

Lista 3 - MAE0317

Guilherme Navarro 8943160, Leonardo 9793436, Diogo 10297481, Eduardo 10263890

Exercício 1

Um estudo foi conduzido para avaliar o efeito de interação entre dois medicamentos (X e Y) usados para estimular o crescimento de crianças acometidas com uma particular síndrome que atinge o desenvolvimento infantil. Sabe-se que o efeito de cada medicamento é modesto, mas o efeito da combinação das duas drogas (X e Y) não tem sido investigado. Os seguintes resultados foram obtidos da avaliação da taxa de crescimento de 16 pacientes:

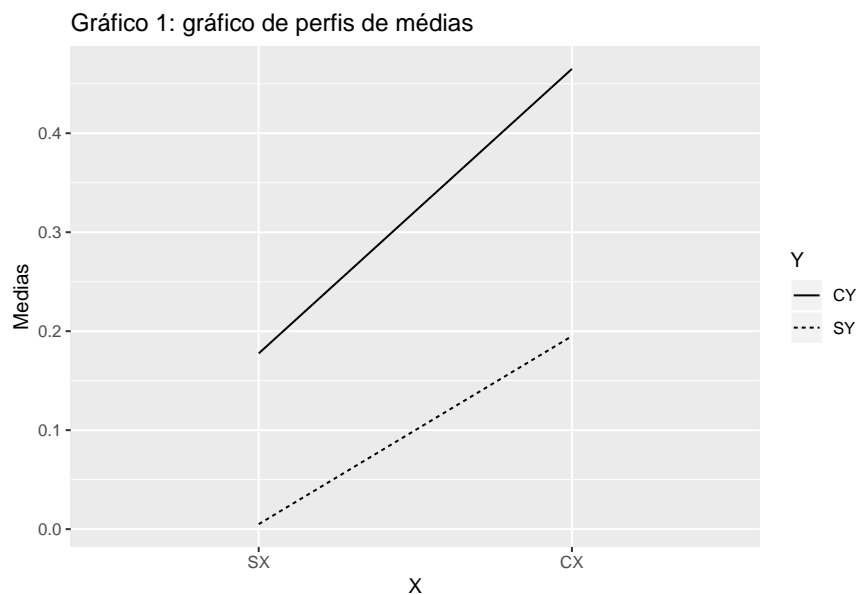
Pacientes			
1	2	3	4
0.02 (A)	0.15 (B)	0.45 (D)	0.18 (C)
0.27 (B)	0.24 (C)	-0.01 (A)	0.58 (D)
0.11 (C)	0.35 (D)	0.14 (B)	-0.03 (A)
0.48 (D)	0.04 (A)	0.18 (C)	0.22 (B)

A=Placebo B= Droga X C=Droga Y D=Drogas X e Y

- (a) Construa o gráfico de perfis de médias. Há indicação efeito de interação entre as drogas X e Y?

Resolução

Seja SX = Sem a utilização do medicamento X; SY = Sem a utilização do medicamento Y; CX = Com a utilização do medicamento X; CY = Com a utilização do medicamento Y



Podemos visualizar pelo gráfico que quase não há interação pois as retas estão praticamente paralelas entre si.

- (b) Ajuste um modelo de ANOVA para estes dados. Apresente o modelo estrutural e distribucional adotados. Realize uma análise de diagnóstico das premissas do modelo. Discuta os resultados.

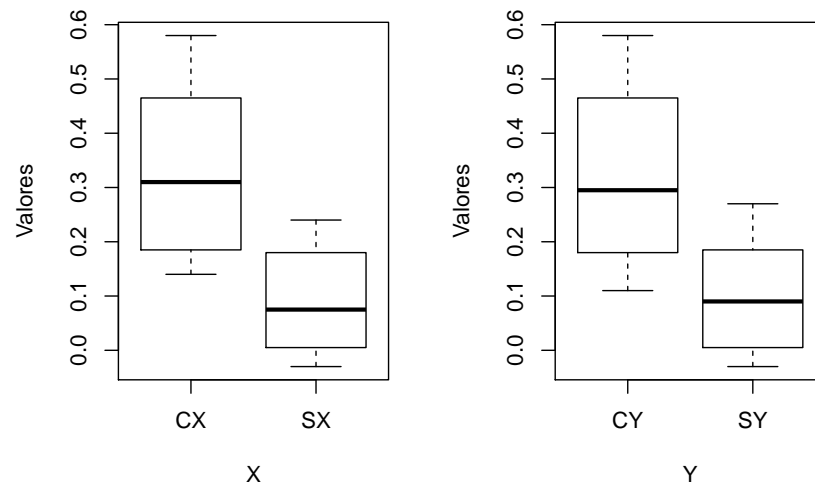
Resolução

Tabelas 1 e 2: Médias da taxa de crescimento com a presença ou ausência dos medicamentos X e Y.

```
##      CY      SY
## 0.32125 0.10000
```

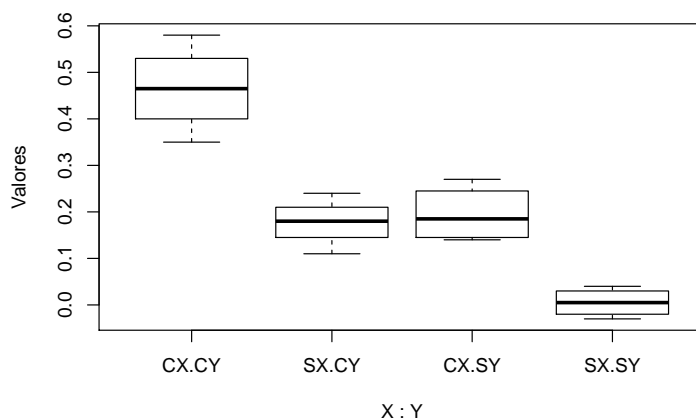
```
##      CX      SX
## 0.33000 0.09125
```

Gráficos 2 e 3: Boxplot's da taxa de crescimento com a presença ou ausência dos medicamentos X ou Y



Pelos gráficos 2 e 3 podemos notar que os medicamentos tem indícios que causam efeito positivo no crescimento.

Gráfico 4: Boxplot da taxa de crescimento com a presença ou ausência dos medicamentos X e Y



No gráfico 4, já podemos notar que quando é utilizado os dois medicamentos a média de crescimento é maior.

Table 1: Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	0.2280063	0.2280063	55.190620	0.0000080
Y	1	0.1958063	0.1958063	47.396369	0.0000169
X:Y	1	0.0095062	0.0095062	2.301059	0.1551793
Residuals	12	0.0495750	0.0041313	NA	NA

O modelo estrutural é: $Y_{ijk} = \mu + T_i + \beta_j + \gamma_{ij} + e_{ijk}$ e o modelo distribuicional é: $e_{ijk} \sim N(0; \sigma^2); Y_{ijk} \sim N(\mu_{ij}; \sigma^2)$.

Gráfico 5:Resíduos vs índice

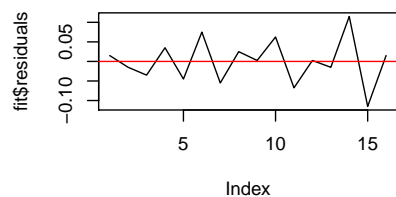


Gráfico 6:Resíduos vs Preditos

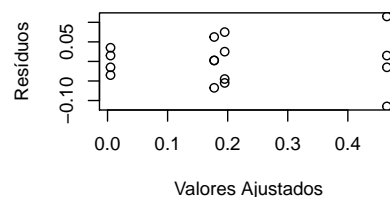


Gráfico 7:Histograma dos resíduos padroni

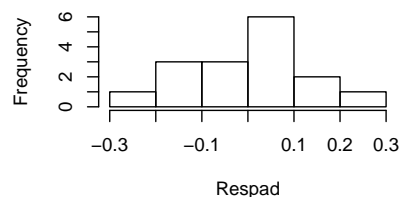
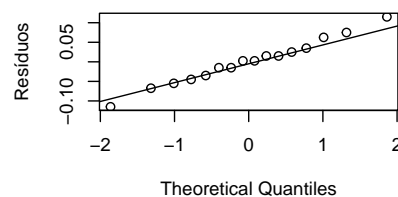


Gráfico 8:QQ-Plot Normal



No gráfico 5 é possível ver a independência dos resíduos. No gráfico 6, existe homocedasticidade nos dados. Nos gráficos 7 e 8 existe normalidade nos dados.

