Oluwakemi Omotunde

IS606

Chapter 7 Homework

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.

It seems that they have a fairly linear relationship.

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

Explanatory variable is defined as the variable used to predict, which in this case, is the calorie content of the menu items.

Response variable is defined as the variable that is predicted and for our case will be amount of carbs in the menu items.

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

To be able to predict the amount of carbs in a particular menu item. From the question, Starbucks only lists they number of calories on the display items.

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

The date is fairly linear, independent, the residuals are nearly normal, and there is constant variability.

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.

$b_1 = (s_y/s_x)*R$ -> (10.37/9.41)*.67 = .74

y – 171.14 = .74(x-107.20) -> y = .74x +91.81

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

The slope says that there is a .74 estimated difference in the mean shoulder girth. The average height when shoulder girth is 0 is 91.81, the intercept.

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

$R^2 = (.67)^{2 =}$ .55 which means that there is about 45% of the height explained by shoulder girth.

(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 = .74(100) + 91.81 = 165.81cm

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

Residual = observed – predicted -> 160-165.81 = -5.81. It means that this model predicted a greater height than observed. It tells us how well the model works.

(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?

No it would not as this would fall outside of regular data set. We could use that point as an extrapolation point, though.

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) Write out the linear model.

b_1 = 4.034

intercept = -.357

y = -.357 + 4.034x -> heart weight(g) = -.357 + 4.034 * body weight(kg)

(b) Interpret the intercept.

The intercept, -.357, indicates the heart weight of a cat weighing 0 kg.

(c) Interpret the slope.

The slope, 4.034, indicates that for every additional kg of body weight, the heart weight would have a change of that much.

(d) Interpret R2.

Our R^2 is 64.66%, which tells us that the percent of the variability seen in heart weight is due to body weight.

(e) Calculate the correlation coefficient.

Sqrt .6466 = .8041

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. 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.

$y-3.9983 = b_1(x-(-.0833)$ is about as far as I got on this one.

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

Yes, the very low t value indicates that there is a relationship between the two.

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

- **Linearity – the plots indicate this criteria is met**
- **Nearly normal residuals - the residuals histogram shows this to be true**
- **Constant variability – from plots, this is true**
- **Independent - met from sample**