Aluna: Camila Braz Soares

Lista 2 - MATD48: Planejamento de Experimentos

(1) $G = \sum_i \sum_j Y_{ij}$ $\quad \bar{y} = \frac{\sum_i \sum_j Y_{ij}}{IJ}$ $\quad \bar{y}_{i.} = \frac{\sum_j Y_{ij}}{J}$ $\quad \bar{y}_{.j} = \frac{\sum_i Y_{ij}}{I}$

$B_j = \sum_i Y_{ij}$ $\quad T_i = \sum_j Y_{ij}$ $\quad Y_{ij} = \mu + T_i + \beta_j + \varepsilon_{ij}$ $\quad i = 1, \dots, I, j = 1, \dots, J$

SQTotal = SQTrat + SQBlocos + SQRes

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i.} - \bar{y})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{.j} - \bar{y})^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} + \bar{y} - \bar{y}_{i.} - \bar{y}_{.j})^2$$

$\rightsquigarrow$ $SQTrat = \sum_i \sum_j (\bar{y}_{i.} - \bar{y})^2 = \sum_i \sum_j \left(\sum_j Y_{ij}/J - \sum_i \sum_j Y_{ij}/IJ\right)^2$

$$= \sum_i \sum_j \left[(\underbrace{\sum_j Y_{ij}}_{=T_i})^2/J^2 - 2\sum_j Y_{ij} \underbrace{\sum_i \sum_j Y_{ij}}_{=G}/IJ^2 + (\underbrace{\sum_i \sum_j Y_{ij}}_{=G}/IJ)^2 \right]$$

$$= \sum_i \sum_j \left[T_i^2/J^2 - 2\sum_j Y_{ij}\, G/IJ^2 + G^2/(IJ)^2 \right]$$

$$= \cancel{J}\sum_i T_i^2/J^{\cancel{2}} - 2\sum_j (\underbrace{\sum_i \sum_j Y_{ij}}_{=G}) G/IJ^2 + \sum_i \sum_j G^2/(IJ)^2]$$

$$= \sum_i T_i^2/J - 2\cancel{J}G^2/IJ^{\cancel{2}} + \cancel{IJ}G^2/(IJ)^{\cancel{2}}]$$

$$= \sum_i T_i^2/J - 2G^2/IJ + G^2/IJ$$

$$\boxed{SQTrat = \sum_i T_i^2/J - G^2/IJ}$$

$\rightsquigarrow$ $SQBlocos = \sum_i \sum_j (\bar{y}_{.j} - \bar{y})^2 = \sum_i \sum_j \left(\sum_i Y_{ij}/I - \sum_i \sum_j Y_{ij}/IJ\right)^2$

$$= \sum_i \sum_j \left[(\underbrace{\sum_i Y_{ij}}_{B_j}/I)^2 - 2\sum_i Y_{ij} \underbrace{\sum_i \sum_j Y_{ij}}_{=G}/I^2J + (\underbrace{\sum_i \sum_j Y_{ij}}_{=G}/IJ)^2 \right]$$

$$= \sum_i \sum_j B_j^2/I^2 - 2\cancel{I} \underbrace{\sum_i \sum_j Y_{ij}}_{=G} G/I^{\cancel{2}}J + \sum_i \sum_j G^2/(IJ)^2]$$

$$= \cancel{I}\sum_j B_j^2/I^{\cancel{2}} - 2G^2/IJ + \cancel{IJ}G^2/(IJ)^{\cancel{2}}$$

$$\boxed{SQBlocos = \sum_j B_j^2/I - G^2/IJ}$$

$\rightarrow$ SQ Total $= \sum_i \sum_j (Y_{ij} - \bar{y})^2 = \sum_i \sum_j \left(Y_{ij} - \sum_i \sum_j Y_{ij} / IJ \right)^2$

$$= \sum_i \sum_j \left[Y_{ij}^2 - 2 Y_{ij} \underbrace{\sum_i \sum_j Y_{ij}}_{=G} / IJ + \left(\underbrace{\sum_i \sum_j Y_{ij}}_{=G} / IJ \right)^2 \right]$$

