

Exercício Prático

Universidade Federal do Ceará
Centro de Ciências
Departamento de Estatística e Matemática Aplicada
Bacharelado em Estatística
Planejamento de Experimentos

Antônio Arthur Silva de Lima
Romulo Barros de Freitas

09 de fevereiro de 2025

Problema

A atividade descrita é baseada na formulação de bebidas a serem experimentadas, onde cada provador será utilizado como grupo controle (bloco), podendo provar cada um até 4 bebidas. Temos um tratamento (bebida) com 6 níveis distintos (sabores). Como primeiro desafio, devemos encontrar o número de blocos ideais a serem utilizados no experimento, isto é, o número ideal de provadores:

```
i = 0
for (i in 100:150) {
  print(i)
  isGYD(find.BIB(6, i, 4))
}
```

Com isso, vemos que o experimento pode ser realizado utilizando 105, 120, 135 ou 150 provadores. Vamos escolher um delineamento com 120 pessoas, criando-o a partir da função `find.BIB()`, do pacote *crossdes*, obtendo uma matriz de dados conforme é mostrado nas 6 primeiras observações da matriz gerada pelo código. Como há 6 bebidas, e considerando 120 provadores, além da restrição de que eles só podem provar até 4 bebidas, passamos os parâmetros 6, 120, 4, nesta ordem, na função.

```
set.seed(521353)
df = find.BIB(6, 120, 4) |> as.tibble()
head(df)
```

```
## # A tibble: 6 x 4
##   V1    V2    V3    V4
##   <dbl> <dbl> <dbl> <dbl>
## 1     1     2     4     6
## 2     1     2     4     6
## 3     1     2     4     5
## 4     1     2     4     5
## 5     3     4     5     6
## 6     1     2     3     6
```

Vamos então organizar a base de acordo com o formato que vem sendo utilizado em sala de aula — blocos, tratamentos e resposta em colunas separadas — aceito por funções de vários pacotes que processam delineamentos experimentais.

```
df = df |>
  mutate(BLOCO = row_number()) |>
  pivot_longer(cols = starts_with("V"),
               names_to = "coluna_aux",
               values_to = "TRATAMENTO") |>
  select(BLOCO, TRATAMENTO)
```

```
df$TRATAMENTO = as.factor(df$TRATAMENTO)
```

```
df
```

```
## # A tibble: 480 x 2
##   BLOCO TRATAMENTO
##   <int> <fct>
## 1     1 1 1
## 2     1 1 2
## 3     1 1 4
## 4     1 1 6
## 5     2 2 1
## 6     2 2 2
## 7     2 2 4
## 8     2 2 6
## 9     3 3 1
## 10    3 3 2
## # i 470 more rows
```

Com isso, devemos agora integrar à base uma variável resposta, com distribuição normal e variância homogênea para cada tratamento.

```
set.seed(521353)
medias = c(3, 5, 6, 7, 9, 10)
variancia = 1

df = df |>
  mutate(RESPOSTA = rnorm(n(),
                           mean = medias[TRATAMENTO],
                           sd = 1))
```

```
df
```

```
## # A tibble: 480 x 3
##   BLOCO TRATAMENTO RESPOSTA
##   <int> <fct>         <dbl>
## 1     1 1 1         2.03
## 2     1 1 2         3.84
## 3     1 1 4         5.94
## 4     1 1 6         9.61
## 5     2 2 1         3.52
## 6     2 2 2         5.87
```

```
## 7      2 4      4.48
## 8      2 6      9.71
## 9      3 1      2.72
## 10     3 2      5.85
## # i 470 more rows
```

Conferindo a suposição de homogeneidade de variâncias entre os tratamentos, realizamos o teste de Bartlett, não tendo evidências suficientes para rejeitar esta suposição.

```
##
## Bartlett test of homogeneity of variances
##
## data: df$RESPOSTA and df$TRATAMENTO
## Bartlett's K-squared = 2.7913, df = 5, p-value = 0.7321
```

Ainda verificando a homogeneidade de variâncias entre os tratamentos, podemos incluir um boxplot, a fim de entender visualmente o comportamento da variação entre os tratamentos, conforme a Figura 1.

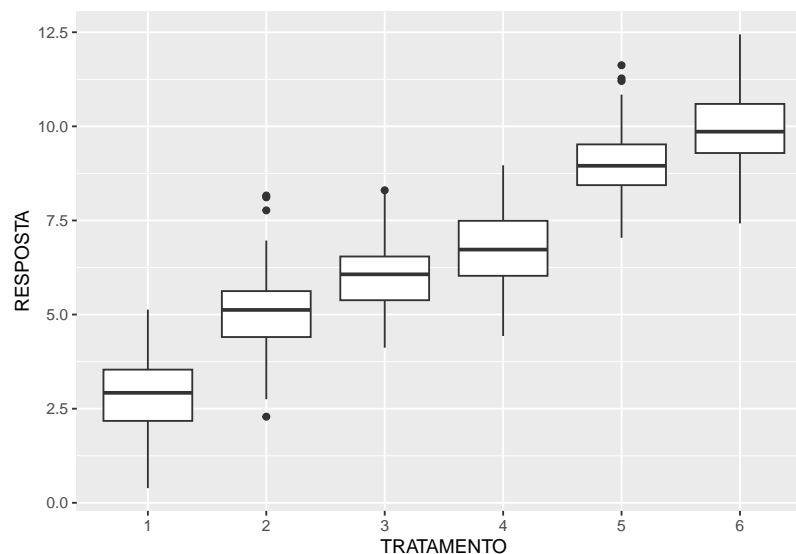


Figura 1: Boxplot de y para cada sabor de bebida

Com isso, observamos o comportamento esperado das caixas, que têm mesma amplitude e apresentam simetria ao redor da média. Vemos também alguns outliers nas bebidas 2, 3 e 5, que talvez possam influenciar os resultados das análises que serão realizadas. De maneira a entender mais a fundo os dados, iremos fazer uma análise descritiva simples. Primeiro, podemos obter as médias das respostas para cada tratamento, esperando que haja uma diferença significativa entre elas, de acordo com o boxplot anterior.

```
tapply(df$RESPOSTA, df$TRATAMENTO, mean)
```

```
##      1      2      3      4      5      6
## 2.888746 5.120140 6.051507 6.854305 9.021180 9.905302
```

