Segundo trabalho de Planejamento de Experimentos

Universidade Federal da Paraíba - CCEN

Gabriel de Jesus Pereira - 20200121424

12 de abril de 2024

Pacotes utilizados

```
library(tidyverse)
library(ggplot2)
library(nortest)
library(rstatix)
library(lmtest)
library(ggpubr)
library(ExpDes)
```

Quadrado latino

O modelo para um quadrado latino

O modelo para um quadrado latino é:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk},\ i,\ j,\ k = 1, ..., p$$

Em que o y_{ijk} é a observação na i-th linha k-th coluna para o j-th tratamento, μ é a média geral, α_i é o efeito na ith linha, τ_j é o efeito do jth tratamento, β_k é o kth efeito na coluna e ϵ_{ijk} é o erro aleatório.

Por fim, chegamos a seguinte tabela de análise de variância:

Variaçao	Soma dos quadrados	Gl	MSE	F_0
Tratamentos	$\frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{..}^2}{N}$	$p-1$	$\frac{SS_{Tratamentos}}{p-1}$	$\frac{MS_{Tratamentos}}{MS_E}$
Linhas	$\frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{Linhas}}{p-1}$	$\frac{MS_{Linhas}}{MS_E}$
Colunas	$\frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{Colunas}}{p-1}$	$\frac{MS_{Colunas}}{MS_E}$
Erro	$S_E = SS_T - SS_{Linhas} - SS_{Colunas} - SS_{Tratamentos}$	$(p-2)(p-1)$	$\frac{SS_E}{(p-2)(p-1)}$	
Total	$\sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$	p^2-1		

Com essa tabela de análise de variância poderemos verificar o efeito entre as linhas, colunas e os tratamentos.

Os dados

The effect of five different ingredients (A, B, C, D, E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately 12 hours, so only five runs can be made in one day. The exper imenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use α 0.05) and draw conclusions.

Os dados são do problema **4.22** do livro do Montgomery, e se trata do efeito de 5 ingredientes no tempo de reação de um processo químico que está sendo estudado.

Tabela 2: Chemical Process Latin Square

	Day 1	Day 2	Day 3	Day 4	Day 5
Batch 1	A=8	B=7	D=1	C=7	E=8
Batch 2	C=11	E=2	A=7	D=3	B=8
Batch 3	B=4	A=9	C=10	E=1	D=5
Batch 4	D=6	C=8	E=6	B=6	A=10
Batch 5	E=4	D=2	B=3	A=8	C=8

Aplicando o quadrado latino

Pelo resultado que se segue abaixo, podemos ver pelo p-valor dos tratamentos (0.00049) que, ao nível de 5% de significância, há um efeito significativo dos ingredientes nos tempos médios de reação dos produtos químicos. É possível ver também que não há um efeito significativo nem na coluna e nem nas linhas, com p-valor 0.347 e 0.455, respectivamente. Ou seja, o dia e o batch não tem um impacto significativo nas observações, ao nível de 5%.

```r
Obs <- c(8,7,1,7,3,11,2,7,3,8,4,9,10,1,5,6,8,6,6,10,4,2,3,8,8)
Batch <- as.factor(rep(1:5, each = 5))
Day <- as.factor(rep(seq(1,5),5))
Ingredient <- as.factor(c(1,2,4,3,5,3,5,1,4,2,2,1,3,5,4,4,3,5,2,1,5,4,2,1,3))
dados1 <- data.frame(Obs, Batch, Day, Ingredient)

QuadradoLatino = latsd(
  treat = Ingredient,
  row = Batch,
  column = Day,
  resp = Obs,
  sigT = 0.05,
  sigF = 0.05
  )
```

```
------------------------------------------------------------------------
Analysis of Variance Table
------------------------------------------------------------------------
           DF     SS     MS      Fc   Pr>Fc
Treatament  4 141.44 35.360 11.3092 0.00049
Row         4  15.44  3.860  1.2345 0.34762
Column      4  12.24  3.060  0.9787 0.45501
Residuals  12  37.52  3.127
Total      24 206.64
------------------------------------------------------------------------
CV = 30.07 %

------------------------------------------------------------------------
Shapiro-Wilk normality test
p-value:  0.5476372
According to Shapiro-Wilk normality test at 5% of significance, residuals can be considered normal.
------------------------------------------------------------------------

Tukey's test
```

```
---------------------------------------------------------------------
Groups Treatments Means
a    3   8.8
a    1   8.4
ab   2   5.6
 b   4   3.4
 b   5   3.2
---------------------------------------------------------------------
```

