22S:30/105, Statistical Methods and Computing
Spring 2008, Instructor: Cowles
PRACTICE PROBLEMS for MIDTERM 2, 2009
Midterm 2

Name: _____ Course no. (30 or 105) _____

Show your work on any problems that involve calculations.

I will grade on a curve and will give partial credit wherever possible.

1. Early childhood education specialists were interested in whether third grade reading scores were different in children who, prior to starting kindergarten, had been in daycare, had attended preschool, or had been cared for exclusively in their own homes. To address this question, researchers obtained data on child care prior to starting school and on current reading scores for a random sample of third graders from a particular school district.

   Was this an experiment or an observational study? Why?

2. According to the Food and Nutrition Board of the National Academy of Sciences, the recommended daily allowance (RDA) of iron for adult females under the age of 51 is 18 milligrams. Iron intakes, in mg, were obtained during a 24-hour period for 45 randomly-selected adult females under the age of 51. These data will be used to carry out a hypothesis test to determine whether the mean iron intake in women in this age group is less than 18 mg.

   To answer some of the following questions, you will need to refer to the accompanying SAS output.

   (a) Is this a one-sided or a two-sided test? Explain in one sentence.

   (b) Write the appropriate null and alternative hypotheses to be tested in this study. Use conventional statistical symbols.

   (c) The investigators performed a t-test at the 0.05 significance level. Which of the following are assumptions that make a t-test appropriate? (Circle all that apply.)
      i. The data are a simple random sample from the population.
      ii. The variable has an approximately normal distribution in the population.
      iii. The value of the population standard deviation is known.
      iv. The sample size is greater than 30.
      v. None of the above.

   (d) Is there anything in the attached SAS output that suggests that a t-test might not be appropriate? Briefly explain.

   (e) Based on the SAS output, what is the p-value for the test of the hypotheses you stated in the first part of this question? Give a numeric answer and a brief explanation.

   (f) Given their chosen significance level of 0.05, should the investigators reject the null hypothesis? Why?

   (g) Briefly explain what your conclusion in the previous part of this problem means regarding adult women's iron intake.

3. Among the demographic variables reported by the U.S. Bureau of Labor Statistics is the age of people in the civilian labor force. Fifty people in the civilian labor force are randomly selected, and their ages are recorded. We will use these data to conduct a test of the alternative hypothesis that mean age of people in the civilian labor force is greater than 40. The formal statement of our hypotheses is:

$$H_0: \quad \mu = 40$$
$$H_A: \quad \mu > 40$$

   For the purposes of this question, assume that the distribution of ages in the civilian labor force is approximate normal, and *pretend* that we know that the exact standard deviation was $\sigma = 12$ years.

   (a) Given the hypotheses as stated, would large values of $\bar{x}$ or small values of $\bar{x}$ cause us to reject the null hypothesis? (large/small)

(b) If we wish to carry out our hypothesis test at significance level $\alpha = 0.05$, with a sample size of 50 and $\sigma = 12$ years, what value of $\bar{x}$ would be required in order to reject H_0? Numeric answer. Show your work, including drawing a picture.

(c) What would the power of the hypothesis test be if the true value of μ is 45? Numeric answer. Again, draw a picture and show your work.

4. Data collected on 50 infants produced the following 95% confidence interval for the population mean age in weeks at which babies begin to crawl:
(29.2, 31.8)

(a) Which quantity do we have 95% confidence lies in the interval given? (Circle one).

i. s
ii. σ
iii. $\bar{x}$
iv. μ
v. none of the above

(b) Which of the following statements would be most likely to be true if the sample had consisted of 100 infants instead of 50? (Circle one.)

i. The endpoints of the confidence interval would be exactly the same as those obtained from the sample of size 50.
ii. The confidence interval most likely would be narrower if the sample size were 100 instead of 50.
iii. The confidence interval most likely would be wider if the sample size were 100 instead of 50.

5. A study was done to evaluate the effects of subliminal messages on performance on academic tests. Informed consent was obtained from the 300 students in an introductory psychology course. The students then were randomly assigned to three different groups. All groups took the same computerized test on psychology. For one group, just before the first question was displayed, the message "You are brilliant" appeared on the screen for 1/20th of a second. For the second group, the message "Apples are nutritious" appeared for 1/20th of a second. For the third group, no message was displayed. After all the students exams were complete and scored, the average scores in the three groups were compared.

(a) (1) Identify the response variable in this study.

(b) (1) Identify the experimental units in this study.

(c) (3) Identify the factor(s) and levels in this study.

(d) (1) Identify the treatments in this study.

Basic Statistical Measures

Location		Variability	
Mean	14.68000	Std Deviation	3.08321
Median	14.70000	Variance	9.50618
Mode	12.50000	Range	14.40000
		Interquartile Range	4.30000

Tests for Location: Mu0=18

Test	-Statistic-		-----p Value------	
Student's t	t	-7.22339	Pr > |t|	<.0001
Sign	M	-13.5	Pr >= |M|	<.0001
Signed Rank	S	-449	Pr >= |S|	<.0001

Stem	Leaf	#	Boxplot
20	7	1	
19	58	2	
18	112336	6	
17	03	2	
16	0334668	7	+-----+
15	0036	4	| |
14	45667	5	*--+--*
13	1	1	| |
12	14555678	8	+-----+
11	05569	5	
10	79	2	
9	4	1	
8			
7			
6	3	1	