232	two.sided	1.20	15.4

Olhando no infer graficamente

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
# calculate the observed statistic
media_turmas <- dfe %>%
  dplyr::filter(turma %in% c("3joan", "3joad")) %>%
  specify(media ~ turma) %>%
  calculate(stat = "t", order = c("3joan", "3joad"))

# generate the null distribution with the theoretical t
distribuicao_teorica <- dfe %>%
  dplyr::filter(turma %in% c("3joan", "3joad")) %>%
  specify(media ~ turma) %>%
  hypothesize(null = "independence") %>%
  calculate(stat = "t", order = c("3joan", "3joad"))
```

Visualizando

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

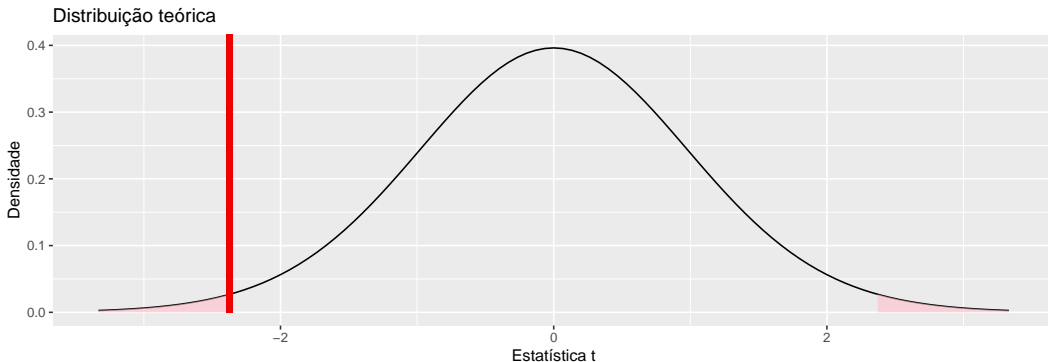
```
# visualize the randomization-based null distribution and test statistic!
```

```
distribuicao_teorica %>%
```

```
  visualize(method = "theoretical") +
```

```
  shade_p_value(media_turmas, direction = "two-sided") +
```

```
  labs(title = "Distribuição teórica", x = "Estatística t", y = "Densidade")
```



Usando ggpubr

Análise de
variância e
correlação

Frederico
Bertholini

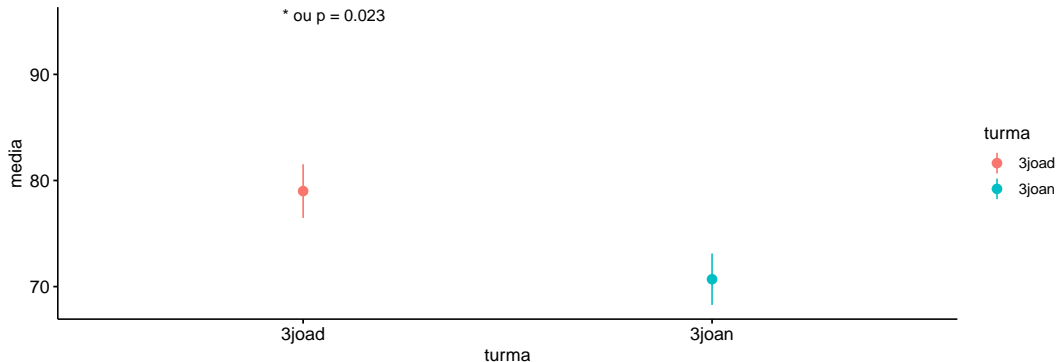
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
library(ggpubr)
dfe %>% dplyr::filter(turma %in% c("3joan", "3joad")) %>%
  ggerrorplot(x = "turma", y = "media", color = "turma",
              position = position_dodge(0.5)) +
  stat_compare_means(aes(label = paste0(..p.signif.., " ou p = ", ..p.format..)),
                    method = "t.test") + theme(legend.position = "right")
```



Tamanho do efeito

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

O d de Cohen pode ser usado como uma estatística de tamanho de efeito para um teste t de duas amostras.

É calculado como a diferença entre as médias de cada grupo, dividido pelo desvio padrão agrupado dos dados.

