Lista 1 Planejamento

Ex 2

$$E\left[\sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})\right] = \sum_{i=1}^{a} E\left[n_i\left(\bar{y}_{i.}^2 - N\bar{y}_{..}^2\right)\right] =$$

$$= \sum_{i=1}^{a} n_i\left[\left(Var[\bar{y}_{i.}] + E^2[\bar{y}_{i.}]\right)\right] - N\left(Var[\bar{y}_{..}] +\right.$$

$$\left. + E^2[\bar{y}_{..}]\right)$$

Como $\bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, $E[\bar{y}_{i.}] = \mu + \tau_i$

e $Var[\bar{y}_{i.}] = \sigma^2/n_i$, e como $\bar{y}_{..} =$

$= \frac{1}{N}\sum_{i=1}^{a} y_{i.}$, $E[\bar{y}_{..}] = \frac{1}{N}\sum_{i=1}^{a} E[\bar{y}_{i.}] =$

$$= \frac{1}{N}\sum_{i=1}^{a} n_i(\mu + \tau_i) = \mu + \sum_{i=1}^{a} n_i\tau_i/N$$

$$Var[\bar{y}_{..}] = \frac{1}{N^2} Var\left[\sum_{i=1}^{a} y_{i.}\right] = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}$$

$$\Rightarrow E[SQTrat] = \sum_{i=1}^{a} n_i \cdot \left[\frac{\sigma^2}{n_i} + (\mu + \tau_i)^2\right] - N\left(\frac{\sigma^2}{N} + \right.$$

$$\left. + \left(\sum_{i=1}^{a} n_i\tau_i/N\right)^2\right) = a\sigma^2 + N\mu^2 + 2\mu\sum_{i=1}^{a} n_i\tau_i + \sum_{i=1}^{a} n_i\tau_i^2$$

$$-\sigma^2 - N\mu^2 - 2\mu\sum n_i\tau_i - \left(\sum n_i\tau_i\right)^2/N$$

/ /

$$= (a-1)\left[\sigma^2 + \frac{n}{a-1}\sum_{i=1}^{a} \tau_i^2\right]$$

$$\therefore \mathbb{E}\left[\frac{SQTrat}{a-1}\right] = \sigma^2 + \frac{n}{a-1}\sum_{i=1}^{a} \tau_i^2,$$

quando o modelo é balanceado, é

$$= \sigma^2 + \sum_{i=1}^{a} n_i \left(\tau_i - \sum_{i=1}^{a} n_i \tau_i / n\right)^2 \Big/ a-1$$

quando é desbalanceado

↳ se $y_{ij} \sim N[\mu + \tau_i, \sigma^2] \Rightarrow$

$$\Rightarrow \sum_{j=1}^{n_i} y_{ij} \sim N\left(n_i(\mu + \tau_i), \sigma^2\right)$$

$$\Rightarrow \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{a} y_{i.} \Rightarrow \sum_{i=1}^{a} \mathbb{E}[y_{i.}] = \sum_{i=1}^{a} n_i(\mu + \tau_i)$$

tilibra

Ex 1

$$\sum_i\sum_j e_{ij}^2 = \underbrace{\sum_i\sum_j (y_{ij}-\mu-\tau_i)^2}_{Q(\mu,\tau_i)} = 0$$

$$\frac{\partial Q}{\partial \mu} = -2\sum_i\sum_j (y_{ij}-\mu-\tau_i)$$

$$\sum_i\sum_j y_{ij}-\mu-\tau_i = y_{..} - N\mu - \sum_i n_i\tau_i =$$

$$= 0 \Rightarrow \sum_i y_{i.} - \sum_i n_i\mu - \sum_i n_i\tau_i = 0 \Rightarrow$$

$$\Rightarrow \sum_i y_{i.} = \sum_i n_i\mu + \sum_i n_i\tau_i$$

$\Rightarrow y_{i.} = n_i\mu + n_i\tau_i, \forall i$ pois $\frac{\partial Q}{\partial \mu} = \frac{\partial Q}{\partial \tau_1} =$

$= \dots = \frac{\partial Q}{\partial \tau_a}$

se $\mu = 0 \Rightarrow y_{i.} = n_i\tau_i \Rightarrow \bar{y}_{i.} = \tau_i, \forall i$

se $\sum_i n_i\tau_i = 0 \Rightarrow \sum_i y_{i.} - \sum_i n_i\mu = 0 \Rightarrow$

$\Rightarrow N\mu = y_{..} \Rightarrow \mu = \bar{y}_{..}$ e $y_{i.} - n_i\mu - n_i\tau_i = 0$

então $0 = y_{i.} - n_i\bar{y}_{..} - n_i\tau_i \Rightarrow \tau_i = \bar{y}_{i.} - \bar{y}_{..}$

Ex 3:

↳ Temos então um experimento completamente aleatorizado com os 45 alunos sendo submetidos a 5 diferentes tratamentos. Utilizaremos então o modelo de desvio médio.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, \dots, 5 \\ j = 1, \dots, 9 \end{cases}$$

y_{ij}: valor da variável resposta no i-ésimo tratamento e na j-ésima unidade experimental.

