

Curso do R do GIEU

Alan R. Panosso

2020-10-14

Contents

Chapter 1

INTRODUÇÃO

A Estatística Experimental tem como objetivo o estudo dos experimentos, isto é, seu planejamento, execução, análise dos dados e interpretação dos resultados obtidos.

Vejamos então, alguns conceitos básicos necessários para um bom entendimento da estatística Experimental.

Alguns conceitos básicos

POPULAÇÃO

Boa parte do conhecimento humano está baseado em um número relativamente reduzido de informações. Isto é verdadeiro, tanto no que se refere aos problemas do cotidiano, como no que se refere à pesquisa científica.

Por definição **POPULAÇÃO** é o conjunto de elementos que têm em comum uma determinada característica. Todo o subconjunto não vazio e com menor número de elemento do que o conjunto definido como **população** constitui, por definição, uma **AMOSTRA** desta população.

Uma população em ecologia é o número total de indivíduos de uma determinada espécie em uma área definida. Por exemplo, o número total de lagartas de *Spodoptera frugiperda* em uma cultura de milho constitui uma população. Esta população, embora finita, é considerada para fins de amostragem como uma população infinita.

Uma vez definida a unidade amostral (1 planta, um conjunto de 5 plantas, ou um quadrado no qual será contado o número de lagartas), a população pode ser considerada como um conjunto de unidades amostrais e um subconjunto tomado aleatoriamente deste subconjunto é chamado de **AMOSTRA ALEATÓRIA DE TAMANHO N**.

Assim sendo, as observações são obtidas através de contagens do número de indivíduos em cada unidade amostral. Estas observações são chamadas de **VARIAVEL EM ESTUDO**.

TRATAMENTO

É o método, elemento ou material, cujo efeito se deseja medir ou comparar em um experimento. Por exemplo, um tratamento pode ser: Uma variedade de cana-de-açúcar, um híbrido de sorgo, uma dose de um adubo para a cultura do milho, um espaçamento para cultura do algodão, um recipiente para produção de mudas de eucalipto, um inseticida para controle de pragas, etc.

EXPERIMENTO OU ENSAIO

Experimento é um trabalho previamente planejado, no qual se faz comparação dos efeitos dos tratamentos.

UNIDADE EXPERIMENTAL OU PARCELA

É a unidade na qual o tratamento é aplicado. É na parcela que obtemos os dados que deverão refletir o efeito de cada tratamento no ensaio. A parcela pode ser constituída por uma planta, uma área ou um grupo de plantas, uma placa de petri com um meio de cultura, um animal, um lote de animais, etc.

DELINEAMENTO EXPERIMENTAL

É o plano utilizado na experimentação, e implica na forma como os tratamentos deverão ser distribuídos nas unidades experimentais e como serão analisados os dados a serem obtidos. Como exemplo, temos o delineamento inteiramente casualizado (DIC), o delineamento em blocos casualizados (DBC) o delineamento em quadrado latino (DQL), entre outros.

Chapter 2

Delineamento Inteiramente Casualizado

Caracterização

O delineamento inteiramente casualizado é o mais simples de todos os delineamentos experimentais, e os experimentos instalados de acordo com este delineamento são chamados de experimentos inteiramente casualizados ou experimentos inteiramente ao acaso.

Este delineamento apresenta as seguintes características:

1. Leva em consideração os princípios da **repetição** e da **casualização**, deixando de lado o princípio do controle local e, portanto, as repetições não são organizadas em blocos.
2. Os tratamentos são designados às parcelas de forma inteiramente casual, com qualquer número de repetições.

As principais vantagens desse delineamento são as seguintes:

- 1) Flexibilidade, uma vez que o número de repetições pode variar de um tratamento para outro, sem causar sérios problemas na análise.
- 2) Proporciona o maior número de graus de liberdade possível para o resíduo.

As principais desvantagens desse delineamento são:

- 1) As parcelas experimentais devem ser homogêneas.

- 2) Leva a uma alta estimativa da variância residual QM_{Res} , uma vez que todas as variações, exceto aquela devido ao efeito de tratamentos, são tomadas como variação do acaso.

Para a instalação desse experimento devemos ter certeza da homogeneidade das condições experimentais. Este delineamento é bastante utilizado em ensaios de laboratório e em ensaios com vasos, realizados dentro de casas de vegetação, em que as condições experimentais podem ser perfeitamente controladas.

