PID:		
Last Name, First Name: _		
Section:		

# Homework 7 Math 11, UCSD, Winter 2018 Due on Tuesday, 13th March

Reading: Chapters 22, 23, 25, and 24.

Note: as always, for hypothesis testing problems, you must show how you calculated your test statistic, and write a concluding sentence in which you explain your conclusion in the context of the problem.

### Exercise 1

In the June 2007 issue, Consumer Reports also examined the relative merits of top-loading and front-loading washing machines, testing samples of several different brands of each type. One of the variables the article reported was "cycle time", the number of minutes it took each machine to wash a load of clothes.

Among the machines rated good to excellent, the 98% confidence interval for the difference in mean cycle time  $(\mu_{Top} - \mu_{Front})$  is (-40, -22).

1. The endpoints of this confidence interval are negative numbers. What does that indicate?

2. What does the fact that the confidence interval does not contain 0 indicate?

3. If we use this confidence interval to test the hypothesis that  $\mu_{Top} - \mu_{Front} = 0$ , what's the corresponding alpha level?

The study of the new CPMP Mathematics methodology tested students' abilities to solve word problems. This table shows how the CPMP and traditional groups performed.

Math Program	п	Mean	SD
CPMP	320	57.4	32.1
Traditional	273	53.9	28.5

What do you conclude?

A researcher wanted to see whether there is a significant difference in resting pulse rates for men and women. The data she collected are displayed in the boxplots and summarized in the next column.



1. What do the boxplots suggest about differences between male and female pulse rates?

2. Is it appropriate to analyze these data using the methods of inference discussed in Chapter 22? Explain.

3. Create a 90% confidence interval for the difference in mean pulse rates. (From technology, df = 40.2.)

4. Does the confidence interval confirm your answer to part 1? Explain

Researchers collected samples of water from streams in the Adirondack Mountains to investigate the effects of acid rain. They measured the pH (acidity) of the water and classified the streams with respect to the kind of substrate (type of rock over which they flow). A lower pH means the water is more acidic. Here is a plot of the pH of the streams by substrate (limestone, mixed, or shale), together with selected parts of a software analysis comparing the pH of streams with limestone and shale substrates:



1. State the null and alternative hypotheses for this test.

2. From the information you have, do the assumptions and conditions appear to be met?

3. What conclusion would you draw?

In an investigation of environmental causes of disease, data were collected on the annual mortality rate (deaths per 100,000) for males in 61 large towns in England and Wales. In addition, the water hardness was recorded as the calcium concentration (parts per million, ppm) in the drinking water. The data set also notes, for each town, whether it was south or north of Derby. Is there a significant difference in mortality rates in the two regions? Here are the summary statistics.



1. Test appropriate hypotheses and state your conclusion.

2. The boxplots of the two distributions show an outlier among the data north of Derby. What effect might that have had on your test?

Which of the following scenarios should be analyzed as paired data? Explain.

1. Spouses are asked about the number of hours of sleep they get each night. We want to see if husbands get more sleep than wives.

2. 50 insomnia patients are given a placebo and 50 are given a mild sedative. Which subjects sleep more hours?

3. A group of college freshmen and a group of sophomores are asked about the quality of the university cafeteria. Do students' opinions change during their time at school?

Some students do homework with music playing in their headphones. (Anyone come to mind?) Some researchers want to see if people can work as effectively with as without distraction. The researchers will time some volunteers to see how long it takes them to complete some relatively easy crossword puzzles. During some of the trials, the room will be quiet; during other trials in the same room, subjects will wear headphones and listen to a Pandora channel.

1. Design an experiment that will require a two-sample t procedure to analyze the results.

2. Design an experiment that will require a matched-pairs t procedure to analyze the results.

3. Which experiment would you consider the stronger design? Why?

Many people believe that students gain weight as freshmen. Suppose we plan to conduct a study to see if this is true.

1. Describe a study design that would require a matched-pairs t procedure to analyze the results.

2. Describe a study design that would require a two-sample t procedure to analyze the results.

Simpson, Alsen, and Eden (Technometrics 1975) report the results of trials in which clouds were seeded and the amount of rainfall recorded. The authors report on 26 seeded and 26 unseeded clouds in order of the amount of rainfall, largest amount first. Here are two possible tests to study the question of whether cloud seeding works.

```
Paired t-Test of \mu(1-2)

Test Ho: \mu(1972-1968) = 0 vs Ha: \mu(1972-1968) \neq 0

Mean of Paired Differences = 0.0337

t-Statistic = 2.458 with 18 df

p = 0.0244

2-Sample t-Test of \mu 1 - \mu 2

Ho: \mu 1 - \mu 2 = 0 Ha: \mu 1 - \mu 2 \neq 0

Test Ho: \mu(1972) - \mu(1968) = 0 vs

Ha: \mu(1972) - \mu(1968) \neq 0

Difference Between Means = 0.0337

t-Statistic = 1.496 with 35 df

p = 0.1434
```

