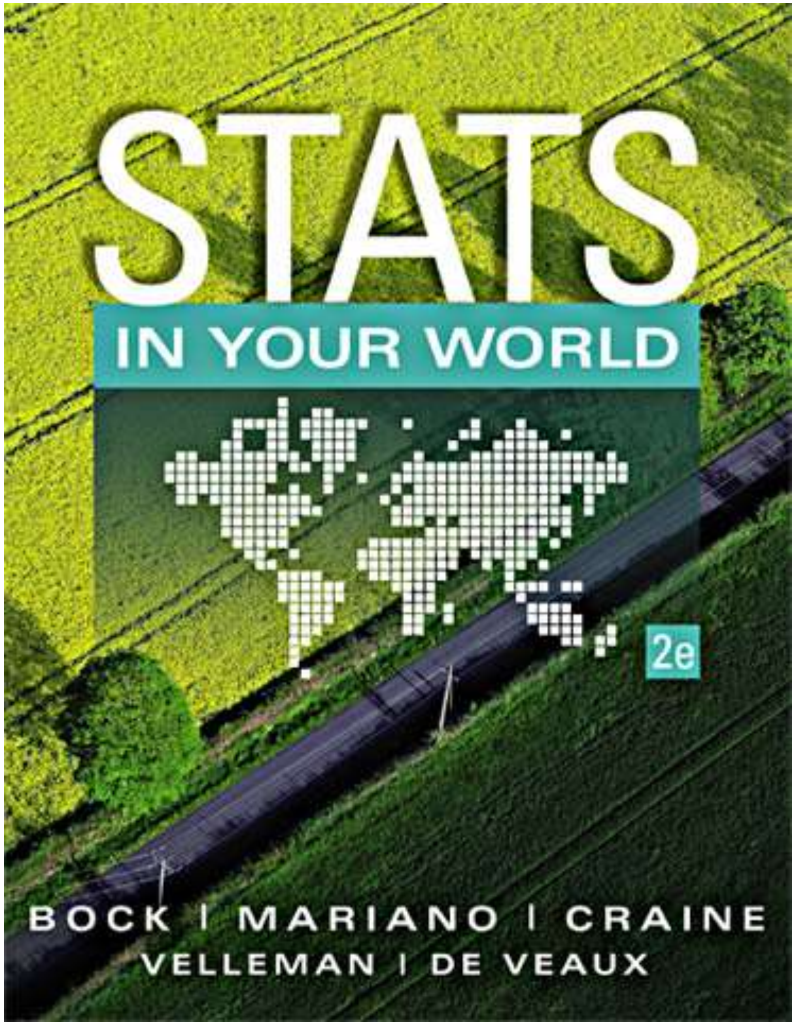
Chapter 8

What's My Curve?



