Chapter 8 – Linear Regression

1. **Cereals.**

   $\widehat{Potassium} = 38 + 27\,Fiber = 38 + 27(9) = 281$ mg. According to the model, we expect cereal with 9 grams of fiber to have 281 milligrams of potassium.

2. **Horsepower.**

   $\widehat{mpg} = 46.87 - 0.084\,HP = 46.87 - 0.084(200) \approx 30.07$ mpg. According to the model, we expect a car with 200 horsepower to get about 30.07 miles per gallon.

3. **More cereal.**

   A negative residual means that the potassium content is actually lower than the model predicts for a cereal with that much fiber.

4. **Horsepower, again.**

   A positive residual means that the car gets better gas mileage than the model predicts for a car with that much horsepower.

5. **Another bowl.**

   The model predicts that cereals will have approximately 27 more milligrams of potassium for each additional gram of fiber.

6. **More horsepower.**

   The model predicts that cars lose an average of 0.84 miles per gallon for each additional 10 horse power.

7. **Cereal again.**

   $R^2 = (0.903)^2 \approx 0.815$. About 81.5% of the variability in potassium content is accounted for by the model.

8. **Another car.**

   $R^2 = (-0.869)^2 \approx 0.75.5$. About 75.5% of the variability in fuel economy is accounted for by the model.

$\hat{a} = 40 + a3(b)$

9. **Last bowl!**

   True potassium content of cereals vary from the predicted values with a standard deviation of 30.77 milligrams.

10. **Last tank!**

    True fuel economy varies from the predicted amount with a standard deviation of 3.287 miles per gallon.

11. **Residuals.**

    a) The scattered residuals plot indicates an appropriate linear model.

b) The curved pattern in the residuals plot indicates that the linear model is not appropriate. The relationship is not linear.

c) The fanned pattern indicates that the linear model is not appropriate. The model's predicting power decreases as the values of the explanatory variable increase.

12. Residuals.

a) The curved pattern in the residuals plot indicates that the linear model is not appropriate. The relationship is not linear.

b) The fanned pattern indicates heteroscedastic data. The models predicting power increases as the value of the explanatory variable increases.

c) The scattered residuals plot indicates an appropriate linear model.

13. What slope?

The only slope that makes sense is 300 pounds per foot. 30 pounds per foot is too small. For example, a Honda Civic is about 14 feet long, and a Cadillac DeVille is about 17 feet long. If the slope of the regression line were 30 pounds per foot, the Cadillac would be predicted to outweigh the Civic by only 90 pounds! (The real difference is about 1500 pounds.) Similarly, 3 pounds per foot is too small. A slope of 3000 pounds per foot would predict a weight difference of 9000 pounds (4.5 tons) between Civic and DeVille. The only answer that is even reasonable is 300 pounds per foot, which predicts a difference of 900 pounds. This isn't very close to the actual difference of 1500 pounds, but at least it is in the right ballpark.

14. What slope?

The only slope that makes sense is 1 foot in height per inch in circumference. 0.1 feet per inch is too small. A trunk would have to increase in circumference by 10 inches for every foot in height. If that were true, pine trees would be all trunk! 10 feet per inch (and, similarly 100 feet per inch) is too large. If pine trees reach a maximum height of 60 feet, for instance, then the variation in circumference of the trunk would only be 6 inches. Pine tree trunks certainly come in more sizes than that. The only slope that is reasonable is 1 foot in height per inch in circumference.

15. Real estate.

a) The explanatory variable (x) is size, measured in square feet, and the response variable (y) is price measured in thousands of dollars.

b) The units of the slope are thousands of dollars per square foot.

c) The slope of the regression line predicting price from size should be positive. Bigger homes are expected to cost more.

16. 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.

22. Last ride.

a) According to the linear model, the duration of a coaster ride is expected to increase by about 0.242 seconds for each additional foot of initial drop.

b) $\hat{Duration} = 91.033 + 0.242(Drop)$

$\hat{Duration} = 91.033 + 0.242(200)$

$\hat{Duration} = 139.433$

According to the linear model, a coaster with a 200 foot initial drop is expected to last 139.433 seconds.

c) $\hat{Duration} = 91.033 + 0.242(Drop)$

$\hat{Duration} = 91.033 + 0.242(150)$

$\hat{Duration} = 127.333$

According to the linear model, a coaster with a 150 foot initial drop is expected to last 127.333 seconds. The advertised duration is shorter, at 120 seconds.
120 seconds – 127.333 seconds = – 7.333 seconds, a negative residual.

23. Misinterpretations.

a) R^2 is an indication of the strength of the model, not the appropriateness of the model. A scattered residuals plot is the indicator of an appropriate model.

b) Regression models give predictions, not actual values. The student should have said, "The model predicts that a bird 10 inches tall is expected to have a wingspan of 17 inches."

24. More misinterpretations.

a) R^2 measures the amount of variation accounted for by the model. Literacy rate determines 64% of *the variability* in life expectancy.

b) Regression models give predictions, not actual values. The student should have said, "The slope of the line shows that an increase of 5% in literacy rate *is associated with an expected* 2-year improvement in life expectancy."

25. ESP.

a) First, since no one has ESP, you must have scored 2 standard deviations above the mean by chance. On your next attempt, you are unlikely to duplicate the extraordinary event of scoring 2 standard deviations above the mean. You will likely "regress" towards the mean on your second try, getting a lower score. If you want to impress your friend, don't take the test again. Let your friend think you can read his mind!

b) Your friend doesn't have ESP, either. No one does. Your friend will likely "regress" towards the mean score on his second attempt, as well, meaning his score will probably go up. If the goal is to get a higher score, your friend should try again.