22S:30/105, Statistical Methods and Computing
Spring 2006, Instructor: Cowles
Midterm 2
PRACTICE PROBLEMS for MIDTERM 2, 2008

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

Show your work on any problems that involve calculations.

There are 40 total points on this midterm. Point values for each question are shown in parentheses. I will grade on a curve and will give partial credit wherever possible.

1. Let the random variable X be the number of rooms in randomly-chosen owner-occupied housing unit in San Jose, California. Here is the distribution of X:

Rooms x	Pr(X = x)
1	0.003
2	0.002
3	0.023
4	0.104
5	0.210
6	0.224
7	0.197
8	0.149
9	0.053
10	0.035

   (a) (2) Is the random variable X discrete or continuous? Briefly justify your answer.

   (b) (1) Express the event "the randomly chosen unit has no more than 3 rooms" as a statement about X.

   (c) (1) What is the probability of the event in the previous question? (numeric answer)

2. The level of pesticides found in the blubber of whales is a measure of pollution of the oceans by runoff from land areas. Investigators studied the concentration of the insecticide dieldrin in a sample of 8 male whales of the minke species in the West Greenland area.

   To answer the questions that follow, assume that the dieldrin concentration in the blubber of all male minke whales in this geographic area follows a normal distribution with standard deviation $\sigma = 50$ nanograms per gram of blubber (ng/g).

   (a) (2) The researchers found the sample mean concentration in the 8 whales to be 357 ng/g. Construct a 90% confidence interval for the population mean concentration of dieldrin. (Numeric answer)

   (b) (1) The number 357 in the previous question is a (circle one):
      i. parameter
      ii. population
      iii. statistic
      iv. sample
      v. standard deviation

   (c) (1) The conventional statistical symbol for the quantity that we are 90% confident lies in the interval you calculated is (circle one):
      i. μ
      ii. $\bar{x}$
      iii. σ
      iv. s
      v. β

   (d) The Food and Drug Administration regulates the amount of dieldrin in raw food. For some foods, no more than 100 ng/g is allowed. We will use the information from the 8 minke whales, and the assumption that $\sigma = 50$ ng/g, to assess whether the mean concentration in whale blubber is above the FDA limit. We will carry out our hypothesis test at significance level $\alpha = .1$.
      i. (1) Write the appropriate null and alternative hypotheses that will be tested. Use conventional statistical symbols.

ii. (2) What is the sampling distribution of $\bar{x}$ if the null hypothesis is true? Give the name of the family of distributions, and the numeric values of the mean and the standard deviation.

iii. (2) Calculate the test statistic (numeric answer).

iv. (2) Find the p-value (numeric answer).

v. (3) The correct interpretation of the p-value is (circle one):
   A. The probability of making Type II error.
   B. The probability that the null hypothesis is true given the data.
   C. A level that we choose before we observe the data that represents the amount of risk we are willing to take of rejecting the null hypothesis when it is true.
   D. The probability, computed assuming the null hypothesis is true, that the test statistic would take on a value as extreme as, or more extreme than, that actually observed.
   E. None of the above.

vi. (2) Should we reject the null hypothesis at our chosen significance level? (yes/no) Briefly justify your answer.

vii. (2) State in words what your answer to the previous question means in terms of this application.

viii. (5) Find the power of this test against the alternative that the population mean is 150 ng/g. (Numeric answer; show all your work.)

ix. (2) `Power` in hypothesis testing (circle all that are true):
   A. is increased if the sample size is increased (all other characteristics of the study being held constant)
   B. is the probability of failing to reject the null hypothesis if it is true
   C. is the probability of rejecting the null hypothesis if it is false
   D. none of the above

(e) (2) How many minke whales would the researchers have had to study to get a 90% confidence interval with width no greater than 20 ng/g? (Numeric answer)

3. The dataset called "Cars" on the course webpage contains measurements on a sample of 38 automobiles. The cars in the sample are considered a simple random sample from all models of cars in the 1978-79 model year. The gas mileage in miles per gallon was measured by Consumers' Union on a test track. Refer to the attached SAS output in answering the following questions.

   (a) (1) `Proc means` was used to calculate a 99% confidence interval. The kind of interval that proc means calculates is (circle one) :
      i. z interval
      ii. t interval
      iii. neither

(b) (2) We would like to use this dataset to test the null hypothesis that mean gas mileage in miles per gallon for all models of cars in the 1978-79 model year is 25 mpg at significance level $\alpha = .01$. Based on the confidence interval, should we reject H_0? Briefly justify your answer.

(c) (4) Circle all of the statements below that reflect correct understanding of this confidence interval.

i. There is 99% probability that the true population mean lies in the interval (21.88, 27.64).

ii. 99% of all possible samples of size 38 from this population would have an $\bar{x}$ in the interval (21.88, 27.64).

iii. This interval was calculated using a method that will produce an interval that traps the true mean when applied in 99% of samples.

iv. We are 99% confident that the gas mileage of any randomly-selected car from the 1978-79 model year would be in the interval (21.88, 27.64).

(d) (2) Does either your substantive knowledge, or anything in the SAS output, suggest that t procedures should *not* be used here? Explain briefly.

4. A statistics professor carries out an experiment to determine the effect of temperature and lighting on the proportion of students that fall asleep during her lectures. She teaches 4 classes per day. For one class, she sets the classroom temperature at 65 and keeps all the lights in the room on during her lectures. For her second class she sets the room temperature at 70 and keeps all the lights on. For her third class she sets the room temperature at 65 and dims the lights. For her fourth class she sets the temperature at 70 and dims the lights. She does this on 5 different days. Following each lecture, she calculates and records the proportion of students that fell asleep at some point during the lecture.

(a) (1) Is this an experiment or an observational study? Briefly justify your answer.

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

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

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

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

(f) (2) Identify two possible confounding variables in this study. Briefly explain.

Analysis Variable : MPG

N	Mean	Lower 99% CL for Mean	Upper 99% CL for Mean
38	24.7605263	21.8764512	27.6446014

Location		Variability	
Mean	24.76053	Std Deviation	6.54731
Median	24.25000	Variance	42.86732
Mode	17.00000	Range	21.80000
		Interquartile Range	12.00000

Tests for Location: Mu0=25

Test	-Statistic-		-----p Value------	
Student's t	t	-0.22547	Pr > |t|	0.8229
Sign	M	0	Pr >= |M|	1.0000
Signed Rank	S	-24	Pr >= |S|	0.7328

Stem	Leaf	#	Boxplot
36	3	1	
34	121	3	
32	5	1	
30	059589	6	+-----+
28	485	3	| |
26	58245	5	| |
24			*--+--*
22	0	1	| |
20	368569	6	| |
18	12562	5	+-----+
16	259006	6	
14	5	1	

----+----+----+----+