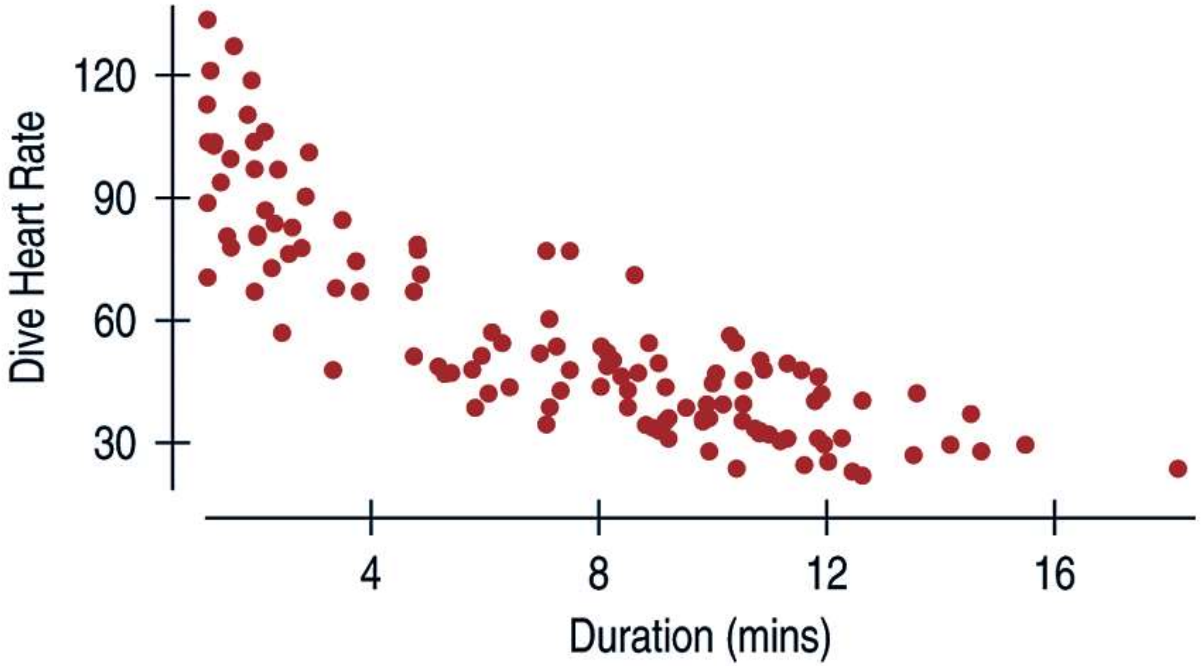
Straight to the Point

- We cannot use a linear model unless the relationship between the two variables is linear.

- How do you tell?

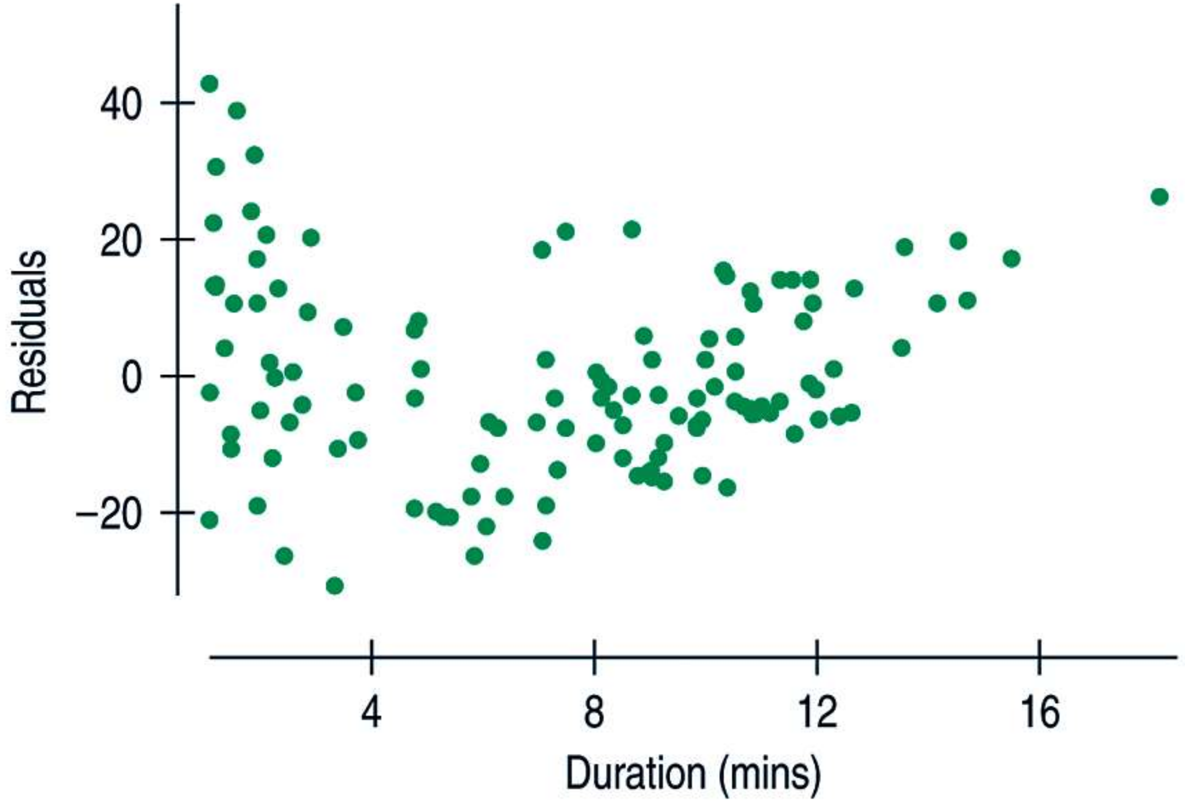
Straight to the Point (cont.)

