Name:_____

Lab Time: _____

ST 411/511, Winter 2012

PRACTICE MIDTERM EXAM

- There are 4 pages of questions, worth a total of 50 points, plus one page of formulas.
- Please (i) do not look at the exam until I tell you to, and (ii) stop writing when I announce that the exam is over.
- The exam is closed-book. You may bring a calculator. You may separate the formula sheet from the exam if you want to.
- Read the questions carefully.
- Read the questions carefully.

1. **[2pts each, 20 total]** TRUE or FALSE

The t-tools are more robust with respect to equal SD's for larger sample sizes than for smaller ones. **FALSE**

(Recall the fish oil and blood pressure study from homework 2 & 3) A log transformation would be appropriate for the reduction in blood pressure in the regular oil and fish oil groups. **FALSE**

The two-sample t-test is robust to departures from the independence assumption. **FALSE**

A p-value of 0.9 gives convincing evidence for the null hypothesis. **FALSE**

A 95% confidence interval will contain the true population mean 95% of the times we calculate one. **TRUE**

If a sample size is large, then the shape of a histogram of the sample will be approximately normal, even if the population distribution is not normal. **FALSE**

The standard deviation of the sampling distribution of the sample average is larger than the standard deviation of the population distribution. **FALSE**

The mean of the sampling distribution of the sample average is the same as the mean of the population distribution. **TRUE**

The Wilcoxon signed-rank test makes fewer assumptions than the paired t-test. **TRUE**

A randomization test assumes that the data are normally distributed. **FALSE**

2. **[3 pts]** A study yields statistically significant evidence that the mean salary for tall men (6' or more) is higher than for short men (less than 6'). Whether or not one can infer that tall men in USA have higher salaries than short men depends on:
(circle the best answer)

(a) the magnitude of the test statistic
(b) the size of the p-value
(c) the sample size of the study
(d) whether the men in the study were a random sample from all men in America.
(e) whether the subjects were a random sample from some population

3. **[3 pts]** A p-value gives:
(circle the best answer)

(a) the proportion of test-statistics that are greater than zero.
(b) the probability that the null hypothesis is true.
(c) evidence for the null hypothesis
(d) the probability of seeing test-statistic as or more extreme under the null hypothesis
(e) the probability the alternative hypothesis is true

4. **[2 pts]** Fill in the blanks with: mean difference OR difference in means

A paired t-test tests the null hypothesis that the **mean difference** is zero.

A two sample t-test tests the null hypothesis that the **difference in means** is zero.

5. **[3 pts]** In the cloud-seeding example discussed in Chapter 3, the mean of the log of rainfall amounts on unseeded days was 3.99, and the mean of the log of rainfall amounts on seeded days was 5.13. Circle the best interpretation of this result. Note: 5.13 – 3.99 = 1.14.

(a) Seeding causes a 1.14 unit increase in the mean rainfall.
(b) Seeding causes a 1.14 unit increase in the median rainfall.
(c) The mean rainfall on seeded days is 1.14 times the mean rainfall on unseeded days.
(d) The median rainfall on seeded days is 1.14 times the median rainfall on unseeded days.
(e) The mean rainfall on seeded days is exp(1.14) times the mean rainfall on unseeded days.
(f) The median rainfall on seeded days is exp(1.14) times the median rainfall on un-seeded days.

6. **[7 pts]** Recall the dataset from homework 1, cholesterol is measured in urban and rural Guatemalans. The following is the output of the R function t.test.

```
      Two Sample t-test

data:  Cholesterol by Group
t = -8.0782, df = 92, p-value = 2.476e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -74.58536 -45.14798
sample estimates:
mean in group Rural mean in group Urban
           157.0000            216.8667
```

Use the above output to fill in the blanks in the following statistical summary.

There is **convincing** evidence that the difference in the population mean cholesterol of urban guatemalans compared to rural Guatemalans is **not equal to zero** (**two**-sample t-test, **two**-sided p-value = **2.4×10^{-12}** or **< 0.001**)
It is estimated that the mean mean cholesterol of urban Guatemalans is **59.9** mg/l **higher** than the population mean for rural Guatemalans.

7. **[4pts total]** A number of volunteers were randomly assigned to one of two groups, one of which received daily doses of vitamin C and one of which received daily doses of placebos. It was found that the rate of colds was statistically significantly lower in the vitamin C group than in the placebo group.

TRUE or FALSE
(a) The researchers can conclude that vitamin C causes the reduction in the rate of colds. **TRUE**

(b) The researchers can conclude that vitamin C will reduce the rate of colds in people that didn't volunteer. **FALSE**

8. **[3 pts]** State the null hypothesis that is tested by Levene's test using words only (no symbols or mathematical notation).

The standard deviations of population 1 is the same as the standard deviation of population 2.
or
The population standard deviations are the same.

9. **[2 pts]** Suppose that two drugs are both effective in prolonging length of life after heart attack. Substantial statistical evidence indicates that the mean life for those using Drug A is substantially greater than the mean length of life for those using Drug B, but the variation of life lengths for drug A is substantially greater as well. Explain why it is difficult to conclude that drug A is better even though the mean life length is greater.

(Conceptual Exercise 4.7) **The mean life length for people on drug A might be longer, but due to the higher spread, more people may die sooner on drug A than on drug B.**

10. **[3 pts]** Four students took a test of manual dexterity (measured on a scale of 0 to 100) before and after drinking a cup of tea.
The data:

	Dexterity Score		**Difference**	**Abs.**	
Student	Before tea	After tea	(after - before)	**difference**	**Rank**
1	48	47	-1	1	1.5
2	53	54	1	1	1.5
3	51	56	5	5	3
4	60	68	8	8	4

Calculate the value of the signed-rank statistic, S, based on the differences as defined in the last column.

S = sum of ranks of positive differences = (1.5 + 3 + 4) = 8.5