

Lista 1

Bruno de Castro Paul Schultze ¹

Rubens Santos Andrade Filho ²

Setembro de 2020

Sumário

1	Questão 1	2
1.1	a	2
1.2	b	3
1.3	c	5
1.4	d	6
	Normalidade dos Resíduos	7
	Homogeneidade da Variância	8
	Independência entre Observações	8
	ANOVA	9

¹Número USP: 10736862

²Número USP: 10370336

1 Questão 1

1.1 a

Primeiro, vamos importar os dados.

```
o2 = read.csv("02cons.csv", header = TRUE)
o2$Trat = as.factor(o2$Trat)
o2$Virus = as.factor(o2$Virus)
```

Agora, vamos separar cada tipo de célula, e também retirar a variável “Grup”, já que não será utilizada.

```
t6 = o2[c("Trat", "Virus", "T6")]
t12 = o2[c("Trat", "Virus", "T12")]
t18 = o2[c("Trat", "Virus", "T18")]
```

Faremos então a análise de variância ANOVA (vamos considerar que suas pressuposições são satisfeitas) para as células T6. Utilizaremos, para todos os tipos de célula, o seguinte modelo, de casela de referência:

$$y_{ijk} = \begin{cases} \mu + e_{ijk} & \text{se } j = 0, k = 0 \\ \mu + \tau_j + \beta_k + (\tau\beta)_{jk} + e_{ijk} & \text{caso contrário} \end{cases} \quad (1)$$

Podemos entendê-lo como: y_{ijk} é a i -ésima observação ($i \in \{1, 2, \dots, 12\}$) submetida ao j -ésimo nível de *Tratamento* ($j \in \{0, 1\}$, sendo 0 o tratamento P) e ao k -ésimo nível de *Vírus* ($k \in \{0, 1\}$, sendo 0 a ausência de vírus), e μ é a média do consumo de oxigênio de observações submetidas ao tratamento P e com vírus ausente. Além disso, note também que o modelo considera o efeito da interação entre *Tratamento* e *Vírus*.

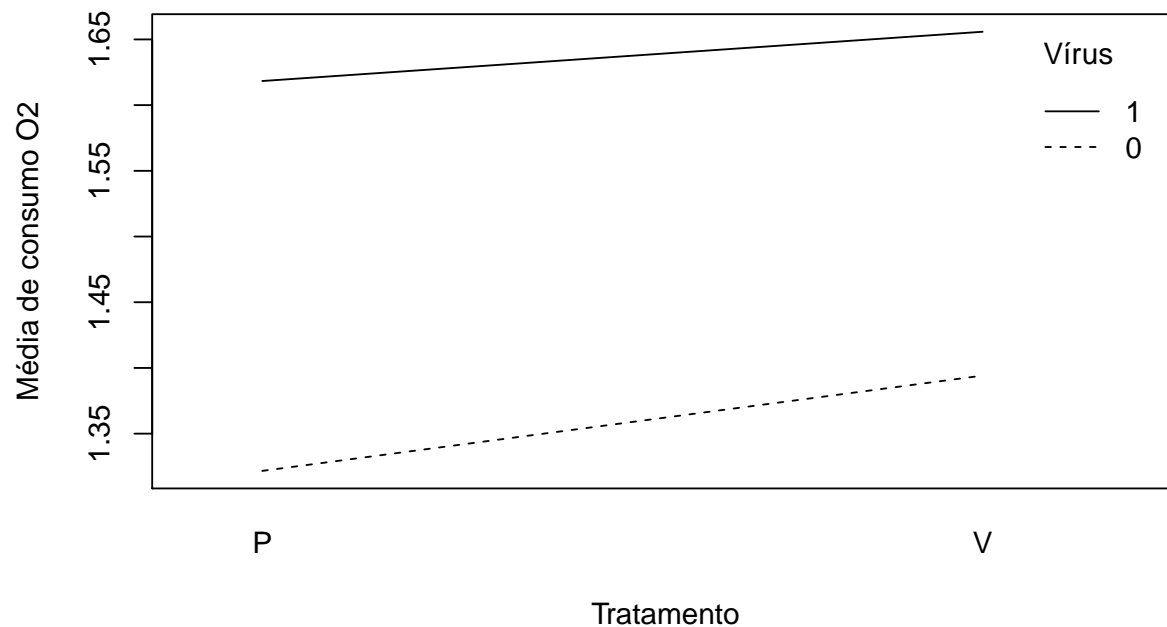
```
t6.aov = aov(T6 ~ Trat * Virus, data = t6)
summary(t6.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Trat         1  0.0363   0.0363    0.955    0.334
## Virus         1  0.9352   0.9352   24.597 1.1e-05 ***
## Trat:Virus     1  0.0037   0.0037    0.097    0.757
## Residuals    44  1.6729   0.0380
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos observar, então, que a um nível de 5% somente o efeito da variável *Vírus* se mostra significativo, então a hipótese que mudanças no tratamento não interferem na média do consumo de oxigênio é mantida. Além disso, uma possível interação entre *Tratamento* e *Vírus* não tem efeito significativo para explicar variações no consumo de oxigênio.

Como um adendo, o gráfico abaixo mostra que a interação entre *Vírus* e *Tratamento* é de fato muito baixa, visto que as retas são quase paralelas entre si.

