

Exercícios - Matemática e Estatística

3 Análise de Variância

AVISO: Os conjuntos de dados de vários Exercícios desta secção encontram-se disponíveis numa área de trabalho associada à disciplina, na máquina `prunus` do ISA. Para disponibilizar estes conjuntos de dados deve-se:

- Montar (*Map network drive*, no menu *Tools* do *My Computer*) a *drive*:
`\\prunus\home\cadeiras`
- Abrir uma sessão do R (na directoria onde está a guardar o seu trabalho).
- A partir da sessão do R, seleccionar a opção *Files*, na barra de menus, e dentro da lista de opções disponibilizada, escolher *Load Workspace*.
- Na janela de diálogo que se abre, seleccionar a nova *drive* (que ficou associada ao `prunus`), depois a directoria `EstDel` e finalmente o ficheiro `exerANOVA.RData`.

Se tudo correu bem, na sessão do R deverão estar agora disponíveis (confirme com o comando `ls()`), entre outros, os objectos `toxicos` (Exercício 1), `C02` (Exercício 3) e `zinco` (Exercício 4).

EXERCÍCIOS

1. Um estudo sobre a presença de substâncias tóxicas no meio ambiente, associadas à utilização de um certo tipo de pesticida envolve análises efectuadas por quatro laboratórios diferentes, relativas à concentração de um determinado composto químico.

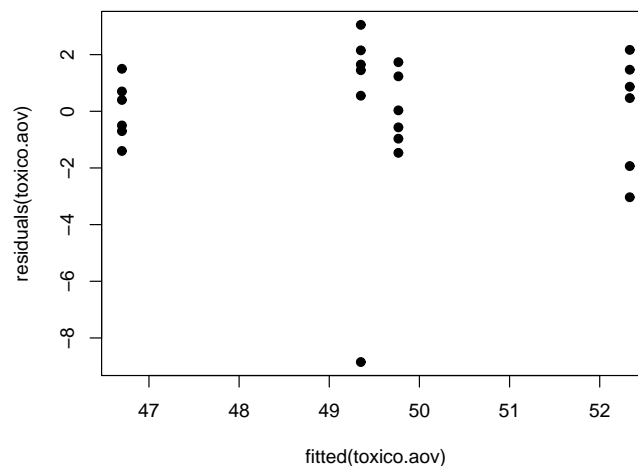
	Laboratórios			
	1	2	3	4
	53.2	51.0	47.4	51.0
	54.5	40.5	46.2	51.5
	52.8	50.8	46.0	48.8
	49.3	51.5	45.3	49.2
	50.4	52.4	48.2	48.3
	53.8	49.9	47.1	49.8
Média ($\bar{y}_{i.}$)	52.33	49.35	46.70	49.77
variância amostral (s_i^2)	4.151	19.480	1.120	1.587

Nota: A média global das observações ($\bar{y}_{..}$) é 49.5375; a variância amostral da totalidade das observações (s_y^2) é 9.887.

Existe alguma preocupação de que os resultados destas análises difram, não apenas em virtude da habitual variabilidade inerente a quaisquer medições, mas também devido à utilização de diferentes técnicas laboratoriais, o que, a ser verdade, poderia exigir maior controlo para verificar se é utilizada uma técnica padrão, comum a todos os laboratórios. A fim de testar essa hipótese, foram entregues

a cada laboratório 6 pequenos contentores com solo recolhido aleatoriamente num terreno que antes fora tratado com o referido pesticida. Os resultados laboratoriais das análises químicas (ppm) são os indicados na tabela (e que estão disponíveis na *data frame* `toxicos`, que pode ser carregada para o \mathbb{R} , seguindo as instruções dadas à cabeça destes Exercícios).

- Indique qual o tipo de delineamento experimental em causa. Explícite o modelo correspondente e todas as hipóteses adicionais que sejam necessárias à consideração do problema em estudo.
- Construa a tabela-resumo da análise de variância correspondente a este caso,
 - Utilizando o comando `aov` do \mathbb{R} .
 - Utilizando apenas uma máquina de calcular e a informação disponível neste enunciado.
- Formalize e efectue o teste de hipóteses adequado ao problema acima referido, com um nível de significância de 5%. Parece-lhe necessário exigir aos laboratórios a adopção de uma técnica padronizada?
- Repita agora o teste da alínea anterior, mas utilizando o nível de significância de 1%. Compare os resultados e comente.
- Utilize a função `model.matrix` do \mathbb{R} para inspeccionar a matriz \mathbf{X} do delineamento. Com base nessa matriz, construa a matriz \mathbf{H} de projecção ortogonal sobre o subespaço gerado pelas colunas da matriz \mathbf{X} . Comente.
- Utilize a função `fitted` do \mathbb{R} para identificar os valores ajustados da variável resposta, nesta Análise de Variância.
- O gráfico dos resíduos (usuais) das observações, contra os valores estimados pelo modelo de análise de variância, é apresentado a seguir. Comente o gráfico e as suas possíveis implicações. Identifique a observação cujo resíduo é, em módulo, mais elevado.



- Caso o teste F indique que o Factor tem efeito sobre a variável resposta, diga quais os laboratórios cujos resultados médios se podem considerar diferentes.

2. Num estudo sobre o café Robusta em Angola¹ analisa-se a frequência de defeitos no grão de café para lotes provenientes de 6 diferentes regiões de Angola. Admita que, em cada região, foram escolhidos

¹ Contribuição para o Estudo das Características dos Cafés em Angola, A. Baião Esteves e J. Santos Oliveira.

aleatoriamente 11 lotes. Na tabela seguinte são apresentadas, para cada região, as médias e desvios-padrão da percentagem do peso total de grãos sem defeito, assim como as variâncias e log-variâncias associadas.

