INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO EXPERIMENTAL– 2024-25

Resoluções dos Exercícios de Análise de Covariância de Efeitos Fixos

1. (a) O comando do *R* para obter a nuvem de pontos pedida:

```
> plot(NLdir ~ NP,col=Casta, data=videiras, pch=16)
```

(b) Os comandos R necessários e os resultados numérios correspondentes ao modelo ANCOVA para modelar o comprimento da nervura lateral direita a partir do comprimento da nervura principal, possibilitando que para cada casta seja ajustada uma recta diferente, são apresentados seguidamente.

```
> videirasancova.lm <- lm(NLdir ~ NP*Casta, data=videiras)
> summary(videirasancova.lm)
---
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.39812    0.32102   4.355 1.57e-05 ***
NP                    0.77780    0.02654  29.305  < 2e-16 ***
CastaFernaoPires     -0.43069    0.48897  -0.881    0.379
CastaVital           -0.66120    0.43788  -1.510    0.132
NP:CastaFernaoPires   0.03395    0.04253   0.798    0.425
NP:CastaVital         0.04100    0.03798   1.079    0.281
---
Residual standard error: 0.8316 on 594 degrees of freedom
Multiple R-squared:  0.8112,Adjusted R-squared:  0.8096
F-statistic: 510.5 on 5 and 594 DF,  p-value: < 2.2e-16
```

A equação do modelo ANCOVA, em notação vectorial, para modelar o comprimento da nervura lateral direita a partir do comprimento da nervura principal, possibilitando que para cada casta seja ajustada uma recta diferente, é descrito da seguinte forma:

$$\vec{\mathbf{Y}} \;=\; \beta_0 \cdot \vec{\mathbf{1}}_n + \beta_1 \cdot \vec{\mathbf{x}} + \alpha_{0:2} \cdot \vec{\mathbf{I}}_2 + \alpha_{0:3} \cdot \vec{\mathbf{I}}_3 + \alpha_{1:2} \cdot \vec{\mathbf{x}} \star \vec{\mathbf{I}}_2 + \alpha_{1:3} \cdot \vec{\mathbf{x}} \star \vec{\mathbf{I}}_3 + \vec{\boldsymbol{\epsilon}}$$

$\vec{\mathbf{Y}}$ é o vector com as $n = 600$ observações da variável resposta (nervura lateral direita);
β_0 representa a ordenada na origem da recta populacional no primeiro nível do factor casta, isto é, casta Água Santa (o nível de referência, logo, não foi explicitado na listagem dos resultados);
$\vec{\mathbf{1}}_n$ é um vector com $n = 600$ uns;
β_1 representa o declive da recta populacional no primeiro nível do factor casta (isto é, da casta Água Santa);
$\vec{\mathbf{x}}$ é o vector com os valores da variável preditora numérica (comprimento da nervura principal);
$\vec{\mathbf{I}}_j$ a variável indicatriz de pertença aos níveis do factor (das observações da casta $j = 2, 3$, Fernão Pires e Vital, respetivamente);
$\alpha_{i:j}$ o desvio no parâmetro β_i $(i = 0, 1)$ na casta $j = 2, 3$ em relação à casta de referência (a Água Santa);
$\vec{\mathbf{x}} \star \vec{\mathbf{I}}_j$ é o vector com os valores do comprimento da nervura principal na casta j $(j > 1)$ e zero noutras posições;
$\vec{\boldsymbol{\epsilon}}$ é o vector com $n = 600$ erros aleatórios, cujos pressupostos são os usuais do modelo linear clássico: $\epsilon_i \cap \mathcal{N}(0, \sigma^2)$, $\forall i$; erros aleatórios independentes. Mas, adicionalmente neste modelo, a homogeneidade das variâncias dos erros aleatórios tem de ser comum às 3 castas.

Interpretando agora as estimativas dos parâmetros, tem-se:

$b_0 = 1.39812$, estimativa da ordenada na origem da recta para a casta Água Santa (o nível de referência);
$b_1 = 0.77780$, estimativa do declive da recta para a casta Água Santa (o nível de referência);
$\hat{\alpha}_{0:2} = -0.43069$, estimativa do desvio da ordenada na origem da recta da casta Fernão Pires relativamente à ordenada na origem da recta da casta Água Santa (o nível de referência);
$\hat{\alpha}_{0:3} = -0.66120$, estimativa do desvio da ordenada na origem da recta da casta Vital relativamente à ordenada na origem da recta da casta Água Santa (o nível de referência);
$\hat{\alpha}_{1:2} = 0.03395$, estimativa do desvio do declive da recta da casta Fernão Pires relativamente ao declive da recta da casta Água Santa (o nível de referência);
$\hat{\alpha}_{1:3} = 0.04100$, estimativa do desvio do declive da recta da casta Vital relativamente ao declive da recta da casta Água Santa (o nível de referência).

Escrevendo a equação da recta ajustada referente a cada casta, tem-se:
A recta para a casta Água Santa (o nível de referência): $y = 1.39812 + 0.77780x$;
A recta para a casta Fernão Pires: $y = (1.39812 - 0.43069) + (0.77780 + 0.03395)x = 0.96743 + 0.81175x$
A recta para a casta Vital: $y = (1.39812 - 0.66120) + (0.77780 + 0.04100)x = 0.73692 + 0.81880x$

(c) A hipótese das rectas das castas Vital e Água Santa terem o mesmo declive é a hipótese $\alpha_{1:3} = 0$. Podemos, portanto, responder a esta questão com o teste t ao parâmetro $\alpha_{1:3}$ seguidamente descrito.