```
interaction.plot(t6$Trat, t6$Virus, t6$T6, xlab = "Tratamento", ylab = "Média de consumo O2", trace.lab = FALSE)
```



1.2 b

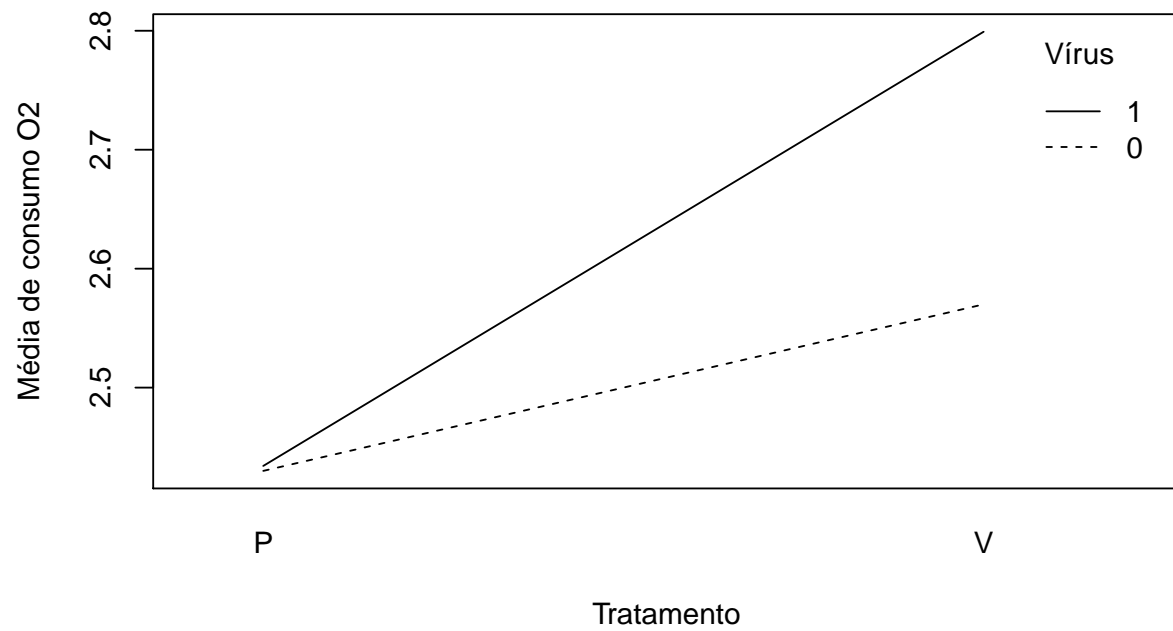
Agora vamos fazer a mesma coisa que na questão anterior, só que para células do tipo T12.