Um d de Cohen de 0,5 sugere que as médias diferem pela metade do desvio padrão dos dados. Um d de Cohen de 1,0 sugere que as médias diferem por um desvio padrão dos dados.

```
dfe %>% dplyr::filter(turma %in% c("3joan", "3joad")) %>%  
  effectsize::cohens_d("media", "turma", data = .)
```

```
Cohen's d |      1e+02% CI  
-----  
0.77 | [0.10, 1.42]
```

**Análise de
variância e
correlação**

**Frederico
Bertholini**

Formalizando
Hipóteses

**Testando
diferentes
hipóteses**

Correlação

Modelo
Linear
Simples

Testando diferentes hipóteses

Amostras pareadas

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Uma mesma medição em dois momentos no tempo para os mesmos indivíduos

$$H_0 : \mu_{t2} - \mu_{t1} = 0 \text{ ou } H_0 : \mu_{\Delta} = 0$$

$$H_A : \mu_{t2} - \mu_{t1} \neq 0 \text{ ou } H_A : \mu_{\Delta} \neq 0$$

Ex.: Nota 1 e Nota 2

Com infer

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dif_mu <- dfe %>%  
  mutate(d=nota2-nota1) %>%  
  specify(response = d) %>%  
  hypothesize(null = "point", mu = 0) %>%  
  calculate(stat = "t")
```

```
hipotetica <- dfe %>%  
  mutate(d=nota2-nota1) %>%  
  specify(response = d) %>%  
  hypothesize(null = "point", mu = 0) %>%  
  calculate(stat = "t")
```

```
hipotetica %>% visualize(method="theoretical") +  
  shade_p_value(dif_mu,direction="two-sided") +  
  labs(title = "Distribuição teórica",x="Estatística t",y="Densidade")  
dfe %>% mutate(d=nota2-nota1) %>% t_test(d ~ NULL)
```

Visualizando

Análise de
variância e
correlação

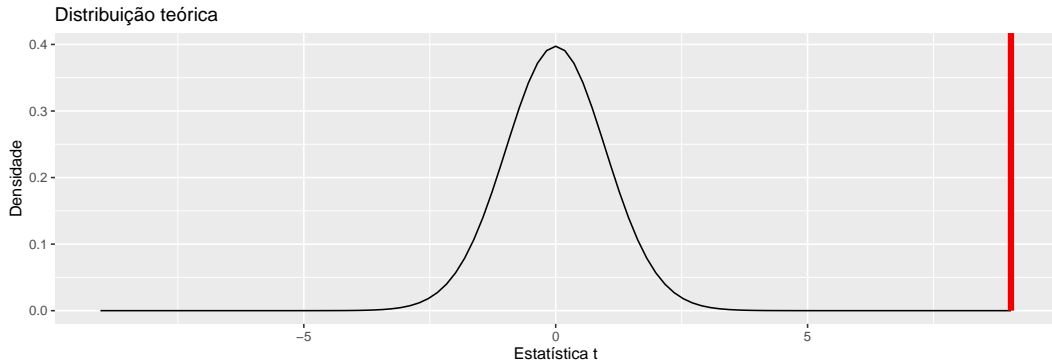
Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples



```
# A tibble: 1 x 6
```

	statistic	t_df	p_value	alternative	lower_ci	upper_ci
	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>
1	9.04	59	9.89e-13	two.sided	3.30	5.17

Por que ANOVA?

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k, \quad H_A : \mu_i \neq \mu_j \text{ para pelo menos um par } i \text{ e } j$$

O que é ANOVA?

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Variabilidade dentro dos grupos = Soma dos Quadrados Dentro (SQD)

$$SQD = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

Variabilidade entre grupos = Soma de Quadrados Entre (SQE)

$$SQE = \sum_{j=1}^c n_j (\bar{X}_j - \bar{\bar{X}})^2$$

Variabilidade total = Soma Total de Quadrados (STQ)

$$STQ = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2$$

ANOVA

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

$$STQ = SQE + SQD$$

$$\text{Fração da variabilidade explicada pelo grupo} = \frac{SQE}{STQ}$$

É possível que, na população, as médias dos grupos sejam iguais e, por acaso, as médias das amostras sejam diferentes.

Quanto maior a variabilidade entre grupos (SQE) e menor a variabilidade dentro dos grupos (SQD), mais evidências teremos que as médias são diferentes na população.

Princípio: Teste F: $\frac{\text{Variância entre grupos}}{\text{Variância dentro dos grupos}}$

$$F = \frac{MQE}{MQD}$$

Na prática

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

H_0 : A média de notas das turmas é igual ou $H_0 : \mu_{3\text{joad}} = \mu_{3\text{joan}} = \mu_{5\text{joan}}$

H_A : A média de notas de pelo menos uma das turmas é diferente ou $H_A : \mu_{3\text{joad}} \neq \mu_{3\text{joan}} \neq \mu_{5\text{joan}}$

Variável dependente: **Notas**

Variável independente: **Turma**

Função aov

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
ANOVAtest <- dfe %>% aov(., formula = media ~ turma)
summary(ANOVAtest)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
turma	2	703	351.5	4.116	0.0214 *
Residuals	57	4867	85.4		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Teste de Tukey

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
Tt <- TukeyHSD(ANOVAtest)
Tt$turma %>% as.data.frame() %>%
  rownames_to_column() %>% mutate(rowname = gsub("\n","",rowname)) %>%
  knitr::kable(col.names = c("", "Dif.", "Lim inf", "Lim sup", "p-valor"),
               digits=3, format = "latex")
```

