Inference for Two Distributions

Use a separate sheet of paper. You must show all work and all steps must be clearly labeled. Submitting just answers will result in a grade of 0! All unexplained numbers will be ignored and final answers must be written in complete sentences.

1. A dietician has developed a diet that is low in fats, carbohydrates, and cholesterol. Although the diet was initially intended to be used by people with heart disease, the dietitian wishes to examine the effect this diet has on the weights of obese people. Two random samples of 100 obese people each are selected, and one group of 100 is placed on the low-fat diet. The other 100 are placed on a diet that contains approximately the same quantity of food but is not as low in fats, carbohydrates and cholesterol. Each person, the amount of weight lost (or gained) in a 3-week period is recorded. The results are below. *(Solutions to a & b posted on Chp 11 powerpoint.)*

	Low-Fat Diet	Other Diet
Mean weight loss	9.3 lbs	7.4 lbs
Sample variance	22.4	16.3

   a. Form and interpret a 95% confidence interval for the difference between the population mean weight losses for the two diets.
   b. Test to see if there is a difference in weight loss for the two diets.

2. *In each of the following situations, explain what is wrong and why
   a. A researcher wants to test H_o: x-bar_1 = x-bar_2 versus the two sided alternative H_a: x-bar_1 is not equal to x-bar_2.
   b. A study recorded the scores of 20 children who were similar in age. The scores of the 10 boys in the study were compared with the scores of all 20 children using the two-sample methods of this chapter.
   c. A two-sample t statistics gave a P-value of 0.96. From this you can reject the null hypothesis with 95% confidence. *(ie: alpha = 0.05).*

3. *College financial aid offices expect students to use summer earnings to help pay for college. But how large are these earnings? One college studied this question by asking a sample of students how much they earned. Omitting the students who were not employed, there were 1296 responses. Here are the data in summary form:

Group	**n**	**x-bar**	**s**
Males	675	$1,884.52	$1,368.37
Females	621	$1,360.39	$1,037.46

   a. The distribution of earnings is strongly skewed to the right. Nevertheless, use of t procedures is justified. Why?
   b. Give a 90% confidence interval for the difference between the mean summer earnings of male and female students.
   c. Once the sample size was decided, the sample was chosen by taking every 20^{th} name from an alphabetical list of all undergraduates. Is it reasonable to consider the samples as SRS's chosen from the male and female undergraduate populations? If not, what type of sampling distribution is this?
   d. What other information about the study would you request before accepting the results as describing all undergraduates?

4. One of the most feared predators in the ocean is the great white shark. It is known that the white shark grows to a mean length of 21 feet; however, one marine biologist believes that great white sharks off the Bermuda coast grow much longer owing to unusual feeding habits. To test this claim, some full-grown great white sharks were captured off the Bermuda coast, measured, and then set free. However, because the capture of sharks is difficult, costly, and very dangerous, only three specimens were sampled. Their lengths were 24, 20 and 22 feet.
   a. Construct a 96% confidence interval for the mean shark length?
   b. The biologist's wants to use $\alpha = .10$, what t-score is this associated with?
   c. Do the data provide sufficient evidence to support the marine biologist's claim?

5. Suppose you wish to compare a new method of teaching reading to "slow learners" to the current standard method. You decide to base this comparison on the results of a reading test given at the end of a learning period of 6 months. Of a random sample of 20 slow learners, 8 are taught by the new method and 12 are taught by the standard method. All 20 children are taught by qualified instructors under similar conditions for a 6-month period. The results of the reading test at the end of this period are given below.

Reading Scores for Slow Learners

New Method	80	80	79	81	76	66	79	76				
Standard	79	62	70	68	73	76	86	73	72	68	75	66

   a. Use a 90% confidence interval, and interpret the interval for the true mean difference between the test scores for the new method and the standard method.
   b. Test to see if the new method's test scores are higher. Let $\alpha = .10$

6. Some college professors make bound lecture notes available to their classes in an effort to improve teaching effectiveness. Because students pay the additional cost, educators want to know whether the students consider the lecture notes to be a good educational value. *Marketing Educational Review* (Fall 1994) published a study of business students' opinions of lecture notes. Two groups of students were surveyed-86 students enrolled in a promotional strategy class that required the purchase of lecture notes, and 35 students enrolled in a sales/retailing elective that did not offer lecture notes. In both courses, the instructor used lectures as the main method of delivery. At the end of the semester, the students were asked to respond to the statement: "Having a copy of the lecture notes was [would be] helpful in understanding the material." Responses were measured on a 9-point semantic difference scale, where 1 = "strongly disagree" and 9 = "strongly agree." A summary of the results is shown in the table below.

Classes Buying Lecture Notes	Classes Not Buying Lecture Notes
mean score = 8.48	mean score = 7.80
standard deviation = 0.94	standard deviation = 2.99

Construct and interpret a 60% confidence interval.

7. *Is red wine better than white wine? Observational studies suggest that moderate use of alcohol reduces heart attacks, and that red wine may have special benefits. One reason may be that red wine contains polyphenols, substances that do good things to cholesterol in the blood and so may reduce the risk of heart attacks. In an experiment, healthy men were assigned at random to drink half a bottle of either red or white wine each day for two weeks. The level of polyphenols in their blood was measured before and after the two-week period. Here are the percent changes in level for the subject in both groups.

Red wine:	3.5	8.1	7.4	4.0	0.7	4.9	8.4	7.0	5.5
White wine:	3.1	0.5	-3.8	4.1	-0.6	2.7	1.9	-5.9	0.1

   a. Is there good evidence that red wine drinkers gain more polyphenols on average than white wine drinkers? Give appropriate statistical justification.
   b. Does this study give reason to think that it is drinking red wine, rather than some lurking variable, that causes the increase in blood polyphenols? Justify your answer.
   c. Construct a 95% confidence interval for the difference in mean polyphenol levels. *(You do not need to rewrite your assumptions.)*

8. *How badly does logging damage tropical rain forests? One study compared forest plots in Borneo that had never been logged with similar plots nearby that had been logged 8 years earlier. The study found that the effect of logging were somewhat less severe than expected. Here are the data on the number of tree species in 12 unlogged plots and 9 logged plots.
   a. The study report says, "Loggers were unaware that the effects of logging would be assessed." Why is this important? The study report also explains why the plots can be considered to be randomly assigned. Why is this important?
   b. Does logging significantly reduce the mean number of species in a plot after 8 years? Give appropriate statistical evidence to support your conclusion.
   c. Construct a 90% confidence interval for the difference in mean number of species between unlogged and logged plots. *(You do not need to rewrite your assumptions.)*

*Yates, Moore, & Starnes. *The Practice of Statistics, 3rd Edition.*