

Test Corrections I

1. The correct answers for this problem are a and c. Because the 90% confidence interval the doctor calculates is between 37.2 and 40.1, it is likely that the prevalence of lung cancer in the area is much higher than the state-wide average, as all values contained within the confidence interval are significantly greater than 31.6. The meaning of a confidence interval is that it contains the true population parameter $1-\alpha\%$ of the time, which is why c is also correct.
2. I wrote that b was a valid interpretation. d is also a valid interpretation because the definition of p-value is that if there is no real difference between the two datasets, the probability we observe a difference as large or larger than observed in this data is the p-value.
5. The question asks which studies involve inference. If a study uses inference, it is either using a hypothesis test or calculating a confidence interval. Out of these options, only c, where the nutritionist is calculating a confidence interval for slope, and d, where the car salesperson is using a t-test to determine if a slope is statistically significant, are using inference, and therefore need to assess the normality & randomness conditions.

- 6
1. I would determine which variable I believe is a predictor & response. In this case, payments made on the account seems to be the predictor variable, and APR at time of payment is the response.
 2. I would next fit the data to a model to get estimates for slope & intercept.
 3. Next, I would create a Fitted values vs residuals plot and check for the linearity & constant variance conditions. Then, I would create a normal quantile plot to determine if the normality conditions are met.
 4. I would use the confidence interval for slope and mean response to answer statistical questions.
- 8 a) Based on these plots, we can conclude that the linearity, constant variance, and normality conditions are met. The residual plot shows even spread above & below the x-axis and no strong funneling, allowing us to conclude that the linearity and constant variance conditions are met, respectively. The normal probability plot shows a relatively linear trend, so the normality condition is met. (cont. →)

The data represents a simple random sample, so we can conclude that the randomness assumption is met, and we have no reason to believe that the size of each diamond is not independent.

9a 30.84 calories

d R^2 , or the coefficient of determination, tells us that the model explains approximately 40-43% of the variation in the data. The p -value for the β_1 test is $0.0023 < 0.05$, which tells us that the correlation we observe is unlikely to be simply from random chance, and β_1 is significantly different from zero.