Archosaurs: Linear Regression Example

jdt

October 2021

Contents

Introduction

Reptiles of the subclass Archosauria include the now extinct dinosaurs and pterosaurs (flying or gliding dinosaurs) as well as birds and crocodiles. For any subclass of animals, although individual species vary greatly in size, it has been found that there is a relatively consistent relationship between body size and brain size.

In "Relative Brain Size and Behavior of Archosaurian Reptiles" (Annual Review of Ecology and Systematics 1977, by James A. Hopson) the author cites a widely observed power law relationship:

$$Brain\ wgt = k \times (Body\ wgt)^{2/3}$$

Using logarithms, this can be transformed into a linear relationship:

$$\log(Brain\ wgt) = \log(k) + 0.67 \times \log(Body\ wgt)$$

It should be noted we would expect "smart" animals, such as the mammals, to differ from "dumb" animals, such as the reptiles, by having a linear relationship with a larger intercept. In other words, "k" would be larger for mammals than reptiles. Also, it does not matter which logarithmic transformation we use (common logs, natural logs or any others), but we will use the natural log.

Power law models are based on the relationship $y = ax^b$, which can be restated as a linear relationship $\log(y) = \log(a) + b \times \log(x)$ using a logarithmic transformation. Power law models represent scale invariant relationships, that is, relationships thought to hold in the same manner for very large and very small observed values. For example, in biology the relationship between habitat area and number of species fits a power law model quite well (see Case 1 and Exercise 22 from Chapter 8 of The Statistical Sleuth, Second Edition, Ramsey and Schafer, 2002). A Web search on power law models will result in dozens of examples, taken mostly from physics and engineering. This is a typical example of a power law model. Although reptiles vary greatly in size, we would expect about the same proportion of their mass to be devoted to brain size. We expect the same for mammals, except we expect the brains of mammals to be bigger than the reptile of the same overall mass, hence the greater value for "k".

The Task

Determine whether a power law model fits the data. In other words:

- Does it make sense to take logarithmic transformations of body weight and brain weight in order to perform a linear regression?
- Are the assumptions of the linear regression met?
- Is our final model a good fit to the data?
- Does the slope of 2/3 actually seem to match the data?

Other questions of interest:

- How do we make predictions in a power law model?
- What do large positive and negative residuals mean in this context?

R

Read data from SAS input file

```r
# this data came from SASHELP.CARS
brain <- read.csv('archosaur.csv', header = TRUE)
summary(brain)
```

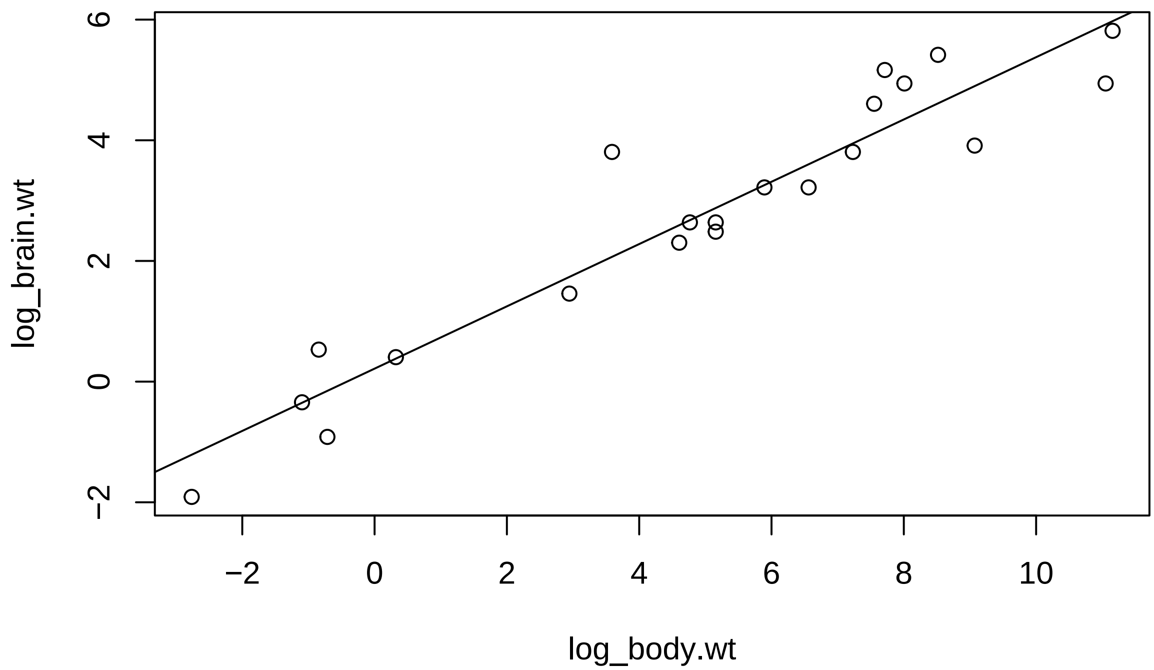
```
##      Type             Details           Body.Weight        Brain.Weight
##  Length:21          Length:21          Min.   :    0.06   Min.   :  0.148
##  Class :character   Class :character   1st Qu.:   19.00   1st Qu.:  4.300
##  Mode  :character   Mode  :character   Median :  173.60   Median : 25.000
##                                        Mean   : 7472.37   Mean   : 64.941
##                                        3rd Qu.: 2236.00   3rd Qu.:100.000
##                                        Max.   :70000.00   Max.   :335.000
```

