Homework Answers:

Making Sense of Linear Models (#2, 3 only)

Textbook pg.185-187 # 1 - 5, ~~17~~ 32

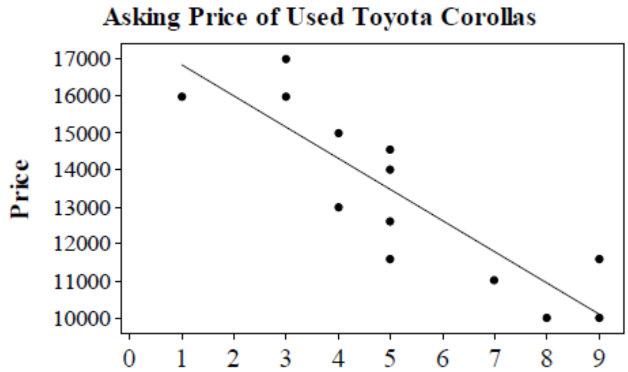
2. The number of wildfires (in thousands) and years since 1982	**Model:** $\widehat{Fires} = -3.4556year + 153.837$ **Slope:** The model predicts that an increase of one year is associated with a decrease of 3455.6 wildfires, on average. **Intercept:** According to the model, 1982 had 153,837 wildfires.	(20, 92) The model predicts -3.4556(20) + 153.837 = 84,725 wildfires in the year 2002. This prediction is 92,000-84,725 = 7275 wildfires too few.
3. Fiber (in grams) and potassium content (in milligrams) of servings of breakfast cereals.	**Model:** $\widehat{Potassium} = 38 + 27Fiber$ **Slope:** The model predicts that an increase of one gram of fiber is associated with an increase of 27 milligrams of potassium. **Intercept:** According to the model, a serving of breakfast cereal without any fiber is predicted to contain 38 grams of potassium. This, of course, is most likely not probable and serves only as a starting point for our model.	(3, 88) The model predicts 38 + 27(3) = 119 grams of potassium when the serving contains 3 grams of fiber. This prediction is 88 – 119 = 31 milligrams of potassium too high.

1. **Typing.** Choice D.
2. **More typing.** Choice D. (B)
3. **Even more typing.** Choice B.
4. **Off to college.** Choice D. $\widehat{gpa\_U} = 0.22 + 0.72 gpa\_HS = 0.22 + 0.72(3.8) = 2.956$ The residual is 3.5 – 2.956 = 0.544.
5. **Linear models and residuals.** Choice D.

Homework Answers:

32. Used cars 2014.

a) We are attempting to predict the price in dollars of used Toyota Corollas from their age in years. A scatterplot of the relationship is at the right.



b) The scatterplot shows that the relationship is straight, so the linear model is appropriate. Using technology, the regression equation to predict the price of a used Toyota Corolla from its age is $\widehat{Price} = 17647 - 844(Years)$.

c) The scatterplot provides evidence that the relationship is Straight Enough. A linear model will likely be an appropriate model.

d) According to the model, for each additional year in age, the car is expected to drop $844 in price.

e) The model predicts that a new Toyota Corolla (0 years old) will cost $17,647.

f) Buy the car with the negative residual. Its actual price is lower than predicted.

g)

$$\widehat{Price} = 17647 - 844(Years)$$
$$\widehat{Price} = 17647 - 844(10)$$
$$\widehat{Price} = 9207$$

According to the model, a 10-year-old Corolla is expected to cost $9207.
The car has an actual price of $8000, so its residual is $8000 – $9207 = – $1207
The car costs $1207 less than predicted.

h) The model would not be useful for predicting the price of a 25-year-old Corolla. The oldest car in the list is 9 years old. Predicting a price after 25 years would be an extrapolation.

17. Roller coaster.

a) The explanatory variable (x) is initial drop, measured in feet, and the response variable (y) is duration, measured in seconds.

b) The units of the slope are seconds per foot.

c) The slope of the regression line predicting duration from initial drop should be positive. Coasters with higher initial drops probably provide longer rides.

