HW7_StephenJones

Stephen Jones

April 13, 2019

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.

Calories and carbohydrates share a positive, linear relationship.

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

The explanatory variable is calories and the response variable is carbohydrates.

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

With a regression line we can predict carbohydrates given calorie count.

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

Data are approximately linearly related but residuals do not appear to vary consistently (not entirely constant variability). Residuals are normally distributed. Overall, I would not fit 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.

$$\hat{y} = \beta_0 + \beta_1 \cdot x$$

where $\beta_1 = \frac{S_y}{S_x} \cdot R$

```r
B_1 <- (9.41/10.37)*0.67

B_0 <- 171.14 - B_1 * 107.2

cat("The equation: y' =",B_0,"+",B_1,"(x)")
```

```
## The equation: y' = 105.9651 + 0.6079749 (x)
```

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

Slope refers to rise over run on the cartesian plane; in this context, it represents .6 cm in additional height for each unit increase in shoulder girth. The intercept indicates the value of y when x = 0, or height in cm when girth is 0.

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

If correlation R is 0.67, R^2=0.4489; this means that roughly 44.9% of variation in height is due to shoulder girth measurement.

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

```r
#use values from previous solutions.
height <- B_0 + B_1 * 100
cat("The estimated height for students with shoulder girth 100 cm is:", round(height,3), "cm.")
```

```
## The estimated height for students with shoulder girth 100 cm is: 166.763 cm.
```

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

The residual is 160 - 166.763 = -6.763, which means that the model is overestimating the height of this student.

(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 data and plots do not include shoulder girth less than approximately 85 cm; shoulder girth of 56 cm is outside the range of the model. It would not be appropriate to predict the height of this child using this model.

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.

```r
B_1 <- 4.034

B_0 <- -0.357

cat("The equation: y' =",B_0,"+",B_1,"(x)")
```

```
## The equation: y' = -0.357 + 4.034 (x)
```

(b) Interpret the intercept.

The intercept indicates the value of y when x = 0, or heart weight when body weight is 0.

(c) Interpret the slope.

Slope refers to rise over run on the cartesian plane; in this context, it represents an additional 4.034 in weight for each unit increase in body weight.

(d) Interpret R^2.

If R^2= 64.66%, roughly 65% of variation in heart weight is related to body weight.

(e) Calculate the correlation coefficient.

The correlation coefficient is 0.8041.

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.

```r
B_0 <- 4.010

#define the slope with the given means at the intercept point
B_1 <- (3.9983-B_0)/-0.0883

cat("The equation: y' =",B_0,"+",B_1,"(x)    The slope:",B_1)
```

```
## The equation: y' = 4.01 + 0.1325028 (x)    The slope: 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.

Yes; the slope is slight but always positive. Although the scatter plot offers dubious evidence for a significant relationship the data table establishes proof that a significant relationship exists.

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

Residuals are randomly and evenly distributed; this indicates a probable linear realtionship.
The residual histogram is slightly skewed but otherwise approximately normal.
Variance is constant, according to the scatterplot. The Q-Q plot supports these assumptions.
By principle, each evaluation is independent.