Transform Data

```r
log_body.wt = log(brain$Body.Weight)
log_brain.wt = log(brain$Brain.Weight)
```

Graph of data

```r
plot(log_brain.wt~log_body.wt)
abline(lm(log_brain.wt~log_body.wt))
```

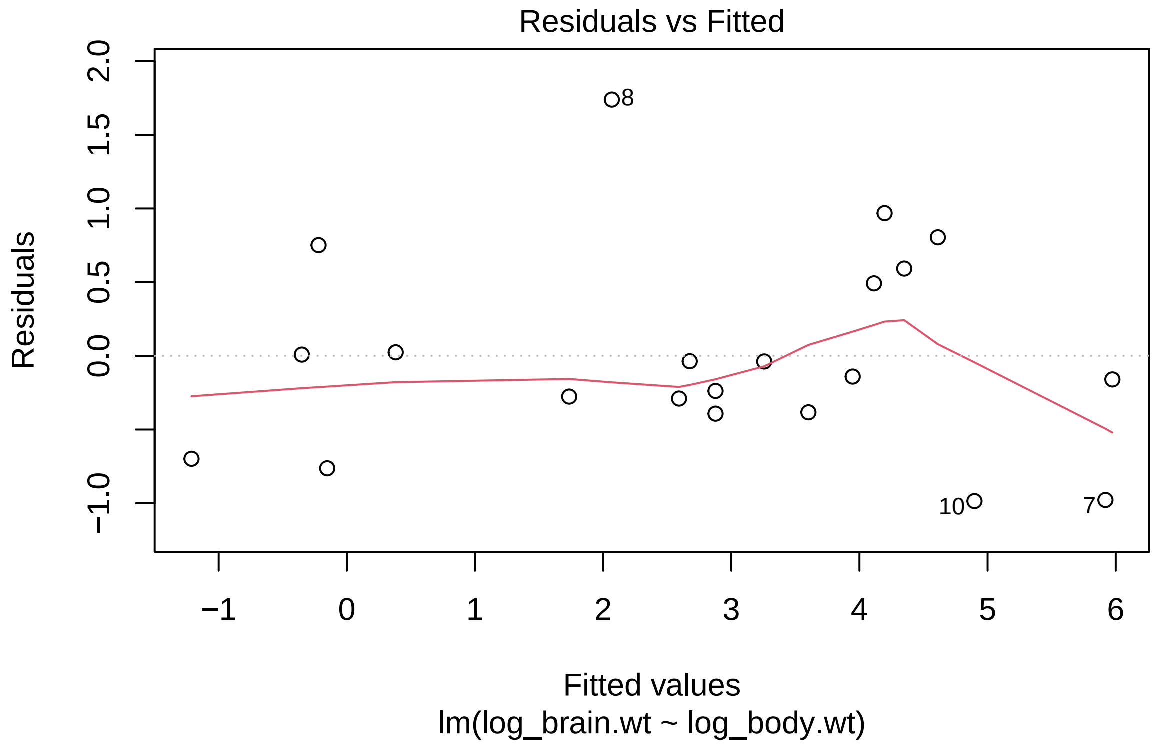


Results

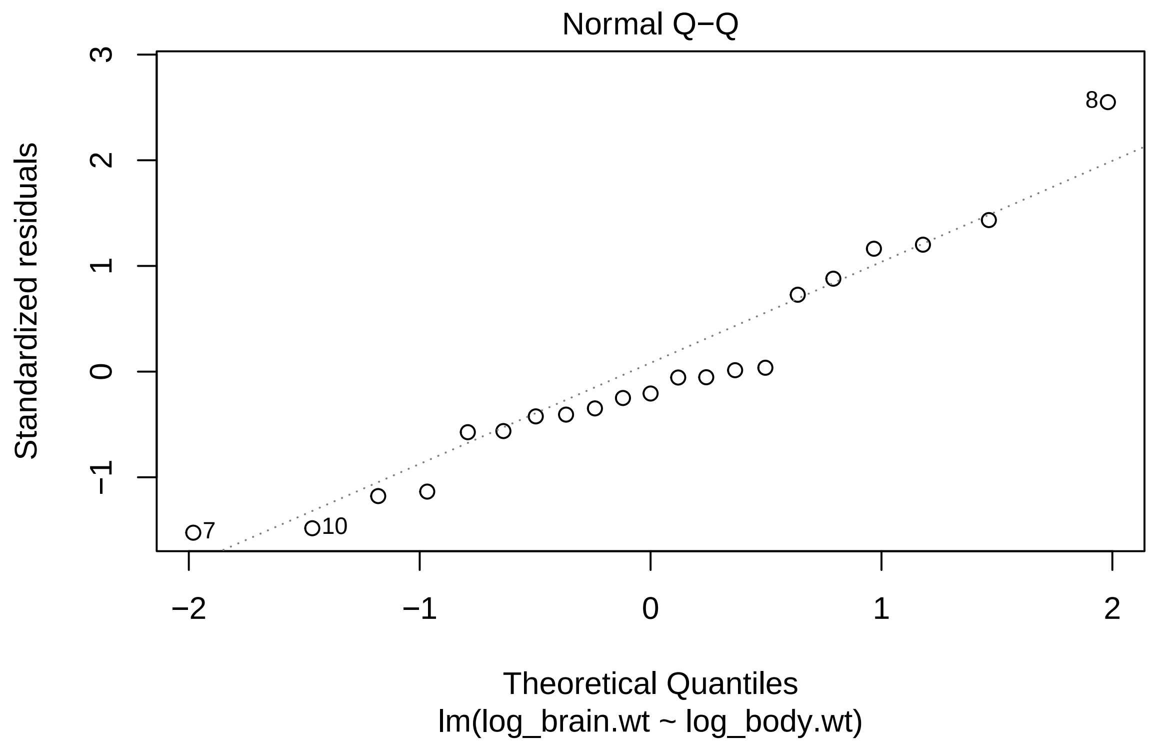
```r
result<-lm(log_brain.wt~log_body.wt)
summary(result)
```

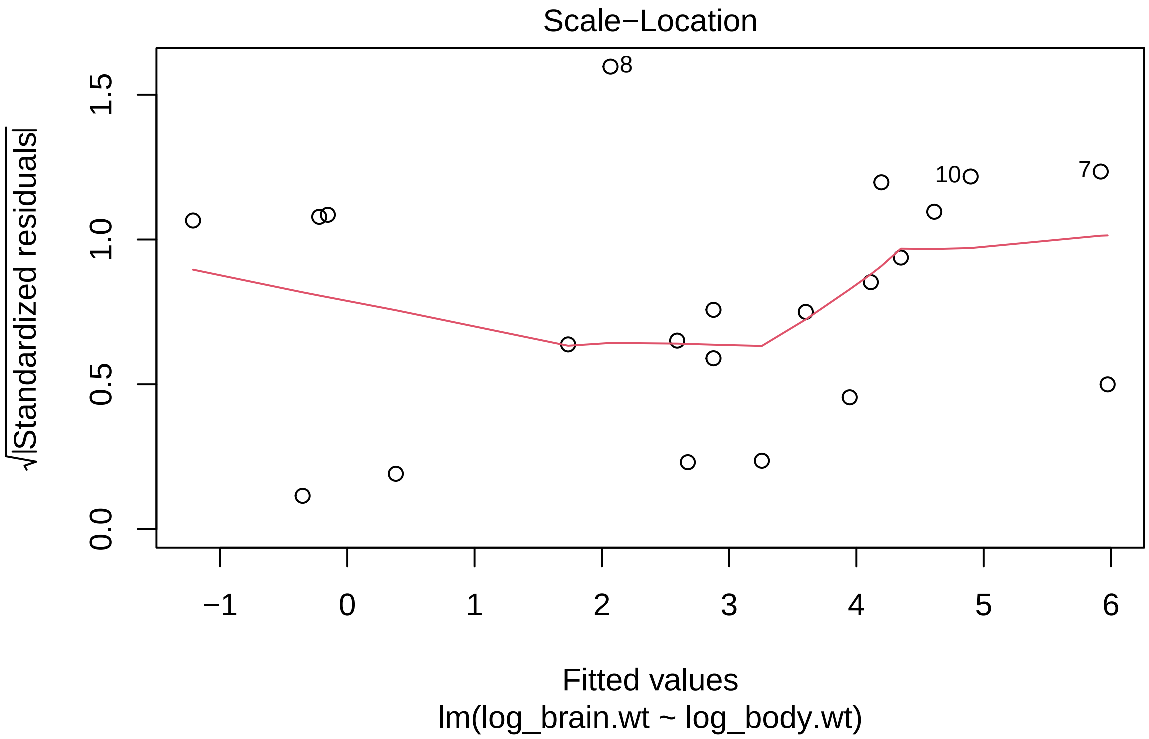
```
##
## Call:
## lm(formula = log_brain.wt ~ log_body.wt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9856 -0.3831 -0.1405  0.4919  1.7389
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.21507    0.24518   0.877    0.391
## log_body.wt  0.51621    0.03874  13.324 4.34e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7008 on 19 degrees of freedom
## Multiple R-squared:  0.9033, Adjusted R-squared:  0.8982
## F-statistic: 177.5 on 1 and 19 DF,  p-value: 4.341e-11
```

