

Estatística e Análise de Dados em Zootecnia 2025/2026

Aplicação 2

Elsa Gonçalves
Secção de Matemática, DCEB, ISA-Ulisboa

Considere o conjunto de dados **Animals** onde se listam pesos médios dos cérebros (em *g*) e dos corpos (em *kg*) para 28 espécies de animais terrestres.

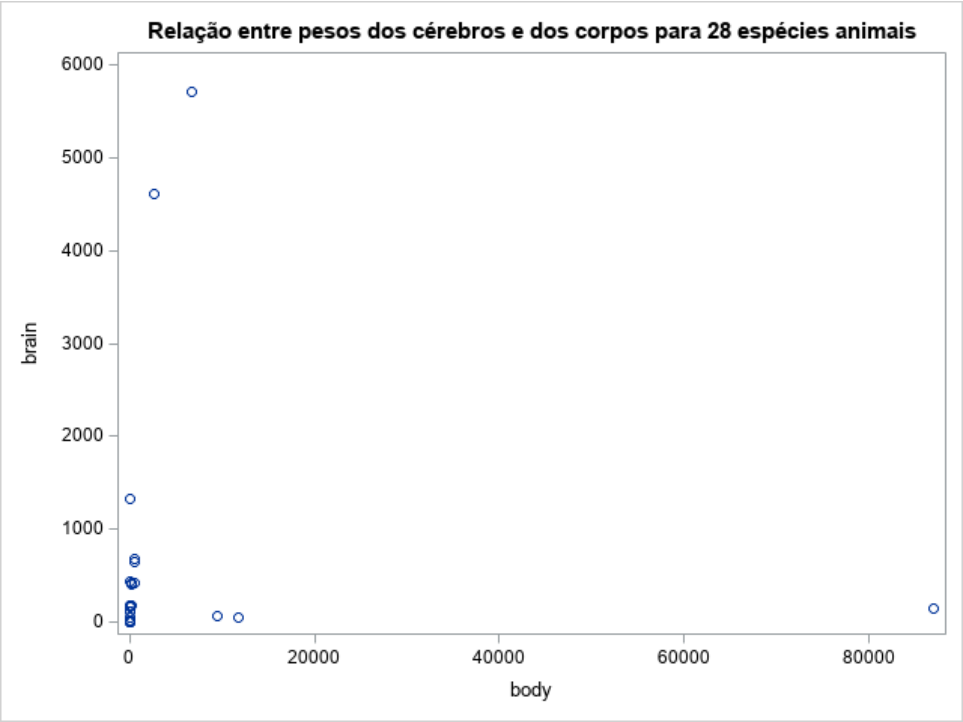
Pretende-se estudar uma relação entre pesos do cérebro (variável resposta *y*) e pesos do corpo (variável preditora *x*).

```
data Animals;  
input animals$ body  brain;  
datalines;  
Mountainbeaver      1.350      8.1  
Cow                  465.000    423.0  
Greywolf             36.330    119.5  
Goat                 27.660    115.0  
Guineapig            1.040      5.5  
Dipliodocus          11700.000    50.0  
Asianelephant        2547.000  4603.0  
Donkey               187.100    419.0  
Horse                521.000    655.0  
Potarmonkey          10.000    115.0  
Cat                  3.300     25.6  
Giraffe              529.000    680.0  
Gorilla              207.000    406.0  
Human                62.000   1320.0  
Africanelephant      6654.000  5712.0  
Triceratops          9400.000    70.0  
Rhesusmonkey         6.800    179.0  
Kangaroo             35.000     56.0  
Goldenhamster        0.120      1.0  
Mouse                0.023      0.4  
Rabbit               2.500     12.1  
Sheep                55.500    175.0  
Jaguar              100.000    157.0  
Chimpanzee           52.160    440.0  
Rat                  0.280      1.9  
Brachiosaurus        87000.000   154.5  
Mole                 0.122      3.0  
Pig                 192.000    180.0  
;  
run;
```

a) Comente a nuvem de pontos de pesos do corpo (eixo horizontal) e pesos do cérebro (eixo vertical), bem como o coeficiente de correlação.

```
proc sgplot data=Animals;
    scatter x=body y=brain;
    title "Relação entre pesos dos cérebros
e dos corpos para 28 espécies animais";
run;

proc corr data=Animals noprob;
    VAR body brain;
    RUN;
```



