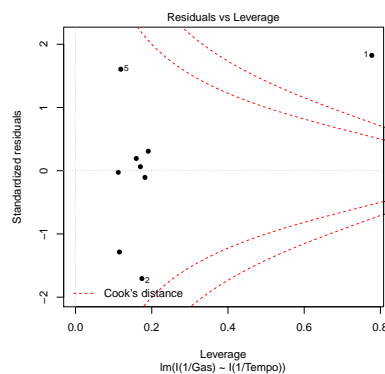




1. Mostre que a linearidade aparente no gráfico da direita, entre  $1/\text{Gas}$  e  $1/\text{Tempo}$ , sugere a adequação do modelo de Michaelis-Menten para a relação entre  $\text{Gas}$  e  $\text{Tempo}$ . Em particular, relacione os coeficientes  $b_0$  e  $b_1$  da recta de regressão na relação linearizada, com as constantes  $c$  e  $d$  no modelo de Michaelis-Menten.
2. Foi ajustado um modelo de regressão linear simples relacionando o recíproco de  $\text{Gas}$  e o recíproco de  $\text{Tempo}$ , com os seguintes resultados:

```
> summary(lm(I(1/Gas) ~ I(1/Tempo), data=Gas.exame))
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.039516   0.006094   6.484 0.000339
I(1/Tempo)   0.638783   0.030089  21.230 1.3e-07
---
Residual standard error: 0.01331 on 7 degrees of freedom
Multiple R-squared:  0.9847, Adjusted R-squared:  0.9825
F-statistic: 450.7 on 1 and 7 DF,  p-value: 1.295e-07
```

- (a) Indique o valor estimado *para a produção de gás* ao fim de 50 horas, a partir do modelo ajustado.
- (b) Construa um intervalo a 95% de confiança para a ordenada na origem da recta de regressão. Diga, com base nesse intervalo, se é admissível a hipótese de que a produção de gás assintótica (quando o tempo tende para infinito) seja 20 unidades (ml de gás por 200 mg de matéria seca), como sugerido pelo gráfico da esquerda.
- (c) Comente o seguinte gráfico de resíduos estandardizados contra valores do efeito alavanca. Em particular, comente o posicionamento da observação no canto superior direito. Indique possíveis causas e consequências desse posicionamento.



### III [5 v.]

Num estudo sobre alfices considerou-se útil modelar a condutividade eléctrica do solo a partir da sua composição química. Em 27 amostras de solo foram medidas, além da condutividade eléctrica (variável  $\text{CE}$ , em microsiemens por centímetro,  $\mu\text{S}/\text{cm}$ ), a concentração de cations de sódio (variável  $\text{Na}$ ), potássio (variável  $\text{K}$ ), cálcio (variável  $\text{Ca}$ ) e magnésio (variável  $\text{Mg}$ ), todas em  $\text{cmol}/\text{kg}$ . As correlações entre cada par de variáveis, são indicados de seguida.

	CE	Na	K	Ca	Mg
CE	1.00000	-0.08618	-0.05741	0.17778	0.55346
Na	-0.08618	1.00000	0.33811	0.20362	0.30437
K	-0.05741	0.33811	1.00000	0.08943	0.38161
Ca	0.17778	0.20362	0.08943	1.00000	0.64198
Mg	0.55346	0.30437	0.38161	0.64198	1.00000