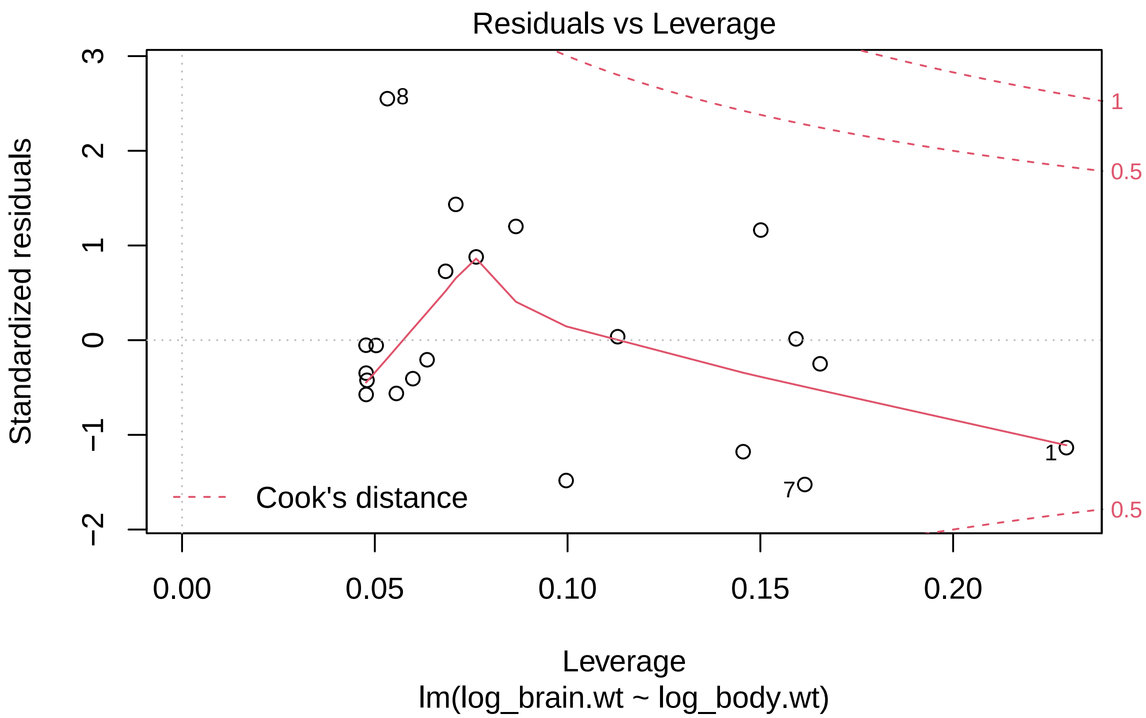
```
plot(result)
```



Residuals vs Fitted
Residuals
2.0
1.5
1.0
0.5
0.0
−1.0
8
10
7
−1
0
1
2
3
4
5
6
Fitted values
lm(log_brain.wt ~ log_body.wt)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log_brain.wt ~ log_body.wt)
−2
−1
0
1
2
3
7
10
8


Scale−Location
√|Standardized residuals|
Fitted values
lm(log_brain.wt ~ log_body.wt)
0.0
0.5
1.0
1.5
−1
0
1
2
3
4
5
6
8
10
7


Residuals vs Leverage
Standardized residuals
Leverage
lm(log_brain.wt ~ log_body.wt)
Cook's distance
1
0.5
0.5
8
7
1
0.00
0.05
0.10
0.15
0.20
3
2
1
0
−1
−2