	Dif.	Lim inf	Lim sup	p-valor
3joan-3joad	-8.306	-15.530	-1.081	0.021
5joan-3joad	-5.818	-12.689	1.052	0.112
5joan-3joan	2.487	-4.580	9.555	0.676

ANOVA com infer

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

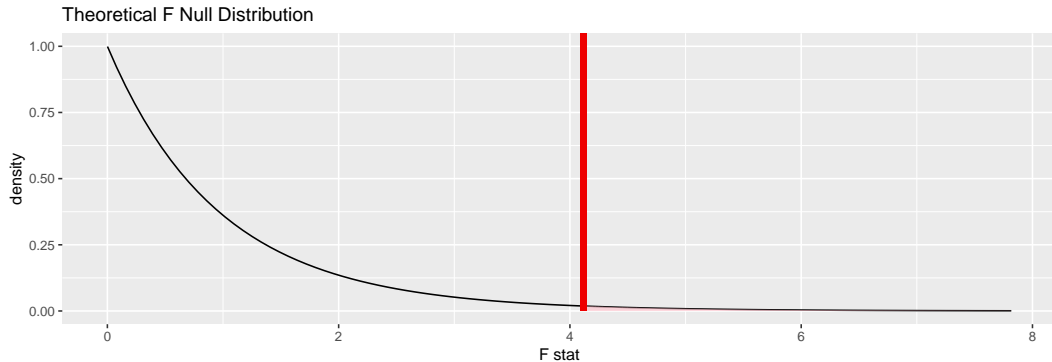
Correlação

Modelo
Linear
Simples

```
observed_f_statistic <- dfe %>%  
  specify(media ~ turma) %>%  
  calculate(stat = "F")
```

```
dfe %>%  
  specify(media ~ turma) %>%  
  hypothesize(null = "independence") %>%  
  visualize(method = "theoretical") +  
  shade_p_value(observed_f_statistic,  
                direction = "greater")
```

```
dfe %>%  
  specify(media ~ turma) %>%  
  hypothesize(null = "independence") %>%  
  generate(reps = 100, type = "permute") %>%  
  calculate(stat = "F") %>%  
  get_p_value(obs_stat = observed_f_statistic,  
              direction = "greater")
```



```
# A tibble: 1 x 1
  p_value
  <dbl>
1    0.01
```

Anova com ggpubr 1

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

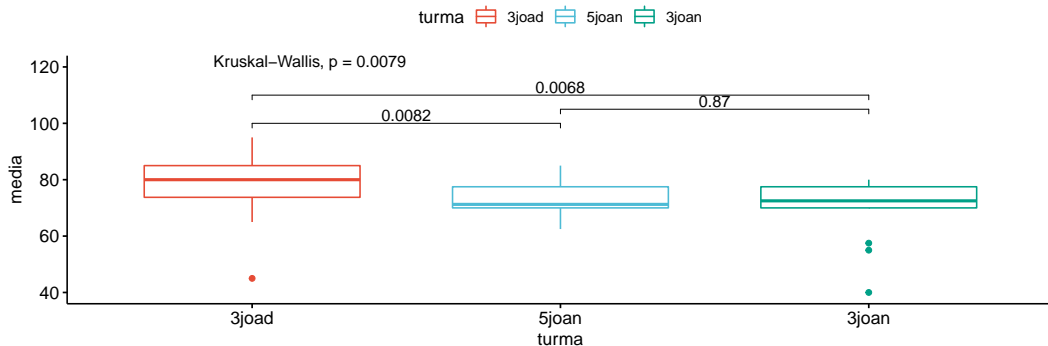
Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>%
```

