Yasharbek Sabitov

ysabitol

Test corrections

1) a and b

to

2) a and c

The p-value of 0.003 doesn't signify whether there is a large or a small difference between the groups but rather how statistically significant the occurrence of white offenders having a smaller incarceration rate is. Meaning, if white offenders do in fact have a smaller incarceration rate, then this difference is probably not caused by mere chance.

5) A and B

Normality and Randomness need to be assessed when conducting hypothesis tests or providing confidence intervals. Neither choices involve that.

6) Step 1: I would use the SLR to model the relationship between my payments, or the predictor, and the APR at the time of my payment, or the response

Step 2: I would then calculate B_0 and B_1 to graph my least squares regression line.

Step 3: I would now plot a residual vs. fitted values plot to test for linearity and constant variance and a normal quantile plot to see if my data comes from a normal distribution

Step 4: Depending on what we want to find, we now can use intervals. For example, we can use a confidence interval to find the mean APR or a prediction interval for a single APR value.

7) The total variation (SSTotal) is explained by the model (SSModel) and consists of an unexplained part due to error (SSE).

$$SSTotal = SSModel + SSE$$

8) a) By looking at the residual vs. predicted plot, we can observe that the data is relatively linear as there is no obvious curvature. Additionally, the points on the residual plot are generally scattered randomly (it seems that some points on the left side are a bit more closely concentrated but generally the spread is random), so there is a relatively consistent variance. Additionally, the normal probability plot shows a fairly consistent linear trend, which supports the normality condition. With these factors in mind, we can conclude that our data meets the independence and randomness assumptions needed for inference

b) I wouldn't expect the overall behavior of any of the plots above to change as this is just a unit conversion.

9)a) The estimate for the standard deviation of the number of calories burned is approximately 30.84 calories.

b) They should expect to burn roughly 80.82 calories per mph

c) Since our job here is to compare the average calories burned of the runner (sample) and the average calories burned for any person within the same age group as our runner (population), we will have to perform a One-Sample t-test

$$t = \frac{m - \mu}{s/\sqrt{n}}$$

where;

t = t-statistic

m = mean of the group

μ = theoretical value or population mean

s = standard deviation of the group

n = group size of sample size.

This would serve as a hypothesis test to either prove or disprove the null hypothesis (there is no difference between runner's rate of burning calories and that of the population).

We could also perform a confidence interval for the sample mean to see if it covers the population mean

d) R^2 tells us the amount of the variance of the response that is explained by our model and if the model is a good fit.

Additionally, the p-value from the ANOVA f test will tell us whether or not we have sufficient evidence to say that there is a statistical significance for our data and whether we should reject the null hypothesis or not.