HW7

Ahsanul Choudhury

November 13, 2016

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. 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.

There is a linear positive relationship between number of calories and amount of carbohydrates (in grams). Carbohydrate increases as calories increase. The residuals seems normally distributed.

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

Calories are explanatory and carb is response variable.

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

We might be able to predict Carbohydrate by looking at calories data in a Starbucks food menu item.

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

Conditions required for fitting a least squares line are:

- Linearity: The data shows week linearity.

- Nearly normal residual: The residual appears to be normal.

- Constant variability: The residual for low calories are smaller and high calories are bigger, so no constant variability.

- Independent observation: The data is independent Starbucks food menu item but they come from same source.

It does not seem like data fully meets the requirement for fitting a least squares line.

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.

Beta estimates are:

```r
b1 <- (9.41 / 10.37) * 0.67
b1
```

```
## [1] 0.6079749
```

```r
b0 <- (b1 * (0 - 107.2)) + 171.14
b0
```

```
## [1] 105.9651
```

Equation

$$\hat{y} = 105.9650878 + 0.6079749 * x$$

x = shoulder girth

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

The slope tells us the change that occurs in height when the the shoulder grith changes, the intercept tells us the height when the shoulde grith is 0 which is impossible.

(c) Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

R^2 = (0.67)^2 = 0.4489 About 45% of variability of height is explained using this model.**

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

166.7625805

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

residual = -6.7625805, the residual indicates our model over estimated the student's 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?

The original data set had a response variable values between 80 and 140 cm. A shoulder girth of 56 cm would be a outlier and would not be ppropriate to use this linear model to predict the height of the child.

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 coeffcients are estimated using a dataset of 144 domestic cats.

(a) Write out the linear model.

$$\hat{y} = -0.357 + 4.034 * bodyweight$$

(b) Interpret the intercept.

The intercept tells us the heart weight when the body weight 0, which is impossible.

(c) Interpret the slope.

The slope tells us the change in heart weight with the body weight changes.*

(d) Interpret R2.

64.66% variation in heart weight is explained by the model.

(e) Calculate the correlation coeffcient.

The correlation coefficient is: 0.8041144

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 e???ectiveness 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 370 CHAPTER 7. INTRODUCTION TO LINEAR REGRESSION 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 evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.

```r
b1 <- (3.9983 - 4.010) / -0.0883
b1
```

```
## [1] 0.1325028
```

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

There is a strong evidence that the relationship between teaching evaluation and beauty is positive, because the p-value is very small. We rejet the null hypothesis and conclude there is a strong evidence that as the teacher's beauty score increases so does the chance of getting a better teaching evaluation score.

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

Conditions required for linear regression are:

- Linearity: Scatter plot shows weak but a positive relation

- Nearly normal residual: Histogram shows a near normal distribution of residuals

- Constant variability: The residual scatter plot shows high but constant variability.

- Independent observation: No information on how the data was collected but we would assume the data are independent observation.