```
ggboxplot(x = "turma", y = "media", color = "turma", palette = "npg") +  
stat_compare_means(comparisons =  
  list(c("3joad", "3joan"), c("3joan", "5joan"), c("3joad", "5joan")),  
  label.y = c(110, 105, 100)) + stat_compare_means(label.y = 120)
```



Anova com ggpubr 2

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

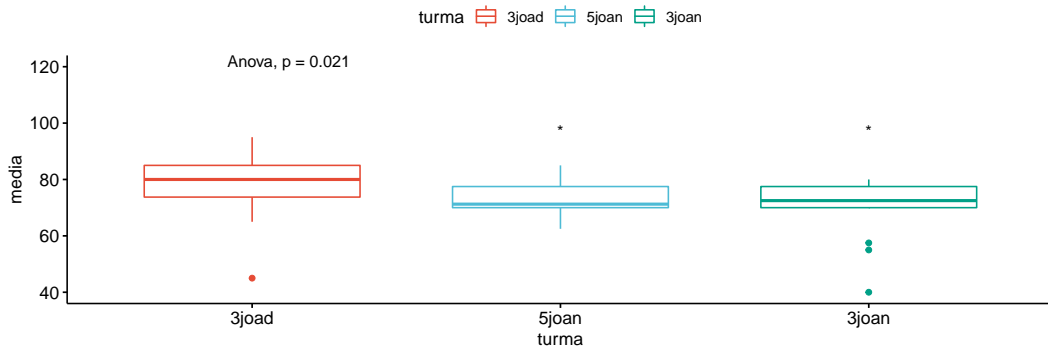
```
# Multiple pairwise test against a reference group
```

```
dfe %>%
```

```
ggboxplot(x = "turma", y = "media", color = "turma", palette = "npg") +
```

```
stat_compare_means(method = "anova", label.y = 120) +
```

```
stat_compare_means(aes(label = ..p.signif..), method = "t.test", ref.group = "3joad")
```



Two-way ANOVA (dois fatores)

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
summary(ANOVAtest2 <- dfe %>% aov(., formula = media ~ turma + interest))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
turma	2	1016	508.1	7.521	0.00131 **
interest	1	2	1.6	0.024	0.87840
Residuals	54	3648	67.6		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2 observations deleted due to missingness

Two-way ANOVA (dois fatores)

```
library(agricolae); HSD.test(ANOVAtest2, trt = c("turma","interest"),console = T)
```

```
Study: ANOVAtest2 ~ c("turma", "interest")
```

```
HSD Test for media
```

```
Mean Square Error: 67.5601
```

```
turma:interest, means
```

	media	std	r	Min	Max
3joad:Principal	82.77778	9.052317	9	70.0	95.0
3joad:Secundário	78.88889	7.817360	9	65.0	90.0
3joan:Principal	68.95833	12.082027	12	40.0	80.0
3joan:Secundário	74.16667	4.082483	6	70.0	80.0
5joan:Principal	73.07692	4.466758	13	67.5	82.5
5joan:Secundário	73.33333	7.071068	9	62.5	85.0

```
Alpha: 0.05 ; DF Error: 54
```

```
Critical Value of Studentized Range: 4.178265
```

Homogeneidade

Análise de
variância e
correlação

Frederico
Bertholini

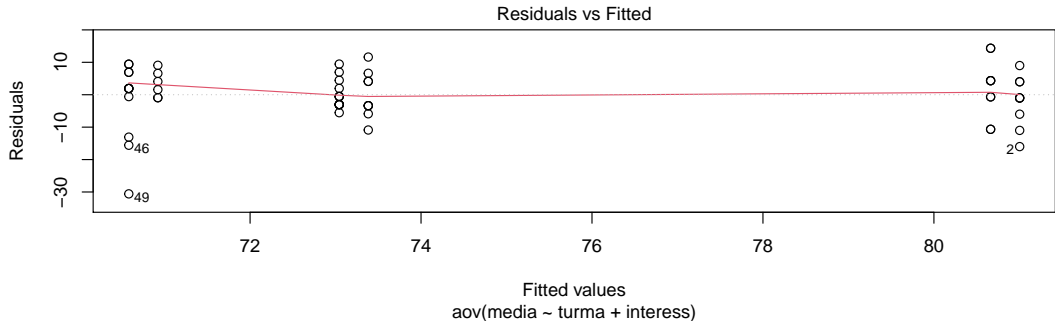
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
plot(ANOVAtest2, 1)
```



```
library(car);leveneTest(media ~ turma * interess, data = dfe)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	5	0.8897	0.4948
	52		

Normalidade

Análise de
variância e
correlação

Frederico
Bertholini

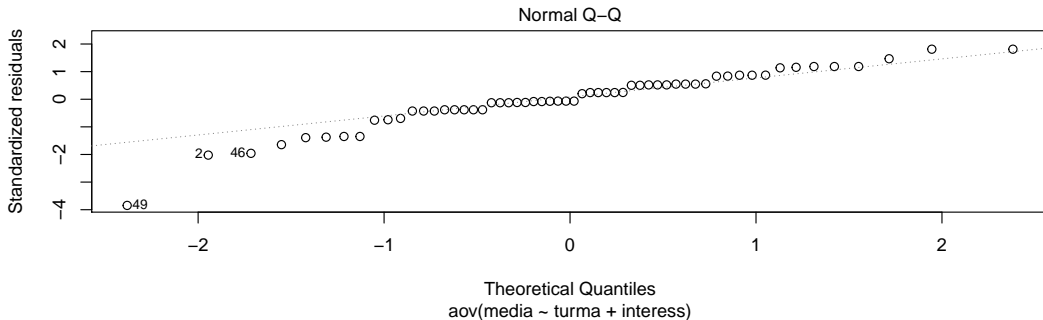
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
plot(ANOVAtest2,2); shapiro.test(x = residuals(object = ANOVAtest2))
```