Região	Média	Desvio-padrão (s_i)	Variância (s_i^2)	$\ln(s_i^2)$
Cabinda	44.19	6.94	48.1636	3.8746
Ambriz	58.87	18.98	360.2404	5.8868
Encoje	51.80	13.92	193.7664	5.2667
Cazongo	60.61	13.29	176.6241	5.1740
Libolo	61.96	24.49	599.7601	6.3965
Amboim	42.11	21.31	454.1161	6.1184

- Descreva o delineamento experimental e o modelo ANOVA que lhe está associado.
 - Construa o quadro-resumo associado a esta ANOVA.
 - Determine a variância da totalidade das $n = 66$ observações.
 - Teste a hipótese de a percentagem de grão sem defeito não depender da região de proveniência do grão. Utilize os níveis de significância 0.05 e 0.01 e comente. Calcule o valor p da estatística calculada.
 - Independentemente do resultado da alínea anterior, verifique quais as regiões cujas médias de grãos com defeito devem ser consideradas diferentes, e quais aquelas em que se pode admitir a igualdade, utilizando um teste de Tukey, com um nível de significância 0.05.
 - Teste a validade da hipótese de homogeneidade de variâncias exigida no modelo que indicou na primeira alínea. Comente o resultado obtido.
3. Sabe-se que o dióxido de carbono tem um efeito crítico no crescimento de populações microbianas; pequenas quantidades de CO_2 podem estimular o crescimento de algumas espécies enquanto que, pelo contrário, grandes concentrações têm de forma geral uma acção inibitória. Este último efeito é usado comercialmente para preservar alimentos armazenados.

Realizou-se um estudo para investigar a acção de diferentes concentrações de CO_2 na taxa de crescimento de *Pseudomonas fragi*; os diferentes níveis (tratamentos) foram pré-fixados e a variável resposta medida foi a percentagem de variação na massa das culturas após uma hora de crescimento nas respectivas condições, originando os seguintes dados (disponíveis na *data frame* C02):

Concentração de CO_2					
	0.0	.083	.29	.50	.86
	62.6	50.9	45.5	29.5	24.9
	59.6	44.3	41.1	22.8	17.2
	64.5	47.5	29.8	19.2	7.8
	59.3	49.5	38.3	20.6	10.5
	58.6	48.5	40.2	29.2	17.8
	64.6	50.4	38.5	24.1	22.1
	50.9	35.2	30.2	22.6	22.6
	56.2	49.9	27.0	32.7	16.8
	52.3	42.6	40.0	24.4	15.9
	62.8	41.6	33.9	29.6	8.8

- Pretende-se testar a hipótese nula $H_0 : \mu_1 = \mu_2 = \dots = \mu_5$, onde μ_i indica a taxa de crescimento esperada para a i -ésima concentração de CO_2 . É sugerida a utilização de uma Análise de Variância. Enuncie as hipóteses necessárias para poder efectuar o Teste referido.

- (b) Haverá evidência suficiente para rejeitar H_0 com uma significância de $\gamma = .05$? Em caso afirmativo, indique as concentrações de CO_2 que têm efeitos diferentes sobre o crescimento do *Pseudomonas fragi*, recorrendo aos intervalos de confiança de Tukey.
- (c) Estude a validade dos pressupostos do modelo ANOVA.
- (d) Dada a natureza da variável preditora, também poderia ser considerada uma regressão linear das taxas de crescimento sobre as concentrações de dióxido de carbono. Ajuste esse modelo e comente a sua utilidade. Compare com o modelo ANOVA e comente.
4. Realizaram-se experiências visando determinar se o processamento de alguns produtos alimentares alterava a concentração de elementos essenciais para o consumo humano. Uma destas experiências pretendia estudar a concentração de zinco no feijão verde. A descrição da experiência disponível indica que 36 lotes de feijão foram aleatoriamente divididos em quatro grupos de 9 lotes e associados à medição do conteúdo de zinco nas seguintes condições: no primeiro grupo o feijão era observado antes de sofrer qualquer tratamento; no segundo grupo era observado antes do branqueamento; no terceiro grupo era observado após o branqueamento; e no quarto grupo era observado após o último passo do processamento. As medições obtidas estão indicadas na Tabela (e na *data frame zinco*).

Concentração de zinco			
Grupo 1	Grupo 2	Grupo 3	Grupo 4
2.23	3.71	2.53	5.46
2.20	4.67	2.87	5.19
2.44	3.45	2.83	5.51
2.11	2.73	2.33	4.82
2.30	2.58	2.19	6.63
1.72	1.85	1.80	2.39
1.78	1.81	1.75	2.09
2.36	2.32	1.83	2.27
2.91	2.50	1.97	2.39

- (a) Construa a tabela de Análise de Variância para o caso em estudo, admitindo a descrição da experiência indicada.
- (b) Explícite as hipóteses nula e alternativa para a Análise de Variância em causa. Efectue o teste de hipóteses ao nível de significância de 5%. Indique quais as conclusões do teste.
- (c) Caso tenha detectado efeitos de grupo sobre a concentração de zinco, indique quais os grupos cuja concentração média deve ser considerada diferente.
- (d) Calcule os resíduos e analise se existe evidência de violação dos pressupostos do modelo.