STAT 300 Midterm Exam

(Time: **50 minutes**, *Friday, October 27, 2017, 12:00 - 12:50pm, II*)

Name ______________________________ **Student ID** ______________________

Lab (circle one): L1A (Tue 8:30am), L1B (Wed, 9am), L1C (Thu 8:30am), L1D (Fri 2pm).

Instructions: This exam is closed notes/books. One "formula/cheat sheet" ($8.5' \times 11'$, two sided) is allowed. Round numerical answers to two decimal points. For Problems 3 and 4, please write on the other side of the page if you need more space, and clearly indicate the problem numbers.

Problem 1 (20 pts, 2 pts each). Circle the (*one*) correct answer: true (T) or false (F).

1. The nonparametric tests in the course are *not* appropriate for large samples. T F
2. The χ^2 test is always "one-sided" in the sense that only large values, not small values, of the test statistic indicate evidence against the null hypothesis H_0. T F
3. When the sample size is large, the null distribution of the Wilcoxon rank sum test statistic can be approximated by a normal distribution based on the Central Limit Theorem. Thus, in this case, the Wilcoxon test is equivalent to the two-sample t-test. T F
4. For some nonparametric methods, we may *not* need to define a test statistic in order to compute p-values. T F
5. Even for large samples, the permutation test does not rely on the Central Limit Theorem. T F
6. The χ^2 test and the QQ plot are not very useful to check the normality of data if the sample size is small. T F
7. ANOVA analysis should always be followed by multiple comparisons. T F
8. A sign test only uses the signs of the data rather than the data values themselves. This will lead to some loss of information even if it is used to test the medians. T F
9. The basic ideas behind the Wilcoxon rank sum test and the Kruskal-Wallis test are similar, although the Wilcoxon test is used for comparing one or two groups while the Kruskal-Wallis test is used for comparing three or more groups. Similarly, the basic ideas behind the t-tests and ANOVA are also similar. T F
10. In a sign test, an exact p-value can be computed without assuming normality and/or large samples. T F

Problem 2 (30 pts, 3 pts each). Clearly circle the (<u>one</u>) most appropriate answer.

1. Which one of the following answers is *not true*: Both ANOVA and t-tests rely on the following assumptions

   A. data are balanced (i.e., roughly equal number of observations in each group)

   B. data are normally distributed or sample size is large

   C. no outliers in the data

   D. none of the above

2. Suppose we wish to compare two programs for weight loss. The data are given below

   ```
   Program I:  1, 3, 4
   Program II: 0, 2
   ```

   The value of the test statistic of a Wilcoxon rank sum test is

   A. 0, B. 1, C. 2, D. 3, D. 4.

3. Referring to the previous question. The value of the test statistic of a permutation test is

   A. 4, B. 6, C. 1.67, D. 3.67

4. In a hypothesis testing problem, if you decide to set the significance level of the test to 0.01 instead of 0.05, then what would you generally expect?

   A. The power of the test will usually increase.

   B. The power of the test will usually decrease.

   C. The power of the test will usually remain the same.

   D. insufficient information to tell.

5. When using the χ^2 test to test the normality of continuous data, if the data are divided into 8 intervals (bins), the degrees of freedom for the χ^2 distribution is

   A. 1 B. 3 C. 5 D. 7

This table is for instructor use only

Problem 1	Problem 2	Problem 3	Problem 4	Total Mark

6. In which one of the following problems, the χ^2 test *cannot* be used?

   A. testing the independence of two discrete random variables

   B. testing the equality of three or more population means (or medians)

   C. testing the normality of continuous data

   D. testing if data follow a discrete distribution such as a binomial distribution.

7. Which one of the following methods will *not* be affected by outliers (i.e., unusually large or small values)

   A. an ANOVA test

   B. the t-test

   C. the permutation test

   D. the Kruskal-Wallis test

8. In a simulation study to evaluate the power of a two-sample t-test for testing equality of two population means, assuming equal variance σ, where many samples are independently generated from models $N(\mu_A, \sigma)$ and $N(\mu_B, \sigma)$ respectively, which one of the following cases may lead to a *decrease* in power (intuitively – no calculation needed)