Tabela 4: Teste de normalidade

```
##
## Shapiro-Wilk normality test
##
## data:  res
## W = 0.99216, p-value = 0.9999
```

O teste da tabela 4 confirma a normalidade dos dados a um nível de significância de 5%.

- (c) Apresente contrastes apropriados entre as médias dos quatro tratamentos (A, B, C e D) que possam ser usados para testar os efeitos principais e de efeito de interação entre as drogas X e Y. Para estes contrastes obtenha intervalos de confiança.

Resolução

Tabela 5: intervalos de confiança

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Valores ~ X * Y, data = tabela)
##
## $X
##          diff          lwr          upr p adj
## SX-CX -0.23875 -0.3087714 -0.1687286 8e-06
##
## $Y
##          diff          lwr          upr    p adj
## SY-CY -0.22125 -0.2912714 -0.1512286 1.69e-05
##
## $`X:Y`
##          diff          lwr          upr    p adj
## SX:CY-CX:CY -0.2875 -0.422434 -0.15256598 0.0001928
## CX:SY-CX:CY -0.2700 -0.404934 -0.13506598 0.0003431
## SX:SY-CX:CY -0.4600 -0.594934 -0.32506598 0.0000017
## CX:SY-SX:CY  0.0175 -0.117434  0.15243402 0.9796870
## SX:SY-SX:CY -0.1725 -0.307434 -0.03756598 0.0117988
## SX:SY-CX:SY -0.1900 -0.324934 -0.05506598 0.0060452
```

Os intervalos de confiança apontam que quando há a presença de pelo menos um medicamento a média de crescimento é maior. Quando comparamos o uso de apenas um deles com o uso conjunto, ao utilizar os dois a média é maior. Mas quando comparamos o uso de apenas do medicamento X com apenas o Y, não a diferença na média.

- (d) Como os medicamentos X e Y devem ser administrados para se obter eficiência máxima no desenvolvimento infantil?

Resolução

A melhor maneira de administrar os medicamentos para se obter eficiência máxima no desenvolvimento infantil é de ingerir os dois medicamentos combinados.

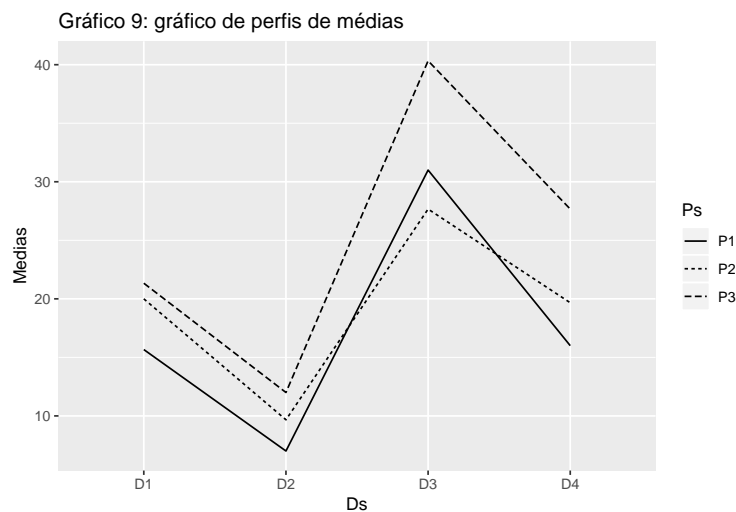
Exercício 2

Uma companhia pretende aumentar suas vendas e para isso adotará um programa de incentivo. Três diferentes programas foram propostos e estão sendo aplicados. A companhia tem 4 departamentos de vendas (D1, D2, D3 e D4) e 9 vendedores de cada departamento foram aleatoriamente selecionados e alocados a participarem de três programas de incentivo (P1, P2 ou P3). Ao final da campanha os vendedores foram avaliados quanto ao número de itens vendidos. Os resultados estão apresentados a seguir:

	P1	P2	P3
D1	16	18	17
	22	27	34
	9	15	13
D2	8	13	14
	7	5	12
	6	11	10
D3	44	31	28
	27	33	54
	22	19	39
D4	17	19	24
	17	23	28
	14	17	31

Realize uma Análise de Variância dos dados, faça suposições necessárias e conclua sobre a eficiência da campanha de incentivo.

Resolução



Podemos verificar que os programas 1 e 3 aparentemente não possuem interação por serem praticamente paralelas entre si.

Tabelas 6 e 7: Número de itens vendidos por programas ou por Departamentos

```
##          P1          P2          P3
## 17.41667 19.25000 25.33333
```

```
##          D1          D2          D3          D4
## 19.000000  9.555556 33.000000 21.111111
```

Gráficos 10 e 11: Boxplot's do número de itens vendidos por programas ou por Departamentos

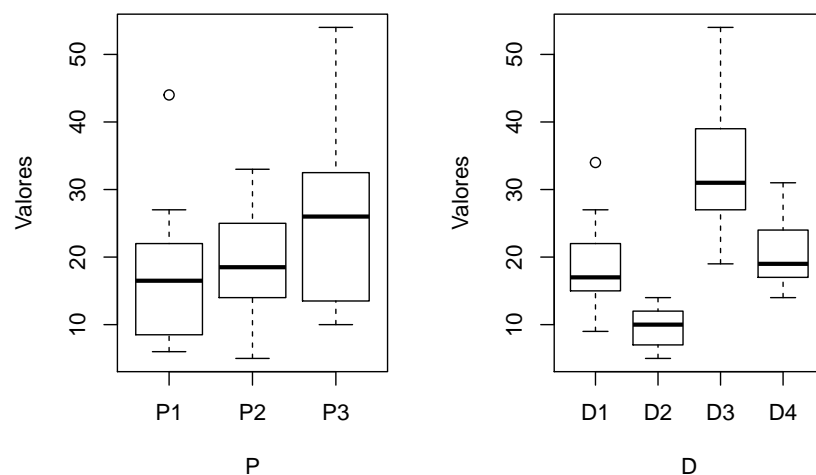
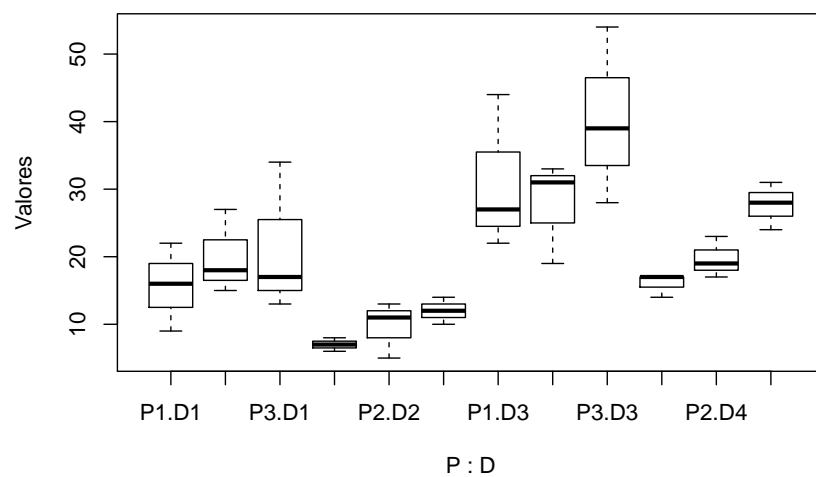


Gráfico 12: Boxplot do número de itens vendidos por programas e por Departamentos

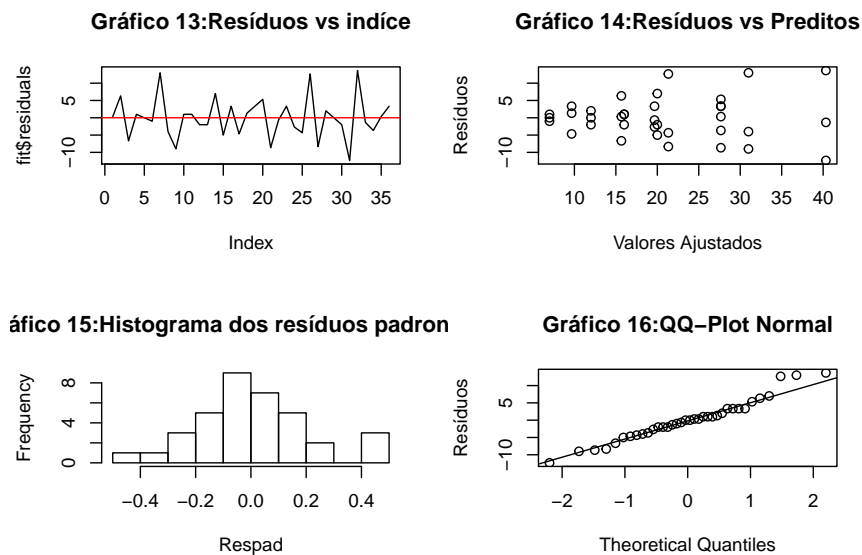


Nos gráficos 11 e 12, vemos que aparentemente o departamento 2 e 3 possuem uma grande diferença de médias.

Table 2: Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
P	2	412.1667	206.08333	4.0320652	0.0309244
D	3	2506.8889	835.62963	16.3492754	0.0000053
P:D	6	150.2778	25.04630	0.4900362	0.8092780
Residuals	24	1226.6667	51.11111	NA	NA

Podemos verificar que não há efeito de interação, a um nível de significância de 5%.



No gráfico 13 é possível ver a independência dos resíduos. No gráfico 14, não existe homocedasticidade nos dados. Nos gráficos 15 e 16 existe normalidade nos dados.

Tabela 9: Teste de normalidade