Shapiro-Wilk normality test

```
data: residuals(object = ANOVAtest2)
```

W = 0.9331, p-value = 0.003273

Qui-quadrado

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Vamos olhar as relações entre:

- 1 Estado civil e interesse na disciplina.
- 2 Interesse na disciplina e turma.

Teste de independência entre variáveis categóricas.

H_0 : Variáveis são independentes

H_A : Variáveis não são independentes

Tabelas de contingência e Qui-quadrado

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Atenção à sua variável de interesse

```
dfe %>% drop_na(estcivil,interest) %>%  
  tabyl(estcivil,interest)
```

estcivil	Secundário	Principal
Casado	5	12
Solteiro	18	22

```
dfe %>% drop_na(estcivil,interest) %>%  
  tabyl(estcivil,interest) %>%  
  janitor::adorn_percentages("col") %>%  
  janitor::adorn_pct_formatting()
```

estcivil	Secundário	Principal
Casado	21.7%	35.3%
Solteiro	78.3%	64.7%

No infer

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
qui_quadrado <- dfe %>% drop_na(estcivil,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  specify(interess ~ estcivil,success = "Principal") %>%  
  calculate(stat = "Chisq")
```

```
teorica_qui_quadrado <- dfe %>% drop_na(estcivil,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  specify(interess ~ estcivil,success = "Principal") %>%  
  hypothesize(null = "independence")
```

Visualizando

Análise de
variância e
correlação

Frederico
Bertholini

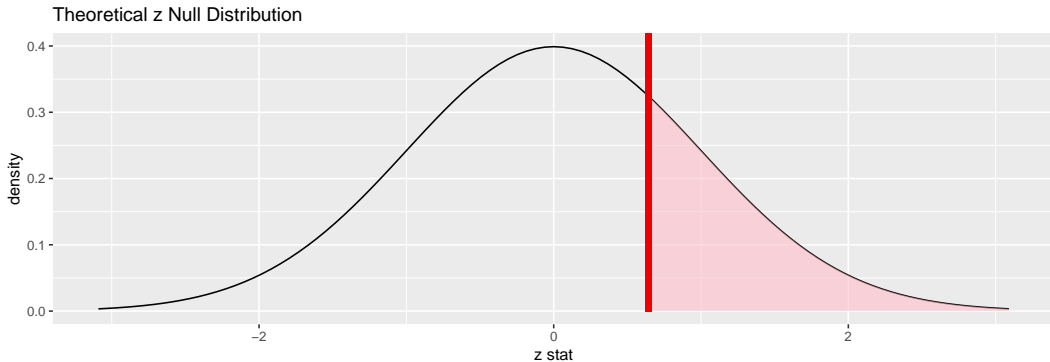
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
teorica_qui_quadrado %>%  
  visualize(method = "theoretical") +  
  shade_p_value(qui_quadrado,  
                direction = "greater")
```



O teste

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% drop_na(estcivil,interest) %>%  
  mutate(estcivil=as.character(estcivil),interest=as.character(interest)) %>%  
  infer::chisq_test(interest ~ estcivil)
```

```
# A tibble: 1 x 3  
  statistic chisq_df p_value  
    <dbl>      <int>   <dbl>  
1     0.644         1     0.422
```


Tabelas de contingência e Qui-quadrado

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simple

```
dfe %>% drop_na(turma,interest) %>%  
  tabyl(turma,interest)
```

turma	Secundário	Principal
3joad	9	9
3joan	6	12
5joan	9	13

No infer

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
qui_quadrado <- dfe %>% drop_na(turma,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  specify(interess ~ turma) %>%  
  calculate(stat = "Chisq")
```

```
teorica_qui_quadrado <- dfe %>% drop_na(turma,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  specify(interess ~ turma) %>%  
  hypothesize(null = "independence")
```

