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

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.

Regression Residuals
- residual = $y - \hat{y}$
- Those observations that do not follow the general trend will have residuals that are far from zero, either positive or negative.
Pattern in the Residuals

- Brain Weight between 0 g and 100 g
  - 25 negative, 10 positive residuals
- Brain Weight between 100 g and 500 g
  - 3 negative, 11 positive residuals

There may be a positive relationship between brain weight and the residuals for brain weights less than 500 g. This pattern could be due to the large brain weight mammal with a relatively short gestation.
Leverage

- A point with an extreme value for the explanatory variable can exert leverage on where the regression line will go.
- The leverage is quantified by something called the “hat” value.

Leverage in SLR

- In simple linear regression there is an equation for the “hat” value, $h$.

$$h = \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x - \bar{x})^2}$$
Comment

- Leverage is a summarizing something about the explanatory variable, \( x \).
- Brain Weight:
  - Mean: \( \bar{x} = 107.2524 \)
  - Std Dev: \( s = 216.35854 \)

Sum of Squares for \( x \)

\[
s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}
\]

\[
(n-1)s^2 = \sum (x - \bar{x})^2
\]

\[
(49)(216.35884)^2 = 2293739.874
\]

Leverage for “Man”

\[
x_0 = 1320
\]

\[
h = \frac{1}{50} \left( \frac{(1320 - 107.2524)^2}{2293739.874} \right)
\]

\[
h = 0.02 + 0.641 = 0.661
\]
Rule of Thumb

- High Leverage Value if
  \[ h > 2 \left( \frac{p + 1}{n} \right) \]
  \[ n = 50, \ p = 1, \]
  \[ 2 \left( \frac{p + 1}{n} \right) = 2 \left( \frac{2}{50} \right) = 0.08 \]

Comment

- The leverage for “Man”, 0.661, is greater than 0.08.
- This indicates that the brain weight for “Man” is unusual and that “Man” is a high leverage point.

Statistical Significance

- Test statistic
  \[ F = \frac{\left( h - \frac{1}{n} \right)/p}{(1-h)/(n-p-1)} \]
- Compute the P-value for an F distribution with p and (n - p - 1) degrees of freedom.
Statistical Significance

“Man”
- \( h = 0.661, p = 1, n = 50 \)

\[
F = \frac{(h - \frac{1}{n})/p}{(1-h)/(n-p-1)} = \frac{(0.661 - 0.02)/1}{(1-0.661)/48} = 90.76
\]

Computing a P-value

- **JMP - Col - Formula**
- \((1 - F \text{ Distribution}(F,p,n-p-1))\)
- For our example
  - \(F=90.76, p=1, n-p-1=48\)
  - P-value = 0.00000000001208

Conclusion – “Man”

- The P-value is much less than 0.001 (the Bonferroni corrected cutoff), therefore “Man” is a statistically significant high leverage point.
Other High Leverage Points

- The only other mammal with leverage above 0.08 is the Okapi.
  - $h = 0.084$
  - $F = 3.354$
  - P-value = 0.07325

Conclusion - Okapi

- The P-value is greater than 0.001, although the Okapi has high leverage, that leverage is not statistically significant.

“Man”

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