Name: Section:

This midterm is due **July 6th, 2016 at 11:59PM Central Time** and must be turned in by that time to receive any credit for the midterm. There is a space to submit your midterm under Unit 7, in 7.4. There will be **no extensions** on this deadline so be sure to plan ahead. If there is an emergency which prevents you from finishing the exam, you should email me, but as I most likely would not give an extension, you should also turn in what you have at the time.

I would prefer you return the midterm as one Word or PDF document. If you cannot use one of these formats, please contact me before the due date and let me know what format you would like to use to ensure that I will be able to open and grade your submitted document.

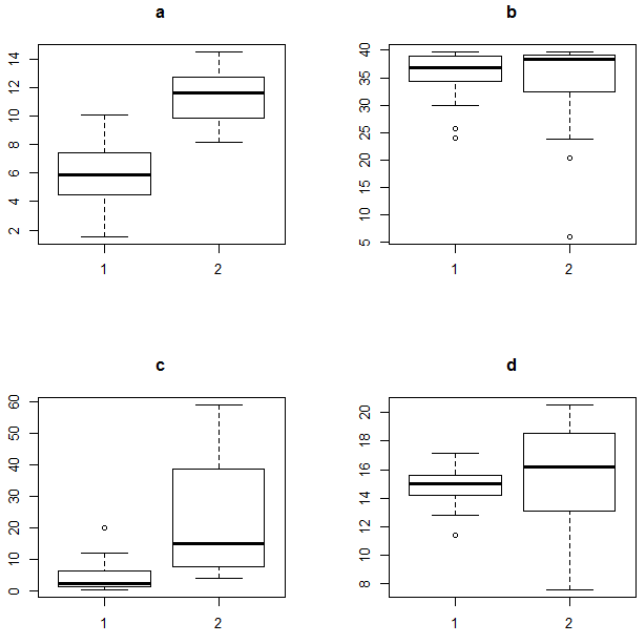
You may not discuss this midterm (or any class material) with other students or anyone else during the midterm**. All questions should be directed to me** (Chelsea: cdallen@smu.edu); the graders and tutor will not answer questions during the midterm.

Turning in this exam is an agreement that you have adhered to the SMU Honor Code and that you have neither given nor received assistance in completing this exam.

Multiple Choice/Multiple Answer (6 points each)

In this section, select the option(s) that best answer(s) the question. On "select all that apply" questions, it is still possible for there to be only one correct answer. There is partial credit available on "select all that apply" questions.

1. The pairs of boxplots on the right correspond to 4 datasets, each with two groups. On each one, we would like to perform a two-sample t-test. For which dataset would a log transformation be appropriate? (Though it is not obvious in this picture, all data points are greater than 0.)
   Select only ONE answer.
   a.
   b.
   c.
   d.



2. Say I am testing the following hypotheses:

$$H_0: \mu_A = \mu_B$$
$$H_a: \mu_A \neq \mu_B$$

   What is the definition of power in this context?
   Select only ONE answer.

a. Power is the probability of failing to reject the null if $\mu_A = \mu_B$.
b. Power is the probability of rejecting the null if $\mu_A = \mu_B$.
c. Power is the probability of failing to reject the null if $\mu_A \neq \mu_B$.
d. Power is the probability of rejecting the null if $\mu_A \neq \mu_B$.

3. Which of the following are true of the rank-sum test?
Select ALL that apply.
   a. It is an alternative to the two-sample t-test
   b. It is an alternative to the one-sample t-test
   c. It is a test for the means of the original data
   d. It is a test for the medians of the original data
   e. It can be used when we have data that we do not have exact values for, but only know to be the greatest or least in the dataset (in other words, we know an observation is more/less than all the other data, but do not know its exact value).
   f. It is robust to outliers.

4. Say we do a hypothesis test and get a p-value of 0.43. Which of the following is/are correct interpretation(s) of the p-value?
Select ALL that apply.
   a. There is a 43% chance that the null hypothesis is true.
   b. There is a 43% chance that the alternative hypothesis is true.
   c. There is a 43% probability that one would get a test statistic as extreme or more extreme than the observed value by chance alone if the null is true.
   d. There is a 43% probability that one would get a test statistic as extreme or more extreme than the observed value by chance alone if the alternative is true.

5. We wish to perform an ANOVA test with a null hypothesis of
$$H_0: \mu_1 = \mu_2 = \mu_3$$
Which is the alternate hypothesis?
Select ALL that apply.
   a. $H_a: \mu_1 \neq \mu_2 \neq \mu_3$
   b. $H_a: \mu_1 \neq \mu_2 \ or \ \mu_1 \neq \mu_3 \ or \ \mu_2 \neq \mu_3$
   c. $H_a: at\ least\ one\ of\ the\ groups\ has\ a\ different\ mean\ than\ the\ others$
   d. $H_a: H_0\ is\ not\ true$

6. The ANOVA form the previous question is performed with significance level $\alpha = 0.05$. The resulting p-value is 0.11. Which of the following is/are appropriate conclusion(s)?
Select ALL that apply.
   a. There is not sufficient evidence (p = 0.11) to reject the null hypothesis. We can conclude that the means are all the same.

b. With a p-value of 0.11, we fail to reject the null hypothesis. There is not sufficient evidence that any of the means are different than the others.
c. There is sufficient evidence (p = 0.11) to accept the null hypothesis; the mean of groups 1, 2, & 3 are the same.
d. With a p-value of 0.11, there is sufficient evidence to reject the null hypothesis. As least one of the means is different than the other two.
e. With a p-value of 0.11, we fail to reject the null hypothesis. Group 1 has the same mean as at least one of the other groups.
f. With a p-value of 0.11, there is not sufficient evidence to reject the null hypothesis. We cannot claim that there is a difference between any of the means

Short Answer (8 points each)

Answer the questions in complete sentences. 1-2 sentences should be enough to answer each question.