Visualizando

Análise de
variância e
correlação

Frederico
Bertholini

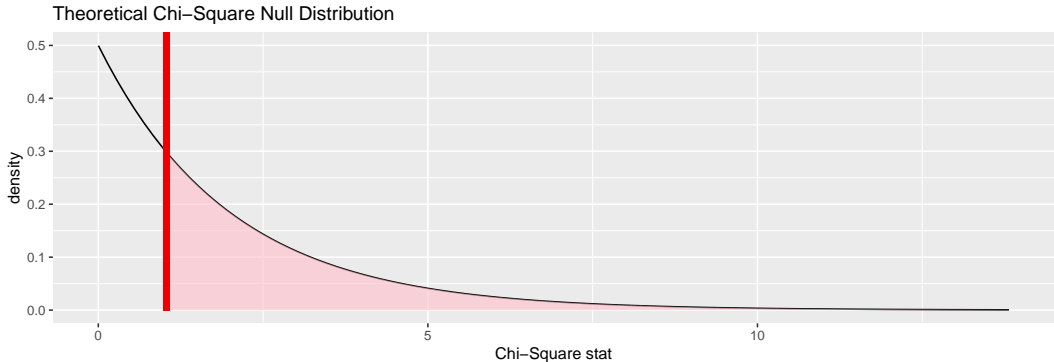
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
teorica_qui_quadrado %>%  
  visualize(method = "theoretical") +  
  shade_p_value(qui_quadrado,  
                direction = "greater")
```



O teste

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% drop_na(turma,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  chisq_test(turma ~ interest)
```

```
# A tibble: 1 x 3  
  statistic chisq_df p_value  
    <dbl>     <int>   <dbl>  
1      1.03         2 0.596
```

```
dfe %>% drop_na(turma,interest) %>%  
  mutate_if(is.factor,as.character) %>%  
  chisq_test(interest ~ turma)
```

```
# A tibble: 1 x 3  
  statistic chisq_df p_value  
    <dbl>     <int>   <dbl>  
1      1.03         2 0.596
```

**Análise de
variância e
correlação**

**Frederico
Bertholini**

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Correlação

Correlação

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

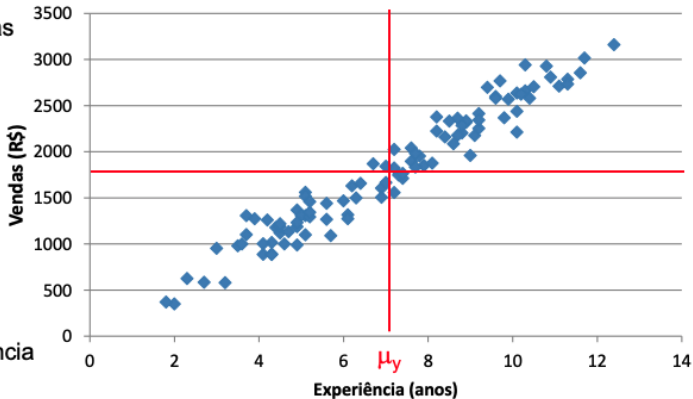
Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Experiência
baixa
Vendas
elevadas

Experiência
elevada
Vendas
elevadas

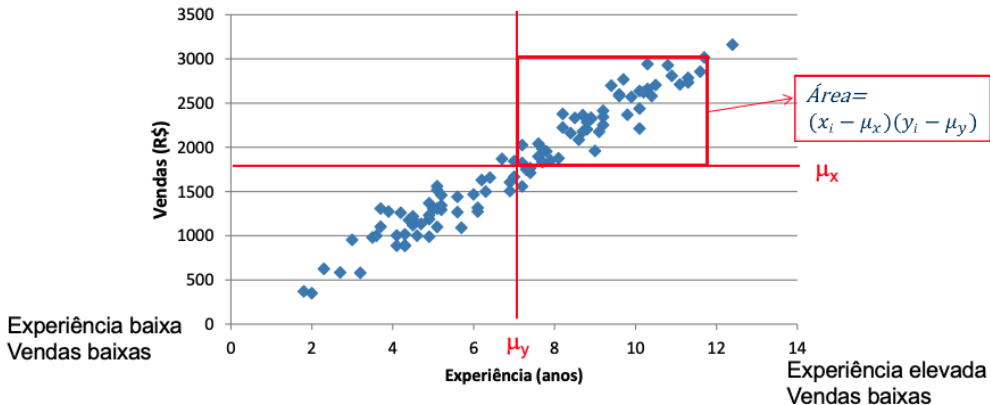


Experiência
baixa
Vendas
baixas

Experiência
elevada
Vendas
baixas

Experiência baixa
Vendas elevadas

Experiência elevada
Vendas elevadas



Áreas com sinal negativo
indicam **associação
negativa** entre
experiência e vendas.

