Chapter 7. (due Nov 15)

* Practice: 7.23, 7.25, 7.29, 7.39
* Graded: 7.24, 7.26, 7.30, 7.40

7.24 Nutrition at Starbucks, Part I. The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain.21 Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content.

(a) Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

**The relationship is positive, although there is a lot of residuals towards the high end. For roughly every 100 calories there are 20 carbs.**

(b) In this scenario, what are the explanatory and response variables?

**The response is carbs, the explanatory is calories.**

(c) Why might we want to \_t a regression line to these data?

**To be able to model and therefore predict the relationship between calories and carbs, which appears to be linear.**

(d) Do these data meet the conditions required for fitting a least squares line?

**Probably, although one could argue that the variability of the data around the line increases with larger values of x, and therefore doesn’t meet the conditions.**

7.26 Body measurements, Part III. Exercise 7.15 introduces data on shoulder girth and

height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation

of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

(a) Write the equation of the regression line for predicting height.

(b) Interpret the slope and the intercept in this context.

(9.41/10.37)(.67) = .61

So the model would predict that height equals .61 times the individual’s shoulder girth on top of 236.32

(c) Calculate R2 of the regression line for predicting height from shoulder girth, and interpret it

in the context of the application.

.67^2 = **.45 So shoulder girth can help to explain about .45 of the variation of height**

(d) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

Y = .61 (100) + 236.32 = **297.32**

(e) The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

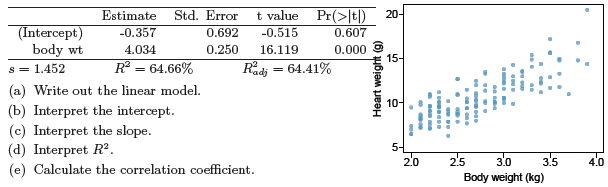
297.32 – 160 = **137.32**, or the model over predicted 137.32 cm of height

(f) A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

**Probably not, the intercept alone would likely overestimate it’s height**

7.30 Cats, Part I. The following regression output is for predicting the heart weight (in g) of cats

from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats.

****

a. **heart weight = -.357 + 4.034x**

b. **Expected heart weight of a cat with no weight is -.357, not a useful number, but serves to adjust the height of the regression line.**

c. **For each additional kg of body weight, the heart weight in grams increases by a multiple of 4.034**

d. **Body weight explains 64.66% of the variability in heart weight**

e. sqrt(64.66%) = **8.04%**

7.40 Rate my professor. Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. Researchers at University of Texas, Austin collected data on teaching evaluation score (higher score means better) and standardized beauty score (a score of 0 means average, negative score means below average, and a positive score means above average) for a sample of 463 professors.24 The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.

(a) Given that the average standardized beauty score is -0.0883 and average teaching valuation

score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the

information provided in the model summary table.

3.9983 = x(-.0883)+4.010

X = (3.9983 – 4.010)/-.0883 = **.1325**

(b) Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.

**Yes, there is convincing evidence since the t-value of 4.13 at n=463 suggests a one-tailed p-value of less than .005**

(c) List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.

1. linearity: I’m not convinced its entirely linear, because there seems to be more residuals at the lower end of beauty

2. Nearly normal residuals: the residuals do appear to be nearly normal, although they have a slight left skew

3. Constant variability: the variability doesn’t seem constant, especially near the lower end of beauty

4. Independent observations: it appears that the data was collected in good order