Foi ajustado um modelo de regressão linear múltipla utilizando todas as concentrações de cátions como preditores. Obtiveram-se os seguintes resultados:

```
Call: lm(formula = CE ~ Na + K + Ca + Mg, data = Alfoces)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-815.76	446.45	-1.827	0.08126 .
Na	-102.21	84.52	-1.209	0.23935
K	-405.78	207.13	-1.959	0.06291 .
Ca	-57.68	29.61	-1.948	0.06432 .
Mg	668.17	145.34	4.597	0.00014 ***

---

Residual standard error: 68.07 on 22 degrees of freedom

Multiple R-squared: 0.5147, Adjusted R-squared: 0.4264

F-statistic: 5.832 on 4 and 22 DF, p-value: 0.002344

1. Teste o ajustamento global do modelo. Discuta a qualidade desse ajustamento. Na sua discussão, tenha também em conta os valores dos coeficientes de determinação usual e modificado.
2. Interprete, no contexto do problema sob estudo, o significado do valor  $-57.68$  na primeira coluna da tabela, indicando as unidades de medida do referido valor.
3. Um investigador afirma que quando aumenta a concentração de cátions de sódio, diminui a condutividade eléctrica média. Admitindo a validade do modelo, e dando o ónus da prova à hipótese do investigador, qual a conclusão a que se pode chegar com base nos dados ( $\alpha = 0.05$ )?
4. Foi afirmado que com base nos resultados do modelo acima ajustado, a única de entre as quatro variáveis preditoras que é importante na modelação da condutividade eléctrica é a variável **Mg**.
  - (a) Sem fazer quaisquer contas, diga se considera legítima esta afirmação.
  - (b) Independentemente da sua resposta na alínea anterior, compare formalmente o modelo acima apresentado com a regressão linear simples em que a variável resposta **CE** tem como único preditor a variável **Mg**. Comente.

#### IV [5 v.]

Um estudo com a variedade de oliveira Cobrançosa envolveu um ensaio em Mirandela. Diferentes génotipos (**clones**) foram analisados, pretendendo-se saber se diferiam em termos de rendimento (*kg/planta*). Uma vez que se desconfiava que os terrenos disponíveis para o ensaio podiam não ser homogéneos, decidiu-se utilizar um delineamento em blocos casualizados. Dividiu-se cada terreno em tantas parcelas quantos os génotipos (**clones**) estudados, associando cada clone a uma dessas parcelas, de forma aleatória. Foi ajustado um modelo ANOVA, tendo sido obtida a seguinte tabela-resumo:

```
> summary(aov(rend ~ clone + terrenos, data=oliveira))
              Df Sum Sq Mean Sq F value    Pr(>F)
clone          6  15.100  2.51671   3.8290 0.008077
terrenos       4   2.384  0.59599   0.9068 0.475746
Residuals    24  15.775  0.65728
```

1. Quantos clones (genótipos) foram estudados? Em quantos terrenos se efectuou a experiência? Justifique as suas respostas.
2. Identifique o tipo de modelo ANOVA ajustado e descreva-o completamente.
3. Teste formalmente ( $\alpha = 0.05$ ) se os clones têm todos igual rendimento médio.
4. Discuta brevemente se há razão para considerar que os terrenos têm efeito sobre o rendimento.
5. Independentemente da sua resposta na alínea anterior, considere um modelo ANOVA que apenas preveja efeitos de clone.
  - (a) Construa a tabela-resumo resultante de ajustar este modelo, justificando como obtém cada valor.
  - (b) As médias amostrais de dois dos clones analisados foram  $\bar{y}_i = 0.8602$  e  $\bar{y}_j = 0.7194$ . Diga se é admissível afirmar que estes rendimentos médios são significativamente diferentes, ao nível  $\alpha = 0.05$ .  
**NOTA:** Caso não tenha conseguido responder à alínea anterior (5a), utilize os valores constantes da tabela-resumo no enunciado geral deste grupo.

### V [3 v.]

1. Mostre que, para qualquer regressão linear múltipla, o coeficiente de determinação modificado  $R_{mod}^2$  é menor do que o coeficiente de determinação usual ( $R^2$ ).
2. Considere o teste  $F$  de ajustamento global numa regressão linear simples ajustada com base em  $n$  observações. Seja  $f_\alpha$  o valor de fronteira da região crítica do teste, ao nível de significância  $\alpha$ .
  - (a) Mostre que não se rejeita a hipótese nula do teste para valores do coeficiente de determinação amostral inferiores a  $\left[1 + \frac{n-2}{f_\alpha}\right]^{-1}$ .
  - (b) Calcule o valor de  $R^2$  para o qual a estatística do teste calculada fica na fronteira entre rejeição e não rejeição, quando se usa um nível de significância  $\alpha = 0.05$  e caso: (i)  $n = 7$ ; (ii)  $n = 15$ ; (iii)  $n = 30$  e (iv)  $n = 1000$ . Comente os seus resultados.
  - (c) Relacione os resultados obtidos na alínea anterior com o papel das hipóteses  $H_0$  e  $H_1$  em qualquer teste de hipóteses.