Problem Set 8

Daniel Deneau

6/28/2021

Problems 8.8, 8.22, 8.30, 8.36

8.8

a) 2
b) 4
c) 3
d) 1

8.22

a) The relationship between calorie count and grams of carbohydrates in Starbucks food menu items is linear and positive, but not very strong. The residual plot shows an increase in the variation of carbs as calories increase. The
histogram of residuals is slightly skewed right, with a center around -5.

b) Here, the number of calories (explanatory) is meant to predict the number of carbohyrdates (response), for a given menu item.

c) Even though the relationship between carbs and calories is not perfectly predictable, the data shows a linear trend. We can model i with a regression line to predict the number of carbs in something, given the number of calories it has, with some error.

d) Linearity: The trend is generally linear. The condition is met. Nearly Normal residuals: The residual histogram does look nearly normal except for different spreads on either side of the center. The condition is barely met. Constant variablity: We actually see an increase in variability of carbs as the calorie count increases. This condition is not met. Independent Observations:Each menu item has its own nutritional information. This condition is met.

Almost all of the coniditons are met except for the constant variability, so we should not fit the data using a least squares line.

8.30

This point at about (53, 28.5) does have high leverage, as the trend line would definitely be pulled downwards towards it, away from the small cloud of points around (70, 30). If we fit a line without that one point, the trend line would run a little higher, and slope downless strongly, since it would not have to account for that one point. Therefore it is an influential point.

8.36

a) The predicted head circumference for a baby whose gestational age is 28 weeks is: 25.75 centimeters.

```
head_circumference <- function(gestational_age) {
  print(3.91 + (0.78 * gestational_age))
}

head_circumference(28)
```

```
## [1] 25.75
```

b) At a significance level of alpha = 0.05, we can reject the null hypothesis and say that there is strong evidence for an signigicant association between gestational age and head circumference

$p = 0.026$ $p < 0.05$

```
null_val = 0 ## Null value for a slope of zero if there is no correlation
b_i = 0.78  ## Point estimate
SE_b_i = 0.35 ## Standard error for point estimate

### Compute the t-value associated with df = 23:
t_star_df = (b_i - null_val) / SE_b_i
2 * (1 - pnorm(t_star_df))
```

```
## [1] 0.02584243
```

8.16

a) y_hat = x_o + 5000
b) y_hat = x_o + 0.25*x_1
c) y_hat = x_o - 0.25*x_1

8.20

It would be an under estimate since our observed value was higher than our predicted value to get a positive residual. ### 8.26 a) y_hat = -0.357 + 4.034 * body_wt b) When the heart weight is 0, the expected body weight is -0.357. c) For every increase in the body_wt, heart weight goes up by 4.034. d) We are 64.66% sure that the values observed are not due to chance, and that the variabilities of the two variables are associated. e) R = 0.804

```
R_sq = 0.6466
r <- sqrt(R_sq)
r
```

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

8.32

a) The relationship between cans and BAC is linear, positive, and strong.
b) y__hat = -0.0127 + 0.0180*beers For every increase in the number of cans, BAC increases by 0.0180
c) Null: Slope of line is 0, there is no association between number of cans and BAC Alternate: Slope of line is not 0. There is some association.
d) There is a 79.21% chance these an association is not due to chance. And the variation in can count explains BAC.
e) The relationship would likely be weaker since the number of drinks would be an underestimate in a lot of cases where people self-report a lower number than observers would record.

```r
r_sq <- 0.89^2
r_sq
```

```
## [1] 0.7921
```