The CORR Procedure

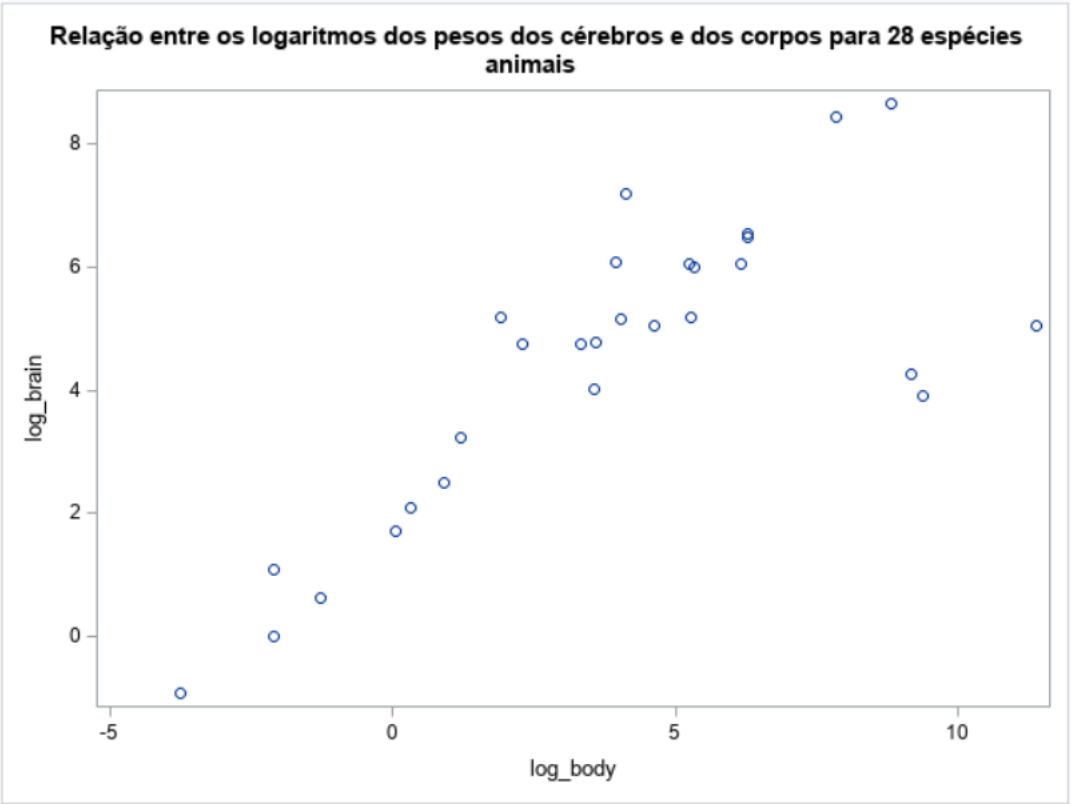
2 Variables: body brain

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
body	28	4278	16480	119796	0.02300	87000
brain	28	574.52143	1335	16087	0.40000	5712

Pearson Correlation Coefficients, N = 28		
	body	brain
body	1.00000	-0.00534
brain	-0.00534	1.00000

b) Foi sugerida uma transformação logarítmica de ambas as variáveis. Comente os resultados obtidos.

```
data logAnimals;  
  set Animals;  
  log_brain = log(brain);  
  log_body = log(body);  
  
run;  
  
proc sgplot data=logAnimals;  
  scatter x=log_body y=log_brain;  
  title "Relação entre os logaritmos dos pesos dos cérebros e dos corpos para 28 espécies animais";  
  
run;
```



```
proc corr data=logAnimals noprob;  
  VAR log_body log_brain;  
  
RUN;
```

2 Variables: log_body log_brain

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
log_body	28	3.77131	3.77066	105.59656	-3.77226	11.37366
log_brain	28	4.42545	2.39928	123.91248	-0.91629	8.65032

Pearson Correlation Coefficients, N = 28		
	log_body	log_brain
log_body	1.00000	0.77949
log_brain	0.77949	1.00000

c) Considere uma relação linear entre $\ln(y)$ e $\ln(x)$. Explícite a relação de base correspondente entre as variáveis originais (não logaritmizadas). Comente.

d) Calcule o coeficiente de determinação associado à relação entre $\ln(x)$ e $\ln(y)$. Interprete o valor obtido. Como se explica que o Coeficiente de Determinação não seja particularmente elevado, sendo evidente a partir da nuvem de pontos que existe uma boa relação entre log-peso do corpo e log-peso do cérebro para a generalidade das espécies?

e) Ajuste a recta de regressão do log-peso do cérebro sobre o log-peso do corpo.