-

+

Áreas com sinal positivo
indicam **associação
positiva** entre
experiência e vendas.

Áreas com sinal positivo
indicam **associação
positiva** entre
experiência e vendas.

+

-

Áreas com sinal negativo
indicam **associação
negativa** entre
experiência e vendas.

μ_x

μ_y

Três conceitos para a mesma idéia:

- Covariância
- Correlação
- Coeficiente de determinação (R^2)

São grandes em valor absoluto se houver forte relação linear

Guess the correlation

Covariância e correlação

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

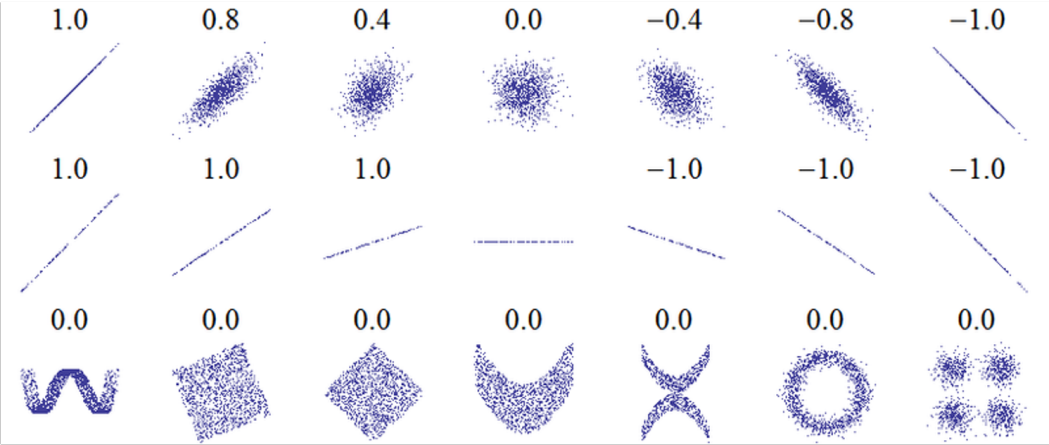
$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^N (X_i - \mu_x)(Y_i - \mu_y)}{N}$$

Uma área positiva indica associação positiva entre as variáveis.

- Mas como saber se é uma associação forte ou fraca?
- Qual a unidade de medida da covariância?

Para eliminar a unidade de medida das variáveis, podemos usar a padronização z . Desta forma, obtemos o coeficiente de correlação, que é a covariância com variáveis padronizadas. Este coeficiente varia de -1 a 1.

$$\text{Corr}(X, Y) = \frac{\sum_{i=1}^N \left(\frac{X_i - \mu_x}{\sigma_x} \right) \left(\frac{Y_i - \mu_y}{\sigma_y} \right)}{N}$$



Faltas e Nota 2

Análise de
variância e
correlação

Frederico
Bertholini

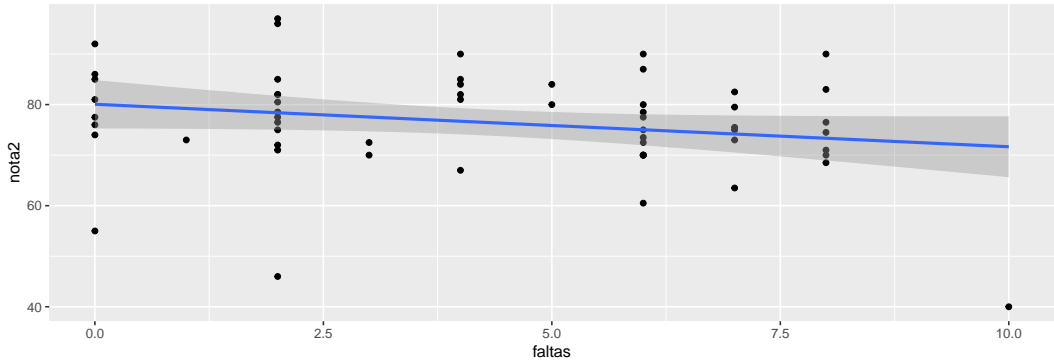
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% ggplot(aes(x=faltas,y=nota2)) +  
  geom_point() + geom_smooth(method="lm")
```



Com ggpubr

Análise de
variância e
correlação

Frederico
Bertholini

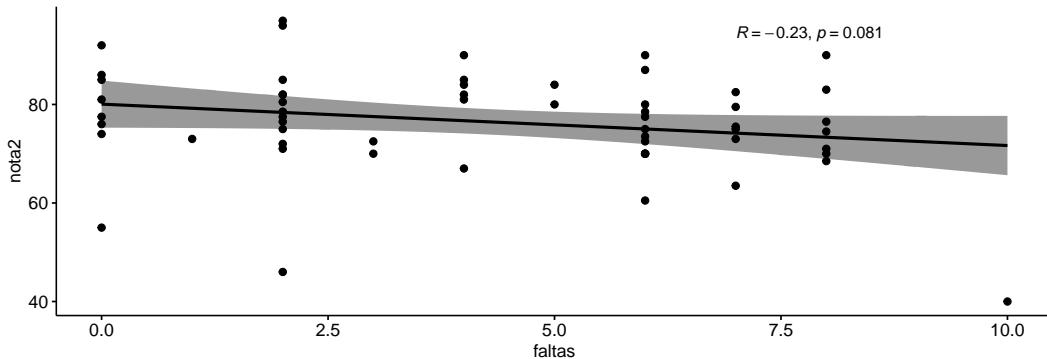
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% ggscatter(x = "faltas", y = "nota2", add = "reg.line", conf.int = TRUE) +  
  stat_cor(method = "pearson", label.x=7)
```