```
t12.aov = aov(T12 ~ Trat * Virus, data = t12)
summary(t12.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Trat       1  0.765   0.7651    9.196 0.00406 **
## Virus      1  0.163   0.1633    1.963 0.16819
## Trat:Virus 1  0.152   0.1519    1.825 0.18358
## Residuals 44  3.661   0.0832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

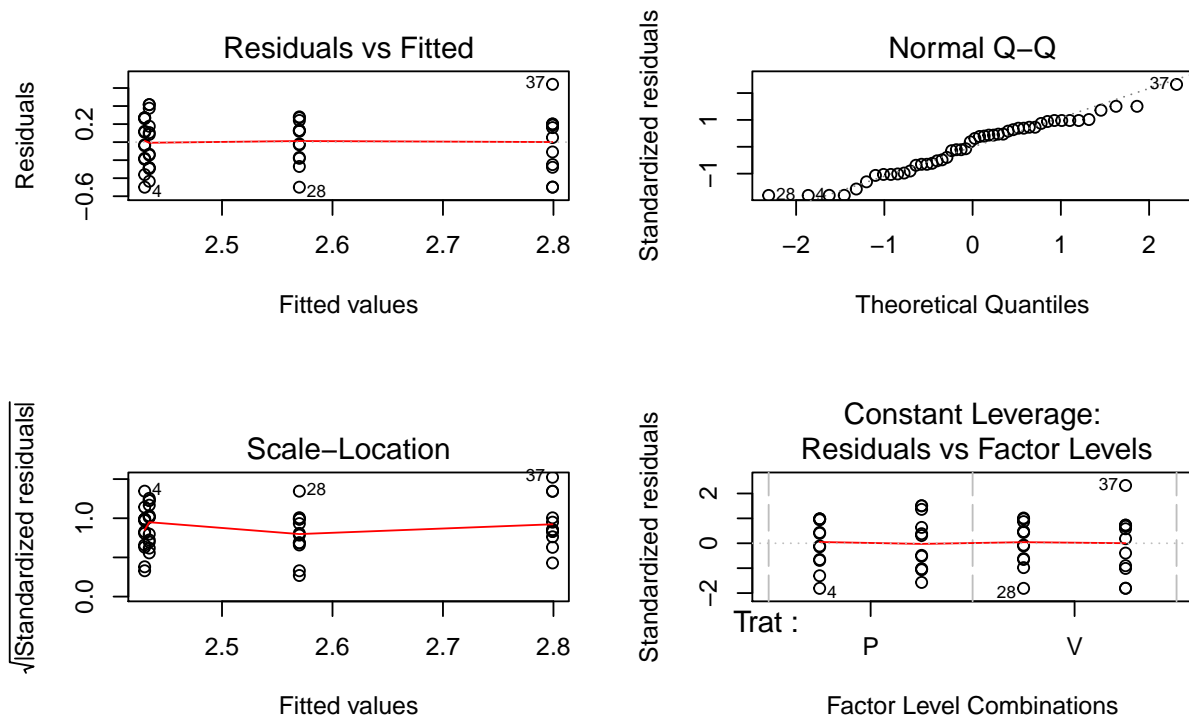
Podemos observar, então, que a um nível de 5% somente o efeito da variável *Tratamento* se mostra significativo, então a hipótese que mudanças no tratamento não interferem na média do consumo de oxigênio é rejeitada. É interessante notar que o algoritmo não mostra significância do efeito da interação entre *Tratamento* e *Vírus*, mas o gráfico de interação abaixo mostra que ele existe. Então, vamos investigar um pouco mais a fundo se tal interação é realmente não-significativa ou se seu efeito está sendo mascarado por algum ponto.

```
interaction.plot(t12$Trat, t12$Virus, t12$T12, xlab = "Tratamento", ylab = "Média de consumo O2", trace
```



Vamos então procurar pontos influentes.

```
par(mfrow = c(2,2))  
plot(t12.aov)
```