$$= \sum_i \sum_j Y_{ij}^2 - 2 \sum_i \sum_j Y_{ij} G / IJ + \sum_i \sum_j G^2 / (IJ)^2$$

$$= \sum_i \sum_j Y_{ij}^2 - 2G^2/IJ + G^2/IJ$$

$$\boxed{\text{SQ Total} = \sum_i \sum_j Y_{ij}^2 - G^2/IJ}$$

$\rightarrow$ SQ Erros = SQ Total - SQ Blocos - SQ Trat

$$= \sum_{ij} Y_{ij}^2 - \cancel{\frac{G^2}{IJ}} - \frac{1}{I} \sum_j B_j^2 + \cancel{\frac{G^2}{IJ}} - \frac{1}{J} \sum_i \tau_i^2 + \frac{G^2}{IJ}$$

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrados Médios	F
Blocos	$J-1$	$\frac{1}{I} \sum_j B_j^2 - \frac{G^2}{IJ}$	$\frac{SQBlocos}{J-1}$	$\frac{QMBlocos}{QMRes}$
Tratamentos	$I-1$	$\frac{1}{J} \sum_i \tau_i^2 - \frac{G^2}{IJ}$	$\frac{SQTrat}{I-1}$	$\frac{QMTrat}{QMRes}$
Resíduo	$(I-1)(J-1)$	$\sum_{ij} Y_{ij}^2 - \frac{\sum_j B_j^2}{I} - \frac{\sum_i \tau_i^2}{J} + \frac{G^2}{IJ}$	$\frac{SQRes}{(I-1)(J-1)}$	
Total	$IJ-1$	$\sum_i \sum_j Y_{ij}^2 - \frac{G^2}{IJ}$		

(2)

$$Z = \|\varepsilon\|^2 = \sum_i \sum_j \varepsilon_{ij} = \sum_i \sum_j [Y_{ij} - E(Y_{ij})]^2 = \sum_i \sum_j [Y_{ij} - \mu - \tau_i - B_j]^2$$

$$\frac{\partial Z}{\partial \mu} = 2 \sum_i \sum_j (Y_{ij} - \hat{\mu} - \tau_i - B_j)(-1) = 0$$

$$\Rightarrow -2\sum_i \sum_j Y_{ij} + IJ\hat{\mu} + J\cancel{\sum_i \tau_i}^{0} + J\cancel{\sum_j B_j}^{0} = 0$$

$$\Rightarrow IJ\hat{\mu} = \sum_i \sum_j Y_{ij} \Rightarrow \hat{\mu} = \sum_i \sum_j Y_{ij} / IJ \Rightarrow \boxed{\hat{\mu} = \bar{y}_{ij}}$$

$$\frac{\partial Z}{\partial \tau_i} = \cancel{2} \sum_i \sum_j (Y_{ij} - \mu - \hat{\tau}_i - B_j)(-1) = 0$$

$$\Rightarrow -\sum_j Y_{ij} + J\mu + \sum_j \hat{\tau}_i + \cancel{\sum_j B_j}^{0} = 0$$

$$\Rightarrow J\hat{\tau}_i = \sum_j Y_{ij} - J\mu$$

$$\Rightarrow \hat{\tau}_i = \sum_j Y_{ij}/J - \mu \Rightarrow \boxed{\hat{\tau}_i = \bar{y}_{i.} - \bar{y}}$$

$$\frac{\partial Z}{\partial B_j} = \cancel{2} \sum_i \sum_j (Y_{ij} - \mu - \tau_i - \hat{B}_j)(-1) = 0$$

$$\Rightarrow -\sum_i Y_{ij} + I\mu + \cancel{\sum_i \tau_i}^{0} + I\hat{B}_j = 0$$

$$\Rightarrow I\hat{B}_j = \sum_i Y_{ij} - I\mu$$

$$\Rightarrow \hat{B}_j = \sum_i Y_{ij}/I - \mu \Rightarrow \boxed{\hat{B}_j = \bar{y}_{.j} - \bar{y}}$$

③ $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$; $\varepsilon_{ij} \sim N(0, \sigma^2)$