A distribuição casual dos tratamentos a todas as parcelas do experimento é a principal característica deste delineamento. Por exemplo, num experimento no delineamento inteiramente casualizado com 5 tratamentos e 5 repetições, a casualização dos tratamentos seria feita sorteando-se para cada uma das 20 parcelas do experimento uma combinação de tratamento e repetição, ou seja:

A1	A2	A3	A4
B1	B2	B3	B4
C1	C2	C3	C4
D1	D2	D3	D4
E1	E2	E3	E4

Assim, um sorteio para distribuição dos tratamento às parcelas poderia ser o seguinte:

B4	D1	B1	A2	D4
D2	A1	C4	D3	B3
E1	E3	B2	C2	C2
A4	C1	E4	E2	E2

Modelo Matemático

Todo delineamento experimental possui um modelo matemático que o representa, que deve ser levado em conta na análise de variância, **aceitando algumas hipóteses básicas necessárias para a validade da análise**. No caso do DIC.

$$x_{ij} = \mu + \tau_i + \epsilon_{ij}$$

onde: x_{ij} : é o valor observado na parcela que recebeu o tratamento i na repetição j ,

μ : é a média geral do experimento,

τ_i é o efeito devido ao tratamento i que foi aplicado à parcela,

ϵ_{ij} é o efeito dos fatores não controlados na parcela que recebeu o tratamento i na repetição j .

Hipóteses básicas para aplicação da análise de variância

As hipóteses básicas que devemos admitir para tornar válida a aplicação da análise de variância são as seguintes:

- 1) **Aditividade:** Os efeitos dos fatores do modelo são aditivos.
- 2) **Independência:** Os erros (desvios) ϵ_{ij} , devido aos efeitos de fatores não controlados ou acaso devem ser independentes.
- 3) **Homocedasticidade (homogeneidade de variâncias):** Os erros (desvios) ϵ_{ij} , devido ao acaso devem possuir variância comum σ^2
- 4) **Normalidade dos desvios:** Os erros ou desvios ϵ_{ij} devido ao acaso devem possuir distribuição normal de probabilidades.

Uma forma resumida de apresentar estas quatro pressuposições é apresentada a seguir:

$$\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$$

Exemplo de aplicação DIC

Num experimento inteiramente casualizado, de competição de variedades de mandioca, realizado numa área perfeitamente homogênea quanto às condições experimentais, foram utilizadas 5 repetições das seguintes variedades:

1 - IAC 5

2 - IAC 7

3 - IAC 11

4 - IRACEMA

5 - MANTIQUEIRA

A designação dos tratamentos às parcelas no campo, juntamente com as produções, expressa em t/ha, foi a seguinte:

20,3 (x ₁₃)	47,8 (x ₅₁)	25,8 (x ₃₃)	28,7 (x ₂₅)	20,9 (x ₂₁)
28,3 (x ₂₄)	43,2 (x ₄₂)	29,3 (x ₁₅)	38,9 (x ₁₁)	41,7 (x ₄₃)
47,8 (x ₅₂)	25,4 (x ₁₂)	50,5 (x ₅₄)	38,7 (x ₄₁)	28,1 (x ₃₁)
27,0 (x ₃₂)	40,3 (x ₄₅)	32,3 (x ₂₃)	26,9 (x ₃₄)	26,2 (x ₂₂)
56,4 (x ₅₅)	25,7 (x ₁₄)	22,3 (x ₃₅)	44,7 (x ₅₃)	39,0 (x ₄₄)

Com estes dados, podemos organizar o quadro seguinte:

Tratamentos	Rep.1	Rep.2	Rep.3	Rep.4	Rep. 5	Total
1 - IAC 5	38,9	25,4	20,3	25,7	29,3	139,6
2 - IAC 7	20,9	26,2	32,3	28,3	28,7	136,4
3 - IAC 11	28,1	27,0	25,8	26,9	22,3	130,1
4 - IRACEMA	38,7	43,2	41,7	39,0	40,3	202,9
5 - MANTIQUEIRA	47,8	47,8	44,7	50,5	56,4	247,2
Total						856,2

As hipóteses que desejamos testar são:

H_0 : As variedades de mandioca testadas não são diferentes entre si quanto à produção.

H_1 : As variedades de mandioca testadas diferem entre si quanto à produção.