- The relationship between *Dive Heart Rate* (in beats per minute) and *Duration* (in minutes) looks fairly linear:



Straight to the Point (cont.)

- A look at the residuals plot shows a problem:



Residual "Bends"

- If there's a clear bend in the residuals, your model has missed something important.
- It's time to look for a better model.
- But hopefully our residual plots will show… nothing. Which indicates our model is strong.

Exponential Models: EXAMPLES

- Populations tend to grow (or just change) exponentially.
- Money such as when inflation increases costs by a given percentage year after year.
- Other things *shrink* exponentially. For example, the way your body metabolizes a certain percentage of a drug may decay exponentially.
- Radioactive decay is another example of an exponential model.

Exponential Models (cont.)

- Exponential model:

$$\hat{y} = a\left(b^{x}\right)$$

Predicted value growth/decay rate

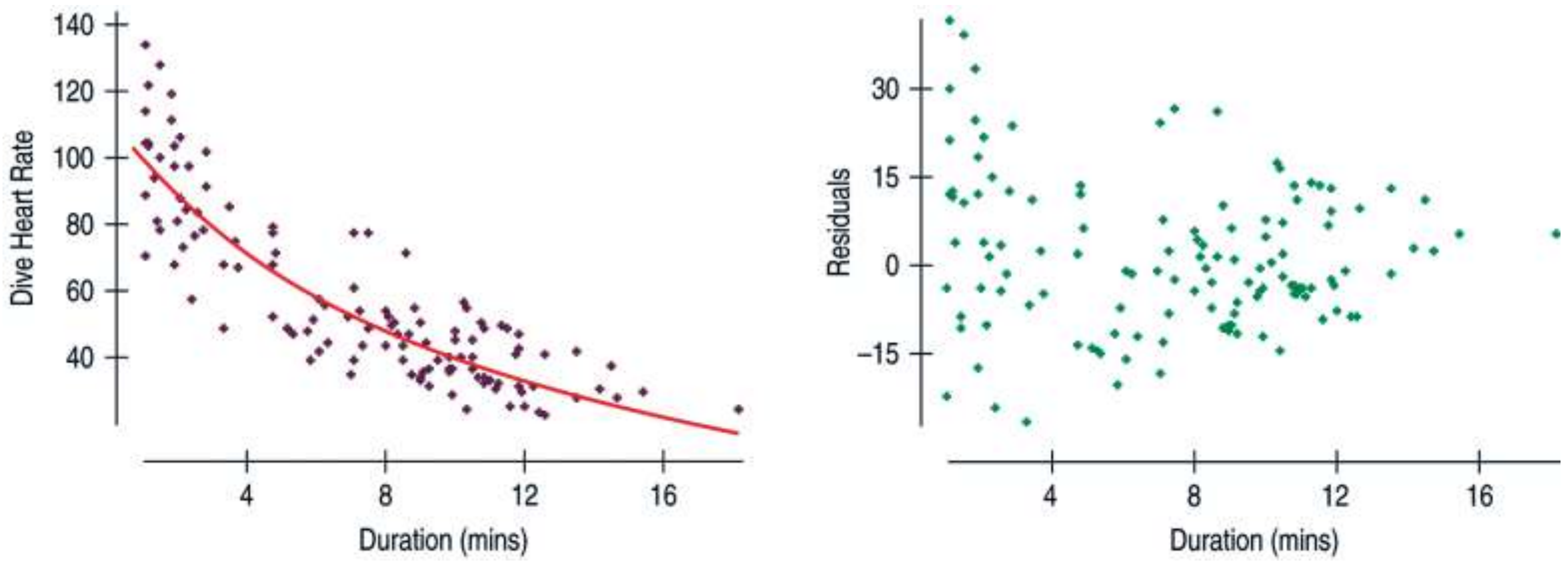
initial value explanatory variable

Exponential Model Parameters

- Recall a parameter is not necessarily from the data. It is used to explain the model.
- "*a*" represents the model's starting value.
- "*b*" represents the growth rate (or decay rate).
  - If $b = 1.02$, that's 102%. A 2% growth rate.
  - If $b = 0.88$, that's 88%. A 12% decrease.
- If the model is decreasing 15% for every 1 unit of change in x, then the model's value of b would be 100% - 15% = 0.85.

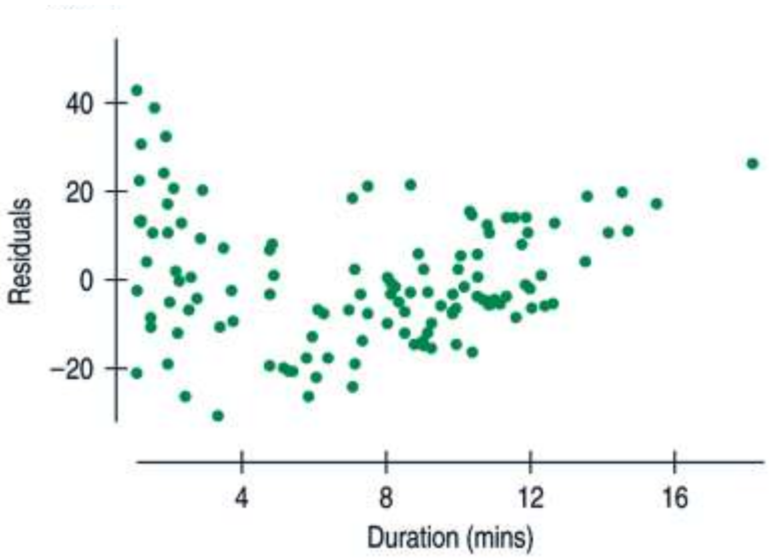
Penguin Dives

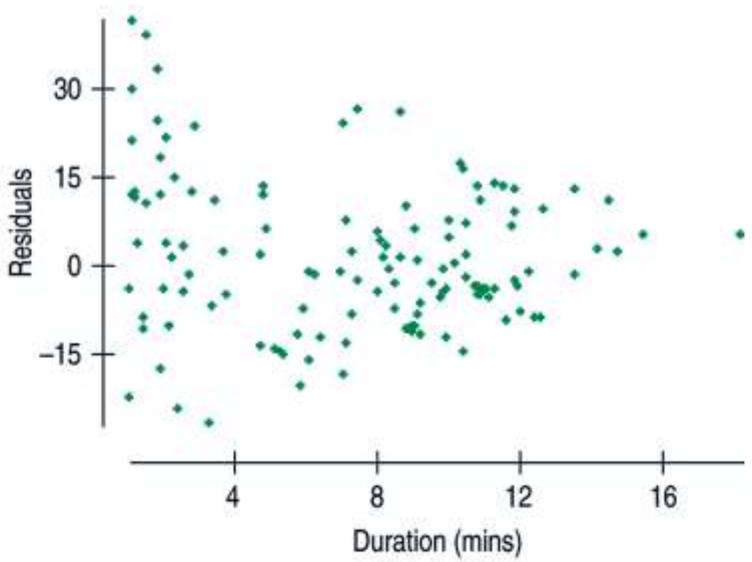
- The researchers studying emperor penguins set out to understand how the duration of a dive is related to the penguin's heart rate.
- Here are the scatterplot showing the curve and the resulting residuals plot:



Residual Plot Comparison

- There's clearly an improvement.
- The residuals now look more random.
  - The slight fanned pattern means the predicting power increases as time increaes





A Model for Penguin Dives

- The equation for our new exponential model is:

$$\widehat{DiveHeartRate} = 102.32\left(0.91\right)^{Duration}$$

A Model for Penguin Dives (cont.)

$$\widehat{DiveHeartRate} = 102.32(0.91)^{Duration}$$

- The value of *a* tells us our model estimates that penguins' heart rates should average around 102.32 beats per minute when they're not diving.

- The value of *b* indicates that heart rates tend to decrease exponentially about 1.00 - 0.91 = 9% for each minute a dive lasts.

Plan B: Power Models

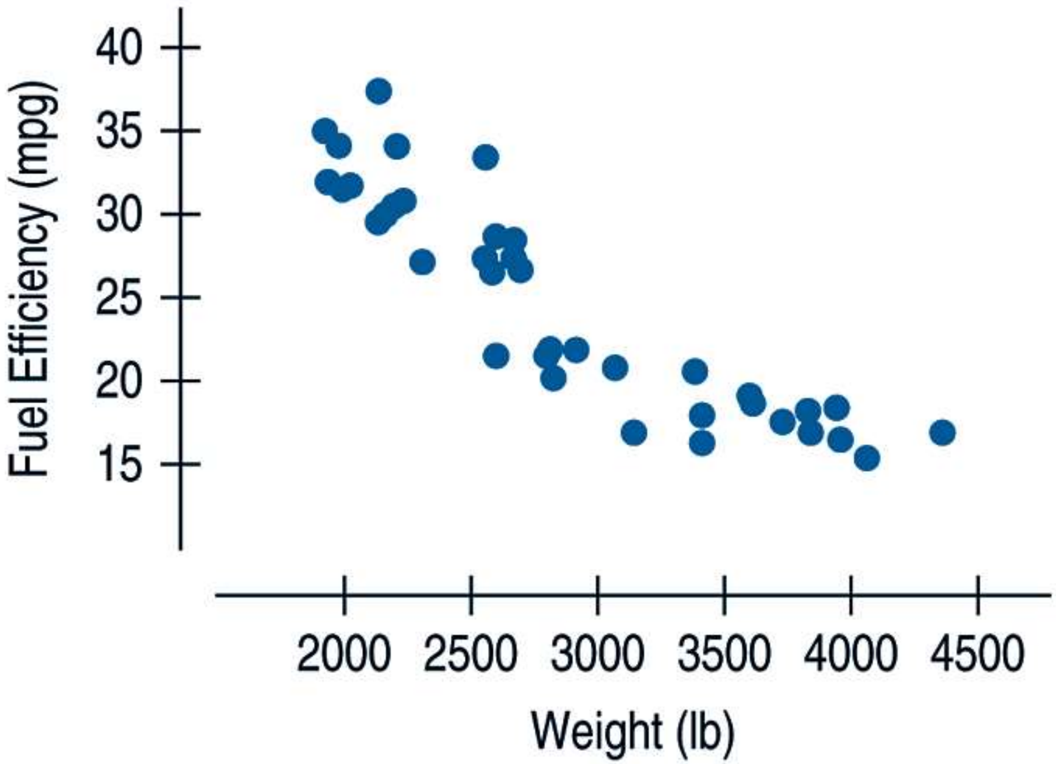
- Not all curves are exponential.
- Sometimes a power model is useful.
- Equations of **power models** look like this:

$$\hat{y} = a\left(x^{b}\right)$$

- This time we are raising *x* to a power.