μ: média global, no problema em questão é a média dos resultados da população de alunos

τ_i: efeito do tratamento i na média global

ϵ_{ij}: erro aleatório referente a j-ésima v.e no i-ésimo tratamento.

- A variável resposta: Resultado na avaliação do j-ésimo aluno no i-ésimo tratamento
- Tratamentos: Diferentes atitudes do professor para com os alunos.
- Unid. Experimentais: Alunos

• As hipóteses

$H_0: \tau_i = 0, \forall i = 1, 2, \ldots, 5$
(As atitudes dos professores com os alunos não influenciam no resultado do teste)

$H_1: \tau_i \neq 0$, p/ pelo menos um i.
(As atitudes dos professores influenciam o resultado do teste dos alunos)

↳ Caso H_0 não seja rejeitada, $\tau_i = 0 \Rightarrow$
$\Rightarrow y_{ij} = \mu + \varepsilon_{ij}$

• Análise gráfica: Pela visualização dos boxplots pode-se notar que as medianas são um pouco distantes umas das outras. Os tratamentos que obteram notas mais simétricas os alunos CENSURADOS e ELOGIADOS, sendo que os alunos elogiados tiraram notas maiores, possuindo assim um media alta com pouca variabilidade. Os alunos ignorados possuem as menores métricas de posição (mediana e média).

• Estatística de teste e sua distribuição sob H_0

$$F_0 = \frac{QMTRAT}{QMErro} \sim F_{a-1, N-a}$$

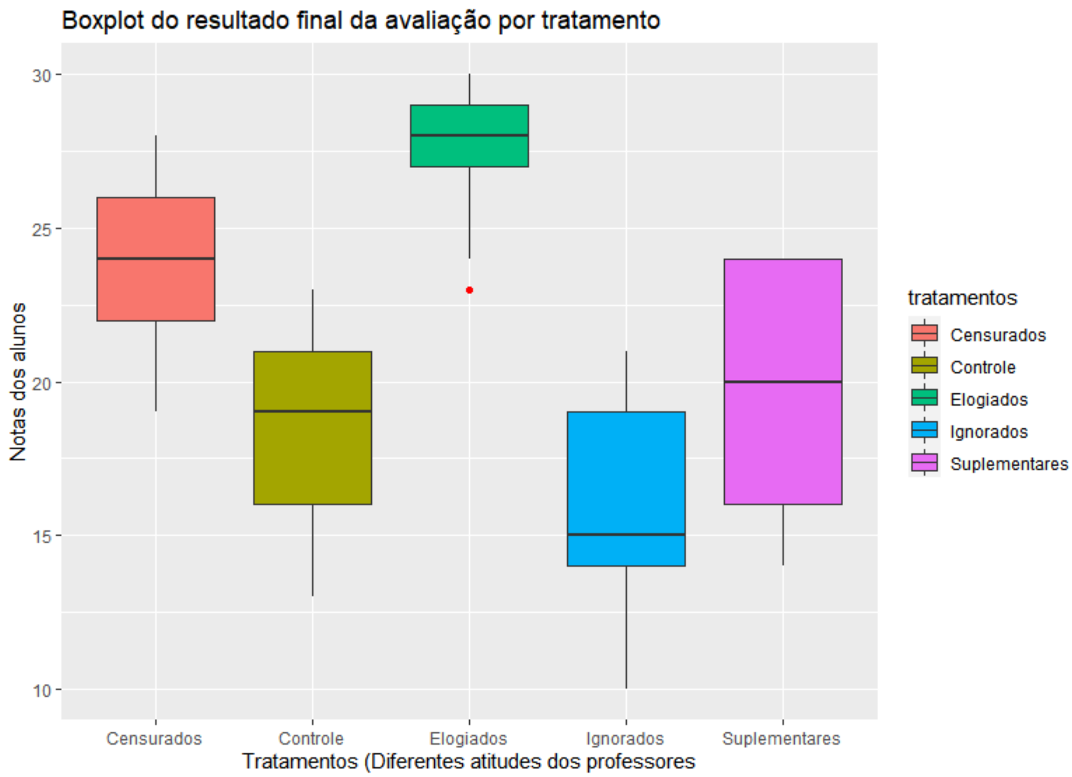
→ Pela ANOVA, p-value = $1,16 \cdot 10^{-7}$
como p-value < α = 1%, rejeita-se H_0.
Sabemos então que os tratamentos influenciam no resultado do aluno na prova.
Para saber quais médias diferem usaremos o teste de tukey

$H_0 : \mu_i = \mu_{i'} \quad \forall (i, i')$
(As notas médias dos alunos na prova final não diferem)

$H_1 : \mu_i \neq \mu_{i'}$ p/ pelo menos um (i, i')
(Há diferenças nas notas dos alunos na prova final por tratamento)

