Test 1 corrections

Question 2: b, a, d

Question 3: c

Question 6:

1. Choose which relation to model

-A simple linear regression model where the predictor is the APR and the response variable is the payments made on the credit card.

The simple linear regression model will summerize the relationship between payments made on the credit card account and the APR at the time of your payment.

Function: $Y = \beta_0 + \beta_1 X + \epsilon$

Slope: on average, the payments made on the credit card either increases or decreases (we don't know value) by a # (β_1) for every additional APR unit.

2. Fit the data to your chosen model.

We want to calculate for the estimates for intercept and slope.

We will use the least squares regression technique to fit the model to the data.

So we will choose estimates to minimize the sum of the squared predication errors.

$\hat{\beta}_0$, $\hat{\beta}_1$: denote estimatted (fitted) values of parameters and $V(\varepsilon)$

3. Assess the fit of the model

Different ways to assess the fit of the model include plotting residuals vs fits plots, normal plots, and R^2

Residuals measures the discrepency in predicting each response. So the residuals vs fits plot will allows us to observe linearity and constant varience. The normal quantile plot (observed data vs expected values) will help us assess normality. If the data points bends away from a straight line, the distribution is either skewed or long tails indicate extreme outliers

The coefficient of determination, R^2, indicates how well the model fits the observed data. We want a high R^2 because it indicates a better fit for the model.

4. Use the estimated model to answer statistically questions

Statistical questions include confidence interval for mean response, confidence interval for slope, and prediction interval for response.

Question 7 :

From the comic, we are given that $R^2 = 0.06$, meaning only 6% of the data lies close to the line. Thus there is no relationship between the two variables. The sum of squared errors, SSE, is important to calculate the coefficient of determination, r^2. Since the r^2 given is very low due to the lack of correlation betwee the predictor and response. Since r^2 is low, SSmod and SSE terms would not have an indication of linear relationship. Thus $SS_{mod} < SSE$.

Question 8:

Normal probability plot : most of the points lie close to the line except for a few points at the ends. The studentized residual help us detect those outliers. So the normal condition is not met

✓ Varience: based on the residual plot shape, it is enough to assume varience

✓ Linearity: based on the residual plot it looks linear

✓ Randomness: They said that the data represents a simple random sample. So we can assume that the randomness condition is met

✓ Independent: If it was indeed random sample (randomness met), then we can assume Independence is met.

Question 9 :

9a) The estimate for the standard deviation of the numbers of calories burned based on this linear model is 30.84 because Residual standard error : 30.84

9c)

Hypothesis:

$H_0: \beta_1 = \#$ ← calories burnt   or $\beta_1 = \beta_2$

$H_A: \beta_1 \neq \#$ $\beta_1 \neq \beta_2$

Where β_2 represents the average rate of burning calories for all people in the age group

β_1 represents our runner's rate of burning calories

\# represents the average rate of burning calories for all people in the age group

Apply a t-test to get a p-value

if p-value is less than $\alpha = 0.50$, then we reject null hypothesis in favor of the alternative, that there is indeed a difference

If we build a 95% confidence interval for the difference in rate of burning calories of our runner & all people in the age group.

If it does not contain zero, then there is a difference