STAT 270 Assignment 4

Upload to Crowdmark by 5pm on Monday, July 13, 2020

Instructions: You must provide any code you use (final computed values are not sufficient). **I strongly recommend that you create a Jupyter Notebook for questions that require data analysis**, embedding your written comments in Markdown cells. Create separate file(s) for each question (in .jpg, .pdf, or .png format) for uploading to Crowdmark.

1. (9 marks) The data in the file `baby.txt` consist of the weights (in grams) of 59 babies born at a local hospital. Assume that the weights are iid and normally distributed.

   a. (4 marks) Construct a 90% confidence interval (CI) for the mean weight of babies in the community. You must show your work, i.e., explicitly compute the components of the CI (sample mean, appropriate quantile, etc.). You may use R to compute these components.
   b. (1 mark) Provide an **informal** interpretation of your CI in a).
   c. (2 marks) Provide a **formal** interpretation of a 90% CI in this setting.
   d. (2 marks) Say you wanted to repeat the study to obtain a 90% CI of maximum width 100 grams. Using the sample SD of weights from the first sample as the population SD (for the purposes of this calculation), how many babies would you need to sample?

2. (9 marks) The data in the file `fishdata.txt` consist of the weights (in grams) and diagonal lengths (in cm) of 159 fish observed at local fish markets. Researchers are interested in whether the mean length of fish is less than the historical mean of 4.5 cm. Assume that lengths are iid and normally distributed. Use a significance level of 5%.

   a. (2 marks) State the null and alternative hypotheses.
   b. (1 mark) What is the observed test statistic?
   c. (1 mark) What is the distribution of the test statistic under the null hypothesis?
   d. (1 mark) What is the p-value?
   e. (2 marks) State your conclusions in the language of the problem.
   f. (2 marks) Provide a formal interpretation of your p-value in d).

3. (7 marks) A chainsaw bar manufacturer is concerned that the mean diameter of the tooling holes in the bars has shifted from its target, 3824 thousandths of an inch. They randomly select 20 bars and measure the diameter of one tooling hole per bar. The diameters are recorded in the file `tooling.txt`. Assume that diameters are iid and normally distributed. Use a significance level of 1%.

   a. (2 marks) State the null and alternative hypotheses.
   b. (1 mark) What is the observed test statistic?
   c. (1 mark) What is the distribution of the test statistic under the null hypothesis?
   d. (1 mark) What is the p-value?
   e. (2 marks) State your conclusions in the language of the problem.

4. (6 marks) An airline has a policy that checked pieces of baggage may weigh up to 50 lbs. The distribution of weights of individual pieces is left-skewed with mean 37 lbs, median 42 lbs, and standard deviation 10 lbs.

   a. (1 mark) What is the probability that one piece of baggage will exceed 42 lbs? If you don't have enough information to answer this question, explain why.
   b. (2 marks) What is the probability that 3 pieces of baggage will exceed $42 \times 3 = 126$ lbs? If you don't have enough information to answer this question, explain why.
   c. (3 marks) What is the probability that 100 pieces of baggage will exceed $42 \times 100 = 4200$ lbs? If you don't have enough information to answer this question, explain why.

5. (3 marks) If $X \sim N(\mu, \sigma^2)$, what is the distribution of $-X$? Explain.

6. (4 marks) On a cold winter's day, let the temperature (in ºC) in a randomly selected SFU classroom be X. Let $E[X] = \mu$ and $Var[X] = \sigma^2$. Recall that X can be converted from ºC to ºF via the following formula:
$$X^* = \frac{9}{5}X + 32.$$

   a. (2 marks) What is the covariance of X and X^*?
   b. (2 marks) What is the correlation of X and X^*?

7. (7 marks) Consider a random selection of 100 patients who undergo treatment for high blood pressure. Let Y_{i1} be the systolic blood pressure of the i^{th} patient prior to treatment, let Y_{i2} be the systolic blood pressure of the i^{th} patient following treatment. Assume that the Y_{i1}'s are independent with mean μ_1 and variance σ^2. Assume that the Y_{i2}'s are independent with mean μ_2 and variance σ^2. (As a result, the random variables $D_i = Y_{i2} - Y_{i1},\ i = 1, \ldots, 100$ are also independent.) Assume that $Corr(Y_{i1}, Y_{i2}) = \rho$ for all i. Let $\bar{D} = \frac{1}{100}\sum_{i=1}^{100} D_i$.

   a. (2 marks) What is the expected value of $\bar{D}$?
   b. (3 marks) What is the variance of $\bar{D}$?
   c. (2 mark) Is $\bar{D}$ approximately normally distributed? Explain.