Primeiro, vamos calcular

- $E\left[\frac{1}{J}\sum_i \tau_i^2\right] = \frac{1}{J} E\left[\sum_i \tau_i^2\right] = \frac{1}{J} E\left[\sum_i \left(\sum_j Y_{ij}\right)^2\right]$

$$= \frac{1}{J} E\left[\sum_i \left(J\mu + (J\tau_i + \cancelto{0}{\sum_j \beta_j} + \sum_j \varepsilon_{ij}\right)^2\right]$$

$$= \frac{1}{J} E\left[\sum_i \left(J^2\mu^2 + 2(J\tau_i + \cancelto{0}{\sum_j \varepsilon_{ij}}) + (J\tau_i + \sum_j \varepsilon_{ij})^2\right)\right]$$

$$= \frac{1}{J} E\left[\sum_i \left(J^2\mu^2 + 2J\tau_i + J^2\tau_i^2 + 2J\tau_i \sum \varepsilon_{ij} + \sum_j \varepsilon_{ij}^2\right)\right]$$

$$= \frac{1}{J} E\left[IJ^2\mu^2 + 2J\cancelto{0}{\sum_i \tau_i} + J^2 \sum_i \tau_i^2 + 2J \cancel{\sum_i \tau_i} \sum_j \varepsilon_{ij} + IJ\sigma^2\right]$$

$$\Rightarrow \boxed{E\left[\frac{1}{J}\sum_i \tau_i^2\right] = IJ\mu^2 + J\sum_i \tau_i^2 + I\sigma^2} \quad (1)$$

- $E\left[\frac{G^2}{IJ}\right] = \frac{1}{IJ} E(G^2) = \frac{1}{IJ} E\left[\left(\sum_i \sum_j Y_{ij}\right)^2\right] = \frac{1}{IJ} E\left[\left(\sum_i \sum_j (\mu + \tau_i + \beta_j + \varepsilon_{ij})\right)^2\right]$

$$= \frac{1}{IJ} E\left[\left(IJ\mu + J\cancelto{0}{\sum \tau_i} + I\cancelto{0}{\sum_j \beta_j} + \sum_i \sum_j \varepsilon_{ij})\right)^2\right]$$

$$= \frac{1}{IJ} E\left[I^2J^2\mu^2 + 2IJ\mu \cancelto{0}{\sum_i \sum_j \varepsilon_{ij}} + \sum_i \sum_j \varepsilon_{ij}^2\right]$$

$$= \frac{1}{IJ}\left(I^2J^2\mu^2 + IJ\underbrace{E(\varepsilon_{ij}^2)}_{=\sigma^2}\right) = IJ\mu^2 + \sigma^2$$

$$\Rightarrow \boxed{E\left[\frac{G^2}{IJ}\right] = IJ\mu^2 + \sigma^2} \quad (2)$$

- $E\left[\frac{1}{I}\sum_j \beta_j^2\right] = \frac{1}{I} E\left[\sum_j \left(\sum_i Y_{ij}\right)^2\right] = \frac{1}{I} E\left[\sum_j \left(\sum_i (\mu + \tau_i + \beta_j + \varepsilon_{ij})\right)^2\right]$

$$= \frac{1}{I} E\left[\sum_j \left(I\mu + \cancelto{0}{\sum_i \tau_i} + I\beta_j + \sum_i \varepsilon_{ij}\right)^2\right]$$

$$= \frac{1}{I} E\left[\sum_j \left(I^2\mu^2 + 2I\mu(I\beta_j + \sum_i \varepsilon_{ij}) + I^2\beta_j^2 + 2I\beta_j \sum_i \varepsilon_{ij} + (\sum_i \varepsilon_{ij})^2\right)\right]$$

$$= \frac{1}{I}\left(JI^2\mu^2 + 2I^2\mu \cancelto{0}{\sum_j \beta_j} + 2I\mu E(\cancelto{0}{\sum_i \varepsilon_{ij}}) + I^2 \sum_j (\beta_j^2) + 2I\sum_j \beta_j \sum_i \varepsilon_{ij} + IJ\sigma^2\right)$$

$$\Rightarrow \boxed{E\left[\frac{1}{I}\sum_j \beta_j^2\right] = IJ\mu^2 + I\sum_j \beta_j^2 + J\sigma^2} \quad (3)$$

