

Planejamento de Experimentos

Fábio N. Demarqui

Análise de variância envolvendo delineamentos com um fator fixo

Exemplo: resistência de saquinhos de papel

Um fabricante de sacos de papel pardo está interessado em melhorar a resistência do produto à tensão. Acredita-se que a resistência à tensão seja uma função da concentração de madeira de lei na polpa, e que a faixa prática de interesse das concentrações da madeira de lei esteja entre 5% e 20%. Quatro níveis de concentração são investigados, a saber: 5%, 10%, 15% e 20%. Seis corpos de prova (unidades experimentais) para cada nível de concentração são testados.

Carregando os pacotes necessários para a realização da análise:

```
library(planex)
library(agricolae)
library(multcomp)

## Loading required package: mvtnorm
## Loading required package: survival
## Loading required package: TH.data
## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##      geyser
library(ggplot2)
```

Na sequência iremos ler/visualizar o banco de dados:

```
# lendo o banco de dados:
data(saquinhos)
saquinhos
```

```
##      resistencia concentracao
## 1             7             5
## 2             8             5
## 3            15             5
## 4            11             5
## 5             9             5
## 6            10             5
## 7            12            10
## 8            17            10
## 9            13            10
## 10           18            10
```

```
## 11      19      10
## 12      15      10
## 13      14      15
## 14      18      15
## 15      19      15
## 16      17      15
## 17      16      15
## 18      18      15
## 19      19      20
## 20      25      20
## 21      22      20
## 22      23      20
## 23      18      20
## 24      20      20
```

Como podemos observar, os dados estão empilhados, ou seja, temos um data.frame cuja primeira coluna contém a resistência dos saquinhos observada, e a segunda coluna contém uma variável indicadora dos níveis (concentração de madeira de lei) do fator de interesse.

A variável concentracao do data.frame dados é uma variável categórica (com 4 níveis). Variáveis categóricas devem ser armazenadas em objetos da classe factor no R. Entretanto, como pode ser visto abaixo, a variável concentracao possui classe integer. Desta forma, antes de procedermos a análise dos dados, precisamos converter o objeto concentracao para um objeto da classe factor:

```
#verificando a classe de cada variável do nosso objeto dados:
sapply(saquinhos, class)
```

```
## resistencia concentracao
##      "integer"      "integer"
```

```
# transformando a variável concentracao em um objeto da classe "factor":
saquinhos$concentracao <- as.factor(saquinhos$concentracao)
sapply(saquinhos, class)
```

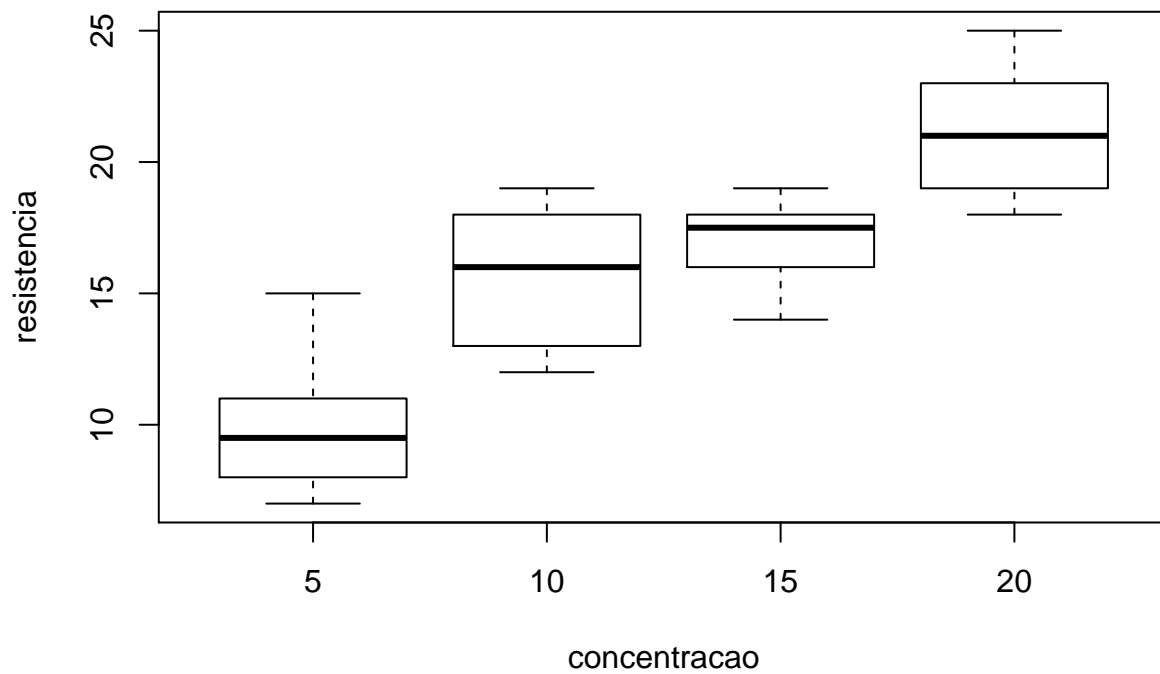
```
## resistencia concentracao
##      "integer"      "factor"
```

