Jake Bohman STAT 021 Test 1v2 Corrections

1) B, C

2) A, C

5) A, B

6)

1. The relationship to model is APR vs payments made, with APR being the predictor and payments made being the response, since we are finding if APR at time of payment has an effect on payments made.

2. Since we are using a linear relationship, we need to find least squares estimates for the slope and intercept, because they are important for the equation

$$\hat{Y} = \beta_0 + \beta_1 X$$, where $\hat{Y}$ is the predicted payment, X is the APR at time of payment, and β_0 is the intercept, and β_1 is the slope.

3. First, we must look at the residual vs fitted values plot to assess linearity, zero mean of errors, and uniform spread of errors. A normal quintile plot is needed to assess the normality of the errors. Assessing the randomness and independence requires critical thinking regarding the process by which the data was collected and the ~~environment~~ population.

4. Given the CLT holds, we can use a confidence interval to find ~~the mean~~ a range of values between which we can have a certain confidence (likely 95%) that the mean response falls inside.

8)

a. The linearity assumption holds because the points are scattered above and below zero at random on the residual plot.
Looking at the normal quintile plot, the normality assumption seems to hold true as the ordered residuals are mostly linear.
The constant variance assumption seems to mostly hold, as the points on the residual plot mostly fall within the same interval, but there are some problematic areas, such as when the predicted price is near 350.
There is no evidence that diamonds have any effect on each other, so the independence assumption probably holds, and the randomness assumption holds because the data was obtained through a SRS.

b. We would not expect the behavior to change, because adjusting to a new scale does not change the shape of the distribution of the residuals

9)

a. $\hat{\sigma} = 30.84$ The standard deviation for the number of calories burned is 30.84 calories.

b. $\hat{\beta}_1 = 80.82$ Our runner can expect to burn 80.82 more calories for each mph increase in running speed.

c. I would use a hypothesis test, assuming that the assumptions of randomness, independence, and the CLT are met for a confidence interval. If the 95% confidence interval for the slope includes 100, then the null hypothesis is not rejected and there won't be enough statistical evidence to conclude a difference, but if it doesn't then the null hypothesis is rejected, meaning that there is enough evidence to determine a difference with 95% confidence.

β_1 is the slope of the regression, or the calories burnt for each mph increase in our runner's speed.

$$H_0: \beta_1 - 100 = 0$$

$$H_A: \beta_1 - 100 \neq 0$$

d. R^2 (0.4313) is useful to determine the predictive value of the model, as it is the proportion of the response variable that can be accounted for by the model.
The p-value for the test of β_1 (.00225) is useful since it shows whether there actually a relationship between the predictor and response as it shows whether the slope is likely zero or not.