

INSTITUTO SUPERIOR DE AGRONOMIA  
**ESTATÍSTICA E DELINEAMENTO EXPERIMENTAL – 2024-25**  
**31 outubro 2024** **Primeiro Teste** **Duração: 1h30**

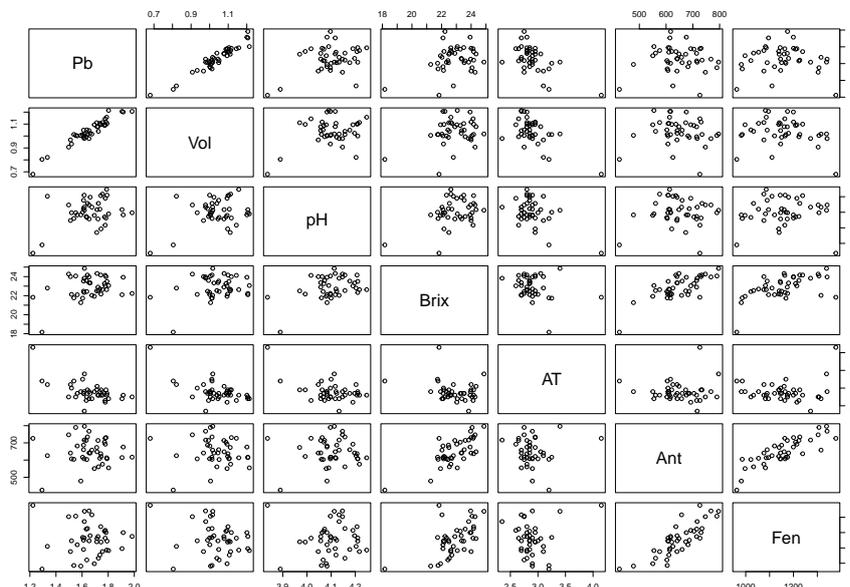
Um estudo sobre a variedade de videira Tinta Roriz em Vila Nova de Foz Coa, pretende modelar o teor de fenóis total no mosto. Trata-se de uma característica cuja análise laboratorial envolve muitos procedimentos. Seria desejável poder modelá-la em função de outras variáveis com medição mais expedita. Foi sugerido modelar o teor de fenóis total (variável **Fen**, em mg/l), a partir do peso do bago (variável **Pb**, em g), do volume do mosto do bago (variável **Vol**, em ml), do pH do mosto (variável **pH**), do teor de sólidos solúveis do mosto (variável **Brix**, grau brix), da acidez total do mosto (variável **AT**, em g/l de ácido tartárico) e do teor antocianinas do mosto (variável **Ant**, em mg/l). Dispõem-se de valores de todas as variáveis para 42 videiras escolhidas ao acaso (**dadosrz**). Eis os seguintes indicadores estatísticos, bem como a matriz de correlações e nuvens de pontos para todos os pares de variáveis.

Variável	Média	Variância
Pb	1.659	0.0229
Vol	1.039	0.0114
pH	4.103	0.0069
Brix	22.90	1.3615
AT	2.898	0.0751
Ant	649.9	5878.6708
Fen	1153.9	10470.8364

```
> round(cor(dadosrz),4)
```

	Pb	Vol	pH	Brix	AT	Ant	Fen
Pb	1.0000	0.9584	0.3642	0.3206	-0.5944	0.0271	-0.0516
Vol	0.9584	1.0000	0.3350	0.2838	-0.5938	-0.0566	-0.2092
pH	0.3642	0.3350	1.0000	0.4057	-0.4979	0.0406	0.0272
Brix	0.3206	0.2838	0.4057	1.0000	-0.2546	0.7365	0.5509
AT	-0.5944	-0.5938	-0.4979	-0.2546	1.0000	-0.0011	0.1104
Ant	0.0271	-0.0566	0.0406	0.7365	-0.0011	1.0000	0.8270
Fen	-0.0516	-0.2092	0.0272	0.5509	0.1104	0.8270	1.0000

```
> plot(dadosrz)
```



I [15.5 valores]

Foi ajustada a regressão linear múltipla do teor de fenóis total (variável **Fen**, em mg/l) sobre as restantes variáveis observadas. Eis os resultados obtidos com o ajustamento deste modelo, bem como a matriz de (co-) variâncias estimadas dos estimadores dos parâmetros do modelo:

```
> rz.lm<-lm(Fen~., data=dadosrz)
> summary(rz.lm)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  236.2844   515.5464   0.458 0.649557
Pb            A         B         3.963 0.000348
Vol          -1153.9381  275.9164  -4.182 0.000184
pH            32.8033   122.9709   0.267 0.791222
Brix           1.6183    12.8047   0.126 0.900152
AT            28.1253    37.8969   0.742 0.462946
Ant           0.9531     0.1817   5.246 7.62e-06
---
Residual standard error: 48.99 on 35 degrees of freedom
Multiple R-squared:  0.8043, Adjusted R-squared:  C
F-statistic: 23.98 on 6 and 35 DF,  p-value: 4.788e-11

> round(vcov(rz.lm), 2)
      (Intercept)      Pb      Vol      pH      Brix      AT      Ant
(Intercept)  265788.12  5405.32 -21630.81 -55594.84 1066.85 -12705.45 -18.04
Pb            5405.32  35978.63 -49075.49 -4295.99  410.56  259.16 -10.21
Vol          -21630.81 -49075.49  76129.87  7245.42 -996.39  1510.37  19.48
pH           -55594.84 -4295.99  7245.42 15121.84 -763.43  1716.75  8.70
Brix          1066.85  410.56  -996.39  -763.43  163.96  -33.49  -1.90
AT           -12705.45  259.16  1510.37  1716.75 -33.49  1436.18  0.41
Ant           -18.04  -10.21  19.48  8.70  -1.90  0.41  0.03
>
```