```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.96308, p-value = 0.2671
```

Para verificar a normalidade fazemos o teste acima.

Tabela 10: intervalos de confiança para os programas

```
##  Tukey multiple comparisons of means
##    95% family-wise confidence level
##
## Fit: aov(formula = Valores ~ P, data = companhia)
##
```

```
## $P
##          diff          lwr          upr          p adj
## P2-P1 1.833333 -9.034323 12.70099 0.9101319
## P3-P1 7.916667 -2.950989 18.78432 0.1894088
## P3-P2 6.083333 -4.784323 16.95099 0.3660849
```

Tabela 11: intervalos de confiança para os departamentos

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Valores ~ D, data = companhia)
##
## $D
##          diff          lwr          upr          p adj
## D2-D1 -9.444444 -18.994467 0.105578 0.0534737
## D3-D1 14.000000 4.449978 23.550022 0.0020467
## D4-D1 2.111111 -7.438911 11.661134 0.9316375
## D3-D2 23.444444 13.894422 32.994467 0.0000010
## D4-D2 11.555556 2.005533 21.105578 0.0127859
## D4-D3 -11.888889 -21.438911 -2.338866 0.0100521
```

Podemos concluir que não há diferença entre os programas. E entre os departamentos: o departamento 1 é igual ao 2 que são piores que o departamento 3, o departamento 1 é igual ao 4 e o departamento 2 é pior que o 4 que por sua vez é pior que o 3.

Exercício 3

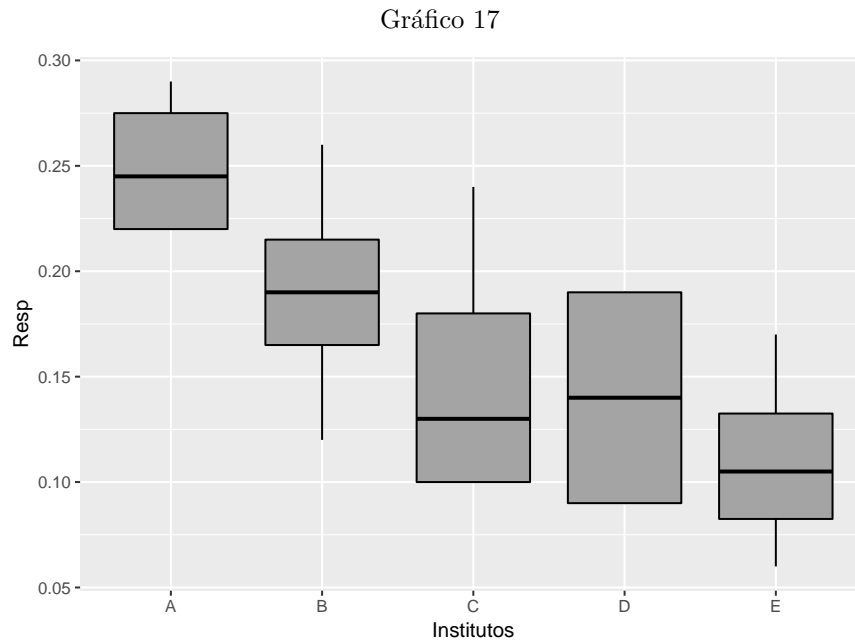
Em uma reunião de negócios, Felipe, vice-presidente da empresa EE, fez a seguinte afirmação: “Eu acredito que nós temos um número exagerado de professores. Precisamos cortar custos”. Foi então, logo contestado: “Temos tantos professores quanto nossos concorrentes de outras instituições”. Felipe continuou: “Sim, mas vocês sabem que a nossa meta de ensino é B e a de outras instituições são diferentes da nossa e isso influencia na demanda de professores”. Frederico concluiu: “Calma. Vou coletar dados para podermos tomar uma decisão mais fundamentada”. Depois de alguns dias Frederico apresentou os seguintes dados da razão do número de professores pelo número de alunos (NP/NA) para 20 instituições de acordo com a meta de ensino:

Instituição	Meta de Ensino	Razão NP/NA	Instituição	Meta de Ensino	Razão NP/NA
1	A	0,29	11	C	0,10
2	A	0,27	12	C	0,10
3	A	0,22	13	D	0,19
4	A	0,22	14	D	0,19
5	B	0,26	15	D	0,09
6	B	0,20	16	D	0,09
7	B	0,18	17	E	0,17
8	B	0,12	18	E	0,12
9	C	0,24	19	E	0,09
10	C	0,16	20	E	0,06

(a) Que decisão deve ser tomada sobre a instituição EE ter ou não um número exagerada de professores?

Resolução

Para tomar a decisão será feito um modelo de ANOVA para o delineamento completamente aleatorizado (DCA) com 1 fator em 5 níveis pois, o que se quer saber é se existe a média de professores da instituição EE é igual a média de outras instituições de acordo com a meta de ensino, apenas com uma breve análise descritiva podemos notar pelo gráfico 1 que existem diferença entre as médias de professores de acordo com a meta de ensino.



Agora contruindo a tabela de ANOVA temos:

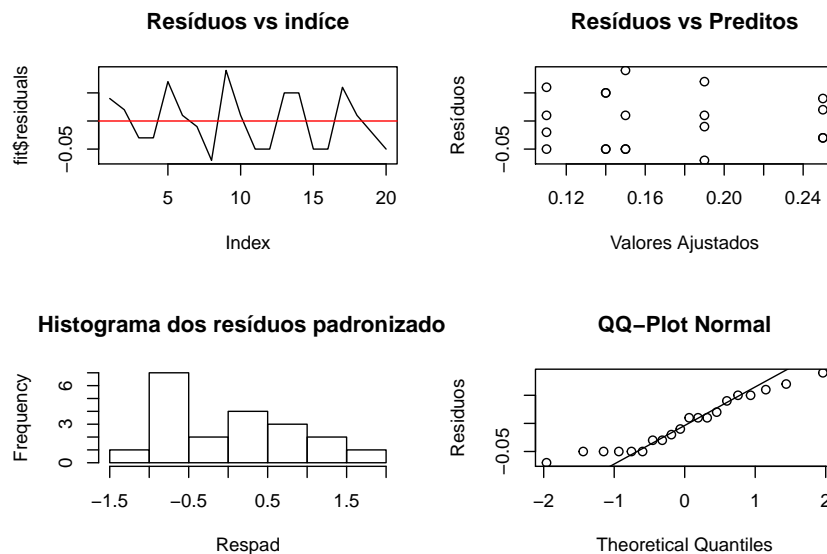
Table 3: Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Institutos	4	0.04672	0.0116800	4.018349	0.0207085
Residuals	15	0.04360	0.0029067	NA	NA

Analizando a tabela de ANOVA, podemos dizer que a um nível de significância de 5% existe diferença entre as médias de professores de acordo com a meta de ensino.

Agora faremos uma análise de diagnóstico para verificar a consistencia do *p-value* obtido na tabela de ANOVA

Gráfico 18: Análise de diagnóstico



```
##
## Shapiro-Wilk normality test
##
## data:  res
## W = 0.93275, p-value = 0.1744

##
## Bartlett test of homogeneity of variances
##
## data:  Resp by Institutos
## Bartlett's K-squared = 1.1135, df = 4, p-value = 0.8921
```

Assim podemos notar que as suposições de Normalidade e variâncias homogêneas estão satisfeitas pois a um nível de significância de 5%, o testes de Shapiro-Wilk para normalidade apresentou um p -value de 0.17, logo não rejeitamos a hipótese nula de que os dados seguem uma distribuição normal, agora para testar se as variâncias são homogêneas utilizamos o teste de Bartlett que apresentou um p -value de 0.89, logo não rejeitamos a hipótese nula de que os dados são homocedásticos, para a hipótese de independência apenas pela análise descritiva do gráfico 2 podemos concluir que as três suposições para o modelo ANOVA estão satisfeitas.