Com esses resultados, é fácil ver que

$$E[QMTrat] = E\left[\frac{SQTrat}{I-1}\right] = \frac{1}{I-1} E[SQTrat] = \frac{1}{I-1} E\left[\underbrace{\frac{1}{J}\sum_i T_i^2}_{(1)} - \underbrace{\frac{G^2}{IJ}}_{(2)}\right]$$

$$= \frac{1}{I-1}\left[\underbrace{E\left[\frac{1}{J}\sum_i T_i^2\right]}_{(1)} - \underbrace{E\left[\frac{G^2}{IJ}\right]}_{(2)}\right]$$ Substituindo (1) e (2) encontrados anteriormente.

$$= \frac{1}{I-1}\left[\left(\cancel{IJ\mu^2} + J\sum_i T_i^2 + I\sigma^2\right) - \left(\cancel{IJ\mu^2} + \sigma^2\right)\right]$$

$$= \frac{1}{I-1}\left[J\sum_i T_i^2 + \sigma^2(I-1)\right]$$

$$= \frac{J\sum_i T_i^2}{I-1} + \sigma^2\frac{\cancel{(I-1)}}{\cancel{(I-1)}}$$

$$\Rightarrow \boxed{E[QMTrat] = \sigma^2 + \frac{J\sum_i T_i^2}{(I-1)}}$$

$$E[QMBlocos] = \frac{1}{J-1} E[SQBlocos] = \frac{1}{J-1} E\left[\frac{1}{I}\sum_j B_j^2 - \frac{G^2}{IJ}\right]$$

$$= \frac{1}{J-1}\left[\underbrace{E\left(\frac{1}{I}\sum_j B_j^2\right)}_{(3)} - \underbrace{E\left(\frac{G^2}{IJ}\right)}_{(2)}\right]$$ Substituindo (2) e (3) encontrados anteriormente, temos

$$= \frac{1}{J-1}\left[\left(\cancel{IJ\mu^2} + IJ\sum_j B_j^2 + J\sigma^2\right) - \left(\cancel{IJ\mu^2} + \sigma^2\right)\right]$$

$$= \frac{1}{J-1}\left[IJ\sum_j B_j^2 + \sigma^2(J-1)\right]$$

$$\Rightarrow \boxed{E[QMBlocos] = \sigma^2 + \frac{IJ\sum_j B_j^2}{IJ}}$$

$$E[QMErro] = E\left(\frac{SQRes}{(I-1)(J-1)}\right) = \frac{1}{(I-1)(J-1)} E(SQRes) \;*$$

$$* \; SQRes = SQTotal - SQBlocos - SQTrat$$

$$= \sum_{ij} Y_{ij}^2 - \frac{1}{I}\sum_j B_j^2 - \frac{1}{J}\sum_i T_i^2 + \frac{G^2}{IJ}$$

?

Agora, vamos encontrar $E\left[\sum_i\sum_j (Y_{ij})^2\right]$

$$E\left[\sum_i\sum_j (Y_{ij})^2\right] = E\left[\sum_i\sum_j (\mu+\tau_i+\beta_j+\varepsilon_{ij})^2\right]$$

$$= E\left[\sum_i\sum_j (\mu^2+\tau_i^2+\beta_j^2+\varepsilon_{ij}^2)\right]$$

$$= IJ\mu^2 + J\sum_i\tau_i^2 + I\sum_j\beta_j^2 + \sum_i\sum_j E(\varepsilon_{ij}^2)$$

$$\boxed{E\left[\sum_i\sum_j (Y_{ij})^2\right] = IJ\mu^2 + J\sum_i\tau_i^2 + I\sum_j\beta_j^2 + IJ\sigma^2} \quad (4)$$

Portanto,

$$E[QMErro] = \frac{1}{(I-1)(J-1)} E(SQRes) = \frac{1}{(I-1)(J-1)} E\left(\sum_{ij} Y_{ij}^2 - \frac{1}{I}\sum_j B_j^2 - \frac{1}{J}\sum_i T_i^2 + \frac{G^2}{IJ}\right)$$

$$= \frac{1}{(I-1)(J-1)}\left[\underbrace{E\left(\sum_{ij} Y_{ij}^2\right)}_{(4)} - \underbrace{E\left(\frac{1}{I}\sum_j B_j^2\right)}_{(3)} - \underbrace{E\left(\frac{1}{J}\sum_i T_i^2\right)}_{(1)} + \underbrace{E\left(\frac{G^2}{IJ}\right)}_{(2)}\right]$$