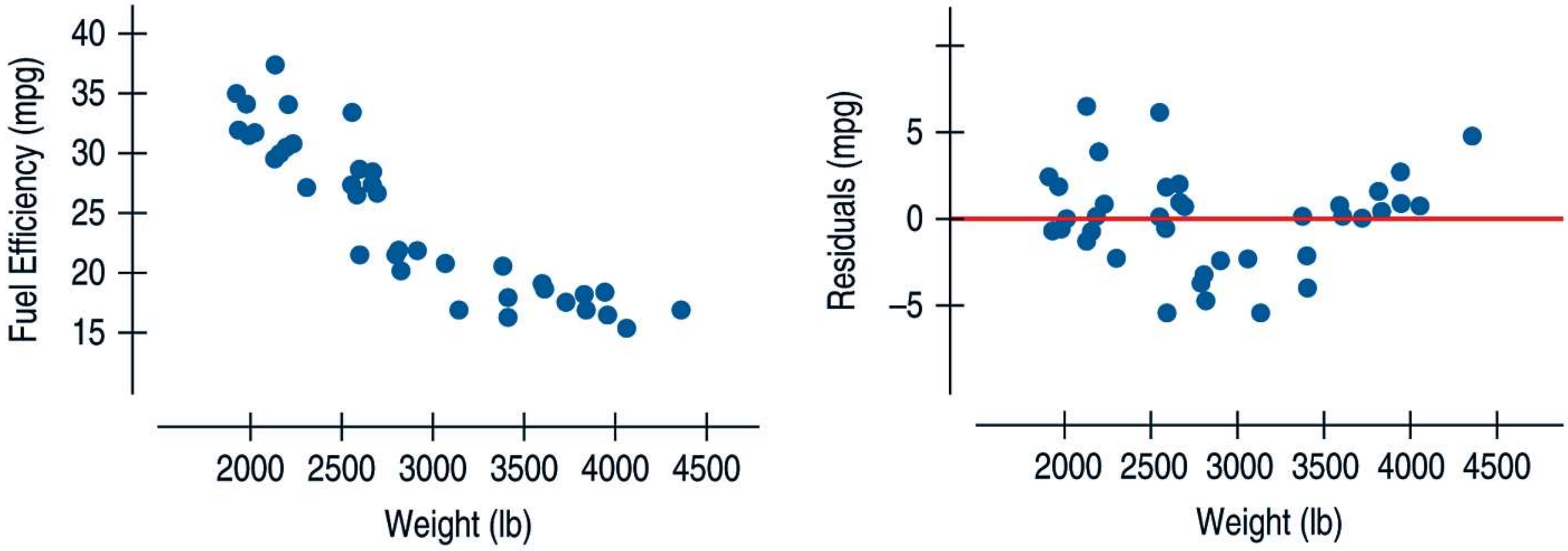
Modeling Fuel Economy

- Describe the association of the scatter plot below



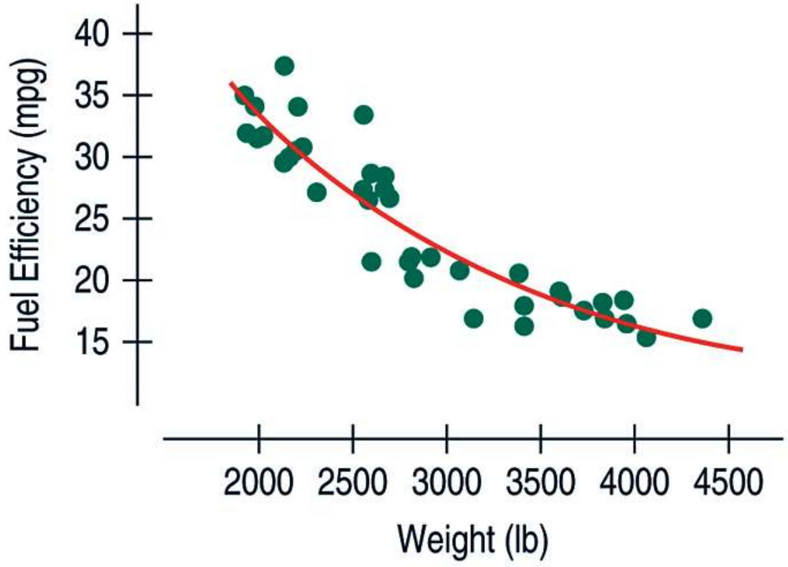
Modeling Fuel Economy

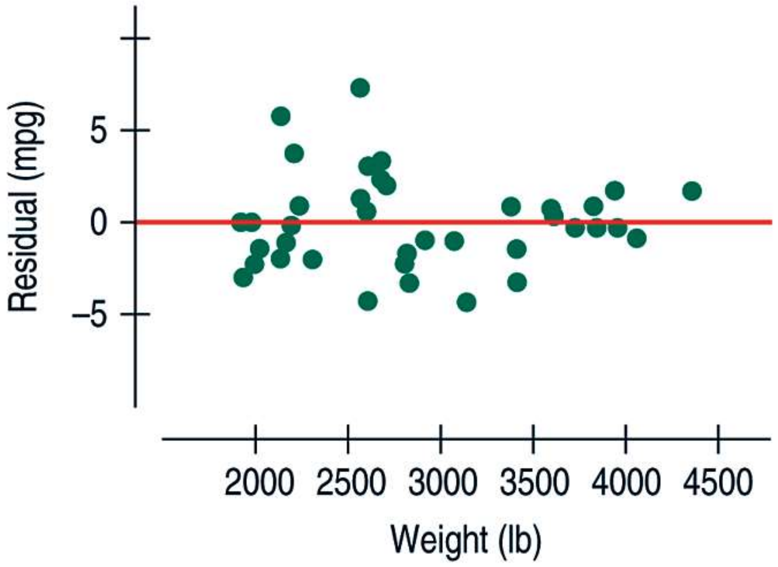
- This scatterplot shows the *Weight* (in pounds) and *Fuel Efficiency* (in mpg) for 38 cars.
- Note that the residual plot for a linear model has a shape that is clearly bent.