SAS

Code

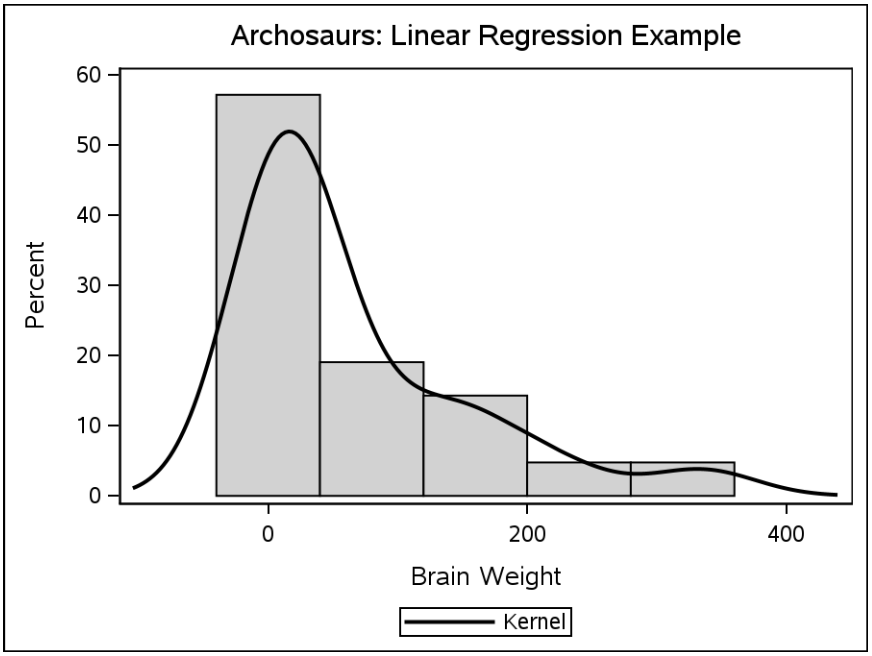
```
options center nodate pagesize=120 ls=80;
libname ldata '/home/jacktubbs/my_shared_file_links/jacktubbs/LaTeX/Class';

/* Simplified LaTeX output that uses plain LaTeX tables  */
ods latex path='/home/jacktubbs/my_shared_file_links/jacktubbs/LaTeX/clean'
 file='archosaur.tex' style=journal
stylesheet="sas.sty"(url="sas");

/*
http://support.sas.com/rnd/base/ods/odsmarkup/latex.html
*/
ods graphics / reset width=4.5in  outputfmt=png
  antialias=on;

/*
This data set is from the JMP Case Study Library
*/

title "Archosaurs: Linear Regression Example";
data brain; set ldata.archosaur;
run;

proc contents data=brain short;
run;

data brain; set brain;
log_brain_wt = log(brain_weight);
log_body_wt = log(body_weight);
run;

proc sgplot data=brain;
histogram brain_weight;
density brain_weight/ type=kernel;
run;

proc sgplot data=brain;
histogram log_brain_wt;
density log_brain_wt/ type=kernel;
run;


proc sgplot data=brain;
scatter y=log_brain_wt x=log_body_wt;
reg y=log_brain_wt x=log_body_wt;
run;
```

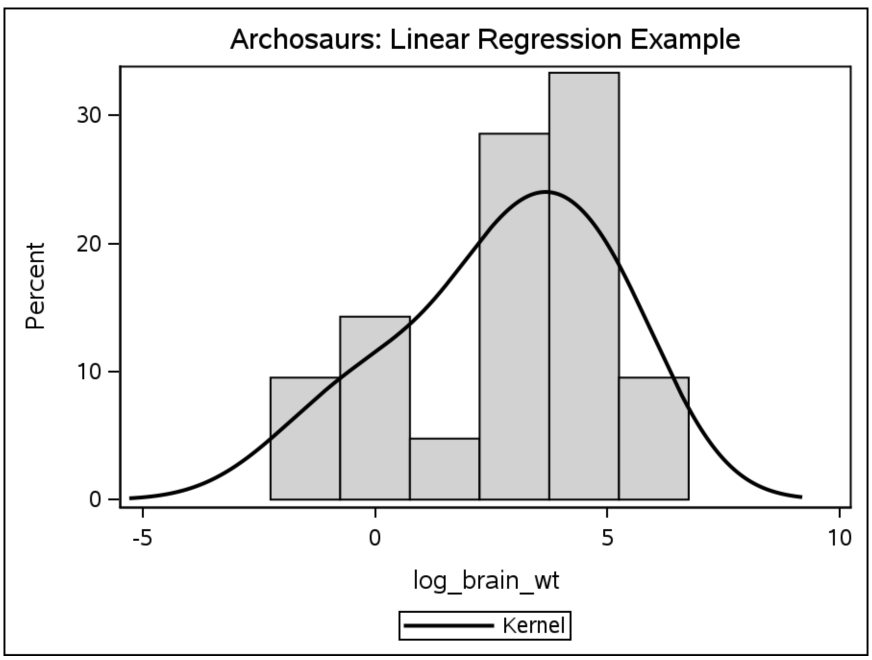
```
proc reg data=brain plots=diagnostics;
model log_brain_wt = log_body_wt;
run;

proc transreg data=brain;
model BoxCox(brain_weight / convenient lambda=-2 to 2 by 0.05) =
         monotone(body_weight);
output out=new;
run;
title3 'Modified brain weight measurement';

proc sgplot data=new;
histogram tbrain_weight;
density tbrain_weight;
run;


ods latex close;

/* Stream a CSV representation of new_bwgt directly to the user's browser. */
/*
proc export data=new_heart
            outfile=_dataout
            dbms=csv replace;
run;

%let _DATAOUT_MIME_TYPE=text/csv;
%let _DATAOUT_NAME=heart.csv;
```