   A. Increase the sample size

   B. Change values $\mu_A = 10$ and $\mu_B = 12$ to $\mu_A = 10$ and $\mu_B = 13$

   C. Change values $\mu_A = 10$ and $\mu_B = 12$ to $\mu_A = 12$ and $\mu_B = 15$

   D. Increase σ

9. Which one of the following methods for constructing test statistics is *not* used in this course so far

   A. Re-sample from the original data (sample) many times to estimate the variability of the data under H_0.

   B. Compare the observed values with expected values under H_0.

   C. Compare different sources of variabilities.

   D. Permute the original data (sample) in all possible ways to estimate the variability of the data under H_0.

   E. None of the above

10. Which one of the following tests may *not* be a good alternative to the two-sample t-test

   A. χ^2 test

   B. Wilcoxon rank sum test

   C. Permutation test

   D. Sign test

Problem 3 (30 pts). A researcher wishes to compare the effectiveness of three different treatments for reducing high blood pressures (systolic blood pressure). The researcher randomly selects 66 subjects who have high blood pressures, and then randomly assign the subjects into one of the three treatment groups, with 22 subjects in each group. Blood pressure reduction of each subject is recorded after two months in the programs. Either choose one correct answer or fill in the blanks in the following questions.

(1) (4 pts) (i) This is an (circle one)   A. experiment   B. observational study

(ii) Which one of the following is *not* used in the study (circle one)

A. randomization   B. replication   C. blocking   D. balance

(2) (10 pts) To compare all three groups, the researcher performed a one-way ANOVA, assuming all assumptions hold. Part of the results are given below (SS denotes Sum of Squares, MS denotes Mean Squares, and df denotes degrees of freedom):

Source of variations	df	SS	MS	Test Statistic	P-value
Between group		13.1			
Within group		28.9			

(i) The value of the between group MS is ____________. (ii) The value of the within group MS is ____________. (iii) The value of the test statistic is ____________. (iv) The null distribution of the test statistic is (clearly specify the parameter values or degrees of freedom as appropriate) ____________________. (v) Assuming equal variances between groups, the value of the common variance is estimated to be ____________.

(3) (4 pts) To perform pair-wise comparisons at 5% level, the significance level for each two-sample comparison should be ____________, because (one sentence)

____

(4) (6 pts) Suppose that each subject in the third group/program took treatment I in the first month and then took treatment II in the second month. To test if treatment I is better than treatment II, suggest two possible methods: Method I ________________________, Method II ________________________. Reason (one sentence):

____.

(5) (3 pts) If the effectiveness of treatment III depends on the genders of the subjects, we say that there is ________________________. In this case, what should we do in data analysis (answer in one sentence)?

____.

(6) (3 pts) If the researcher wishes to compare the average effect of treatment I and II with the effect of treatment III, the null hypothesis should be (in math notation):

____.

Problem 4 (20 pts). An instructor wishes to investigate if activity-based labs help student learning. There are two lab sections (each lab has 20 students) in her course. She asked one teaching assistant to lead Lab I that uses activities and another teaching assistant to lead Lab II that does not use activities. After five lab sessions, an identical quiz is given in each lab section. Two students in Lab I and four students in Lab II failed the quiz. All the other students passed the quiz.

(1) (4 pts). (i) Suggest *two* methods to test the instructor's hypotheses:

____

(ii) Which one of your two methods may be more reliable here? Answer: ______________

(2) (4 pts). (i) If the performances of students do not depend on which lab they attend, the expected number of students who pass the quiz in Lab I should be ______________. (ii) The difference between the observed number and expected number in (i) is (observed – expected) ______________.

(3) (2 pts). Suppose that the result is significant at the 5% significance level (either in this study or in another study). Can we draw a causal relationship between the use of activities in labs and student learning? Circle one: Yes No.

(4) (10 pts) Suppose the instructor also wishes to test if most students (i.e., over 50%) like activity-based labs. The instructor decides to make a general claim that most students like activity-based labs if 14 or more students in Lab I (of 20 students) like their labs in this study. *For questions (i) and (ii) below, you just need to write down the key steps and equations for computing the answers. You do NOT need to actually compute the numbers.*
(i) What is the significance level of the above test? (ii) What is the power of the test if in general 60% students like activity-based labs. (iii) If a normal approximation is used to compute the p-value here, the approximate normal distribution of the test statistic under H_0 is (specify the mean and standard deviation) ______________________________.