Com ggpubr 2

Análise de
variância e
correlação

Frederico
Bertholini

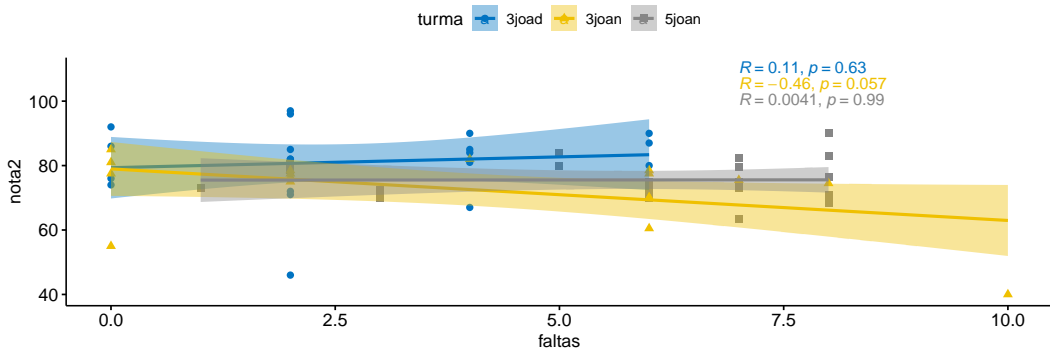
Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

```
dfe %>% ggscatter(x = "faltas", y = "nota2", add = "reg.line", conf.int = TRUE,  
  color = "turma", palette = "jco", shape = "turma") +  
  stat_cor(aes(color = turma), label.x = 7, label.y=c(110,105,100))
```



Correlação

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples

Um bom Guia

Vendo correlações graficamente

```
dfe %>% select_if(is.numeric) %>% cor() %>%  
  corrplot::corrplot(.,method="number",type="upper",diag=FALSE )
```

Vendo correlações graficamente

Análise de
variância e
correlação

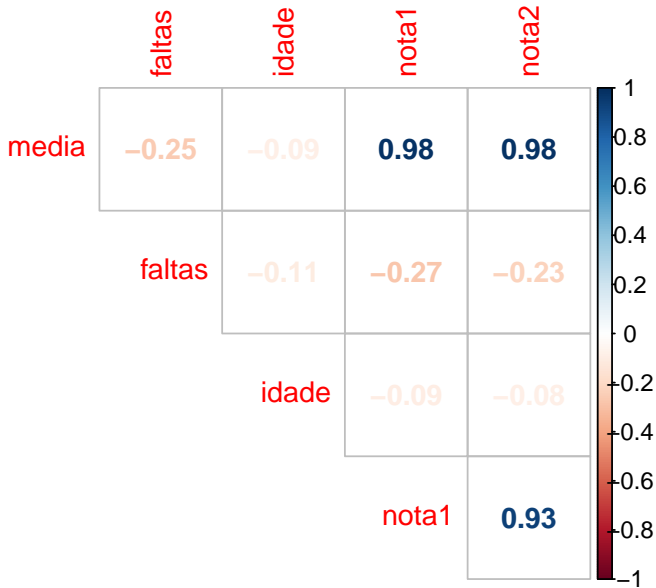
Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

Modelo
Linear
Simples



**Análise de
variância e
correlação**

**Frederico
Bertholini**

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

**Modelo
Linear
Simples**

Modelo Linear Simples

Modelo Linear Simples

Análise de
variância e
correlação

Frederico
Bertholini

Formalizando
Hipóteses

Testando
diferentes
hipóteses

Correlação

**Modelo
Linear
Simples**

Modelo linear simples estimado por mínimos quadrados ordinários (MQO)..