Modeling Fuel Economy (cont.)

- We have eliminated the linear model.
- Because fuel economy is a ratio and reciprocals can be expressed using a negative exponent, we try a power model:





Modeling Fuel Economy (cont.)

- Not an ideal residual plot, but models are rarely perfect!
- This power model should be more useful:

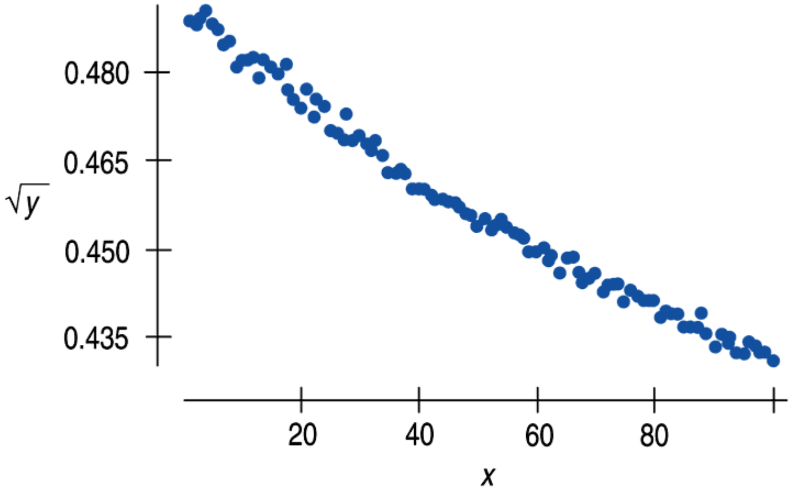
$$\widehat{MPG} = 79157\left(Weight\right)^{-1.022}$$