1. Which of these tests is appropriate for these data? Explain.

2. Using the test you selected, state your conclusion.

Researchers examined the number of people admitted to emergency rooms for vehicular accidents on 12 Friday evenings (6 each on the 6th and 13th). Based on these data, is there evidence that more people are admitted, on average, on Friday the 13th? Here are two possible analyses of the data:

Year	Month	6th	13th
1989	October	9	13
1990	July	6	12
1991	September	11	14
1991	December	11	10
1992	March	3	4
1992	November	5	12

```
Paired t-Test of \mu(1-2) = 0 vs. \mu(1-2) < 0
Mean of Paired Differences = -3.333
t-Statistic = -2.7116 with 5 df
P = 0.0211
2-Sample t-Test of \mu 1 = \mu 2 vs. \mu 1 < \mu 2
Difference Between Means = -3.333
t-Statistic = -1.6644 with 9.940 df
P = 0.0636
```

1. Which of these tests is appropriate for these data? Explain.

2. Using the test you selected, state your conclusion.

3. Are the assumptions and conditions for inference met?

Having done poorly on their Math final exams in June, six students repeat the course in summer school and take another exam in August.

June	54	49	68	66	62	62
Aug.	50	65	74	64	68	72

1. If we consider these students to be representative of all students who might attend this summer school in other years, do these results provide evidence that the program is worth-while?



2. This conclusion, of course, may be incorrect. If so, which type of error was made?

Does a person's cholesterol level tend to change with age? Data collected from 1406 adults aged 45 to 62 produced the regression analysis shown. Assuming that the data satisfy the conditions for inference, examine the association between age and cholesterol level.

Dependent variable is Chol $s = 46.16$						
Variable	Coefficient	SE(Coeff)	t-Ratio	<b>P-Value</b>		
Intercept	194.232	13.55	14.3	≤0.0001		
Age	0.771639	0.2574	3.00	0.0056		

1. State the appropriate hypothesis for the slope.

2. Test your hypothesis and state your conclusion in the proper context.

How strong was the association between student scores on the Math and Verbal sections of the old SAT? Scores on each ranged from 200 to 800 and were widely used by college admissions offices. Here are summaries and plots of the scores for a graduating class at Ithaca High School:



1. Is there evidence of an association between Math and Verbal scores? Write an appropriate hypothesis.

2. Discuss the assumptions for inference.

3. Test your hypothesis and state an appropriate conclusion.

4. Find a 90% confidence interval for the slope of the true line describing the association between Math and Verbal scores.

5. Explain in this context what your confidence interval means.

Does your IQ depend on the size of your brain? A group of female college students took a test that measured their verbal IQs and also underwent an MRI scan to measure the size of their brains (in 1000s of pixels). The scatterplot and regression analysis are shown, and the assumptions for inference were satisfied.



Test an appropriate hypothesis about the association between brain size and IQ.

A business analyst was interested in the relationship between a company's sales and its profits. She collected data (in millions of dollars) from a random sample of Fortune 500 companies and created the regression analysis and summary statistics shown. The assumptions for regression inference appeared to be satisfied.

	Profits	Sales	Dependent variable is Profits		
Count	79	79	R-squared	= 66.2%	s = 466.2
Mean	209.839	4178.29	Variable	Coefficient	SE(Coeff)
Variance	635,172	49,163,000	Intercept	-176.644	61.16
Std Dev	796.977	7011.63	Sales	0.092498	0.0075

1. Is there a statistically significant association between sales and profits? Test an appropriate hypothesis and state your conclusion in context.

2. Do you think that a company's sales serve as a useful predictor of its profits? Use the values of both  $R^2$  and s in your explanation.

3. Find a 95% confidence interval for the slope of the regression line. Interpret your interval in context.

## Exercise 16

The following software outputs provide information about the Size (in square feet) of 18 homes in Ithaca, New York, and the city's assessed Value of those homes.



1. Explain why inference for linear regression is appropriate with these data.

2. What percentage of the variability in assessed Value is explained by this regression?

3. Is there a significant association between the Size of a home and its assessed Value? Test an appropriate hypothesis and state your conclusion.

4. Give a 90% confidence interval for the slope of the true regression line, and explain its meaning in the proper context.

5. From this analysis, can we conclude that adding a room to your house will increase its assessed Value? Why or why not?