Agora para responder a pergunta de que decisão deve ser tomada sobre a instituição EE ter ou não um número exagerada de professores, faremos comparações múltiplas pelo método de Dunnett pois nossa temos um grupo de referência que no caso são os Institutos com meta de ensino B.

Table 4: Comparação pelo método de Dunnett (conf=95%)

	Diff	Lower	Upper	Decision
A-B	0.060122	-0.043907	0.16415	FTR H0
C-B	-0.03985	-0.143795	0.064096	FTR H0
D-B	-0.050035	-0.154057	0.053986	FTR H0
E-B	-0.080113	-0.184209	0.023982	FTR H0

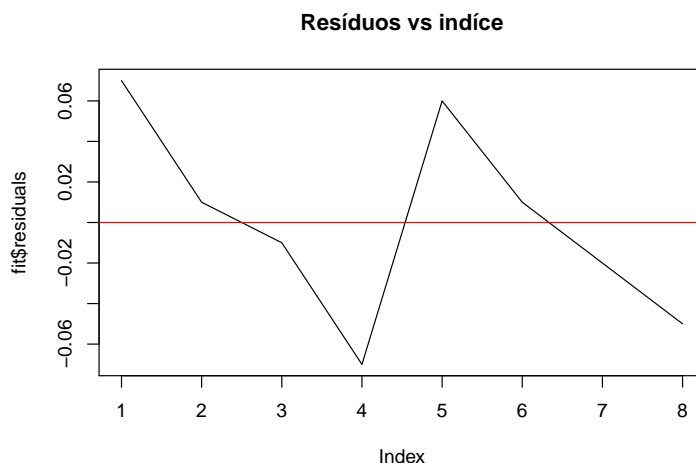
Com isso nós podemos responder a pergunta inicial sobre a instituição EE ter ou não um número exagerada de professores, de fato a instituição EE tem a mesma quantidade de professores do que as outras instituições independente da meta de ensino.

- (b) Na reunião, Felipe questionou: Mas, e se compararmos apenas aquelas que o usam a Meta B com E? Como essa informação pode ser usada na análise? Sua inclusão modifica a decisão sugerida em a)? Faça suposições necessárias.

Resolução

Se compararmos apenas aquelas que o usam a Meta B com E, o modelo de ANOVA se resume a um teste T de comparação de médias, assim mantendo as suposições de normalidade, independência e homocedasticidade, fazendo os testes temos que:

Gráfico 19: Verificação da hipótese de independência



```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.95464, p-value = 0.7578

##
##  Bartlett test of homogeneity of variances
##
## data:  Resp by Institutos
## Bartlett's K-squared = 0.1102, df = 1, p-value = 0.7399
```

Assim podemos concluir que a hipótese de independência está satisfeita pela análise do gráfico 4, fixando um nível de significância de 5% a hipótese de normalidade também pois obtivemos um *p-value* de 0.76, portanto não rejeitamos a hipótese de que os dados tem distribuição normal, para a homocedasticidade o teste de Bartlett nos retornou um *p-value* de 0.74 o que nos faz não rejeitar a hipótese de variâncias homogêneas, satisfeitas todas as hipóteses podemos prosseguir com o teste T de comparação de médias, onde as hipóteses podem ser definidas como:

$$\begin{cases} H_0 : \mu_B - \mu_E = 0 \\ H_1 : \mu_B - \mu_E \neq 0 \end{cases}$$

```
##
## Two Sample t-test
##
## data: base.r1$Resp by base.r1$Institutos
## t = 2.1509, df = 6, p-value = 0.07501
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01100846 0.17100846
## sample estimates:
## mean in group B mean in group E
## 0.19 0.11
```

Assim a um nível de significância de 5% não rejeitamos a hipótese nula, logo assim como no item a, não obtivemos diferença entre os Institutos com metas de ensino B e E. Essa informação pode ser usada na análise apenas para frisar a ideia de que não diferença entre as médias de Institutos, provando que vice presidente da EE estava errado.

Exercício 4

Um experimento tem por finalidade estabelecer condições ótimas para o funcionamento de um sistema de tratamento de água. Os seguintes fatores (e seus níveis) estão envolvidos na operação:

A: Concentração de reagente de biomassa (3000 e 6000 mg/l)

B: Concentração de clarificador de biomassa (8000 e 12000mg/l)

C: Taxa de fluxo (78.5 e 940 m³/d)

Uma variável Y, que mede a qualidade da água, é adotada no estudo. Suponha que o interesse do experimento seja encontrar os níveis dos fatores que maximizam a qualidade da água. Os resultados são mostrados a seguir:

A	B	C	Y
-1	-1	-1	195
-1	-1	1	496
-1	1	-1	87
-1	1	1	1371
1	-1	-1	102
1	-1	1	1001
1	1	-1	354
1	1	1	775

(a) Que delineamento experimental foi adotado no estudo? Que limitações podem ser apontadas?

Resolução

O delineamento experimental adotado foi o completamente aleatorizado com 2 fatores (2 x 2 x 2). As limitações que podem ser observadas são é que temos uma observação por fator e nível, logo não teremos não temos réplicas, assim não podemos ajustar um modelo completo com todas as interações, pois não teríamos graus de liberdade para os resíduos.

(b) Construa gráficos apropriados para representar os efeitos principais e de interação de primeira (aos pares de tratamentos) e segunda ordem (para os três tratamentos) entre os fatores sob estudo. Comente.

Resolução

Gráfico 20: Perfis de média (A com B)

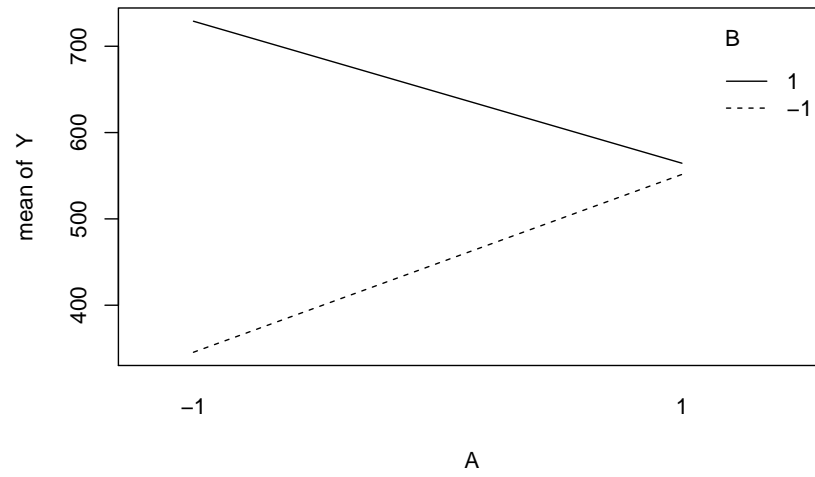


Gráfico 21: Perfis de média (A com C)

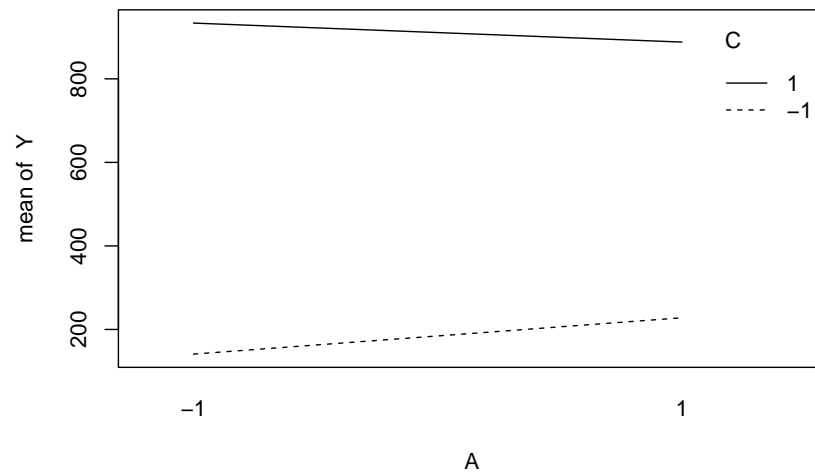


Gráfico 22: Perfis de média (B com C)