Aplicação no R - DIC

Utilizando o R para obtermos o quadro da análise de variância, os dados estão disponíveis online em: Mandioca.



```
# Carregando o pacote para a análise
library(ExpDes.pt)
library(tidyverse)
```

```
# Caminho dos dados
caminho<-"https://raw.githubusercontent.com/arpanosso/curso_GIEU/master/dados/mandioca.txt"

# Lendo o arquivo de dados
dados<-read.table(caminho,h=T,sep="\t")

# verificando os 6 primeiros registros
head(dados)
```

```
##   Trat Rep    Y
## 1    1  1 38.9
## 2    1  2 25.4
## 3    1  3 20.3
## 4    1  4 25.7
## 5    1  5 29.3
## 6    2  1 20.9
```

```
# Análise de variância e teste de Tukey com a função dic
trat <- dados$Trat # Criando o vetor de tratamentos
prod <- dados$Y # Criando o vetor com a variável resposta

# Utilizando a função
dic(trat,prod,mcomp = "tukey")
```

```
## -----
## Quadro da analise de variancia
## -----
##           GL      SQ QM      Fc      Pr>Fc
## Tratamento  4 2135.94  3 28.592 5.0773e-08
## Residuo      20  373.52  2
## Total        24 2509.46  1
## -----
## CV = 12.62 %
##
## -----
## Teste de normalidade dos residuos
## Valor-p:  0.2156065
## De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser considerados
## -----
##
## -----
## Teste de homogeneidade de variancia
## valor-p:  0.1115615
## De acordo com o teste de bartlett a 5% de significancia, as variancias podem ser consideradas
```

```
## -----  
##  
## Teste de Tukey  
## -----  
## Grupos Tratamentos Medias  
## a      5    49.44  
## b      4    40.58  
## c      1    27.92  
## c      2    27.28  
## c      3    26.02  
## -----
```

Chapter 3

Delineamento em Blocos Casualizados

Caracterização

O delineamento em blocos casualizados (DBC) é o mais utilizado dos delineamentos experimentais, e os experimentos instalados de acordo com esse delineamento são denominados de **experimentos em blocos casualizados** ou **experimentos em blocos ao acaso**. Além dos princípios da *repetição* e da *casualização*, leva em conta também o princípio do *controle local*.

Sempre que houver dúvidas a respeito da homogeneidade das condições experimentais é conveniente usar o princípio do controle local, estabelecendo blocos com parcelas homogêneas.

As principais características deste delineamento são:

1. As parcelas são distribuídas em grupos ou blocos (princípio do controle local) de tal forma que elas sejam o mais uniforme possível dentro de cada bloco.
2. O número de parcelas por bloco deve ser igual ao número de tratamentos (blocos completos casualizados).
3. Os tratamentos são designados às parcelas de forma casual, sendo essa casualização feita dentro de cada bloco.

Esse delineamento é mais eficiente que o delineamento inteiramente casualizado, e essa eficiência depende da uniformidade das parcelas dentro de cada bloco, podendo inclusive, haver diferenças acentuadas de um bloco para outro.

Então, por exemplo, se temos um experimento em blocos casualizados em que desejamos estudar o efeito de 4 variedades (V1,V2,V3 e V4), sendo cada uma delas repetidas 5 vezes, teremos o seguinte plano experimental:

DBC

Bloco 1	V ₃	V ₁	V ₄	V ₂
Bloco 2	V ₂	V ₄	V ₃	V ₁
Bloco 3	V ₄	V ₁	V ₂	V ₃
Bloco 4	V ₃	V ₂	V ₄	V ₁
Bloco 5	V ₁	V ₃	V ₂	V ₄

Note que dentro de cada bloco temos as variedades sorteadas ao acaso. Caso este mesmo ensaio fosse montado no delineamento inteiramente casualizados, o sorteio seria feito em todas as parcelas do experimento, e os tratamentos não seriam agrupados em blocos. Por exemplo, no delineamento inteiramente casualizado teríamos:

DIC

V ₁	V ₄	V ₂	V ₁
V ₃	V ₂	V ₁	V ₄
V ₄	V ₃	V ₂	V ₄
V ₃	V ₁	V ₄	V ₃
V ₂	V ₃	V ₁	V ₂

As principais vantagens desse delineamento são:

1. Controla as diferenças que ocorrem nas condições experimentais de um bloco para outro.
2. Permite, dentro de certos limites, utilizar qualquer número de tratamentos e blocos.
3. Nos conduz a uma estimativa mais exata para a variância residual.
4. A análise de variância é relativamente simples, sendo apenas um pouco mais demorada que a do delineamento inteiramente casualizado, pois possui uma causa de variação a mais (Blocos).