Substituindo (1), (2), (3) e (4), encontramos

$$E[QMErro] = \frac{1}{(I-1)(J-1)} E\Big[(\cancel{IJ\mu^2} + \cancel{J\sum_i\tau_i^2} + \cancel{I\sum_j\beta_j^2} + IJ\sigma^2) - (\cancel{IJ\mu^2} + \cancel{I\sum_j\beta_j^2} + J\sigma^2)$$

$$- (\cancel{IJ\mu^2} + \cancel{J\sum_i\tau_i^2} + I\sigma^2) + (\cancel{IJ\mu^2} + \sigma^2)\Big]$$

$$= \frac{1}{(I-1)(J-1)}\left[IJ\sigma^2 - J\sigma^2 - I\sigma^2 + \sigma^2\right]$$

$$= \frac{1}{(I-1)(J-1)}\left[J\sigma^2(I-1) - \sigma^2(I-1)\right]$$

$$= \frac{1}{\cancel{(I-1)}\cancel{(J-1)}}\left[\sigma^2\cancel{(I-1)}\cancel{(J-1)}\right] = \sigma^2$$

$$\Rightarrow \boxed{E[QMErro] = \sigma^2}$$

Questão 4

LETRA A

O conjunto de dados *PimentelEg5.2* refere-se ao experimento de competição de variedades de batatinha feito pelo Engenheiro Agrônomo Oscar A. Garay em Balcare, Argentina. O experimento foi realizado em blocos casualizados. O conjunto de dados possui 32 observações e 3 variáveis, são elas: o bloco, a variedade e a produção.

FORMATO:

Um *data.frame* com 32 observações e 3 variáveis, em que:

- BLOCO

Fator de 4 níveis qualitativos, usado para controle local.

- VARIEDADE

Fator de 8 níveis qualitativos que são as variedades de batatinha.

- PRODUÇÃO

Produção de batatinha, em ton ha^{-1}, nas unidades experimentais.

LETRA B

** Hipóteses para os tratamentos:**

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$ H_1 : pelo menos dois tratamentos possuem médias diferentes.

** Hipóteses para os blocos:**

$H_0 : B_1 = B_2 = B_3 = B_4$ H_1 : pelo menos um dos blocos possuem médias diferentes.

LETRA C

Erros para os tratamentos

Erro tipo I: Afirmar que existem pelo menos dois tratamentos com efeitos diferentes, quando na realidade, todos os tratamentos possuem o mesmo efeito.

Erro tipo II: Afirmar que todos os tratamentos são iguais, quando na realidade existem pelo menos dois tratamentos diferentes.

Erros para os tratamentos

Erro tipo I: Afirmar que existem pelo menos dois blocos com efeitos diferentes, quando na realidade, todos os blocos possuem o mesmo efeito.

Erro tipo II: Afirmar que todos os blocos são iguais, quando na realidade existem pelo menos dois blocos diferentes.

LETRA D

Fonte de variação	g.l.	Soma Quad.	Média Quad.	F	P-valor
Tratamento	7	919,72	131,389	15,3744	5,723e-07
Bloco	3	50,53	16,843	1,9709	0,1493
Resíduos	21	179,46	8,546		
——————-	——	—————	—————-	———	———-
Total	31	1.149,71			

Como o p-valor do tratamento é muito pequeno (p-valor < 0,01), temos evidências para rejeitar a hipótese nula ao nível de 1%, portanto, podemos afirmar que existem pelo menos dois tratamentos com efeitos diferentes.

Para os blocos, temos que p-valor = 0,1493 > 0,10, não temos evidências para rejeitar a hipótese nula, portanto, podemos afirmar que os efeitos dos blocos são iguais.