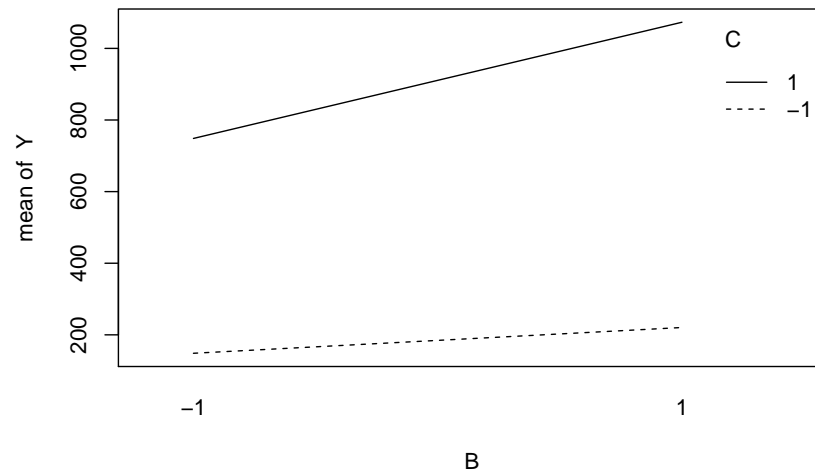


Gráfico 23: Perfis de média (A:B com C)

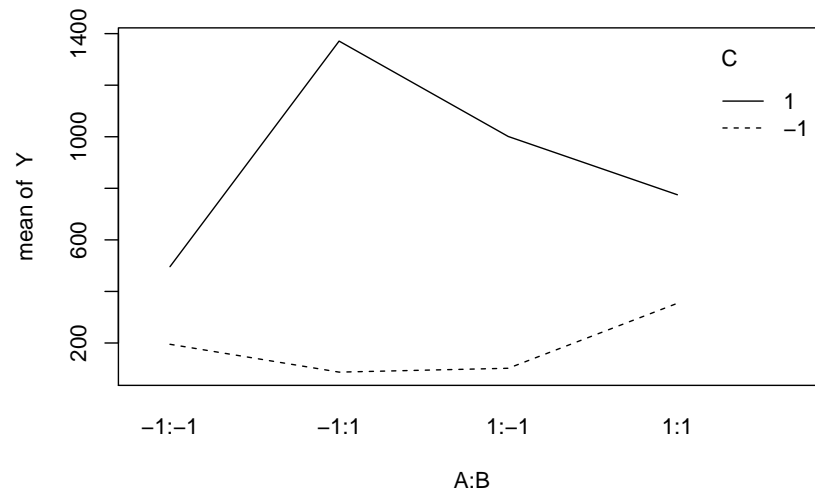
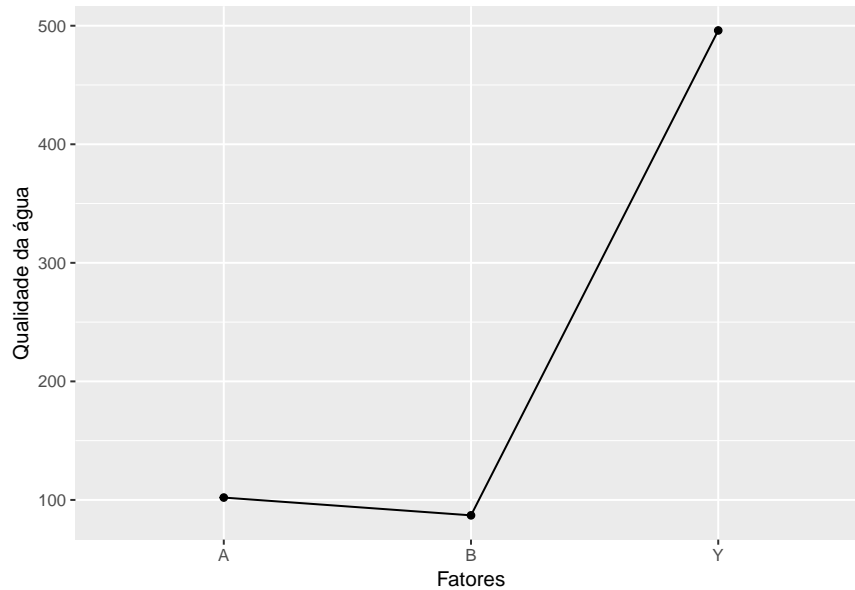


Gráfico 24: Perfis de média (Fatores)



Podemos notar descritivamente pelos gráficos 21 e 22 que quase não há interação pois as retas estão praticamente paralelas entre si, porém nos gráficos 20 e 23 podemos notar uma diferença significativa na interação dos fatores A com B com C, pois as retas tendem a se cruzar, e no gráfico 24 notamos uma diferença significativa da média com dos fatores A e B (praticamente iguais), porém a média do fator C é muito maior sendo o componente principal para a qualidade da água.

- (c) Proponha uma modelo de ANOVA (reduzido) para estes dados. Como os fatores influenciam na qualidade da água? Quais níveis dos fatores maximizam Y?

Resolução

Table 5: Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	861.125	861.125	0.0091583	0.9283623
B	1	78606.125	78606.125	0.8359986	0.4122732
C	1	1054878.125	1054878.125	11.2189300	0.0285830
Residuals	4	376106.500	94026.625	NA	NA

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Y ~ A + B + C, data = tabela)
##
## $A
##      diff      lwr      upr    p adj
## 1--1 20.75 -581.2568 622.7568 0.9283623
##
## $B
##      diff      lwr      upr    p adj
## 1--1 198.25 -403.7568 800.2568 0.4122734
```