As principais desvantagens desse delineamento são:

1. Pela utilização do princípio do controle local há uma diminuição no número de graus de liberdade do resíduo.
2. A exigência de homogeneidade dentro do bloco limita o número de tratamentos, que não pode ser muito grande.

Modelo matemático

Para podermos analisar um experimento em qualquer delineamento, necessitamos conhecer o modelo matemático do mesmo, e aceitar algumas hipóteses básicas necessárias para a validade da análise de variância.

O modelo matemático do delineamento em blocos casualizados é o seguinte:

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

onde, x_{ij} representa o valor esperado na parcela que recebeu o tratamento i e que se encontra no bloco j .

μ é a média geral do experimento.

τ_i é o efeito devido ao tratamento i que foi aplicado à parcela.

β_j é o efeito devido ao bloco j em que se encontra a parcela.

ϵ_{ij} é o efeito dos fatores não controlados ou acaso na parcela que recebeu o tratamento i e que se encontra no bloco j .

CRITÉRIO DO TESTE:

Para Tratamentos:

Comparamos o valor F calculado para tratamentos com o valor de F tabelado em função do número de GL de Tratamentos e GL do resíduo, ao nível α de significância.

Se $F_{T_{rat}} > F_{Tab}$, concluímos que o teste é significativo, portanto, rejeitamos H_0 e devemos concluir que existe diferença significativa entre os efeitos dos tratamentos testados em relação à variáveis (característica) em estudo.

Para Blocos:

Pra Blocos, a comparação é feita entre o valor de F calculado com o F tabelado em função do número de GL de Blocos e GL do resíduo, ao nível α de significância

Se $F_{B_{locos}} > F_{Tab}$, concluímos que o teste é significativo, portanto, rejeitamos H_0 e devemos concluir que os blocos possuem efeitos diferentes em relação à característica em estudo, ou seja, os blocos foram eficientes no controle da heterogeneidade local.

Exemplo de aplicação do DBC

No trabalho “Influência do genótipo e da adubação sobre algumas características fenotípicas de *Zea mays* L (Milho)”, realizado por BARBOSA (1976), foram utilizadas 4 cultivares de milho:

C_1 = OPACO 2

C_2 = PIRANÃO

C_3 = COMPOSTO FLINT

C_4 = AGROCERES AG-152

O ensaio foi montado de acordo com o delineamento em blocos casualizados, sendo utilizados 5 blocos para controlar as diferenças de fertilidade do solo entre terraços.

Os resultados obtidos para a produção em kg/ha, foram os seguintes e podem ser encontrados online em milho.

Tratamentos	Bloco1	Bloco2	Bloco3	Bloco4	Bloco5	Total
OPACO2	2812	2296	3501	3301	3691	15601
PIRANAO	3728	3588	4418	4544	5084	21362
COMP.FLINT	5359	5106	7477	8007	7956	33905
AG152	4482	4510	5236	5930	5025	25183
Total	16381	15500	20632	21782	21756	96051

As hipóteses que desejamos testar são as seguintes:

$$\begin{cases} H_0 : \text{As cultivares testadas no diferem entre si em relao produo da cultura do milho.} \\ H_1 : \text{As cultivares testadas diferem entre si em relao produo da cultura do milho.} \end{cases}$$

Aplicação em R - DBC



```
dados<-read.table(caminho,h=T,sep="\t")
head(dados)
```

```
##      trat bloco    y
## 1  OPACO2      1 2812
```



```
## 2  OPAC02      2 2296
## 3  OPAC02      3 3501
## 4  OPAC02      4 3301
## 5  OPAC02      5 3691
## 6  PIRANAO     1 3728
```

```
# Extraindo os fatores e a variável resposta
```

```
trat<-as.factor(dados$trat)
bloco<-as.factor(dados$bloco)
y<-dados$y
```

```
# Definindo o modelo matemático
```

```
modelo<-aov(y~trat+bloco)
anova(modelo)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: y
```

```
##          Df    Sum Sq Mean Sq F value    Pr(>F)
## trat        3 35402022 11800674  44.3450 9.068e-07 ***
## bloco        4  9221681  2305420   8.6634  0.00158 **
## Residuals   12  3193330   266111
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Comparação de Médias pelo teste de Tukey
```

```
require("agricolae")
```

```
## Loading required package: agricolae
```

```
##
```

```
## Attaching package: 'agricolae'
```

```
## The following objects are masked from 'package:ExpDes.pt':
```

```
##
```

```
##      lastC, order.group, tapply.stat
```

```
glRes<-df.residual(modelo)
```

```
QMres<-deviance(modelo)/glRes
```

```
tukey <- HSD.test(modelo,"trat", group=TRUE,console=TRUE)
```