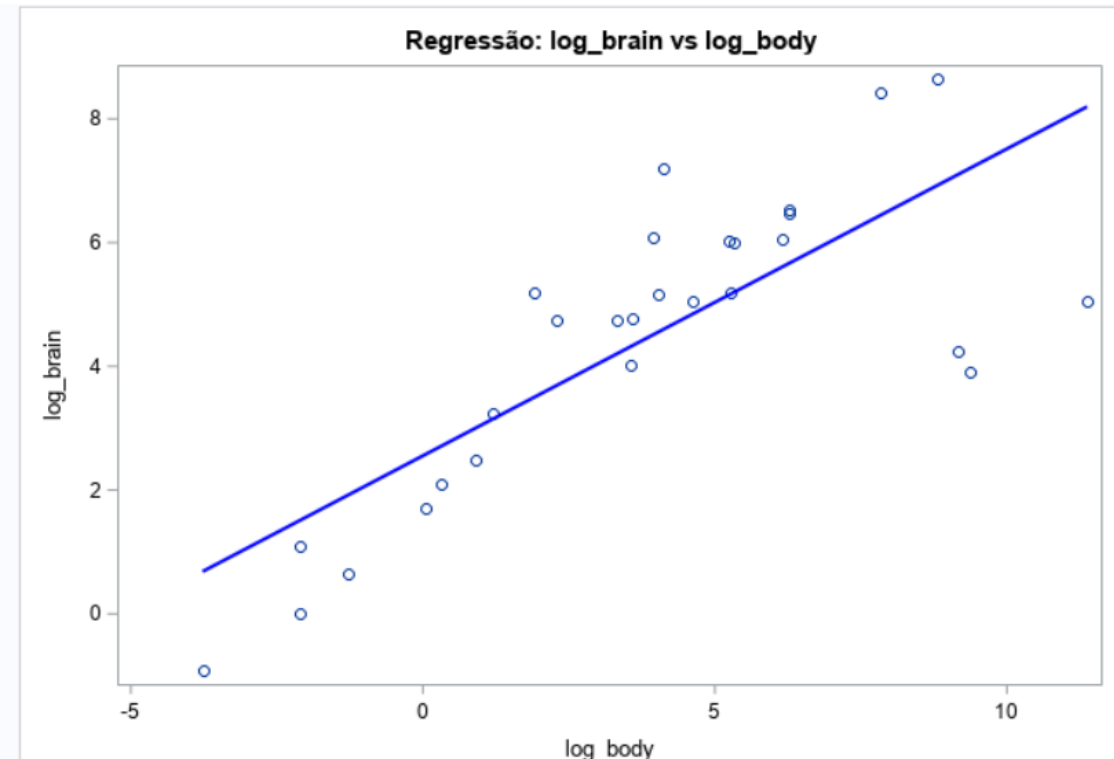
```

proc reg data=logAnimals;
  model log_brain = log_body ;
  output out=resultados_reg p=valores_ajustados;
RUN;
proc sgplot data=resultados_reg;
  scatter x=log_body y=log_brain;
  series x=log_body y=valores_ajustados /
lineattrs=(color=blue thickness=2); /* recta ajustada */
  title "Regressão: log_brain vs log_body";
run;