Reminders...

Assumptions and Conditions

- Quantitative Variables Condition:
  - Regression can only be done on two quantitative variables (and not two categorical variables), so make sure to check this condition.
- Straight Enough Condition:
  - The linear model assumes that the relationship between the variables is linear.
  - A scatterplot will let you check that the assumption is reasonable.

Assumptions and Conditions (cont.)

- If the scatterplot is not straight enough, stop here.
  - You can't use a linear model for *any* two variables, even if they are related.
  - They must have a *linear* association or the model won't mean a thing.
- Some nonlinear relationships can be saved by re-expressing the data to make the scatterplot more linear.

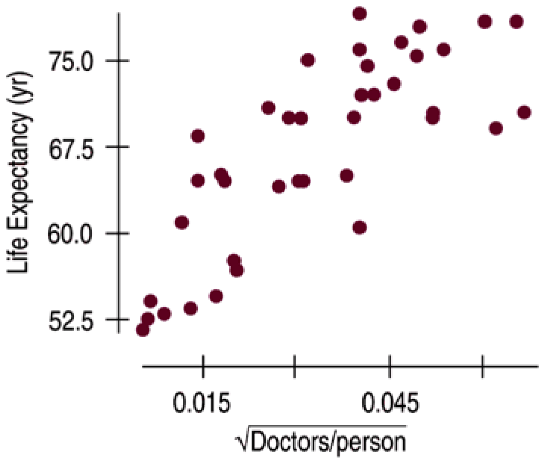
 Chapter 7, Slide 23

Lurking Variables and Causation

- No matter how strong the association, no matter how straight the line, there is no way to conclude from a regression alone that one variable *causes* the other.
  - There's always the possibility that some third variable is driving both of the variables you have observed.
- With observational data, as opposed to data from a designed experiment, there is no way to be sure that a lurking variable is not the cause of any apparent association.

Lurking Variables and Causation (cont.)

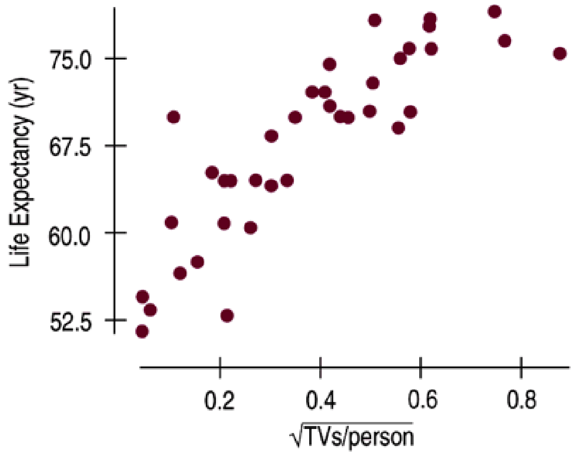
- The following scatterplot shows that the average *life expectancy* for a country is related to the number of *doctors* per person in that country:



 Chapter 7, Slide 30

Lurking Variables and Causation (cont.)

- This new scatterplot shows that the average *life expectancy* for a country is related to the number of *televisions* per person in that country:



 Chapter 7, Slide 31

Lurking Variables and Causation (cont.)

- Since televisions are cheaper than doctors, send TVs to countries with low life expectancies in order to extend lifetimes. Right?
- How about considering a lurking variable? That makes more sense…
  - Countries with higher standards of living have both longer life expectancies *and* more doctors (and TVs!).
  - If higher living standards *cause* changes in these other variables, improving living standards might be expected to prolong lives and increase the numbers of doctors and TVs.

 Chapter 7, Slide 32

Homework:

Distance and Ticket Price Worksheet

(use calculator for #1 and #2 not formula)

~~Quiz Tomorrow!!~~