

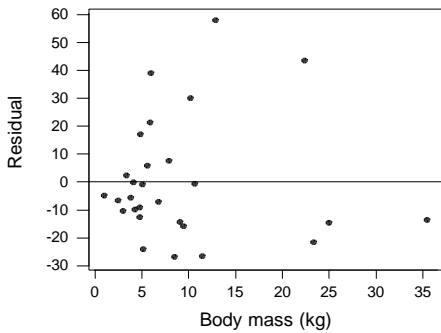
Stat 104 – Lecture 8

Residuals

- Residuals help us see if the linear model makes sense.
- Plot residuals versus the explanatory variable.
 - If the plot is a random scatter of points, then the linear model is the best we can do.

1

Plot of Residuals vs Body Mass



Interpretation of the Plot

- The residuals are scattered randomly. This indicates that the linear model is an appropriate model for the relationship between body mass and bite force for *Canidae*.

3

Stat 104 – Lecture 8

$(r)^2$ or R^2

- The square of the correlation coefficient gives the amount of variation in y , that is accounted for or explained by the linear relationship with x .

4

Body mass and Bite force

- $r = 0.9807$
- $(r)^2 = (0.9807)^2 = 0.962$ or 96.2%
- 96.2% of the variation in bite force can be explained by the linear relationship with body mass.

5

Regression Conditions

- Quantitative variables – both variables should be quantitative.
- Linear model – does the scatter diagram show a reasonably straight line?
- Outliers – watch out for outliers as they can be very influential.

6

Stat 104 – Lecture 8

Regression Cautions

- Beware of extraordinary points.
- Don't extrapolate beyond the data.
- Don't infer x causes y just because there is a good linear model relating the two variables.

7

Extraordinary Points

- <https://netfiles.uiuc.edu/jimarden/www/cuwu/datalist.html>
 - Scatter Plots
 - Check – Blank Plot and click Update
 - Add point

8

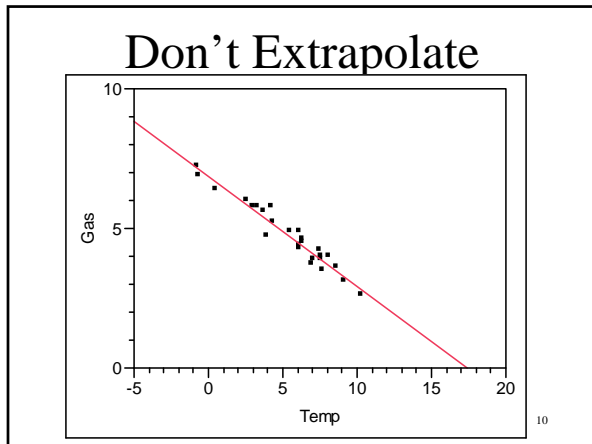
Don't Extrapolate

- Explanatory (x) – Average outdoor temperature ($^{\circ}$ C).
- Response (y) – Amount of natural gas used (1000 cu ft).

$$\hat{y} = 6.85 - 0.393x$$

9

Stat 104 – Lecture 8



Don't Extrapolate

- Explanatory ($x = 20$) – Average outdoor temperature ($^{\circ}$ C).
- Response (y) – Amount of natural gas used (1000 cu ft).

$$\hat{y} = 6.85 - 0.393(20)$$
$$\hat{y} = -1.01$$

11

Correlation \neq Causation

- Don't confuse correlation with causation.
 - There is a strong positive correlation between the number of crimes committed in communities and the number of 2nd graders in those communities.
- Beware of lurking variables.

12