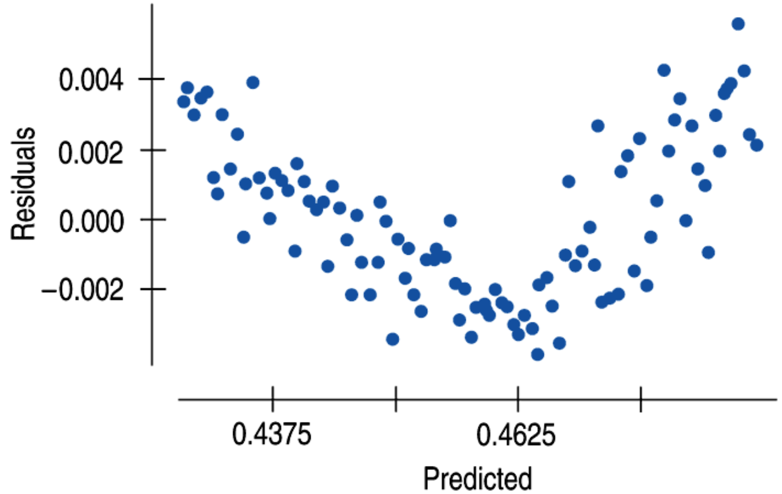
R-squared

- R-squared is a measure of “goodness of fit”
  - Similar to “r” for correlation, it is a number between 0 and 1.
    - 1 means perfect fit, closer to zero means less good fit
- There is no threshold for “good” fit. Context will determine the strength of the measurement

What Can Go Wrong?

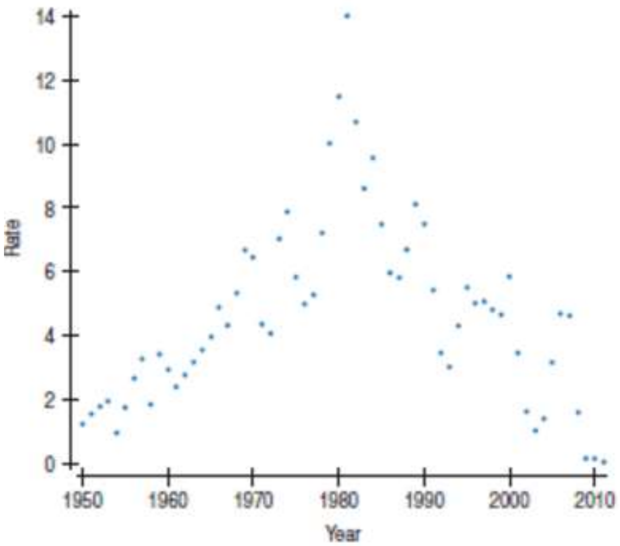
- Remember the basic rule of data analysis: Make a Picture!
- Don't expect your model to be perfect.
- Even a "fairly linear" association might be modeled much better by a curve.





What Can Go Wrong? (cont).

- Watch out for scatterplots that turn around (bend, neither purely increasing nor purely decreasing). We won't learn see how to create models that deal with this.



- Don't round too much! When dealing with exponents, even minor changes can make big differences.
  - Example: $12^4 = 20736$ $12^{4.01} = 21257.7255\ldots$

What Have We Learned?

- When the Straight Enough Condition fails, we may be able to fit a curved model.
  - We can try an exponential or power model.
  - We decide whether the model is appropriate by looking for random scatter in the residuals plot.
- Our models won't be perfect, but may be useful.
  - Some curvature may be okay provided the residuals are very small.
  - Some relationships are too complex to be described by these simple models.