Agora que os saquinhos estão devidamente preparados, vamos proceder a uma análise descritiva:

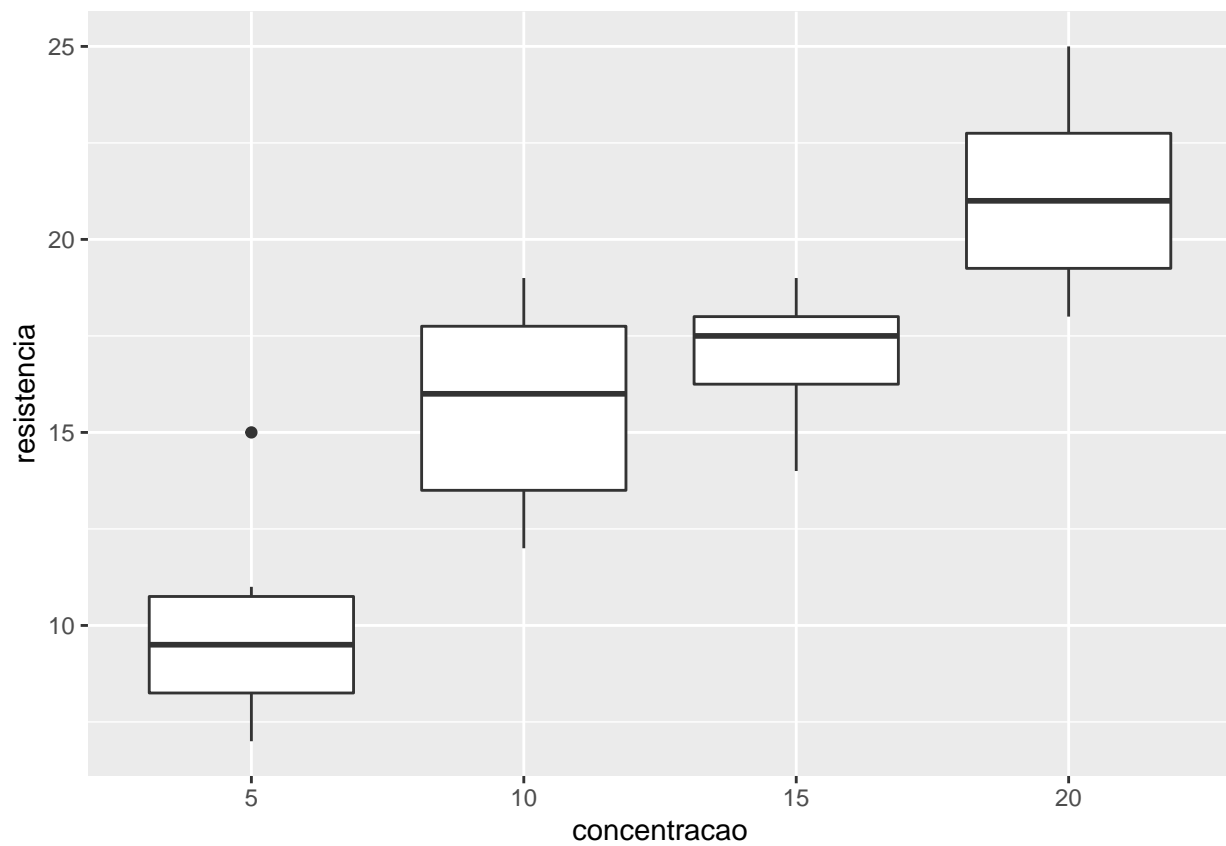
```
# obtendo as médias amostrais associadas a cada nível do fator de interesse:
medias <- tapply(saquinhos$resistencia, saquinhos$concentracao, mean)
medias
```

```
##      5      10      15      20
## 10.00000 15.66667 17.00000 21.16667
```

```
# explorando graficamente:
plot(resistencia~concentracao,data=saquinhos)
```



```
# alternativamente, utilizando o pacote ggplot2:
g <- ggplot(saquinhos, aes(concentracao, resistencia))
g + geom_boxplot()
```