LETRA E

```r
dados = PimentelEg5.2
names(dados) = c("bloco", "trat", "resposta")


modelo.aov = aov(resposta~trat+bloco, data = dados)
TukeyC(modelo.aov)
```

```
## Results
##              Means G1 G2 G3
## S. Rafalela 25.45  a
## Huinkul     25.05  a
## B 72-53 A   22.80  a   b
## B 116-51    22.50  a   b
## B 1-52      22.27  a   b
## B 25-50 E   16.50      b  c
## Buena Vista 12.42         c
## Kennebec    10.70         c
##
## Sig.level
##  0.05
##
## Diff_Prob
##             S. Rafalela Huinkul B 72-53 A B 116-51 B 1-52 B 25-50 E Buena Vista
## S. Rafalela       0.000   0.400     2.650    2.950  3.175     8.950      13.025
## Huinkul           1.000   0.000     2.250    2.550  2.775     8.550      12.625
## B 72-53 A         0.896   0.952     0.000    0.300  0.525     6.300      10.375
## B 116-51          0.835   0.912     1.000    0.000  0.225     6.000      10.075
## B 1-52            0.780   0.872     1.000    1.000  0.000     5.775       9.850
## B 25-50 E         0.006   0.009     0.093    0.122  0.150     0.000       4.075
## Buena Vista       0.000   0.000     0.001    0.002  0.002     0.522       0.000
## Kennebec          0.000   0.000     0.000    0.000  0.000     0.146       0.989
##             Kennebec
## S. Rafalela   14.750
## Huinkul       14.350
## B 72-53 A     12.100
## B 116-51      11.800
## B 1-52        11.575
## B 25-50 E      5.800
## Buena Vista    1.725
## Kennebec       0.000
##
## MSD
##             S. Rafalela Huinkul B 72-53 A B 116-51 B 1-52 B 25-50 E Buena Vista
## S. Rafalela       0.000   6.933     6.933    6.933  6.933     6.933       6.933
## Huinkul           6.933   0.000     6.933    6.933  6.933     6.933       6.933
## B 72-53 A         6.933   6.933     0.000    6.933  6.933     6.933       6.933
```

```
## B 116-51          6.933   6.933     6.933    0.000  6.933     6.933       6.933
## B 1-52            6.933   6.933     6.933    6.933  0.000     6.933       6.933
## B 25-50 E         6.933   6.933     6.933    6.933  6.933     0.000       6.933
## Buena Vista       6.933   6.933     6.933    6.933  6.933     6.933       0.000
## Kennebec          6.933   6.933     6.933    6.933  6.933     6.933       6.933
##             Kennebec
## S. Rafalela    6.933
## Huinkul        6.933
## B 72-53 A      6.933
## B 116-51       6.933
## B 1-52         6.933
## B 25-50 E      6.933
## Buena Vista    6.933
## Kennebec       0.000
```

De acordo com o Teste de Tukey temos que os melhores testes são o S. Rafalela e o Huinkul ao nível de 5%.

LETRA F

Teste de Normalidade

Para testar se os resíduos possuem distribuição Normal, vamos utilizar os teste de Shapiro-Wilks, cujas hipóteses são:

H_0 : Os dados seguem uma distribuição Normal.

H_1 : Os dados não seguem uma distribuição Normal.

```
modelo = lm(resposta~trat+bloco, data = dados)

shapiro.test(rstandard(modelo))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rstandard(modelo)
## W = 0.96895, p-value = 0.471
```

Como p-valor = 0,471 > 0,05, não temos evidências para rejeitar a hipótese de normalidade, portanto podemos afirmar que os resíduos seguem uma distribuição Normal.

Teste de homocedasticidade

Agora vamos testar se os resíduos possuem homocedasticidade utilizando o teste de Breusch-Pagan, cujas hipóteses são:

H_0 : Os dados são homocedásticos.

H_1 : Os dados não são homocedásticos.

```
bptest(modelo)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo
## BP = 11.955, df = 10, p-value = 0.2881
```

A partir do teste de Breusch-Pagan obtemos um p-valor = 0,2881 > 0,05, portanto, não temos evidências para rejeitar a hipótese de homocedasticidade, portanto, podemos afirmar ao nível de 5% de significância os resíduos são homocedásticos.