```
##
## $C
##      diff      lwr      upr      p adj
## 1--1 726.25 124.2432 1328.257 0.0285837
```

Assim como esperado, os fatores A com B tem médias iguais a um nível de significância de 5%, porém o fator C não tem média igual os outros, no caso é maior que dos fatores A e B, mas também podemos notar uma interação entre os fatores A, B e C ao mesmo tempo. Os níveis que maximizam Y são do fator A é concentração de reagente de biomassa 3000 mg/l do fator B concentração de clarificador de biomassa 12000mg/l e do fator C taxa de fluxo 940 m3/d.

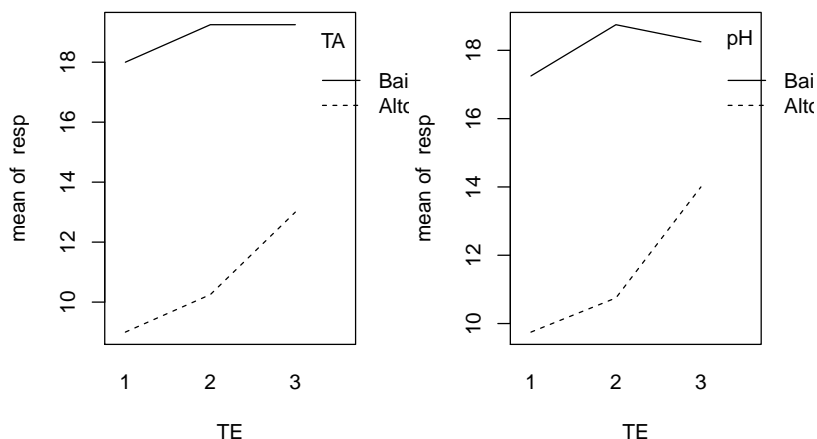
Exercício 6

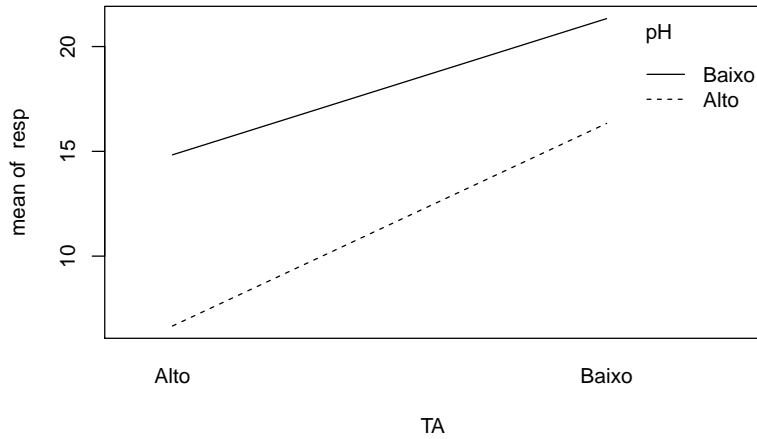
Quando a polpa de frutas secas é usada na gastronomia, a presença de granulações afeta negativamente qualidades sensoriais dos alimentos. Um experimento foi conduzido para determinar como essa sensação é afetada pela temperatura de secagem (TE), acidez (pH) da polpa e teor de açúcar (TA). Duas replicações foram consideradas. As medidas dessa sensação são apresentadas a seguir. Analise os dados para determinar quais fatores influenciam a sensação de granulações no alimento e qual combinação de fatores leva a uma menor sensação.

TE	Rep	TA Baixo		TA Alto	
		pH Baixo	pH Alto	pH Baixo	pH Alto
1	1	21	12	13	1
	2	21	18	14	8
2	1	23	14	13	1
	2	23	17	16	11
3	1	17	20	16	14
	2	23	17	17	5

Resolução

Grafico 25: perfis de média





Pelos gráficos plotados, temos que não há interação entre TA e Ph. Temos também que pode não haver uma interação entre os fatores TA-TE e Ph-TE, visto que as retas nos gráficos aparentam estar quase paralelas.

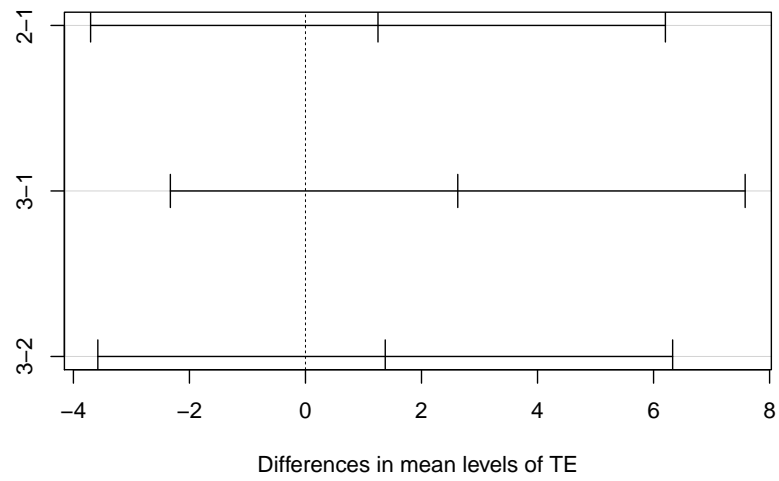
Table 6: Tabela de ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TE	2	27.583333	13.791667	1.0000000	0.3965695
TA	1	392.041667	392.041667	28.4259819	0.0001789
pH	1	260.041667	260.041667	18.8549849	0.0009578
TE:TA	2	10.083333	5.041667	0.3655589	0.7012748
TE:pH	2	16.583333	8.291667	0.6012085	0.5638542
TA:pH	1	15.041667	15.041667	1.0906344	0.3169127
TE:TA:pH	2	5.083333	2.541667	0.1842900	0.8340045
Residuals	12	165.500000	13.791667	NA	NA

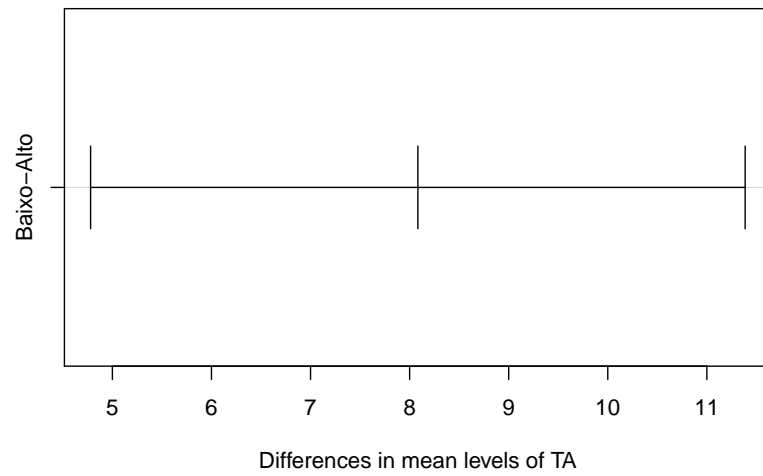
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = resp ~ TA + pH, data = aovbase.b)
##
## $TA
##          diff      lwr      upr    p adj
## Baixo-Alto 8.083333 5.213941 10.95273 8.2e-06
##
## $pH
##          diff      lwr      upr    p adj
## Baixo-Alto 6.583333 3.713941 9.452725 0.000103
```

