

Estatística e Análise de Dados em Zootecnia 2025/2026

Aplicação 3

Elsa Gonçalves
Secção de Matemática, DCEB, ISA-Ulisboa

Considere os seguintes dados referentes a 8 variáveis medidas em suínos.

```
data porcos;
input Area Gordurasubcut Peso Rendimento Gordurarenalpel Comprimento LarguraAnca Musculo
;
datalines;
8.8      3.5      13.9      50      200      72      25      58
9.2      3.0      15.0      47      170      68      24      57
8.6      3.5      13.4      48      180      73      23      57
8.7      4.0      14.2      48      150      74      25      56
8.5      3.5      13.0      51      160      69      22      56
8.2      4.0      14.8      49      190      70      21      56
8.0      4.5      12.8      46      210      71      27      55
7.9      5.0      15.1      48      200      76      23      55
7.6      4.5      13.6      47      190      65      20      55
7.5      5.0      14.1      50      210      66      28      55
7.6      4.5      13.7      49      250      65      22      54
7.4      4.0      12.2      48      280      74      21      54
7.3      6.0      14.7      49      230      68      20      54
7.0      5.5      13.1      52      280      62      26      53
7.5      5.0      14.0      50      250      72      21      53
6.8      6.0      12.9      43      300      75      29      53
6.5      6.5      14.2      53      310      71      23      52
6.8      7.0      12.8      45      260      65      22      52
7.0      6.5      13.5      47      290      68      20      51
6.8      7.0      12.9      48      330      69      21      50
;
run;
```

Eis os seguintes indicadores estatísticos, bem como a matriz de correlações.

a) Comente os resultados obtidos.

8 Variables: Area Gordurasubcut Peso Rendimento Gordurarenalpel Comprimento LarguraAnca Musculo

b) A partir da matriz de correlações entre as variáveis sob estudo, diga qual o melhor modelo de regressão linear simples para prever a percentagem de músculo na carcaça? Justifique.

c) Para a regressão linear simples que escolheu, determine o coeficiente de determinação e realize a correspondente decomposição da soma dos quadrados total.

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Area	20	7.68500	0.77750	153.70000	6.50000	9.20000
Gordurasubcut	20	4.92500	1.23837	98.50000	3.00000	7.00000
Peso	20	13.69500	0.82045	273.90000	12.20000	15.10000
Rendimento	20	48.40000	2.32605	968.00000	43.00000	53.00000
Gordurarenalpel	20	232.00000	53.66563	4640	150.00000	330.00000
Comprimento	20	69.65000	3.82891	1393	62.00000	76.00000
LarguraAnca	20	23.15000	2.71981	463.00000	20.00000	29.00000
Musculo	20	54.30000	2.10513	1086	50.00000	58.00000

Pearson Correlation Coefficients, N = 20								
	Area	Gordurasubcut	Peso	Rendimento	Gordurarenalpel	Comprimento	LarguraAnca	Musculo
Area	1.00000	-0.90591	0.37364	0.02968	-0.88474	0.25096	0.10068	0.90649
Gordurasubcut	-0.90591	1.00000	-0.21019	-0.12607	0.79433	-0.22783	-0.11368	-0.90952
Peso	0.37364	-0.21019	1.00000	0.26034	-0.43726	0.09324	-0.09871	0.36964
Rendimento	0.02968	-0.12607	0.26034	1.00000	-0.02783	-0.21983	-0.15141	0.08169
Gordurarenalpel	-0.88474	0.79433	-0.43726	-0.02783	1.00000	-0.11424	-0.07068	-0.88144
Comprimento	0.25096	-0.22783	0.09324	-0.21983	-0.11424	1.00000	0.14682	0.23572
LarguraAnca	0.10068	-0.11368	-0.09871	-0.15141	-0.07068	0.14682	1.00000	0.23992
Musculo	0.90649	-0.90952	0.36964	0.08169	-0.88144	0.23572	0.23992	1.00000

Nota: o resultado obtido com o ajustamento do melhor modelo de regressão linear simples é:

```
proc reg data=porcos;  
  model Musculo = Gordurasubcut;  
run;
```

Root MSE	0.89899	R-Square	0.8272
Dependent Mean	54.30000	Adj R-Sq	0.8176
Coeff Var	1.65560		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	61.91463	0.84451	73.31	<.0001
Gordurasubcut	1	-1.54612	0.16654	-9.28	<.0001

O resultado do ajustamento da regressão linear múltipla da percentagem de músculo na carcaça sobre todas as restantes variáveis é o seguinte:

```
proc reg data=porcos ;
model Musculo = Area Gordurasubcut Peso
Rendimento Gordurarenalpel Comprimento
LarguraAnca ;
RUN;
```

d) Escreva a equação do hiperlano ajustado.

e) Use o valor do coeficiente de determinação obtido para determinar a decomposição da soma dos quadrados total. Comente os resultados.

f) Compare o coeficiente estimado da variável preditora **“Gordurasubcut”** com o valor correspondente obtido no modelo anterior. Comente.

Root MSE	0.71981	R-Square	0.9262
Dependent Mean	54.30000	Adj R-Sq	0.8831
Coeff Var	1.32561		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	54.11487	9.27061	5.84	<.0001
Area	1	0.06200	0.70162	0.09	0.9310
Gordurasubcut	1	-0.93861	0.36030	-2.61	0.0230
Peso	1	0.24489	0.26196	0.93	0.3683
Rendimento	1	0.00623	0.08323	0.07	0.9416
Gordurarenalpel	1	-0.01436	0.00714	-2.01	0.0673
Comprimento	1	0.01774	0.04832	0.37	0.7199
LarguraAnca	1	0.11974	0.06255	1.91	0.0797