A tabela ANOVA é obtida da seguinte forma:

```
# ajustando o modelo:
modelo <- aov(resistencia~concentracao,data=saquinhos)
# obtendo a tabela ANOVA:
summary(modelo)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## concentracao  3  382.8  127.60   19.61 3.59e-06 ***
## Residuals    20  130.2    6.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como podemos observar, considerando um nível de significância de 5%, temos evidências para rejeitar a hipótese de igualdade das médias. Consequentemente, devemos proceder às comparações múltiplas, isto é, verificar quais pares de médias diferem significativamente.

Procedimentos de comparações múltiplas

No nosso curso estudaremos 4 procedimentos de comparações múltiplas, a saber: mínima diferença significativa de Fisher, Bonferroni, Tukey, e Dunnet.

Os 3 primeiros procedimentos de comparações múltiplas devem ser empregados quando temos interesse em realizar comparações envolvendo todos os possíveis pares de médias. Já os testes de Dunnet devem ser utilizados quando temos interesse em realizar comparações envolvendo um grupo controle. Neste caso, dizemos que um dos níveis do fator de interesse é o nível de referência (grupo controle), e temos interesse em verificar se existem diferenças apenas entre a média associada ao nível de referência e os demais níveis do fator de interesse.

Teste da mínima diferença significativa de Fisher

A função `LSD.test` do pacote `agricolae` realiza o procedimento de comparações múltiplas da mínima diferença significativa de Fisher. Quando a opção `group=FALSE` é utilizada, a função `LSD.test` fornece as diferenças estimadas dos pares de médias, juntamente com os respectivos intervalos de confiança de 95% para as diferenças das médias, e os p-valores correspondentes.

```
# Acessando o help da função LSD.test:
? LSD.test
print(LSD.test(modelo, "concentracao", group=FALSE))
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
##  6.508333 20 15.95833 15.98628 2.085963 3.072423
##
## $parameters
##      test p.adjusted      name.t ntr alpha
##  Fisher-LSD      none concentracao  4 0.05
##
## $means
##      resistencia      std r      LCL      UCL Min Max  Q25  Q50  Q75
## 10      15.66667 2.804758 6 13.494136 17.83920 12 19 13.50 16.0 17.75
## 15      17.00000 1.788854 6 14.827469 19.17253 14 19 16.25 17.5 18.00
## 20      21.16667 2.639444 6 18.994136 23.33920 18 25 19.25 21.0 22.75
## 5       10.00000 2.828427 6 7.827469 12.17253 7 15 8.25 9.5 10.75
##
## $comparison
##      difference pvalue signif.      LCL      UCL
## 10 - 15      -1.333333 0.3761      -4.405756 1.739089
```

```
## 10 - 20 -5.500000 0.0013      ** -8.572423 -2.427577
## 10 - 5   5.666667 0.0010      ***  2.594244  8.739089
## 15 - 20 -4.166667 0.0104      *  -7.239089 -1.094244
## 15 - 5   7.000000 0.0001      ***  3.927577 10.072423
## 20 - 5   11.166667 0.0000      ***  8.094244 14.239089
##
## $groups
## NULL
##
## attr(,"class")
## [1] "group"
```

A partir da saída fornecida acima podemos concluir que não existe diferença significativa apenas entre as resistências médias associadas aos níveis de concentração de madeira de lei de 10% e 15% (p-valor > 0.05; intervalo de confiança contém o valor zero). Todos os demais pares de médias diferem significativamente.

Quando utilizamos o argumento group=TRUE temos a seguinte saída:

```
print(LSD.test(modelo, "concentracao", group=TRUE))
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
##      6.508333 20 15.95833 15.98628 2.085963 3.072423
##
## $parameters
##      test p.adjusted      name.t ntr alpha
##      Fisher-LSD      none concentracao  4  0.05
##
## $means
##      resistencia      std r      LCL      UCL Min Max  Q25  Q50  Q75
## 10      15.66667  2.804758 6 13.494136 17.83920 12 19 13.50 16.0 17.75
## 15      17.00000  1.788854 6 14.827469 19.17253 14 19 16.25 17.5 18.00
## 20      21.16667  2.639444 6 18.994136 23.33920 18 25 19.25 21.0 22.75
## 5       10.00000  2.828427 6  7.827469 12.17253  7 15  8.25  9.5 10.75
##
## $comparison
## NULL
##
## $groups
##      resistencia groups
## 20      21.16667      a
## 15      17.00000      b
## 10      15.66667      b
## 5       10.00000      c
##
## attr(,"class")
## [1] "group"
```

Neste caso, a função LSD.test o valor tabelado (t.value) da distribuição de referência (t de Student), juntamente com o valor da mínima diferença significativa (LDS). A comparação das médias neste caso é feita da seguinte forma: níveis que comparilham a mesma letra não possuem diferenças significativas considerando um nível de significância de 5% (para mudar o nível de significância para 0.01, por exemplo, basta usar o argumento alpha=0.01 da LSD.test).

Teste de Bonferroni

A função `LSD.test` também é usada para realizarmos comparações múltiplas utilizando a correção de Bonferroni. Para tanto, basta passar o argumento `p.adj="bonferroni"` para função `LSD.test`, como ilustrado a seguir:

```
print(LSD.test(modelo, "concentracao", group=FALSE, p.adj="bonferroni"))
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      MSD
##    6.508333 20 15.95833 15.98628 2.927119 4.311364
##
## $parameters
##      test  p.adjusted      name.t ntr alpha
## Fisher-LSD bonferroni concentracao 4 0.05
##
## $means
##      resistencia      std r      LCL      UCL Min Max  Q25  Q50  Q75
## 10      15.66667 2.804758 6 13.494136 17.83920 12 19 13.50 16.0 17.75
## 15      17.00000 1.788854 6 14.827469 19.17253 14 19 16.25 17.5 18.00
## 20      21.16667 2.639444 6 18.994136 23.33920 18 25 19.25 21.0 22.75
## 5       10.00000 2.828427 6 7.827469 12.17253 7 15 8.25 9.5 10.75
##
## $comparison
##      difference pvalue signif.      LCL      UCL
## 10 - 15 -1.333333 1.0000      -5.644697 2.9780304
## 10 - 20 -5.500000 0.0079      ** -9.811364 -1.1886363
## 10 - 5 5.666667 0.0060      ** 1.355303 9.9780304
## 15 - 20 -4.166667 0.0622      . -8.478030 0.1446971
## 15 - 5 7.000000 0.0007      *** 2.688636 11.3113637
## 20 - 5 11.166667 0.0000      *** 6.855303 15.4780304
##
## $groups
## NULL
##
## attr(,"class")
## [1] "group"
```

```
print(LSD.test(modelo, "concentracao", group=TRUE, p.adj="bonferroni"))
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      MSD
##    6.508333 20 15.95833 15.98628 2.927119 4.311364
##
## $parameters
##      test  p.adjusted      name.t ntr alpha
## Fisher-LSD bonferroni concentracao 4 0.05
##
## $means
##      resistencia      std r      LCL      UCL Min Max  Q25  Q50  Q75
## 10      15.66667 2.804758 6 13.494136 17.83920 12 19 13.50 16.0 17.75
## 15      17.00000 1.788854 6 14.827469 19.17253 14 19 16.25 17.5 18.00
## 20      21.16667 2.639444 6 18.994136 23.33920 18 25 19.25 21.0 22.75
## 5       10.00000 2.828427 6 7.827469 12.17253 7 15 8.25 9.5 10.75
##
## $comparison
## NULL
##
```

```
## $groups
##      resistencia groups
## 20      21.16667      a
## 15      17.00000     ab
## 10      15.66667      b
## 5       10.00000      c
##
## attr(,"class")
## [1] "group"
```

A análise dos resultados é similar àquela discutida para o procedimento de comparações múltiplas da mínima diferença significativa de Fisher. Entretanto, dada a mudança do procedimento de comparações múltiplas empregado, agora temos as seguintes conclusões: considerando $\alpha = 0.05$, não existem diferenças significativas entre os níveis 10% e 15%, e 15% e 20%.

Teste de Tukey

Para realizar o procedimento de comparações múltiplas de Tukey utilizamos a função `HSD.test`, conforme ilustrado a seguir:

```
print(HSD.test(modelo, "concentracao", group=FALSE))
```

```
## $statistics
##      MSerror Df      Mean      CV      MSD
## 6.508333 20 15.95833 15.98628 4.122563
##
## $parameters
##      test      name.t ntr StudentizedRange alpha
## Tukey concentracao 4      3.958293 0.05
##
## $means
##      resistencia      std r Min Max   Q25   Q50   Q75
## 10      15.66667 2.804758 6 12 19 13.50 16.0 17.75
## 15      17.00000 1.788854 6 14 19 16.25 17.5 18.00
## 20      21.16667 2.639444 6 18 25 19.25 21.0 22.75
## 5       10.00000 2.828427 6 7 15 8.25 9.5 10.75
##
## $comparison
##      difference pvalue signif.      LCL      UCL
## 10 - 15 -1.333333 0.8022      -5.455896 2.78922925
## 10 - 20 -5.500000 0.0066      ** -9.622563 -1.37743741
## 10 - 5 5.666667 0.0051      ** 1.544104 9.78922925
## 15 - 20 -4.166667 0.0470      * -8.289229 -0.04410408
## 15 - 5 7.000000 0.0007      *** 2.877437 11.12256259
## 20 - 5 11.166667 0.0000      *** 7.044104 15.28922925
##
## $groups
## NULL
##
## attr(,"class")
## [1] "group"
```

```
print(HSD.test(modelo, "concentracao", group=TRUE))
```

```
## $statistics
##      MSerror Df      Mean      CV      MSD
```

```
## 6.508333 20 15.95833 15.98628 4.122563
##
## $parameters
## test name.t ntr StudentizedRange alpha
## Tukey concentracao 4 3.958293 0.05
##
## $means
## resistencia std r Min Max Q25 Q50 Q75
## 10 15.66667 2.804758 6 12 19 13.50 16.0 17.75
## 15 17.00000 1.788854 6 14 19 16.25 17.5 18.00
## 20 21.16667 2.639444 6 18 25 19.25 21.0 22.75
## 5 10.00000 2.828427 6 7 15 8.25 9.5 10.75
##
## $comparison
## NULL
##
## $groups
## resistencia groups
## 20 21.16667 a
## 15 17.00000 b
## 10 15.66667 b
## 5 10.00000 c
##
## attr(,"class")
## [1] "group"
```

A saída da função `HSD.test`, bem como a interpretação dos resultados, é similar àquela fornecida pela função `LSD.test`. Como podemos observar, o procedimento de comparações múltiplas de Tukey forneceu, neste caso, os mesmos resultados de que o procedimento da mínima diferença significativa de Fisher.

Teste de Dunnet:

Suponha agora que o fabricante de saquinhos de papel já produz saquinhos utilizando a concentração de madeira de lei de 20%, e ele deseja reduzir tal concentração de modo a baratear o processo de produção, sem prejuízos quanto a resistência dos saquinhos produzidos. Neste caso devemos considerar que o nível 20% como o nível de referência (grupo controle). Note que neste novo contexto apenas comparações de pares de médias envolvendo o grupo controle são de interesse prático. Para a realização das comparações múltiplas vamos usar a função `glht` do pacote `multcomp`. Antes disso, precisamos informar ao R que o nível de referência da nossa variável categórica é o nível 20%, e reajustar nosso modelo:

```
# verificando os níveis da variável concentracao:
levels(saquinhos$concentracao)

## [1] "5" "10" "15" "20"

# alterando o nível de referência:
saquinhos$concentracao <- relevel(saquinhos$concentracao, ref="20")
# conferindo a alteração:
levels(saquinhos$concentracao)

## [1] "20" "5" "10" "15"

modelo <- aov(resistencia~concentracao, data=saquinhos)
comp <- glht(modelo, linfct = mcp(concentracao = "Dunnet"))
summary(comp)

##
```



```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: aov(formula = resistencia ~ concentracao, data = saquinhos)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## 5 - 20 == 0   -11.167      1.473  -7.581  < 0.001 ***
## 10 - 20 == 0    -5.500      1.473  -3.734  0.00369 **
## 15 - 20 == 0    -4.167      1.473  -2.829  0.02710 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Note que a saída da função `glht` é deferente das saídas discutidas anteriormente, sendo fornecidos apenas os p-valores. Em termos práticos, considerando um nível de significância de 5%, podemos concluir o grupo controle (nível 20%) difere significativamente dos demais níveis. Consequentemente, podemos concluir que a redução do nível de concentração de madeira de lei acarretará uma redução significativa na resistência média dos saquinhos de papel.