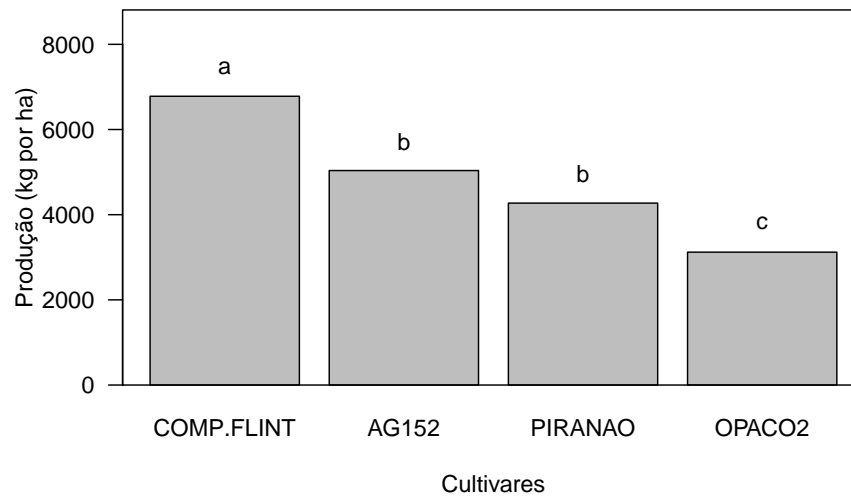
```
##
```

```
## Study: modelo ~ "trat"
```

```
##
## HSD Test for y
##
## Mean Square Error: 266110.8
##
## trat, means
##
##          y          std r  Min  Max
## AG152      5036.6   596.4368 5 4482 5930
## COMP.FLINT 6781.0 1431.4177 5 5106 8007
## OPAC02      3120.2   565.1997 5 2296 3691
## PIRANAO     4272.4   616.1240 5 3588 5084
##
## Alpha: 0.05 ; DF Error: 12
## Critical Value of Studentized Range: 4.19866
##
## Minimum Significant Difference: 968.628
##
## Treatments with the same letter are not significantly different.
##
##          y groups
## COMP.FLINT 6781.0      a
## AG152      5036.6      b
## PIRANAO     4272.4      b
## OPAC02      3120.2      c
```

```
bar.group(tukey$groups,
          las=1,
          ylim=c(0,max(y)*1.10),
          xlab="Cultivares",
          ylab="Produção (kg por ha)",
          main="Teste de Tukey (5%)");box()
```

Teste de Tukey (5%)



```
# Cálculo do CV
cv<-100*sqrt(QMres)/mean(y)
paste(round(cv,2), "%", sep="")
```

```
## [1] "10.74%"
```

Utilizando o pacote “ExpDes.pt”

```
# Carregando o pacote para a análise
library(ExpDes.pt)

# verificando os 6 primeiros registros
head(dados)
```

```
##      trat bloco    y
## 1  OPACO2     1 2812
## 2  OPACO2     2 2296
## 3  OPACO2     3 3501
## 4  OPACO2     4 3301
## 5  OPACO2     5 3691
## 6 PIRANAO     1 3728
```

```
# Análise de variância e teste de Tukey com a função dbc
trat <- dados$trat # Criando o vetor de tratamentos
bloco <- dados$bloco # Criando o vetor dos blocos
y <- dados$y # Criando o vetor com a variável resposta

# Utilizando a função
dbc(trat,bloco,y,mcomp = "tukey")
```

```
## -----
## Quadro da analise de variancia
## -----
##          GL          SQ QM          Fc          Pr>Fc
## Tratamento  3 35402022  2 44.345 0.00000091
## Bloco        4  9221681  3  8.663 0.00158020
## Residuo      12 3193330  4
## Total        19 47817033  1
## -----
## CV = 10.74 %
##
## -----
## Teste de normalidade dos residuos
## valor-p:  0.1786606
## De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser c
## -----
##
## -----
## Teste de homogeneidade de variancia
## valor-p:  0.03461775
## ATENCAO: a 5% de significancia, as variancias nao podem ser consideradas homogeneas
## -----
##
## Teste de Tukey
## -----
## Grupos Tratamentos Medias
## a    COMP.FLINT      6781
## b    AG152    5036.6
## b    PIRANAO    4272.4
## c    OPACO2      3120.2
## -----
```