Exam 1 Test Corrections – Oliver Clackson (11:30)

1) B and D

5) B and A

6)

- Choose – Decide if a linear model is the best type of model for the relationship in question, given the nature of the variables (are they categorical / quantitative).
- Fit – Which variable is our predictor? Which is the response variable? Once decided, calculate the least-squares regression line for the predictor against the response, *specifically determining the 3 parameters involved in a SLR model: Beta-0, the least-squares intercept, Beta-1, the slope of the least-squares line, and sigma-hat-squared, the variance of the error within the supposed model.*
- Assess – Check the conditions of the model. Does the relationship appear linear; are the residuals centered about 0; are they independent, normally distributed; was there sampling bias? Further, utilize statistics such as R^2 to assess how well our model explains the variability in the response data.
- Use – What conclusions can we make using the model? Is there a relationship between predictor and response? What kind of relationship?

7)

Relationships that contain a large amount of error cannot be fit well with a linear regression model. When the error within a model is large, the explanatory power of the model is small (i.e., it accounts for less of the total variability in the response variable). In this way, there is an inverse relationship between the SSM and SSE terms within a given model.

In the comic and its supposed dataset, the presented model does not account for much of the variation in the response variable, and thus the residuals would be quite large if one were to compare the model's predicted values to the accordant actual values. SSM directly corresponds to r^2, which represents the variability in the response variable that can be explained by the model, which in this case is low. SSE represents the size of the residuals, which in this case are large. Knowing that there is an inverse relationship between SSM and SSE, we can conclude that, in this comic model, SSM would either be equal to or less than the model SSE.

8a)

- Linearity: Appears to be a linear relationship.
- 0 Mean: Residuals centered about 0.
- Constant Variance: Generally random distribution. Some instances of largely differing residual variances at different predicted prices, but sufficient to accept.
- Independence: No reason to suspect bias.

- Residual Normality: Quantile plot looks normal, but histogram is bimodal and has gaps between center values and outer values. Cannot accept.
- SRS: Yes-- given as an SRS.

9a) 30.84 calories

9c) Use a 1-sample T-test for means

Use this type of test because we are looking to see if an observed numeric value (i.e., not a proportion) provides significant evidence to suggest that the parameter is different than a presupposed value. In this case, we are looking to see if our runner's observed rate of caloric burn per mph increase in speed is significant evidence to suggest that the runner's actual rate of caloric burn is different than the populational increase in caloric burn per mph increase, 100 calories/mile increase.

H0: mu = 100
HA: mu ≠ 100

Where mu represents the actual average increase in caloric burn per 1 mph increase for *our runner.*

9d)

- R^2 represents the proportion of the total variability in the response variable that can be explained by the explanatory variable (and thus, our model). The greater the R^2 value, the better the model.
- The p-value of the t-test for slope performed on Beta-1-hat (aveSpeed) is significant at alpha = 0.05. This is sufficient evidence to suggest that there is a relationship between the explanatory variable and predictor variable, and thus that our linear model is a good fit for the data.