Tobias Le

Stat21 Midterm Corrections

1. b, d
2. a, c
3. b
4. c, d
5. a, b
6. Step 1: The predictor variable is the APR payments made on my credit card. The response variable is the APR at the time of my payments. We model their relationship.

Step 2: We derive our model by computing least squares estimates for the intercept and slope of

$\hat{l} = \beta_0 + \beta_1 x + \varepsilon$,

where l: APR rate, x: payments on credit card, ε: random error

Step 3: To assess the fit of the model, we compute residuals and investigate the residual vs. fitted-values plot as well as the Normal quantile plot (we look for normality).

Step 4: With the estimated model, we can determine confidence intervals for our slope, mean response. For instance, we can test if there truly is a statistically significant linear relationship between our variables ($H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$)

7. Full points received.

8. a) We observe that evidence for linearity of the variable's relationship in the first test graph. Using the second, we conclude that our error component has constant variance since the data points look uniformly scattered. Lastly, the studentized residuals plot shows a Normal distribution of residuals across the quantiles.

Scanned with CamScanner

h) Full points received.

9 a) The estimated standard deviation is 30.84 calories (derived as residual standard error in R).

b) Our runner expects to burn ~~80.~~ 80.82 more calories for each additional mph increase in average running speed.

c) We use a 2-sample t-test to determine if there's a difference in our sample means. Define μ_{runner} and $\mu_{same\text{-}age}$ as the mean rate of burning calories of our runner and of a population of individuals in the same age group as our runner, respectively. Thus, we test:

$H_0: \mu_{runner} = \mu_{same\text{-}age}$ against $H_1: \mu_{runner} \neq \mu_{same\text{-}age}$

Since $[80.82 - 22.51, 80.82 + 22.51] = [58.31, 103.33]$ ~~contai~~ contains values that fall within 1 standard deviation of the estimated average rate of burning calories of our runner (this interval includes 100), we can say that there is not a significant difference in means.

d) The R^2 value of this model is 0.4313. This coefficient of determination measures how close our data is to our fitted values. In this case, ~~there is not a strong~~ our model does not predict very well.

~~Furthermore~~ However, the p-value of our slope is 0.00225, which is much smaller than 0.05, indicating that there's a high probability that our model correctly depicts the linear relationship between our variables.