Gráfico 26: Comparações múltiplas

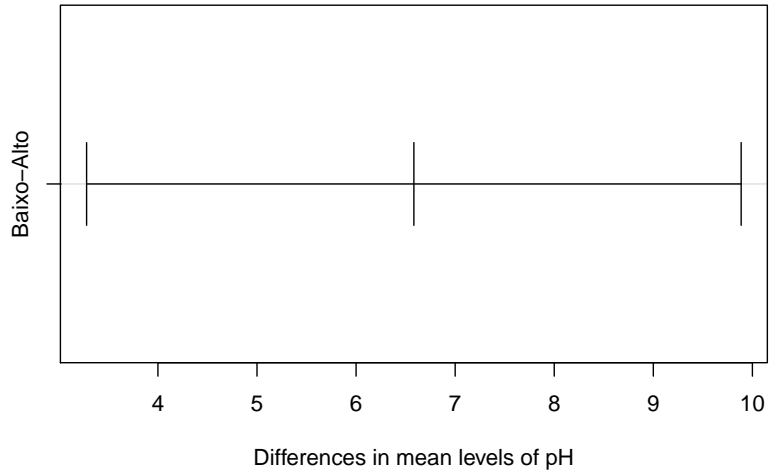
95% family-wise confidence level



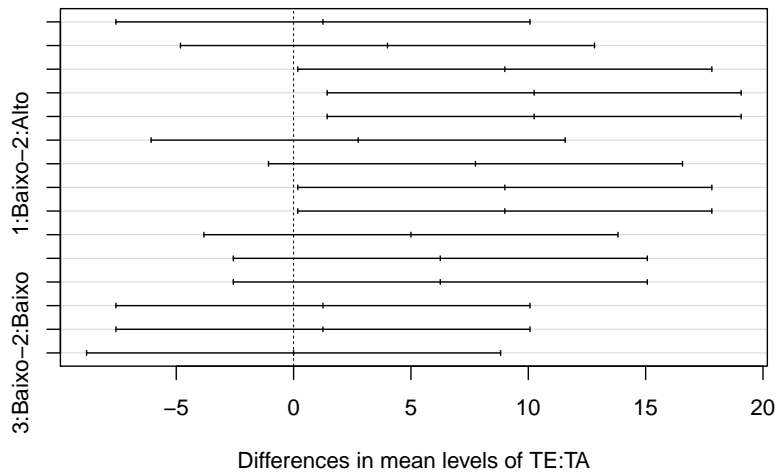
95% family-wise confidence level



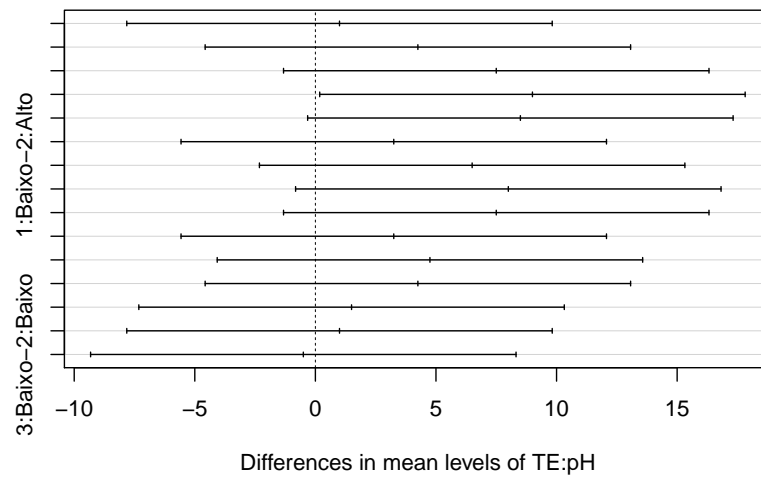
95% family-wise confidence level



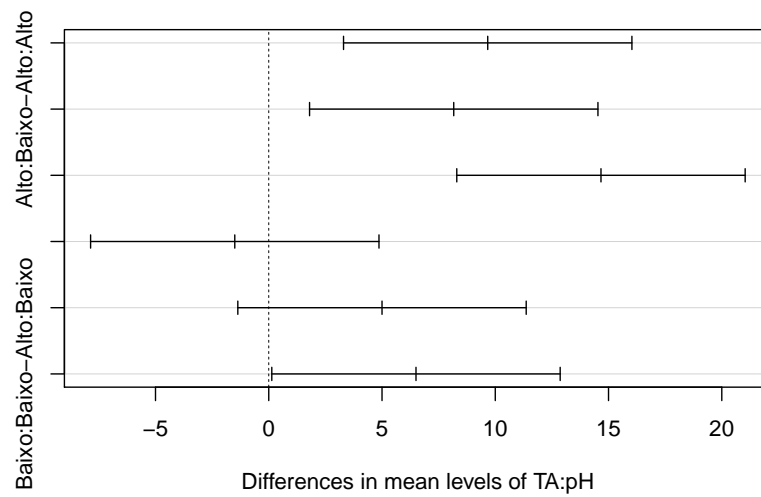
95% family-wise confidence level

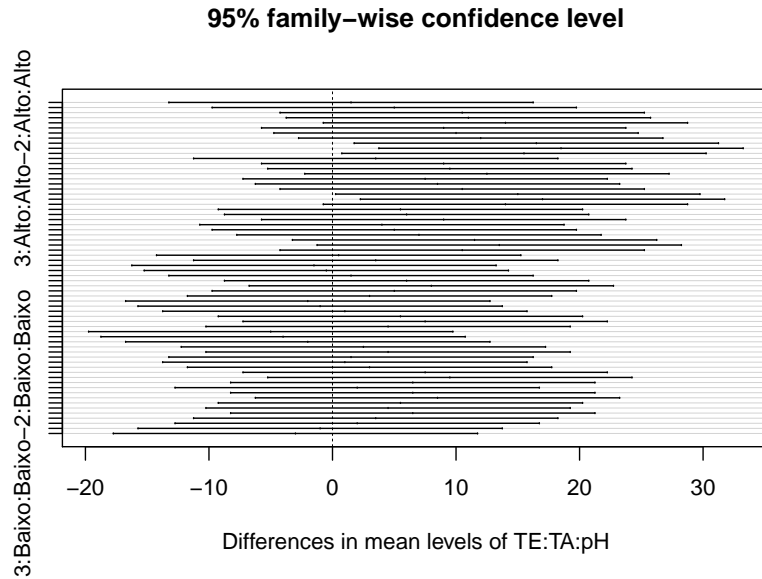


95% family-wise confidence level



95% family-wise confidence level

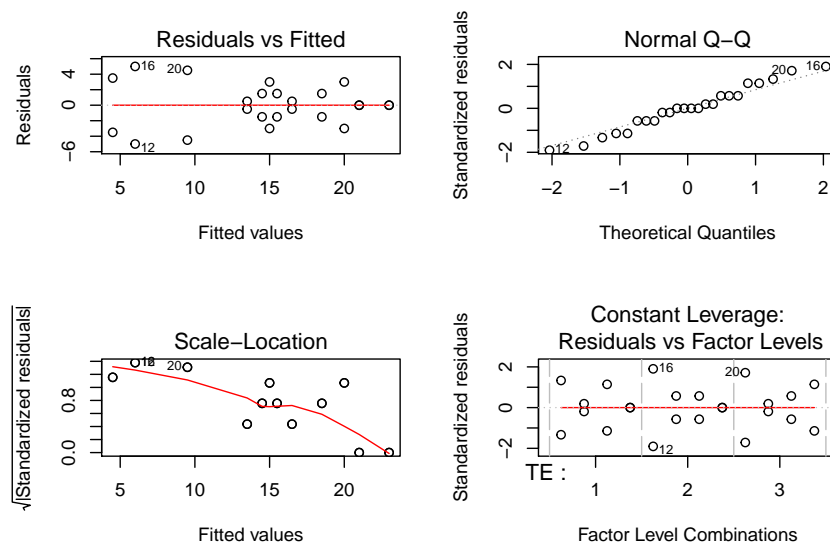


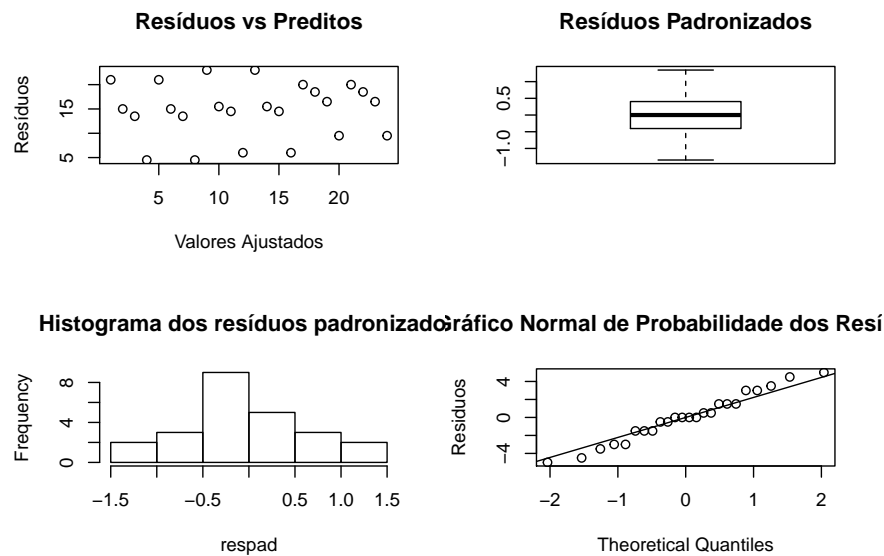


Os resultados da tabela de anova confirmam que não há significância para considerarmos interação com nenhuma combinação de fatores. Os fatores que influenciam na granulação são TA e pH. As comparações múltiplas indicam que as combinações ALTO:ALTO(TA:Ph) são as que produzem os menor sensação de granulação.

Para verificar a significância dos testes realizados acima iremos realizar uma análise de diagnósticos

Gráfico 27: Análise de Diagnóstico





```
##
## Shapiro-Wilk normality test
##
## data: avres
## W = 0.97507, p-value = 0.7907
```

A um nível de significância de 5%, não rejeitamos a hipótese de que os resíduos tem distribuição normal. Os gráficos e o teste acima indicam que as suposições para a aplicação de anova estão satisfeitas.

Códigos

```
# Exercício 1

library(reshape2)
library(dplyr)
library(ggplot2)
X <- c("SX", "SX", "SX", "SX", "CX", "CX", "CX", "CX", "SX", "SX", "SX", "SX", "CX", "CX", "CX", "CX")
Y <- c("SY", "SY", "SY", "SY", "SY", "SY", "SY", "SY", "CY", "CY", "CY", "CY", "CY", "CY", "CY", "CY")
Valores <- c(0.02, -0.01, -0.03, 0.04, 0.15, 0.27, 0.14, 0.22, 0.18, 0.24, 0.11, 0.18, 0.45, 0.58, 0.35, 0.48)
tabela <- data.frame(X, Y, Valores)

SCY = c("SY", "CY")
SX = c(0.005, 0.1775)
CX = c(0.195, 0.465)
media <- data.frame(SCY, SX, CX)

data2 <- melt(media, SCY.vars = "SCY")
data2 <- group_by(data2, SCY)
names(data2) <- c("Y", "X", "Medias")
ggplot(data2) +
```

[illegible]

item b

```
tapply(tabela$Valores,tabela$X, mean)
```

```
par(mfrow = c(1,2))
```

```
boxplot(Valores ~ Y, data = tabela)
```

```
par(mfrow = c(1,1))
```

```
m1 = lm(Valores ~ X*Y, data = tabela)
```

```
knitr::kable(caption = "Tabela de ANOVA", an)
```

```
fit <- aov(Valores~X*Y, tabela)
```

```
s2 <- anov$"Mean Sq"[2]
```

```
plot(fit$residuals, type="l")
```

```
abline(h=0, col="red")
```

```
plot(fit$fit, fit$res, xlab="Valores Ajustados", ylab="Resíduos")
```

```
res <- fit$res
```

```
hist(Respad, main=NULL)
```

```
qqnorm(res,ylab="Resíduos", main=NULL)
```

```
title("Gráfico8:QQ-Plot Normal")
```

```
shapiro.test(res)
```

```
# item c
```

TukeyHSD(a1)