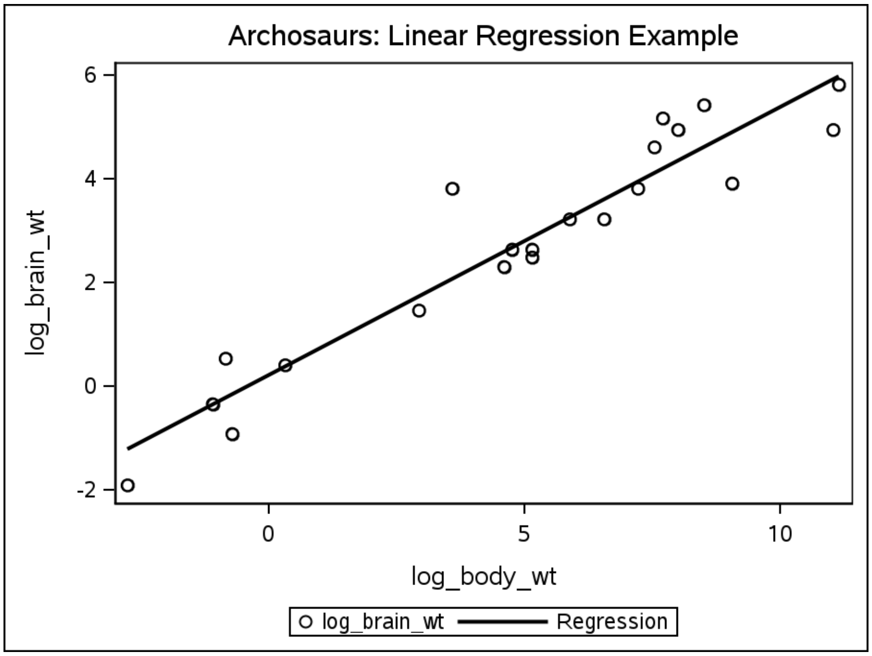
Archosaurs: Linear Regression Example

The CONTENTS Procedure

Alphabetic List of Variables for WORK.BRAIN
Body_Weight Brain_Weight Details Type


Archosaurs: Linear Regression Example
60
50
40
30
20
10
0
Percent
0
200
400
Brain Weight
Kernel


Archosaurs: Linear Regression Example
30
20
10
0
Percent
-5
0
5
10
log_brain_wt
Kernel


Archosaurs: Linear Regression Example
6
4
2
0
-2
log_brain_wt
0
5
10
log_body_wt
log_brain_wt
Regression

Archosaurs: Linear Regression Example

The REG Procedure

Model: MODEL1

Dependent Variable: log_brain_wt

Number of Observations Read	21
Number of Observations Used	21

Analysis of Variance					
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	1	87.17316	87.17316	177.52	<.0001
Error	19	9.33022	0.49106		
Corrected Total	20	96.50338			

Root MSE	0.70076	*R-Square*	0.9033
Dependent Mean	2.76846	*Adj R-Sq*	0.8982
Coeff Var	25.31230		