Exercício 3 - Continuação

A interpretação da imagem abaixo está na descrição do exercício 3,



Abaixo estará a tabela ANOVA,

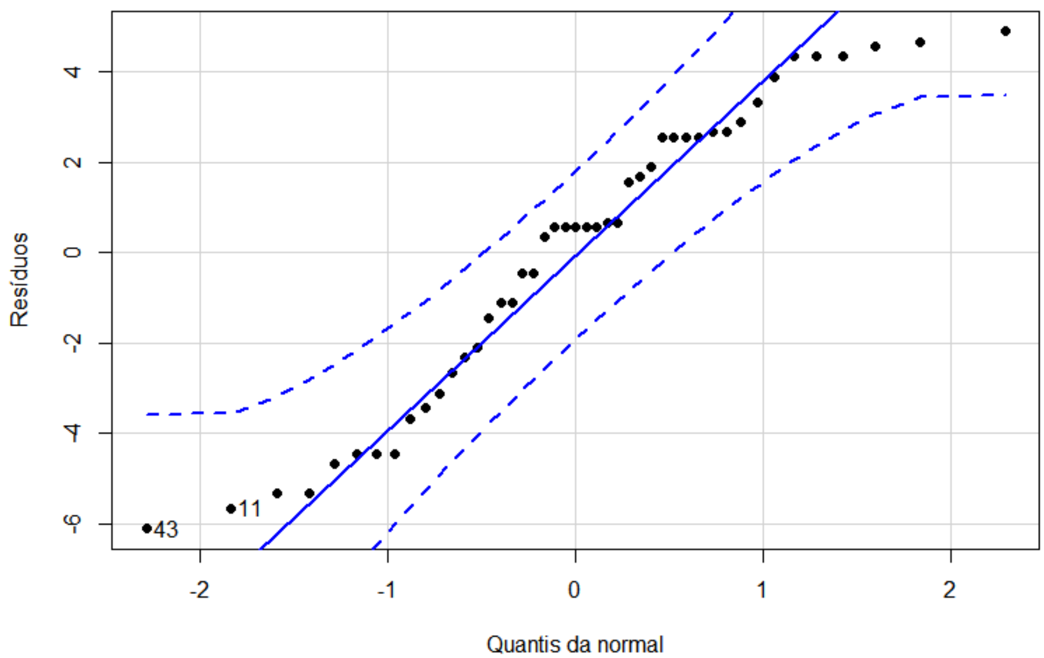
```
            Df Sum Sq Mean Sq F value   Pr(>F)    
tratamentos  4  722.7  180.67   15.27 1.16e-07 ***
Residuals   40  473.3   11.83                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Podemos ver na tabela que SQTrat = 722,7 e SQErros = 473,3. Na coluna ao lado das somas de quadrados temos os quadrados médios e em seguida o valor-p, mostrando o menor grau de significancia para rejeição de H0. Como nosso teste é de apenas 1% de nível se significância, podemos ver que a hipótese de que as diferentes atitudos do professor com aluno influencia no seu desempenho da prova final. Como vimos então, existe um efeito no modelo linear (desvio em relação a média global).

Verificando as suposições

Neste tipo de experimentos temos que verificar as suposições feitas no começo do estudo, essas são normalidade, independência e homocedasticidade dos erros.

Pelo gráfico de normalidade podemos ver que os erros quase se alinham com a reta normal,



Olhando para o teste de hipóteses de normalidade, podemos ver que a hipótese nula (os erros são amostra aleatória de uma distribuição normal) não será rejeitada, adotando $\alpha = 0.01$.

```
        Shapiro-Wilk normality test

data:  res
W = 0.94086, p-value = 0.02313
```

O teste de homogeneidade de variância de barlett nos retorna um valor-p maior do que 1%, portanto os erros são homocedasticos.

```
        Bartlett test of homogeneity of variances

data:  res by dados$tratamentos
Bartlett's K-squared = 2.3515, df = 4, p-value = 0.6714
```

	diff	lwr	upr	p adj
Controle-Censurados	-5.111111	-9.7425867	-0.4796356	0.0241651
Elogiados-Censurados	4.000000	-0.6314756	8.6314756	0.1189108
Ignorados-Censurados	-7.333333	-11.9648089	-2.7018578	0.0004900
Suplementares-Censurados	-3.777778	-8.4092533	0.8536978	0.1567932
Elogiados-Controle	9.111111	4.4796356	13.7425867	0.0000155
Ignorados-Controle	-2.222222	-6.8536978	2.4092533	0.6496469
Suplementares-Controle	1.333333	-3.2981422	5.9648089	0.9222024
Ignorados-Elogiados	-11.333333	-15.9648089	-6.7018578	0.0000002
Suplementares-Elogiados	-7.777778	-12.4092533	-3.1463022	0.0002101
Suplementares-Ignorados	3.555556	-1.0759200	8.1870311	0.2033828

Para saber quais são as médias que diferem uma da outra, é necessário o teste de tukey.

Neste teste de tukey, podemos ver na ultima coluna os valores p, concluímos então que o rendimento na prova final diferem entre os alunos foram Ignorados-Censurados, Elogiados-Controle, Suplementares-Elogiados e Ignorados-Elogiados. Entre as outras notas médias o comportamento é próximo.