Exercício 2

[illegible]

```
Valores <- c(16,22,9,8,7,6,44,27,22,17,17,14,18,27,15,13,5,11,31,33,19,19,23,17,17,34,13,14,12,10,28,54)
```

```
companhia <- data.frame(P,D,Valores)
```

```
Ps <- c("P1", "P2", "P3")
```

```

D1 <- c(mean(Valores[1:3]),mean(Valores[13:15]),mean(Valores[25:27]))
D2 <- c(mean(Valores[4:6]),mean(Valores[16:18]),mean(Valores[28:30]))
D3 <- c(mean(Valores[7:9]),mean(Valores[19:21]),mean(Valores[31:33]))
D4 <- c(mean(Valores[10:12]),mean(Valores[22:24]),mean(Valores[34:36]))
Medias <- data.frame(Ps,D1,D2,D3,D4)

data2 <- melt(Medias, Ps.vars = "Ps")
data2 <- group_by(data2, Ps)
names(data2) <- c("Ps","Ds","Medias")
ggplot(data2) +
  geom_line(aes(Ds, Medias, group =Ps,
                linetype = Ps))+ggtitle("Gráfico 9: gráfico de perfis de médias")

tapply(companhia$Valores,companhia$P, mean)
tapply(companhia$Valores,companhia$D, mean)

par(mfrow = c(1,2))
boxplot(Valores ~ P, data = companhia)
boxplot(Valores ~ D, data = companhia)

boxplot(Valores ~ P:D, data = companhia)

m1 = lm(Valores ~ P*D, data = companhia)
a <- anova(m1)
knitr::kable(caption = "Tabela de ANOVA",a)

fit <- aov(Valores~P*D, companhia)
anov <- anova(fit)
s2 <- anov$"Mean Sq"[2]
par(mfrow=c(2,2))
plot(fit$residuals, type="l")
title("Gráfico 13:Resíduos vs índice")
abline(h=0, col="red")

plot(fit$fit, fit$res, xlab="Valores Ajustados", ylab="Resíduos")
title("Gráfico 14:Resíduos vs Preditos")

res <- fit$res
Respad <- (res/sqrt(s2))
hist(Respad, main=NULL)
title("Gráfico 15:Histograma dos resíduos padronizados")

qqnorm(res,ylab="Resíduos", main=NULL)
qqline(res)
title("Gráfico 16:QQ-Plot Normal")

shapiro.test(res)

a1 = aov(Valores ~ P, data = companhia)
TukeyHSD(a1)

a1 = aov(Valores ~ D, data = companhia)
TukeyHSD(a1)

```



```

# Exercício 3

# item a

# Preparação dos dados

library(car)
library(ggplot2)
library(asbio)

base.r <- c(0.29,0.27,0.22,0.22,
            0.26,0.20,0.18,0.12,
            0.24,0.16,0.10,0.10,
            0.19,0.19,0.09,0.09,
            0.17,0.12,0.09,0.06)
base.r <- as.data.frame((matrix(base.r, 4, 5)))
colnames(base.r) <- c("A", "B", "C", "D", "E")

base.r <- stack(base.r)
colnames(base.r) <- c("Resp", "Institutos")

# Análise dos dados

base <- base.r

n <- dim(base)[1]
K <- length(levels(base$Institutos))

# box-plots
# verificar visualmente problemas:
# pontos atípicos, assimetria e variâncias heterogêneas
ggplot(base.r, aes(x=Institutos, y=Resp)) +
  geom_boxplot(fill='#A4A4A4', color="black")

#####
# Ajuste do modelo de ANOVA (DCA com 1 fator em K níveis)
#####

fit <- aov(Resp~Institutos, base)

anov <- anova(fit)
names(anov)
anov

# O que podemos concluir dessa análise? Qual é a hipótese em teste?

# Diagnóstico - Análise de resíduos

s2 <- anov$"Mean Sq"[2] # estimativa da variância residual

par(mfrow=c(2,2))

# Verificação de independência

```

```

plot(fit$residuals, type="l") # res x índice das observações
title("Resíduos vs índice")
abline(h=0, col="red")

plot(fit$fit, fit$res, xlab="Valores Ajustados", ylab="Resíduos")
title("Resíduos vs Preditos")

res <- fit$res # extraindo resíduos
Respad <- (res/sqrt(s2)) # resíduos padronizados
hist(Respad, main=NULL)
title("Histograma dos resíduos padronizados")

qqnorm(res, ylab="Resíduos", main=NULL)
qqline(res)
title("QQ-Plot Normal")

# Teste de normalidade de Shapiro-Wilk
# Qual é a hipótese H0
shapiro.test(res) #conclusão?

# Teste de homogeneidade de variâncias
bartlett.test(Resp~Institutos, data=base)

# Comparações Múltiplas
fit.du <- with(base, pairw.anova(y=Resp, x=Institutos, control="B", method="dunnett"))
d<-as.data.frame(fit.du$summary)
row.names(d) <- fit.du$comp

# item b

base.r1 <-c(0.26,0.20,0.18,0.12,
           0.17,0.12,0.09,0.06)
base.r1 <- as.data.frame((matrix(base.r1, 4, 2)))
colnames(base.r1) <- c("B", "E")

base.r1 <- stack(base.r1)
colnames(base.r1) <- c("Resp", "Institutos")

fit <- aov(Resp~Institutos, base.r1)

res <- fit$res # extraindo resíduos
# Teste de normalidade de Shapiro-Wilk
shapiro.test(res)

# Teste de homogeneidade de variâncias
bartlett.test(Resp~Institutos, data=base.r1)

# Verificação de independência
plot(fit$residuals, type="l") # res x índice das observações
title("Resíduos vs índice")
abline(h=0, col="red")
t.test(base.r1$Resp~base.r1$Institutos, var.equal=T)

```

Exercício 4

#item b

```
A <- c(rep("-1",4),rep("1",4))
B <- c("-1","-1","1","1","-1","-1","1","1")
C <- c("-1","1","-1","1","-1","1","-1","1")
Y <- c(195,496,87,1371,102,1001,354,775)
tabela <- data.frame(A,B,C,Y)
```

```
with(tabela, {
  interaction.plot(A, B, Y)
})
```

```
with(tabela, {
  interaction.plot(A, C, Y)
})
```

```
with(tabela, {
  interaction.plot(B, C, Y)
})
```

```
with(tabela, {
  interaction.plot(A:B, C, Y)
})
```

```
CA <- 102
CB <- 87
CC <- 496
x <- c("A", "B", "Y")
y <- c(CA,CB,CC)
d <- data.frame(x,y)
```

```
ggplot(d, aes(x=x, y=y, group=1)) +  
  geom_line() +  
  geom_point() +  
  labs(x="Fatores", y="Qualidade da água")
```

```
# item c
```

```
fit <- aov(Y ~ A+B+C+A*B*C-A:C, tabela)
fit1 <- anova(fit)
knitr::kable(caption = "Tabela de ANOVA", fit1)
TukeyHSD(fit)
```

Exercício 6

```
base.b <- c(21, '1', 'Baixo', 'Baixo', 12, '1', 'Baixo', 'Alto', 13, '1', 'Alto', 'Baixo', 1, '1', 'Alto')

aovbase.b <- as.data.frame(t(matrix(base.b, 4, 24)))
colnames(aovbase.b) <- c("resp", "TE", "TA", "pH")
aovbase.b$resp <- as.numeric(levels(aovbase.b$resp))[aovbase.b$resp]
```

```

# GRÁFICO PERFIL DE MÉDIAS.

par(mfrow=c(1,2))
with(aovbase.b, {
  interaction.plot(TE, TA, resp)
  interaction.plot(TE, pH, resp)
})
par(mfrow=c(1,1))
with(aovbase.b, {
  interaction.plot(TA, pH, resp)
})

#ANOVA TE*TA*pH E COMPARAÇÃO DE MÉDIAS

fit <- aov(resp ~ TE * TA * pH, aovbase.b)
fit1 <- anova(fit)
knitr::kable(caption = "Tabela de ANOVA", fit1)

fit1 <- aov(resp ~ TA + pH, aovbase.b)
TukeyHSD(fit1)

plot(TukeyHSD(fit))

ava<-anova(fit)
s2 <- ava$"Mean Sq"[8] # estimativa da variância residual

par(mfrow=c(2,2))
plot(fit)

plot(fit$fit, fit$resp, xlab="Valores Ajustados", ylab="Resíduos")
title("Resíduos vs Preditos")

avres <- fit$residuals # extraindo resíduos
respad <- (avres/sqrt(s2)) # resíduos padronizados
boxplot(respad)
title("Resíduos Padronizados")

hist(respad, main=NULL)
title("Histograma dos resíduos padronizados")

qqnorm(avres,ylab="Resíduos", main=NULL)
qqline(avres)
title("Gráfico Normal de Probabilidade dos Resíduos")

shapiro.test(avres)

```