Verificando normalidade

Considerando um nível de significância de 5% foram realizados 3 testes para verificar normalidade dos resíduos: Shapiro-Wilk, Lilliefors e Anderson-Darling. Os três testes confirmam a suposição de normalidade dos resíduos.

```r
shapiro.test(QuadradoLatino$residuals)
```

```

    Shapiro-Wilk normality test

data:  QuadradoLatino$residuals
W = 0.96606, p-value = 0.5476
```

```r
lillie.test(QuadradoLatino$residuals)
```

```

    Lilliefors (Kolmogorov-Smirnov) normality test

data:  QuadradoLatino$residuals
D = 0.11934, p-value = 0.4746
```

```r
ad.test(QuadradoLatino$residuals)
```

```

    Anderson-Darling normality test

data:  QuadradoLatino$residuals
A = 0.34711, p-value = 0.4515
```

Verificando homogeneidade

Agora considerendo o teste de Levene e o de Bartlett e um nível de 5% de confiança para testar a homogeneidade das variâncias, vemos que pelos dois testes de hipóteses é indicado a homogeneidade das variâncias.

```r
dados1 |>
  levene_test(Obs ~ Ingredient)
```

```
# A tibble: 1 x 4
    df1   df2 statistic     p
  <int> <int>     <dbl> <dbl>
1     4    20     0.444 0.775
```

```r
bartlett.test(Obs ~ Ingredient, data = dados1)
```

```

	Bartlett test of homogeneity of variances

data:  Obs by Ingredient
Bartlett's K-squared = 1.5544, df = 4, p-value = 0.817
```

Anova com medidas repetidas

Sobre medidas repetidas

O modelo para uma análise de variância com medidas repetidas é:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

Em que y_{ij} representa a resposta do sujeito j ao tratamento i. Em que τ_i é o efeito do ith tratamento e β_j é um parâmetro associado com o jth sujeito.

Assim, chegamos a seguinte tabela de análise de variância:

Variaçao	Soma dos quadrados	Gl	MSE	F_0
Dentre sujeitos	$\sum_{j=1}^{n} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{an}$	$n-1$		$\frac{MS_{Tratamentos}}{MS_E}$

Variaçao	Soma dos quadrados	Gl	MSE	F_0
Intra sujeitos	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \sum j = 1^n \frac{y_{.j}^2}{a}$	$n\,(a-1)$		$\frac{MS_{Linhas}}{MS_E}$
Tratamentos	$\sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$	$a-1$	$\frac{SS_{Tratamentos}}{a-1}$	$\frac{MS_{Colunas}}{MS_E}$
Erro	(2) − (3)	$(a-1)\,(n-1)$	$\frac{SS_E}{(a-1)(n-1)}$	
Total	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{an}$	$an-1$		

Os dados

15.21. Three different Pinot Noir wines were evaluated by a panel of eight judges. The judges are considered a random panel of all possible judges. The wines are evaluated on a 100 point scale. The wines were presented in random order to each judge, and the following results obtained.

Analyze the data from this experiment. Is there a difference in wine quality? Analyze the residuals and comment on model adequacy.

Os dados para a Anova com medidas repetidas foi retirado do livro do Montgomery, exercício **15.21**. O que o enunciado da questão nos pede é para testar se há diferença na qualidade do vinho que foi pontuado pelos jurados.

```
dados2 <- tibble(
  juiz = as.factor(rep(1:8, each = 3)),
  vinho = as.factor(rep(1:3, 8)),
  obs = c(
    85, 88, 93,
    90, 89, 94,
    88, 90, 98,
    91, 93, 96,
    92, 92, 95,
    89, 90, 95,
    90, 91, 97,
    91, 89, 98
  )
)
```