1. No *output* do comando *summary* do R identifique e calcule, justificando, os valores em falta A, B e C.
2. Efectue em pormenor o teste *F* de ajustamento global, explicando o seu significado, e interprete o valor do coeficiente de determinação  $R^2$ .
3. Será admissível afirmar que quando o teor de antocianas aumenta 1 mg/l, mantendo os resultantes preditores constantes, o teor de fenóis total aumenta, em média, 1 mg/l? Responda através do teste de hipóteses adequado explicando, em pormenor, todos os seus passos.
4. Considere o modelo de regressão linear múltipla ajustado:
  - (a) Descreva, em pormenor, esse modelo em notação matricial;
  - (b) Prove que a distribuição das observações  $\vec{Y}$  da variável resposta é:

$$\vec{Y} \sim \mathcal{N}_n(\mathbf{X}\vec{\beta}, \sigma^2 \mathbf{I}_n).$$

5. Num algoritmo de exclusão sequencial com base em testes T aos parâmetros do modelo (para  $\alpha=0.10$ ), qual a primeiro preditor a excluir do modelo? Justifique a sua resposta.
6. Após a aplicação de um algoritmo de exclusão sequencial, chegou-se a um submdelo final com 3 preditores (Pb, Vol e Ant). O Teste F parcial de comparação do modelo completo (com 6 preditores ) com esse seu submodelo (com 3 preditores), forneceu os seguintes resultados:

```
> anova(rz.lmsub,rz.lm)
...
  Res.Df  RSS Df Sum of Sq      F Pr(>F)
1      38 85421
2      35 84002  3   1418.9 0.1971 0.8977
```

- A qualidade do ajustamento do modelo completo e do seu submodelo difere significativamente ( $\alpha=0.05$ )? Descreva, em pormenor, o teste efectuado para responder à questão.
- Qual a estimativa da variância dos erros aleatórios do submodelo com os 3 preditores (Pb, Vol e Ant)? Para esse submodelo, determine o valor das três somas de quadrados associadas. Apresente os cálculos efetuados.
- Dado um Modelo de Regressão Linear Múltipla com  $p$  preditores, e um seu submodelo com apenas  $k < p$  preditores, a estatística do teste para comparar modelo completo e submodelo (estatística do teste  $F$  parcial) é dada por:

$$F = \frac{\frac{SQRE_S - SQRE_C}{p-k}}{\frac{SQRE_C}{n-(p+1)}}$$

onde  $SQRE_C$  e  $SQRE_S$  indicam as Somas de Quadrados Residuais, respectivamente, do modelo completo e do submodelo. Prove que esta estatística pode ser escrita numa forma alternativa, envolvendo os Coeficientes de Determinação do modelo completo ( $R_C^2$ ) e do submodelo ( $R_S^2$ ), isto é:

$$F = \frac{n - (p + 1)}{p - k} \cdot \frac{R_C^2 - R_S^2}{1 - R_C^2}.$$

## II [4.5 valores]

Seguidamente o responsável pela análise dos dados decide modelar o teor de fenóis total (variável **Fen**, em mg/l) usando uma regressão linear simples com o preditor teor de antocianas (variável **Ant**, em mg/l).

- Qual o coeficiente de determinação da regressão linear simples do teor de fenóis total sobre o teor de antocianas?
- Após a observação da nuvem de pontos entre as duas variáveis, o analista decidiu ajustar uma regressão linear simples usando a transformação logarítmica da variável resposta. Eis os resultados obtidos.

```
> summary(lm(log(Fen)~Ant, dadosrz))
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.4237374  0.0659109   97.461 < 2e-16
Ant          0.0009631  0.0001007    9.561 6.93e-12
---
Residual standard error: 0.04945 on 40 degrees of freedom
Multiple R-squared:  0.6956, Adjusted R-squared:  0.688
F-statistic: 91.42 on 1 and 40 DF,  p-value: 6.927e-12
```

```
> confint(lm(log(Fen)~Ant, dadosrz))
      2.5 %      97.5 %
(Intercept) 6.2905265791 6.556948248
Ant         0.0007595435      ?
```

- Calcule um intervalo a 95% de confiança para o declive da reta populacional. Interprete o seu significado.
- Deduza a relação não linear entre o teor de fenóis total e o teor de antocianas que corresponde à regressão linear ajustada. Escreva a equação da curva ajustada.