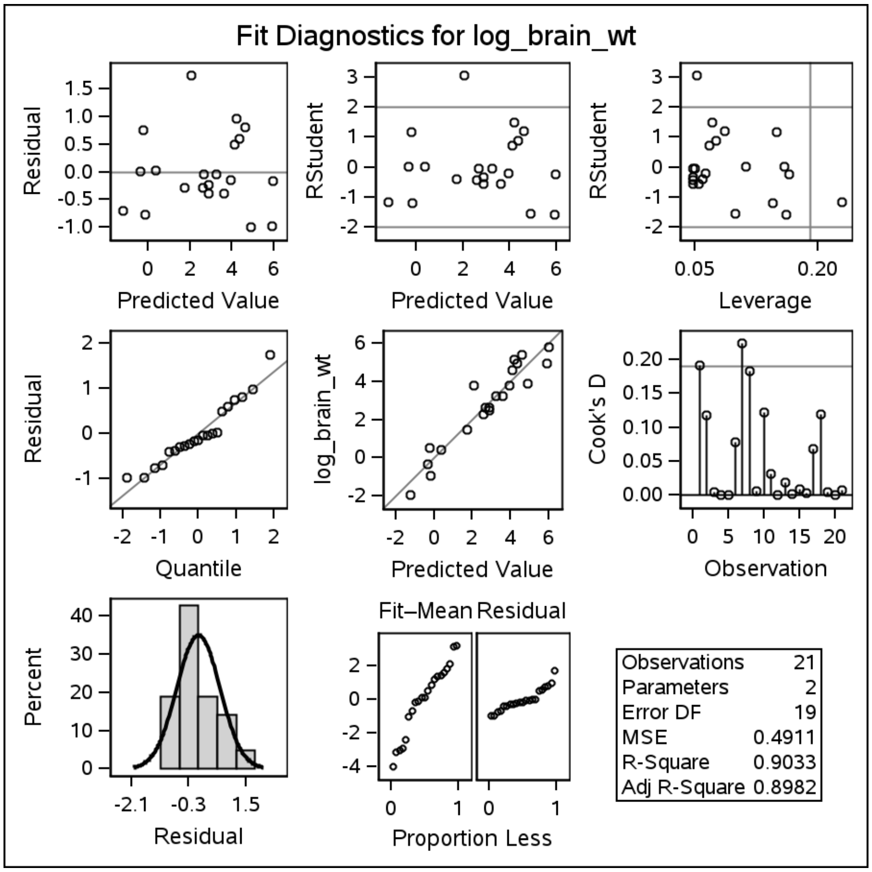
Parameter Estimates					
Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*
Intercept	1	0.21507	0.24518	0.88	0.3913
log_body_wt	1	0.51621	0.03874	13.32	<.0001

