Chapter 3: Linear Regression

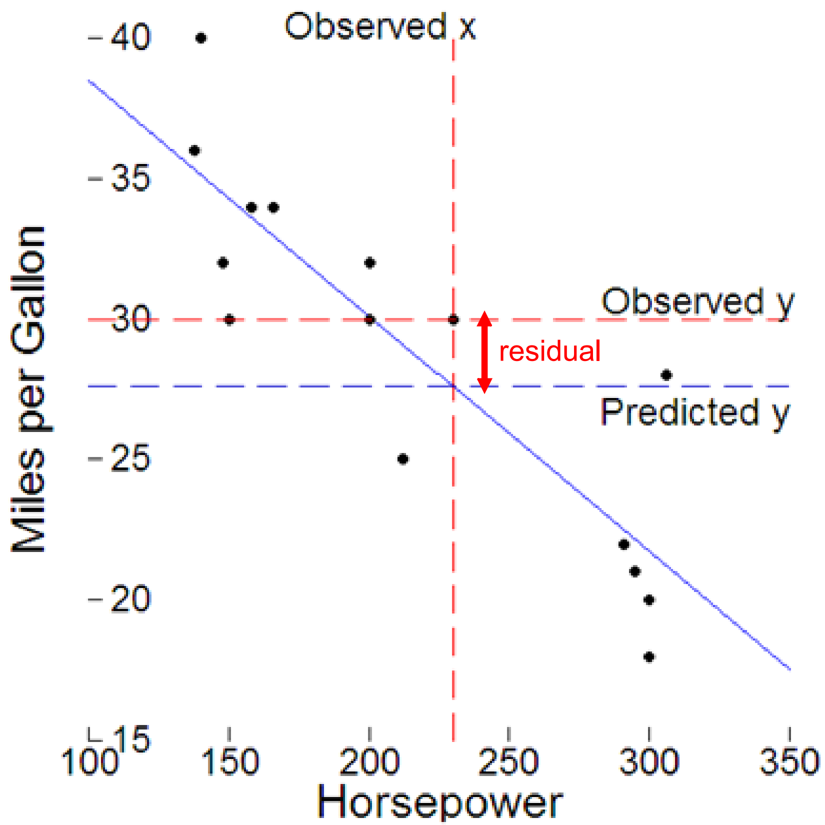
Skills

- Understand the relationship between observed values, predicted values and residuals
- Calculate the Least Squares Regression line
- Calculate and interpret the coefficient of determination
- Construct a histogram of the residuals
- Construct a residual plot
- Read computer output for regression

The Point

Our goal is to find a model (an equation) that fits the data. There are many ways to do this!
The most popular method is called **Least Squares Regression**.
(some people erroneously call this the *Line of Best Fit*)

Terminology



Least Squares Facts

The algebra that determines the Least Squares Regression Line guarantees that it will pass through at least one point: $\left(\bar{x}, \bar{y}\right)$

More algebra shows that the slope of the line must be: $r\dfrac{s_y}{s_x}$

Important to note:

x is the observed x-value

y is the observed y-value

$\hat{y}$ is the predicted y-value

Making Predictions

To make a prediction...plug into the model!
You should only plug in *x*; you should NOT plug in a *y* and back-solve for *x*.
To store the model in the calculator:
LinReg (a + bx) L_1, L_2, Y_1
To use the model:
Y_1 (<plug in *x*>)

Example

Six male gray kangaroos were sampled. The nasal length and width were measured (in millimeters). Researchers would like to predict the nasal width from the nasal length.

Example

Length	Width
609	241
629	222
620	233
564	207
645	247
493	189

Diagnostics

There are many ways to determine if the linear model is a good fit for the data.

- Visual appraisal
- r and r^2
- Residual Plot
- Histogram of the Residuals

Visually

Just look at it!

Does the line do a good job of following the trend in the data?

This is one way to see if there is any curvature in the data.

Interpreting r and r^2

r measures the strength and direction of a linear relationship.

r^2 measures the proportion of variation in the response variable that can be explained by a linear relationship with the explanatory variable.

Residual Plot

This is a scatterplot of residuals against predicted values (residuals on the *y*-axis, predicted response on the *x*-axis).

The calculator automatically stores the residuals for you when you do a regression!

To generate the predicted values: Y_1 $(L_1) \rightarrow L_4$

Look For…

In a residual plot, you want no patterns.
The points should make a sort of random cloud centered on the x-axis.
A triangle shape indicates the response variable has more variation for some predictions than others (thus, some predictions will be more reliable than others)

Histogram of Residuals

Easy enough!

This histogram should be fairly normal, centered at zero.

Example

DOT officials are trying to estimate the static weight of trucks based on a system that weighs them as they are moving...

Motion	Static
26	27.9
29.9	29.1
39.5	38
25.1	27
31.6	30.3
36.2	34.5
25.1	27.8
31	29.6
35.6	33.1
40.2	35.5