Os respectivos desvios padrão para cada tratamento são obtidos.

```
tapply(df$RESPOSTA, df$TRATAMENTO, sd)
```

```
##          1          2          3          4          5          6
## 0.9415814 1.1183928 0.9766549 1.0009076 0.9880182 0.9877175
```

Vemos que a bebida que menos apresentou variabilidade nas respostas foi a bebida 1, enquanto a mais heterogênea foi a bebida 2. Analisando conjuntamente, vemos que a bebida com menor coeficiente de variação foi, na verdade, a bebida 6, enquanto o maior foi para a bebida 1.

```
CV = function(x){
  cv = sd(x)/mean(x) * 100
  return(round(cv, 2) |> str_c("%"))
}
```

```
tapply(df$RESPOSTA, df$TRATAMENTO, CV)
```

```
##          1          2          3          4          5          6
## "32.59%" "21.84%" "16.14%" "14.6%" "10.95%" "9.97%"
```

Por fim, iremos ajustar um delineamento em blocos incompletos balanceados, por meio da função `BIB.test` do pacote *agricolae*, utilizando o teste de Tukey para realizar as comparações múltiplas de médias entre as bebidas.

```
BIB.test(df$BLOCO, df$TRATAMENTO, df$RESPOSTA, console = T)
```

```
##
## ANALYSIS BIB: df$RESPOSTA
## Class level information
##
## Block:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
## Trt  :  1 2 4 6 5 3
##
## Number of observations: 480
##
## Analysis of Variance Table
##
## Response: df$RESPOSTA
##          Df Sum Sq Mean Sq F value    Pr(>F)
## block.unadj 119  408.95     3.44   3.2248 < 2.2e-16 ***
## trt.adj       5 2338.79   467.76 438.9413 < 2.2e-16 ***
## Residuals   355  378.31     1.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## coefficient of variation: 15.5 %
## df$RESPOSTA Means: 6.640197
##
## df$TRATAMENTO, statistics
##
## df$RESPOSTA mean.adj      SE r      std      Min      Max
## 1      2.888746 2.922328 0.1206402 80 0.9415814 0.3891188 5.131725
```

```

## 2      5.120140 5.111399 0.1206402 80 1.1183928 2.2879444 8.162164
## 3      6.051507 6.104940 0.1206402 80 0.9766549 4.1196248 8.301633
## 4      6.854305 6.832420 0.1206402 80 1.0009076 4.4309868 8.965167
## 5      9.021180 9.000506 0.1206402 80 0.9880182 7.0388008 11.622377
## 6      9.905302 9.869587 0.1206402 80 0.9877175 7.4213939 12.444157
##
## LSD test
## Std.diff   : 0.1720507
## Alpha      : 0.05
## LSD        : 0.3383668
## Parameters BIB
## Lambda     : 48
## treatmeans : 6
## Block size : 4
## Blocks     : 120
## Replication: 80
##
## Efficiency factor 0.9
##
## <<< Book >>>
##
## Comparison between treatments means
##      Difference pvalue sig.
## 1 - 2 -2.1890710      0 ***
## 1 - 3 -3.1826119      0 ***
## 1 - 4 -3.9100916      0 ***
## 1 - 5 -6.0781784      0 ***
## 1 - 6 -6.9472594      0 ***
## 2 - 3 -0.9935409      0 ***
## 2 - 4 -1.7210206      0 ***
## 2 - 5 -3.8891074      0 ***
## 2 - 6 -4.7581884      0 ***
## 3 - 4 -0.7274797      0 ***
## 3 - 5 -2.8955666      0 ***
## 3 - 6 -3.7646475      0 ***
## 4 - 5 -2.1680868      0 ***
## 4 - 6 -3.0371678      0 ***
## 5 - 6 -0.8690810      0 ***
##
## Treatments with the same letter are not significantly different.
##
##      df$RESPOSTA groups
## 6      9.869587      a
## 5      9.000506      b
## 4      6.832420      c
## 3      6.104940      d
## 2      5.111399      e
## 1      2.922328      f

```

Pelo resultado da ANOVA, podemos ver que há significância para os tratamentos e também para os blocos, indicando que a variável resposta seja influenciada pelo tipo de bebida. Para atestar que a ANOVA foi realizada de maneira adequada, verificamos os graus de liberdade e os valores de λ e r . Nos resíduos, esperamos $N - a - b + 1$ graus de liberdade, onde a e b são o número de tratamentos e blocos respectivamente. Para os graus de liberdade dos tratamentos (ajustados) e blocos, esperamos $a - 1$ e $b - 1$, respectivamente. O valor de

λ esperado é de $\frac{r(k-1)}{a-1}$, onde r é o número de repetições de cada tratamento, e k o número de tratamentos por bloco.

Ainda da saída da ANOVA, observamos um coeficiente de variação relativamente baixo, de 15,5%, e uma resposta média geral de 6,64 aproximadamente. Com o teste de Tukey, vemos que todas as médias das bebidas diferem entre si, sendo as bebida 6 e 1 as de maior e menor médias, respectivamente.