Stat 21 Test 1 Corrections.

3/4

Multiple Choice:

2: A, C

5: A, C

Short Answer

6.

1. We are looking at the relationship of payments made on the CC and APR. In this case, APR is the predictor variable and payments are the response.

2. To fit the data to the model, we would want to collect an SRS of CC payments and APR at time of payments from a population of interest. We want to make sure data follows a normal distribution when fitting it.

3. ~~To assess fit, we want to check the 6 conditions needed~~ We first want to create a residuals vs fitted values plot to assess fit/see if our linearity and constant variance conditions hold true. We also need to make sure that we do a normal quantile plot to see if there is normality w/ the errors.

4. W/ all of our conditions being met and the model being ~~ass~~ already assessed for fit, we could then construct a confidence interval for slope, which would help us answer statistical questions based on the CI bounds.

7.

We know $SS_{mod} < SSE$, and that $SSE = SS_{TOT} - SS_{MOD}$. We also know $R^2 = \frac{SS_{mod}}{SS_{tot}} = .06$, which tells us $SS_{TOT} > SS_{mod}$ by a large margin because R^2 is so small. This therefore tells us that SSE is much larger than SS_{mod} b/c ~~SStot~~ SS_{TOT} is so much greater than SS_{mod}. As $SS_{TOT}\uparrow$, $SSE\uparrow$ w/ SS_{mod} constant

8.

a. We know that the two conditions for inference with an SLR is that the data is collected randomly and that it approximates a normal distribution. We can check the normality condition by checking the plot above ~~as it takes~~ which shows that it takes on an approx. normal distribution, and the normal probability plot shows that the data points lie along the line and there is therefore normality. As the data is from a SRS, we know that the randomness condition is met. We also need to check linearity condition, which can be done by observing that there isn't a distinct pattern on the residual plot. We also see that the constant variance condition holds from the residual plot, as the data points are distributed above and below its average. We know that there is independence amongst observations, as the size of one diamond does not determine the size of another (unless they are cut from the same stone).

9.

c. You could conduct a one-sample t-test for means to see if a sample of ppl the same age as our runner have the same average rate of calories burned for each MPH increase as our runner. ~~$H_0: M = M_0$, H_A~~ $H_0: M - M_0 = 0$, $H_A: M - M_0 \neq 0$, which says our null hypothesis is that there is no difference in average calorie burning, and our alternative says that there is a difference. Run a hypothesis test and calculate the t-score, then find and compare it to its critical value, and if t-score > critical value you can reject H_0.

d. The R^2 value of .4313 helps us determine if the model is a good fit, and specifically shows us that 43.1% of the variation in the burned calories can be explained by the difference in MPH ran. The p-value of .002255 can also help us determine if the model is a good fit, and specifically shows us that there is a very low probability of our test statistic being a result of pure chance.