Stat 401 B – Lecture 30

Simple Linear Regression

- Example - mammals
- Response variable: gestation (length of pregnancy) days
- Explanatory: brain weight

1

"Man"

- Extreme negative residual but that residual is not statistically significant.
- The extreme brain weight of "man" creates high leverage that is statistically significant.

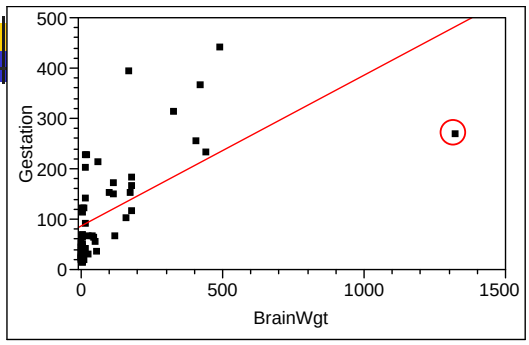
2

"Man"

- Is the point for "Man" influencing where the simple linear regression line is going?
- Is this influence statistically significant?

3

Stat 401 B – Lecture 30



4

Simple Linear Regression

- Predicted Gestation = 85.25 + 0.30*Brain Weight
- R^2 = 0.372, so only 37.2% of the variation in gestation is explained by the linear relationship with brain weight.

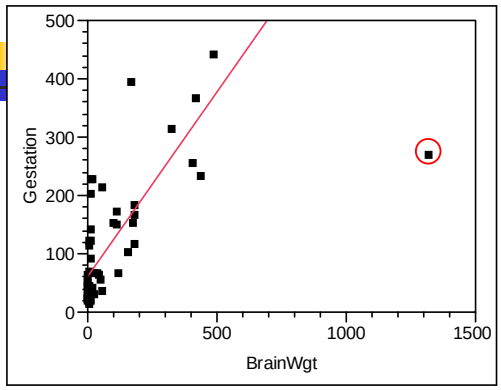
5

Exclude "Man"

- What happens to the simple linear regression line if we exclude "Man" from the data?
- Do the estimated intercept and estimated slope change?

6

Stat 401 B – Lecture 30



7

Simple Linear Regression

- Predicted Gestation = 62.05 + 0.634*Brain Weight
- R^2 = 0.600, 60% of the variation in gestation is explained by the linear relationship with brain weight.

8

Changes

- The estimated slope has more than doubled once “Man” is removed.
- The estimated intercept has decreased by over 20 days.

9

Stat 401 B – Lecture 30

Influence

- It appears that the point associated with "Man" influences where the simple linear regression line goes.
- Is this influence statistically significant?

10

Influence Measures

- Quantifying influence involves how much the point differs in the response direction as well as in the explanatory direction.
- Combine information on the residual and the leverage.

11

Cook's D

$$d = \left(\frac{h}{p+1}\right)\left(\frac{z}{(1-h)}\right)^2$$

- where z is the standardized residual and p is the number of explanatory variables in the model.

12

Stat 401 B – Lecture 30

Cook's D

- If $d > 1$, then the point is considered to have high influence.

13

Cook's D for "Man"

$$d = \left(\frac{h}{p+1}\right)\left(\frac{z}{(1-h)}\right)^2$$

$$d = \left(\frac{0.6612}{2}\right)\left(\frac{-2.516}{(1-0.6612)}\right)^2$$

$$d = 18.23$$

14

Cook's D for "Man"

- Because the d value for "Man" is greater than 1, it is considered to exert high influence on where the regression line goes.

15

Stat 401 B – Lecture 30

Cook's D

- There are no other mammals with a value of *d* greater than 1.
- The okapi has *d* = 0.30
- The Brazilian Tapir has *d* = 0.10

16

Studentized Residuals

- The studentized residual is the standardized residual adjusted for the leverage.

$$r_s = \frac{z}{\sqrt{(1-h)}}$$

17

Studentized Residuals

	z	h	r_s
Brazilian Tapir	3.010	0.0217	3.043
"Man"	–2.516	0.6612	–4.323
Okapi	2.443	0.0839	2.552

18

Stat 401 B – Lecture 30

Studentized Residuals

- If the conditions for the errors are met, then studentized residuals have an approximate *t*-distribution with degrees of freedom equal to $n - p - 1$.

19

Computing a P-value

- JMP – Col – Formula
- (1 – t Distribution($|r_s|$, n-p-1))*2
- For our example
  - r_s = 3.043, n-p-1=48
  - P-value = 0.0038

20

Studentized Residuals

	z	h	r_s	P-value
Brazilian Tapir	3.010	0.0217	3.043	0.0038
"Man"	–2.516	0.6612	–4.323	<0.0001
Okapi	2.443	0.0839	2.552	0.0139

21

Conclusion – "Man"

- The P-value is much less than 0.001 (the Bonferroni corrected cutoff), therefore "Man" has statistically significant influence on where the regression line is going.

Other Mammals

- The Brazilian Tapir has the most extreme standardized residual but not much leverage and so is not influential according to either Cook's D or the Studentized Residual value.

Other Mammals

- The Okapi has high leverage, greater than 0.08, but it's standardized residual is not that extreme and so is not influential according to either Cook's D or the Studentized Residual value.