7. What is the main advantage of a randomized experiment over an observational study?

8. What would be an example of a situation that required an observational study instead of an experiment? Give reasons why an experiment would not be appropriate or possible.

9. Say we are doing a test with the following hypotheses:

$$H_0: \mu_A = \mu_B$$
$$H_a: \mu_A \neq \mu_B$$

Also, based on the plots of the samples, we determine that a permutation test is the best method. We first find the difference in our sample means: $\bar{x}_A - \bar{x}_B = 5.3$. We then do 1000 permutations of the data and find the difference for each of those samples' means. Here is a table categorizing those differences:

$\bar{x}_A - \bar{x}_B$ range	Number of samples with $\bar{x}_A - \bar{x}_B$ in that range
Less than -5.3	15
Between -5.3 and 0	477
Between 0 and 5.3	486
Greater than 5.3	22

This means, for example, that 15 of these 1000 differences are less than -5.3.
Do you reject or fail to reject? State the conclusion of the test, including the p-value.

Analysis (5 points for each part)

The questions below all use two datasets, which is in the same zip file as this midterm. Use the dataset and your software of choice (and whatever hand calculations you may feel are necessary) to answer the following questions. For each question, include your code or a description of your calculations and a

copy of any tables you refer to. **Please note that I do <u>NOT</u> want a print-out of the data (or the datalines if you choose to put your dataset in that way) or any other tables or plots that you do not refer to in your answers.**

At the beginning and end of most Live Sessions, I ask how confident you feel about that week's material. Let's say I save all that data from my four classes and for each person I take the difference (after class rating minus before class rating). Each dataset contains 60 observations.

Note: this is not real data from any of the classes as I did not want any students to feel their information had been used inappropriately; this dataset is entirely fabricated and not representative of your classes.

10. Naturally, I wish to know if students feel more confident with the material after the Live Session than before (in other words, whether the difference is larger than 0). Use the dataset "Rating.csv" to perform an appropriate t-test and answer the following questions. This dataset contains the following variables (though you may not need them all):
    - Before: the student's rating before the Live Session begins on a scale of 1-10
    - After: the student's rating after the Live Session is over on a scale of 1-10
    - Diff: the difference in the student's rating (after – before)

    a. The assumption for this confidence interval is the same as the assumption for a t-test. What are these assumptions? By looking at the data, do you think this procedure is appropriate? Why or why not?

    b. **Regardless of your previous answer**, find the 95% confidence interval in the software of your choice. Give the confidence interval and its interpretation in the context of the problem. (This interpretation would likely start with "We are 95% confident that…")

    c. As I mentioned, I am specifically curious as to whether the ratings are increasing from the beginning to the end of class. Based on your confidence interval, do you think this is true. Why or why not? (You do not need to run another test or give a p-value. I am looking for a conclusion based solely on your confidence interval from the previous part.)

11. Despite my efforts, each class' Live Session is slightly different. Let's say that I now want to know whether the 4 different classes have different amounts of improvement. To do this, I should perform Analysis of Variance (ANOVA). The new dataset, "RatingClass.csv" includes the following variables (though you may not need them all):
    - Section: a variable giving the class day and whether it is the first or second class that day (M1 = first class on Monday, M2 = second class on Monday, T1 = first class on Tuesday, and T2 = second class on Tuesday)
    - Before: the student's rating before the Live Session begins on a scale of 1-10

- After: the student's rating after the Live Session is over on a scale of 1-10
- Diff: the difference in the student's rating (after – before)

a. Write down the hypotheses for this test. The alternative hypothesis can be in a sentence if you prefer.

b. What are the assumptions for ANOVA? Do you think ANOVA is appropriate for this data? Why or why not?

c. **Regardless of your previous answer**, use ANOVA (with $\alpha = 0.05$) to perform the test in the software of your choice. Give the associated p-value.

d. Give a conclusion. Be sure to include your p-value, whether you reject or fail to reject the null, and your findings in the context of the problem.

e. Use the procedure of your choice to determine which means are different from each other (be sure to tell me which procedure you used) and give appropriate conclusions. You can summarize.

f. **Extra Credit**: Answer the below questions for extra credit points on the test!

i. The next two questions require you to perform two contrasts. Since you want to do two tests, you should adjust the individual significance levels accordingly in order to keep the overall significance level at 0.05. Using the Bonferroni Adjustment to correct for the fact that you are doing 2 tests, what is the individual significance level for each test (in other words, what is the benchmark with which you will compare your p-values)?

ii. Is there a difference between the combined mean of the Monday classes and the combined mean of the Tuesday classes? Also, give the 95% (simultaneous) confidence interval for $\mu_{Mondays} - \mu_{Tuesdays}$.

iii. Is there a difference between the combined mean of the first classes (M1 & T1) and the combined second classes (M2 & T2)? Also, give the 95% (simultaneous) confidence interval for $\mu_{first} - \mu_{second}$.