5. A and C

6. Step 1:

Choose which variable is the predictor and response variables. Here, the APR at time of payment is the predictor (X) and the payments are the response (Y)

Step 2:

Start by putting the data into a scatterplot (two quant variables) to determine if theres a linear relationship. Using ~~[illegible]~~ least-squares error method, fit a regression line to the data, which gives us $Y = \beta_0 + \beta_1 X$, where Y is payments and X is APR.

3. Create a fitted vs. residual plot to assess the fit and check linearity and constant variance assumptions hold. Use a normal quantile plot to determine normality of errors (should be on straight line)

4. Assuming the conditions are met, we can examine the relationship between these two variables and make informed decisions about credit card spending. We can construct confidence intervals for the slope of our regression equation or a confidence interval for the mean response.

Q7:

We know that $SSE = SS_{Tot} - SS_{mod}$,

Therefore, $SS_{mod} < SSE$ because

$R^2 = \frac{SS_{mod}}{SS_{tot}}$ and $R^2 = 1 - \frac{SSE}{SS_{tot}}$.

Therefore, we know SSE is much bigger than SS_{mod} because R^2 is so small

Q8: a.) The normal probability plot tells us that the ~~[illegible]~~ normality condition is met because the points fall along the straight line. The linearity and constant variance conditions are met because there are points above and below the ~~[illegible]~~ mean of the residual plot and there is no clear pattern, such as a fan shape. The data is a simple random sample so the randomness condition is met. The standardized residual plot also confirms the normality condition. The independence condition should be met because the size of one diamond should not have an effect on the size others. However, we can't be 100% sure (maybe they are cut from same rock)

b. No, because assuming the data transformation is constant and the zero-mean condition is met, the plots will reflect the same data patterns with a new form of measurement

Q9: a. The estimate for the standard deviation on # of calories burned based on the SLR is 30.84 calories.

b. We could conduct one-sample t-test to compare the mean rate of burning calories of our runner to the average for all people in the age group.

The $H_0: \mu - \mu_0 = 0$

$H_a: \mu - \mu_0 \neq 0$

The null states that the means are equal and the alternative hypothesis states that the means are different. We will calculate a t-score, and compare it to the critical value. If the t-score is greater than the critical value, we will reject the null hypothesis and accept H_a.