Observemos então que os pontos 28, 37 e 4 se destacam. Vamos retirá-los e refazer a análise de variância.

```
t12.aov2 = aov(T12 ~ Trat * Virus, data = t12[-c(37, 28, 4),])
summary(t12.aov2)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Trat       1  0.5656   0.5656    8.693 0.00525 **
## Virus      1  0.0183   0.0183    0.281 0.59910
## Trat:Virus  1  0.0781   0.0781    1.200 0.27966
## Residuals 41  2.6673   0.0651
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vemos então que não, os pontos influentes não estavam mascarando um possível efeito significativo da interação entre *Vírus* e *Tratamento*, tal interação existe mas seu efeito não é significativo.

1.3 c

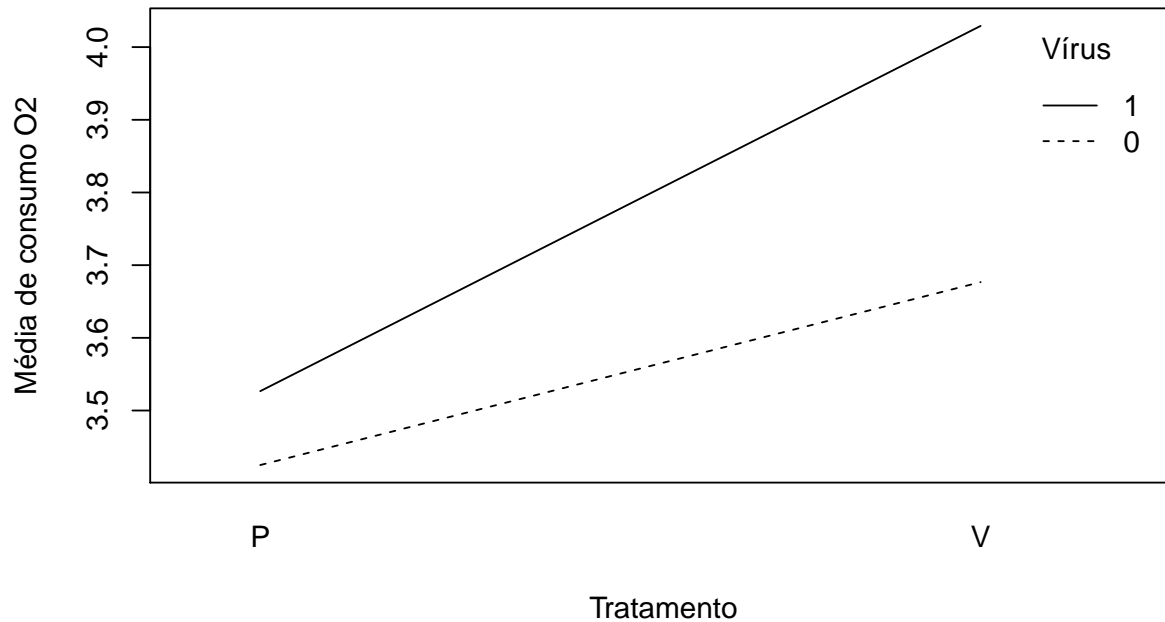
Faremos então o mesmo que nas questões anteriores, mas para células do tipo T18.

```
t18.aov = aov(T18 ~ Trat * Virus, data = t18)
summary(t18.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Trat       1  1.706   1.7063   18.083 0.000109 ***
## Virus      1  0.619   0.6188    6.558 0.013950 *
## Trat:Virus  1  0.189   0.1888    2.000 0.164306
## Residuals 44  4.152   0.0944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para esse tipo de célula tanto *Tratamento* quanto *Vírus* têm efeitos significativos para explicar a variação do consumo médio de oxigênio a um nível de 5%. Novamente, a interação entre as duas não parece significativa.

```
interaction.plot(t18$Trat, t18$Virus, t18$T18, xlab = "Tratamento", ylab = "Média de consumo O2", trace
```



Notemos, novamente, que até existe certa interação, mas graficamente ela também não parece significativa.

1.4 d

Agora o objetivo é construir um novo tipo de célula a partir de uma simulação. Construiremos então uma experimento DCA Fatorial 2x2 com Efeito de Interação, além disso o modelo estrutural é o mesmo utilizado anteriormente, a saber:

$$y_{ijk} = \mu + \tau_{tratamento} + \beta_{virus} + (\tau\beta)_{tratamento:virus} + e_{ij}$$

Estamos considerando também que as observações tem distribuição Normal. Agora, vamos definir as variáveis arbitrariamente, para construir as simulações e conseguir o efeito desejado de interação.

```
set.seed(42)
mu = 10
tau = -5
beta = -5
inter11 = 11
inter01 = -3
inter10 = -3
n = 12
sigma = 2
```