```

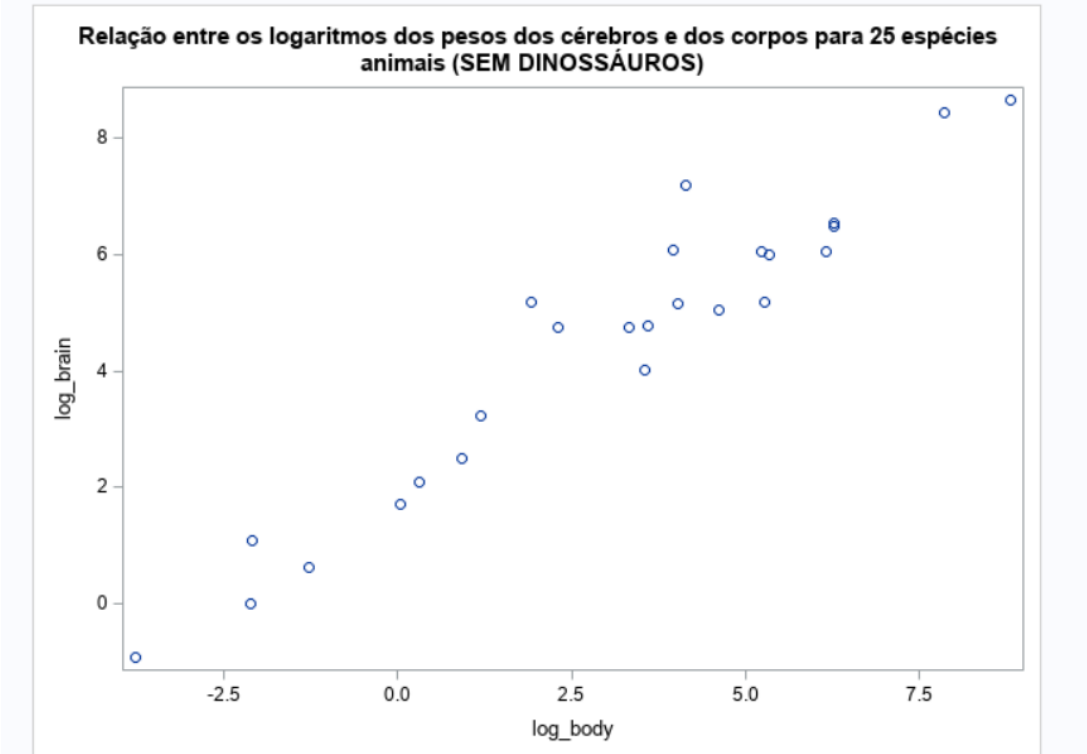
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.55490	0.41314	6.18	<.0001
log_body	1	0.49599	0.07817	6.35	<.0001

f) Considere agora a estimativa para o declive da recta, $b_1=0.49599$. Qual o significado biológico deste valor, quer na relação entre variáveis logaritmizadas, quer na relação entre as variáveis originais (não logaritmizadas)?



g) Os três pontos que se destacam na parte inferior da nuvem de pontos são referentes a dinossauros (observações 6, 16 e 26). Comente agora o resultado do ajustamento da recta de regressão de log-peso do cérebro sobre log-peso do corpo considerando apenas os dados (logaritmizados) referentes a espécies que não sejam dinossauros. Em particular, como se explica a elevação considerável no valor do coeficiente de determinação?

```
data logAnimalsSEMDINOSSAUROS;  
  set AnimalsSEMDINOSSAUROS;  
  log_brain = log(brain);  
  log_body = log(body);  
  
run;  
  
proc sgplot data=logAnimalsSEMDINOSSAUROS;  
  scatter x=log_body y=log_brain;  
  title "Relação entre os logaritmos dos  
  pesos dos cérebros e dos corpos para 25  
  espécies animais (SEM DINOSSÁUROS)";  
run;
```



Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
log_body	25	3.02828	3.24072	75.70708	-3.77226	8.80297
log_brain	25	4.42847	2.53931	110.71177	-0.91629	8.65032

Pearson Correlation Coefficients, N = 25		
	log_body	log_brain
log_body	1.00000	0.96005
log_brain	0.96005	1.00000

```
proc reg data=logAnimalsSEMDINOSSAUROS;
    model log_brain = log_body ;
    output out=resultados_regSEMDINOSSAUROS
p=valores_ajustadosSEMDINOSSAUROS;
RUN;
```

```
proc sgplot
data=resultados_regSEMDINOSSAUROS;
    scatter x=log_body y=log_brain;
    series x=log_body
y=valores_ajustadosSEMDINOSSAUROS /
lineattrs=(color=blue thickness=2); /*
recta ajustada */
    title "Regressão: log_brain vs
log_body";
run;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.15041	0.20060	10.72	<.0001
log_body	1	0.75226	0.04572	16.45	<.0001

