Lin-HW7

Bin Lin

2016-11-13

7.24 Nutrition at Starbucks, Part I. (a) Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

There is a positive linear association between number of calories and amount of carbohydrates.

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

Explanatory: number of calories Response: amount of carbohydrates

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

We can predict amount of carbohydrates from certain number of calories.

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

No, because the residuals appear to be fan shaped, indicating non-constant variance. Therefore, these data do not meet the conditions required for fitting a least squares line.

7.26 Body measurements, Part III. (a) Write the equation of the regression line for predicting height.

y = 105.9651 + 0.608 * x

```
r <-0.67
x_mean <- 107.2
Sx <- 10.37
y_mean <- 171.14
Sy <- 9.41
slope <- r * (Sy / Sx)
slope
```

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

```
intercept <- y_mean - slope * x_mean
intercept
```

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

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

The slope means for each additional cm increase of shoulder girth, the height of that individual will increase by 0.608 cm.
The intercept means when should girth is 0 cm, the individual’s height should be 105.9651 cm. It does not make sense, since shoulder girth can never be 0 cm. The y-intercept serves only to adjust the height of the regression line.

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

R^2 equals 0.4489. It means about 44.89% of the variability in height is explained by shoulder girth.

```
r ^ 2
```

```
## [1] 0.4489
```

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

166.7651 cm

```
y = 105.9651 + 0.608 * 100
y
```

```
## [1] 166.7651
```

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

residual = 160 - 166.7651 = -6.7651 cm. A negative residual means that the model overestimates the height of individuals.

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, this calculation would require extrapolation.

7.30 Cats, Part I. (a) Write out the linear model.

y = -0.357 + 4.034 * x

b. Interpret the intercept.

The expected heart weight is -0.357 g for someone whose body weight is 0 kg. This number does not make sense, it is just serves to adjust the height of the regression line.

c. Interpret the slope.

For each additional kg increase of body weight, we expect the heart weight to increase by 4.034g.

d. Interpret R2.

R^2 means the heart weight explains 64.66% of the variability of body weight.

e. Calculate the correlation coefficient.

0.804

```
R2 <- 0.6466
R <- sqrt (R2)
R
```

```
## [1] 0.8041144
```

7.40 Rate my professor. (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.

Slope = 0.1325

```
x_mean <- -0.0883
Sx <- 0.0255
y_mean <- 3.9983
Sy <- 0.0322
intercept <- 4.010
slope <- (y_mean - intercept) / x_mean
slope
```

```
## [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.

Yes, because the probability that we ramdonly obtain a t value greater than 4.13 at degrees of freedom 461 (463 - 2) is closed to 0. Therefore, The result, the slope of the relationshio between teaching evaluation and beauty is positive, is statistically significant at alpha = 0.01 level.

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 scatterplot shows linear relationship, and the residual plot shows no patten, and all the data points are evenly distributed on both side of the 0 line.

Nearly normal residuals: The histogram of the residual shows unimodel and bell shaped distribution that is quite symmetric. The normal probability plot also indicate normal distribution of residuals.

Constant variability: The variability of residuals around the 0 line is roughly constant. No pattern or fan shape observed.