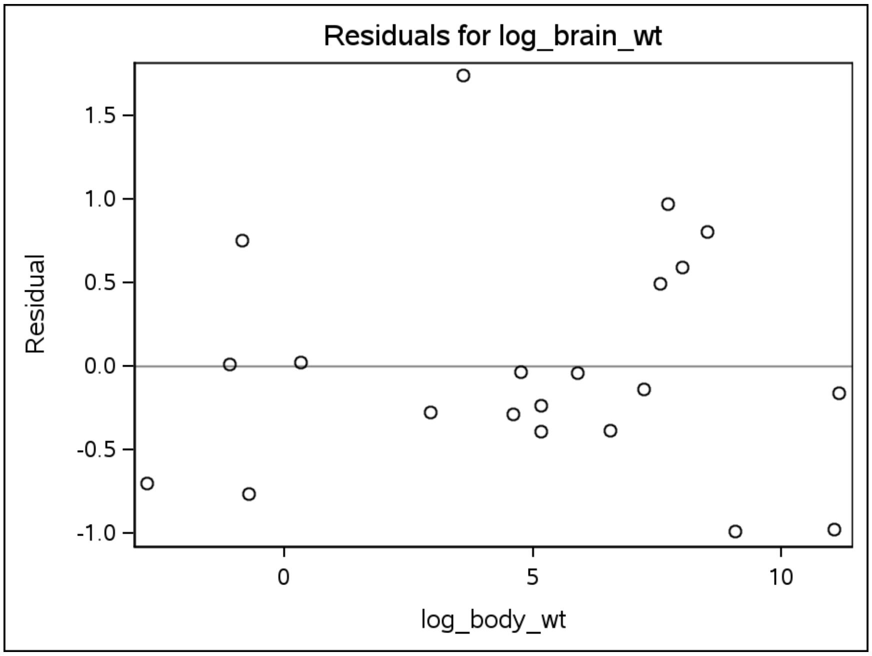
Archosaurs: Linear Regression Example

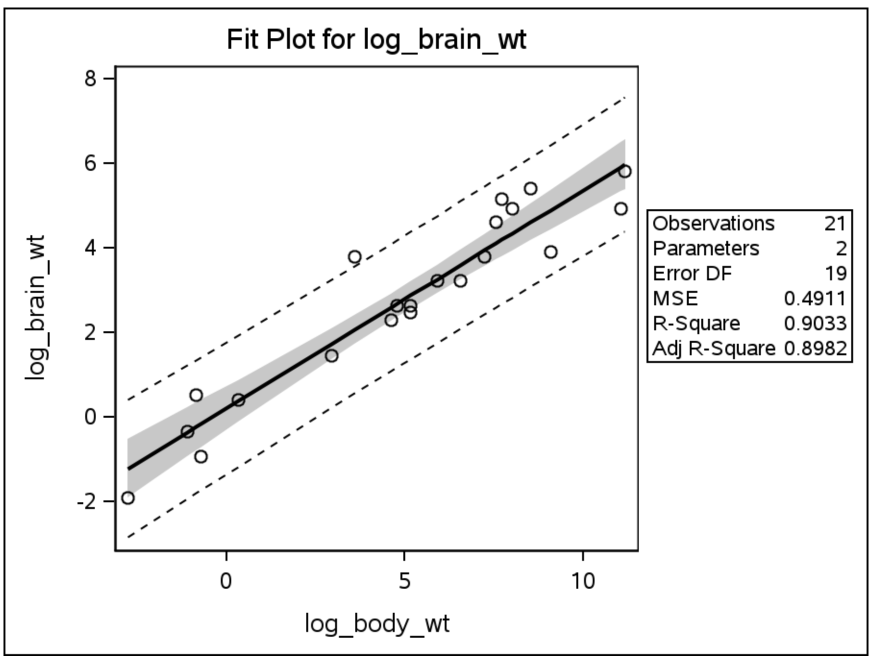
The REG Procedure

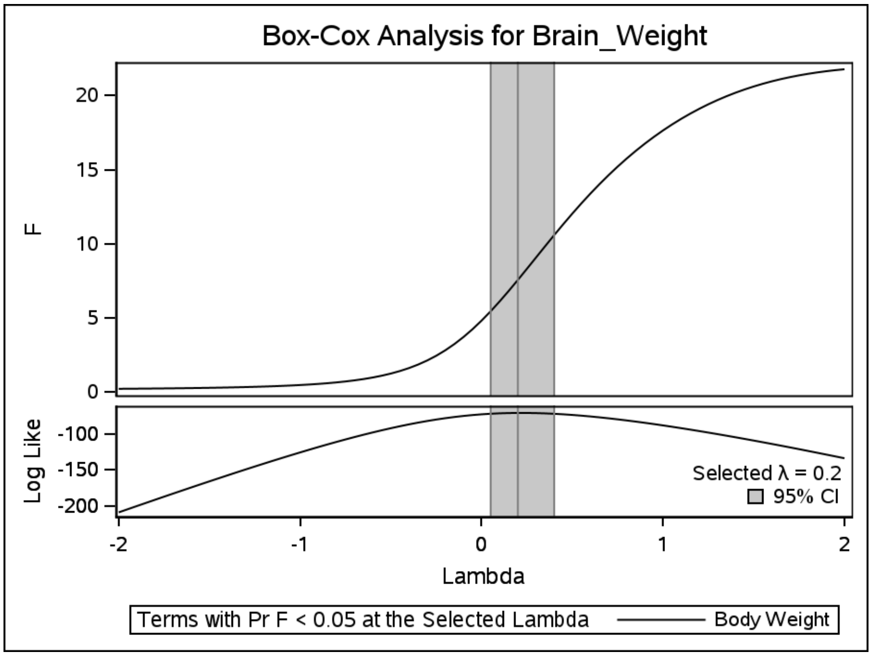
Model: MODEL1

Dependent Variable: log_brain_wt




Residuals for log_brain_wt
Residual
1.5
1.0
0.5
0.0
-0.5
-1.0
0
5
10
log_body_wt


Fit Plot for log_brain_wt
log_brain_wt
8
6
4
2
0
-2
0
5
10
log_body_wt
Observations 21
Parameters 2
Error DF 19
MSE 0.4911
R-Square 0.9033
Adj R-Square 0.8982


Box-Cox Analysis for Brain_Weight
F
20
15
10
5
0
Log Like
-100
-150
-200
Selected λ = 0.2
95% CI
-2
-1
0
1
2
Lambda
Terms with Pr F < 0.05 at the Selected Lambda
Body Weight

Archosaurs: Linear Regression Example

The TRANSREG Procedure

TRANSREG MORALS Algorithm Iteration History for BoxCox(Brain_Weight)					
Iteration Number	*Average Change*	*Maximum Change*	*R-Square*	*Criterion Change*	*Note*
1	0.78669	1.69925	0.28464		
2	0.00000	0.00000	0.94741	0.66277	Converged

Algorithm converged.