Definimos as variáveis arbitrárias, agora vamos de fato simular as observações.

```

y.00 = rnorm(n, mu, sigma )
y.00 = as.data.frame(y.00)
y.00 = cbind(y.00, rep(0,12), rep(0,12))
colnames(y.00) = c("TSim", "Trat", "Virus")
y.01 = rnorm(n, mu + beta + inter01, sigma)
y.01 = as.data.frame(y.01)
y.01 = cbind(y.01, rep(0,12), rep(1,12))
colnames(y.01) = c("TSim", "Trat", "Virus")
y.10 = rnorm(n, mu + tau + inter10)
y.10 = as.data.frame(y.10)
y.10 = cbind(y.10, rep(1,12), rep(0,12))
colnames(y.10) = c("TSim", "Trat", "Virus")
y.11 = rnorm(n, mu + tau + beta + inter11, sigma)
y.11 = as.data.frame(y.11)
y.11 = cbind(y.11, rep(1,12), rep(1,12))
colnames(y.11) = c("TSim", "Trat", "Virus")

tsim = rbind(y.00, y.01, y.10, y.11)
tsim$Trat[tsim$Trat ==0] = "p"
tsim$Trat[tsim$Trat ==1] = "v"
tsim$Trat = as.factor(tsim$Trat)
tsim$Virus = as.factor(tsim$Virus)
rm(y.00,y.01,y.10,y.11)

```

Agora, com a simulação construída, vamos checar as suposições da ANOVA.

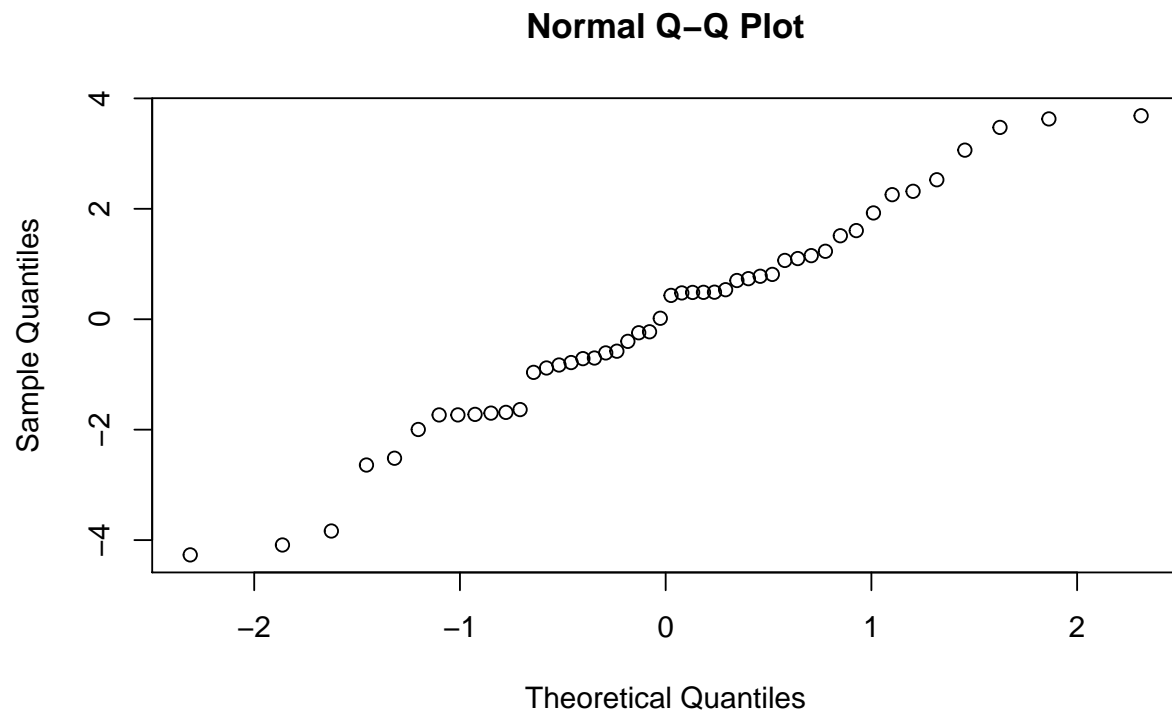
Normalidade dos Resíduos

Para essa checagem, vamos usar duas abordagens. Primeiro um gráfico QQ comparando com a Distribuição Normal, e também o Teste de Shapiro Wilk.

```

tsim.aov = aov(TSim ~ Trat * Virus, data = tsim)
qqnorm(tsim.aov$residuals)

```



A distribuição parece de fato se adequar à normalidade, mas de qualquer forma utilizaremos o Teste de Shapiro Wilk para confirmar.

```
shapiro.test(tsim.aov$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tsim.aov$residuals
## W = 0.9801, p-value = 0.5828
```

Como o p-valor é bem maior que 5%, não rejeitamos a hipótese de normalidade a um nível de 5%, então podemos continuar a análise.

