Fall 2020 **Statistics 330** **Practice Exam 3**

(– points)

Please write your first and last name here:

Name ____

Instructions:

This practice exam contains the type and style of questions you will find on Exam 3. There are many questions that could be asked, so this is not exhaustive. Refer back to Module 5 homeworks for more type's of questions that could be asked.

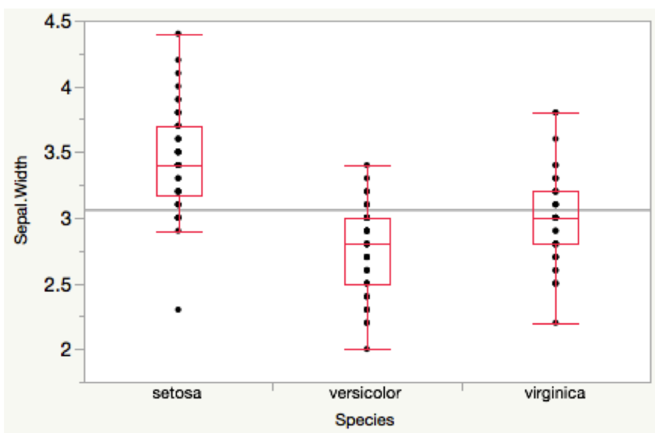
1. Consider the data set: 7 18 25 20 32 24 24

   (a) Calculate the mean, median and IQR (6 points)

   (b) Check for outliers using the 1.5(IQR) rule, and indicate which data points are outliers. (4 points)

   (c) Remove any outliers and recalculate the mean, median, IQR. If there are no outliers, then you can say "answer same as (a)" (6 points)

2. The boxplots below show the sepal width for the three different species of *Iris*: Setosa, Versicolor, and Virginica. Answer (a)-(c) about the boxplots. (4 points)



   (a) Which of the 3 species has largest maximum sepal width?

   (b) Which of the 3 species has the largest range?

   (c) Which of the 3 species has the smallest IQR?

   (d) Which of the 3 species has the smallest median sepal width?

3. Suppose $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ for $i = 1, 2, 3$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Consider two estimators: (1) X_1 and (2) $\bar{X} = \frac{X_1+X_2+X_3}{3}$

   (a) Calculate the Bias for each estimator. (4 points)

   (b) Calculate the MSE (mean squared error) for each estimator. (4 points)

   (c) Which estimator is preferred? Use MSE in your argument. (4 points)

4. Let $X_1, \ldots, X_n$ be a random sample from the Gamma distribution with $\alpha = 4$ and λ unknown. The probability density function (pdf) for X_i is:

$$f_X(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x}, \quad x \geq 0.$$

   (a) Find an estimator of the parameter λ using the method of moments. (8 points)

   (b) Find the maximum likelihood estimator of λ. (8 points)

5. Suppose engineers are designing a new golf club to replace a company's older model. An experiment was done where shots were hit independently with the new model and the old model using a robotic arm to simulate "standard" conditions. Let μ_1 be the true mean distance of the new model and μ_2 be the true mean distance of the old model. The following is the collected data of the shot distances (in yds). (Assume Sample sizes are large for normal based tests)

Model	mean	std. dev	n
New	$\bar{x}_1 = 293.3$	$s_1 = 8.1$	$n_1 = 45$
Old	$\bar{x}_2 = 288.2$	$s_2 = 10.2$	$n_2 = 40$

(a) Consider a hypothesis test to test whether the true mean distance of the new model is different than 290 yds.

i. Give the null and alternative hypotheses. (4 points)

ii. Calculate the test statistic. (3 points)

iii. Calculate the $p-$value. Give your decision and conclusion. (3 points)

(b) Consider a hypothesis test to test whether new model hits farther than the old model on average.

i. State the null and alternative hypotheses. (4 points)

ii. Calculate the test statistic. (3 points)

iii. Calculate the $p-$value. Give your decision and conclusion. (3 points)

6. Researchers comparing the effectiveness of two pain medications randomly selected a group of patients who had been complaining of a certain kind of joint pain. They randomly divided these people into two groups, then administered the pain killers. Of the 112 people in the group who receive medication A, 84 said this pain reliever was effective. Of the 108 people in the other group who received medication B, 66 reported that pain reliever was effective.

   (a) Let p_A = true proportion of people who say pain reliever A is effective and
   p_B = true proportion of people who say pain reliever B is effective

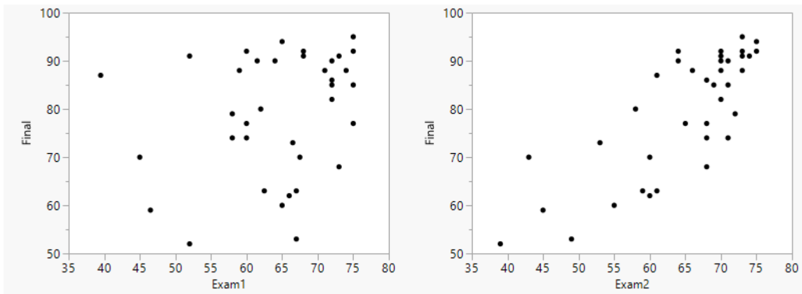
   If a hypothesis test was to be conducted to see if there is evidence that pain reliever A is "better" than pain reliever B, what should the Null and Alternative hypotheses be? (4 points)

   (b) Construct a 98% confidence interval for the difference in the proportion of people who find the two medications effective. Write the interpretation of this interval. (8 points)

   (c) Based on the confidence interval you made, is there evidence of a difference in the effectiveness of the two medications? Explain.

A statistics professor wants to see if they can predict a student's final exam score based on their exam score from an earlier exam. Data from a past class was gathered and the variables from $n = 40$ students were: x_1 = exam1 score, x_2 = exam2 score, and y = final exam score.

(a) The professor first makes some scatterplots to see which x variable might be a better predictor of final exam score. Based on the scatterplots, which x appears to be a better predictor of y? Why? (x_1 vs y on the left, x_2 vs y on the right) (2 points)



(b) The least squares regression line using x_1 as a predictor of y is $\hat{y} = 48.35 + .492x_1$. Based on the following statistics, give the least squares regression line for using x_2 as a predictor of y. (3 points)

$$\sum_{i=1}^{40} x_2 = 2616, \quad \sum_{i=1}^{40} y = 3221.2, \quad \sum_{i=1}^{40} (x_2 - \overline{x}_2)^2 = 3336.94, \quad \sum_{i=1}^{40} (x_2 - \overline{x}_2)(y - \overline{y}) = 3637.26$$

(c) A new student has the following data:

x_1	x_2
68	74

Using both equations, give the corresponding predicted final exam scores for this student. (3 points)