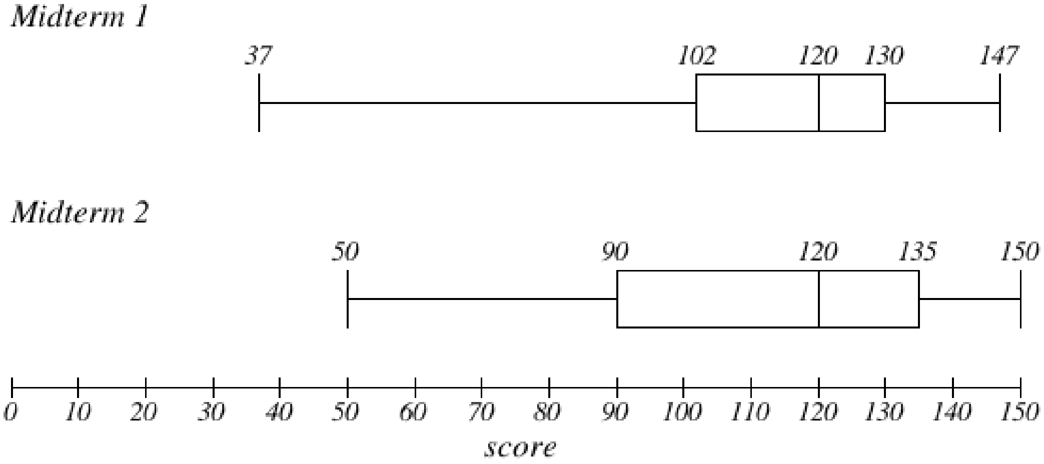
MSDS 6371 Exam 1 – Summer 2015

Answer the questions that follow to the best of your ability. On questions that require hand calculations, please show the formula used and the formula with the correct numbers in the correct places in order to get full credit for the problem. For multiple choice questions 1,2,4, you may simply highlight the correct answer or make it a different color. For short answer questions, answers should be in the 2-4 sentences range.

Use the boxplots below to answer questions 1 and 2.



1. (4 points) The boxplot above shows the grades of students in a statistics class on two midterms. Which midterm has a greater percentage of students with scores at or above 120?
    a. Midterm 1
    b. Midterm 2
    c. Both Midterms are about equal
    d. It is impossible to tell this level of detail from a boxplot

2. (4 points) Refer again to the boxplot above. Which of the following is correct?
    a. The means of both midterms are larger than their medians
    b. The means of both midterms are smaller than their medians
    c. The means of both midterms are about the same as their medians
    d. There is no way to tell the relationship between mean and median from a boxplot

3. (4 points) What does it mean to say that a result is statistically significant?

4. (4 points) An agricultural researcher plant 25 plots with a new variety of corn that is drought resistant and hence potentially more profitable. The average yield for these plots is 150 bushels per acre. Assume that the yield per acre for the new variety of corn follows a normal distribution with unknown mean μ and that a 95% confidence interval for μ is found to be 150 +/- 3.29. Which of the following is true?
    a. A test of the hypotheses H0: μ = 150, Ha: μ < 150 would be rejected at the 0.05 level
    b. A test of the hypotheses H0: μ = 150, Ha: μ ≠150 would be rejected at the 0.05 level
    c. A test of the hypotheses H0: μ = 150, Ha: μ >150 would be rejected at the 0.05 level
    d. A test of the hypotheses H0: μ = 160, Ha: μ ≠160 would be rejected at the 0.05 level

5. (12 points) You have recently taken a job at a research facility, and your first duty is to calculate a sample size for a study. You type the following program into SAS

```
proc power;
onesamplemeans
mean = 3
nullmean = 0
ntotal = .
stddev = 10
power = .8; run;
```

   a. What is the value of the probability of Type II error?
   b. What is the value of the probability of Type I error?
   c. Suppose the standard deviation is decreased to 8. What will happen to the number of subjects, all else staying the same?

6. (4 points) Suppose a researcher writes in a journal article that “the obtained p was p = 0.032; thus, there is only a 3.2% chance that the null hypothesis is correct.” Is this a correct or incorrect statement? Explain your answer.

7. Presentation counts! Keep your analysis to 2 pages (Single sided) including graphs, plots and charts. There should be 1 page max for each statistical test (2 tests total.) Include all statistical symbols such as μ and σ. Finally, remember that a “Test” includes addressing the assumptions, doing the necessary statistical analysis, and writing a meaningful conclusion.
   The data consist of 3974 happiness values recorded from a random sample of 3974 people from Missouri. The higher the score the happier the subject is reported to be. 1996 of the people had a dog (and were thus put in the ‘dog’ group) and 1978 did not have a dog (and were thus put into the “No Dog” Group.) The researcher would like to know if the happiness scores of the dog owners is significantly bigger than that of the non-dog owners.
   a. Obtain the data from Section 7.4 in the Coursework area. The csv file is called “Exam1DogData.csv”.
   b. (10 points) Test (if possible) to see if the mean of the happiness scores of the dog owners is significantly greater than that of the non-dog owners. Test at the alpha = .01 level of significance.
   c. (10 points) Test (if possible) to see if the median of the happiness scores of the dog owners is significantly greater than that of the non-dog owners. Test at the alpha = .01 level of significance.
   d. (5 points) Explain how you could use a permutation test to test if the population median of the happiness scores of dog owners is significantly greater than that of the non-dog owners (Do not do any calculations to answer this part).
   e. (5 points) Which analysis do you feel is more appropriate and why?

8. (2 points) You are more than halfway done! Take a break and check out this website:

   http://en.wikipedia.org/wiki/William_Sealy_Gosset

   List one interesting thing about the man who discovered the Student t distribution. Do not spend a lot of time on this question …you answer should be a very short sentence!

9. (10 points) For this problem you will need to use the cityrate.csv file located in Section 7.4. We are analyzing the interest of auto loans in 5 different cities: Chicago, Dallas, LA, NY, and Phoenix. We want to investigate if there is a difference in mean interest rate between the north and the south. Let Chicago and NY represent the north and let Dallas, LA and Phoenix represent the South.

   Use a contrast to test the claim at the alpha = .05 level of significance that the north has a different mean auto interest rate than the south. Be sure and clearly state Ho and Ha. You may describe the Ho and Ha with respect to μi's or γ or show it both ways. Perform a complete analysis: 1) State the problem 2) Address the assumptions. 3) Conduct the Test 4) Clearly state the conclusion in the context of the problem. Also, provide the SAS proc glm statement for this problem.

10. Still using the city auto interest rate data, we now want to simply compare Dallas and Chicago. Specifically we would like to test if Dallas has a different mean auto interest rate than Chicago.
    a. (5 points) We of course would like to perform the most powerful test available. Describe whether you would use a simple two sample t-test using only the data for the two cities or a contrast to compare these two means and why.
    b. (10 points) Now test the same claim using a contrast by hand. You do not need to actually write it with your hand … but clearly show the calculations you made to carry out the contrast (typing the equations.) You may skip the assumptions checking and simply show your work in finding the 'g', SE(g), t-statistic and p-value. And of course write a short but complete conclusion.

11. (10 points) Let's take a step back now and pretend we had no idea going into this analysis which pairs of cities might be different; so we wish to test all the pairs and see which ones are statistically significant. Use confidence intervals or hypothesis tests to determine which pairs of cities are statistically different. Be sure and address why you chose the methods you chose and defend (if any) assumptions you needed to utilize those methods.