Homogeneidade da Variância

Para isso utilizaremos o Teste de Levene

```
leveneTest(TSim ~ Trat * Virus, data = tsim)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  1.4563 0.2395
##      44
```

Pelo p-valor não rejeitamos a hipótese de homogeneidade da variância.

Independência entre Observações

Pela forma que a simulação foi construída, as observações são de fato independentes entre sim.

ANOVA

Agora que comprovamos que as suposições da ANOVA estão satisfeitas pela simulação, vamos aplicá-la.

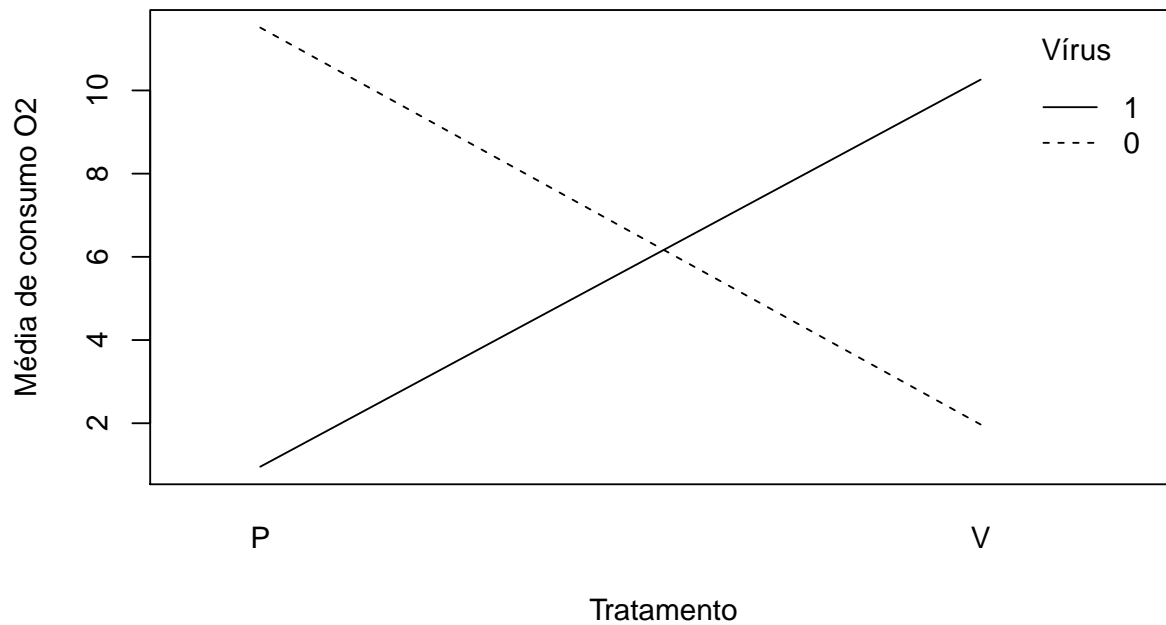
```
tsim.aov = aov(TSim ~ Trat * Virus, data = tsim)
summary(tsim.aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Trat       1    0.2      0.2    0.043 0.8373
## Virus      1   15.4     15.4    3.944 0.0533 .
## Trat:Virus  1 1065.5   1065.5  272.864 <2e-16 ***
## Residuals 44   171.8      3.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos ver que, a um nível de 5%, somente o efeito da interação entre *Tratamento* e *Vírus* mostrou-se significativo para explicar a variância do consumo médio de oxigênio.

Agora vamos checar através do gráfico de interação se essa satisfaz a interação sugerida no enunciado da questão.

```
interaction.plot(tsim$Trat, tsim$Virus, tsim$TSim, xlab = "Tratamento", ylab = "Média de consumo O2", t
```



De fato satisfaz.