Hipóteses: $H_0 : \alpha_{1:3} = 0 \quad \text{vs.} \quad H_1 : \alpha_{1:3} \neq 0$

Estatística do Teste: $T = \frac{\hat{\alpha}_{1:3} - 0}{\hat{\sigma}_{\hat{\alpha}_{1:3}}} \cap t_{(n-(numero\ de\ parametros\ do\ modelo))}$, sob H_0.

Nível de significância: $\alpha = 0.05$.

Região Crítica: (Bilateral) Rejeitar H_0 se $|T_{\text{calc}}| > t_{0.025(594)} \approx 1.96$.

Conclusões: Tem-se $T_{\text{calc}} = 1.079 < 1.96$. Assim, não se rejeita a hipótese nula ao nível de significância de 0.05, concluindo-se que os declives das rectas para as castas Vital e Água Santa não são significativamente diferentes. Uma vez que é disponibilizado o valor do $p\text{-}value$ do teste, também se chegaria facilmente a este conclusão, pois o $p - value = 0.281$ é maior que $\alpha = 0.05$.

(d) Trata-se um Teste F de comparação de um modelo com 3 rectas de regressão linear diferentes (modelo completo, descrito na alínea b) e o submodelo de recta única ($\overrightarrow{\mathbf{Y}} = \beta_0 \cdot \overrightarrow{\mathbf{1}_n} + \beta_1 \cdot \overrightarrow{\mathbf{x}} + \overrightarrow{\boldsymbol{\epsilon}}$). Caso os parâmetros de acréscimo (desvios) $\alpha_{i:j}$ sejam *todos* iguais a zero, a recta de regressão é a mesma para as três castas (isto é, $\alpha_{0:2} = \alpha_{0:3} = \alpha_{1:2} = \alpha_{1:3} = 0$).

Hipóteses: $H_0 : \alpha_{i:j} = 0\,,\ \forall\, i = 0,1; j = 2,3 \quad \text{vs.} \quad H_1 : \exists\,(i,j)$ t.q. $\alpha_{i:j} \neq 0$.

Estatística do Teste

(ver formulário RLM):
$F = \frac{(SQRE_S - SQRE_C)/(diferenca\ entre\ o\ numero\ de\ parametros\ dos\ 2\ modelos)}{SQRE_C/(g.l.SQRE_C)} \cap F_{(4,594)}$, sob H_0
ou
$F = \frac{g.l.SQRE_C}{diferenca\ entre\ o\ numero\ de\ parametros\ dos\ 2\ modelos} \cdot \frac{R_c^2 - R_s^2}{1 - R_c^2} \cap F_{(4,594)}$, sob H_0

Nível de significância do teste: $\alpha = 0.05$

Região Crítica (Região de Rejeição) : Unilateral direita

Rejeitar H_0 se $F_{calc} > f_{0.05(4\,,\,594)} \approx 2.39$

O resultado do ajustamento do modelo completo encontra-se na alínea b. O resultado do ajustamento do submodelo é:

```
> videiras.lm <- lm(NLdir ~ NP, data=videiras)
> summary(videiras.lm)
```

```
Call:
lm(formula = NLdir ~ NP, data = videiras)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1650 -0.5262 -0.0379  0.5369  3.7009

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.96218    0.18309   5.255 2.06e-07 ***
NP           0.80841    0.01607  50.314  < 2e-16 ***
---

Residual standard error: 0.8339 on 598 degrees of freedom
Multiple R-squared:  0.8089,Adjusted R-squared:  0.8086
F-statistic:  2532 on 1 and 598 DF,  p-value: < 2.2e-16
```

Calculando a estatística do teste com base nos coeficientes de determinação amostrais obtém-se: $F_{calc} = \frac{594}{4} \cdot \frac{0.8112-0.8089}{1-0.8112} = 1.81$. Este valor é menor que $f_{0.05(4\,,\,594)} \approx 2.39$ Assim, não se rejeita H_0 ao nível de significância de 0.05. Isto é, a qualidade do ajustamento do modelo ANCOVA não é significatibamente diferente da qualidade do ajustamento do submodelo que considera uma única recta para as três castas, optando-se, assim, pelo modelo mais parcimonioso (modelo RLS).

Este teste $Fparcial$ obtém-se no R com o seguinte comando:

```
> anova(videiras.lm, videirasancova.lm)
Analysis of Variance Table

Model 1: NLdir ~ NP
Model 2: NLdir ~ NP * Casta
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    598 415.80
2    594 410.81  4    4.9948 1.8055 0.1262
```

(e) A matriz X do modelo ANCOVA ajustado na alínea b é obtida com o comando:

```
> model.matrix(videirasancova.lm)
    (Intercept)   NP CastaFernaoPires CastaVital NP:CastaFernaoPires NP:CastaVital
1             1 13.8                1          0                13.8           0.0
2             1  9.1                1          0                 9.1           0.0
3             1 14.5                1          0                14.5           0.0
4             1 13.8                1          0                13.8           0.0
5             1 12.0                1          0                12.0           0.0
6             1 11.5                1          0                11.5           0.0
7             1 12.5                1          0                12.5           0.0
8             1  9.4                1          0                 9.4           0.0
9             1 12.5                1          0                12.5           0.0
10            1 10.3                1          0                10.3           0.0
11            1  9.6                1          0                 9.6           0.0
12            1 10.5                1          0                10.5           0.0
13            1  8.3                1          0                 8.3           0.0
....
```

2. A resolução deste exercício é análoga à do exercício anterior e ao exemplo feito nas aulas teóricas (slides 373 a 376).