Suposições

Normalidade

Para a suposição de normalidade foram realizados 3 testes: Shapiro-Wilk, Lilliefors e Anderson-Darling. Todos os testes escolhidos não rejeitaram a suposição de normalidade, ao nível de 5% de significância.

```r
modelo2 <- aov(obs ~ Error(juiz) + vinho, data = dados2)

dados2 |>
  group_by(vinho) |>
  do(tidy(shapiro.test(.$obs))) |>
  knitr::kable()
```

vinho	statistic	p.value	method
1	0.9009201	0.2945123	Shapiro-Wilk normality test
2	0.9590109	0.8006264	Shapiro-Wilk normality test
3	0.9388881	0.6001739	Shapiro-Wilk normality test

```r
dados2 |>
  group_by(vinho) |>
  do(tidy(lillie.test(.$obs))) |>
  knitr::kable()
```

vinho	statistic	p.value	method
1	0.2147381	0.3406769	Lilliefors (Kolmogorov-Smirnov) normality test
2	0.1845333	0.5846848	Lilliefors (Kolmogorov-Smirnov) normality test
3	0.1588522	0.7990819	Lilliefors (Kolmogorov-Smirnov) normality test

```r
dados2 |>
  group_by(vinho) |>
  do(tidy(ad.test(.$obs))) |>
  knitr::kable()
```

vinho	statistic	p.value	method
1	0.3950347	0.2822387	Anderson-Darling normality test

vinho	statistic	p.value	method
2	0.2245596	0.7331001	Anderson-Darling normality test
3	0.2377478	0.6831905	Anderson-Darling normality test

Homogeneidade

Para a suposição de homogeneidade, ao nível de 5% de significância, foram realizados dois testes de hipóteses: Levene e Bartlett. Os dois testes não rejeitam a hipótese de homogeneidade entre as variâncias. O p-valro do teste de Bartlett foi de 0.764 e o de Levene foi de 0.895.

```r
dados2 |>
  levene_test(obs ~ vinho)
```

```
# A tibble: 1 x 4
    df1   df2 statistic     p
  <int> <int>     <dbl> <dbl>
1     2    21     0.111 0.895
```

```r
bartlett.test(obs ~ vinho, data = dados2)
```

```

	Bartlett test of homogeneity of variances

data:  obs by vinho
Bartlett's K-squared = 0.53844, df = 2, p-value = 0.764
```

Cronbach's Alpha

O código abaixo calcula o Cronbach's Alpha. Podemos ver que chegamos num alpha de 69.8%, o que indica uma medida de confiabilidade razoável. Dessa forma, a consistência interna dessa pesquisa de vinhos é razoável.

```r
conb_data <- dados2 |>
  pivot_wider(names_from = vinho, values_from = obs)
ltm::cronbach.alpha(conb_data[2:4], CI = TRUE, standardized = FALSE, B = 1000)
```

```
Cronbach's alpha for the 'conb_data[2:4]' data-set

Items: 3
Sample units: 8
alpha: 0.698

Bootstrap 95% CI based on 1000 samples
  2.5%  97.5%
-2.106  0.904
```

Resultado do modelo de medidas repetidas

Dessa forma, ao nível de 5% de significância, chegamos ao resultado de que a qualidade do vinho é significantemente diferente pela pontuação dada por cada juiz aos vinhos.

```r
summary(modelo2)
```

```
Error: juiz
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  7     48   6.857               

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)    
vinho      2  186.3   93.17   44.98 8.04e-07 ***
Residuals 14   29.0    2.07                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teste de Bonferroni

Pelo teste de Bonferroni, podemos ver que o grupo 1 e 2 são os únicos grupos não significantemente diferentes.

```r
dados2 |>
  pairwise_t_test(
    obs ~ vinho, paired = TRUE,
    p.adjust.method = "bonferroni"
```

```
    )
```

```
# A tibble: 3 x 10
  .y.   group1 group2    n1    n2 statistic    df        p    p.adj p.adj.signif
* <chr> <chr>  <chr>  <int> <int>     <dbl> <dbl>    <dbl>    <dbl> <chr>
1 obs   1      2          8     8     -1.27     7 0.244    0.732    ns
2 obs   1      3          8     8     -7.85     7 0.000103 0.000309 ***
3 obs   2      3          8     8     -7.28     7 0.000166 0.000498 ***
```