Statistics 571: Statistical Methods
Summer 2003
University of Tennessee, Knoxville
Instructor: Ramón V. León

Name:

This is a take-home exam. You are to work alone. You cannot ask questions except from the instructor. Please type your answers or paste any supporting JMP output to this MS Word document. Note that you do not have to retype the data in this exam since you can copy it from this document and paste it into JMP. Use the Equations editor of MS Word to type any mathematics. This exam is due at noon Thursday, July 3. Please email the completed MS Word document as an attachment to rleon@utk.edu.

1. The following is the one-year percentage returns for a sample of 17 stock funds:

32.2
29.5
29.9
32.4
30.5
50.1
32.1
35.2
28.6
9
20.6
28.6
30.5
45
33
29.4
37.1

a. Do these data seem to have any outliers? Why? Paste supporting JMP output.

b. What is interquartile range?

c. What are the values of the upper and lower inner fences?

d. What are the values of the upper and lower outer fences?

e. What is the five point summary?

f. Do these data seem to be normally distributed? Why? Paste supporting JMP output

g. Do a stem and leaf plot for these data. What does it indicate?

h. What conclusions can you make about this data?

2. Suppose that 80 % of people obtain news from television, 63% from newspapers, 47% from radio, 45% from television and newspapers, 29% from television and radio, 21% for newspapers and radio, and 4 % from television, newspapers, and radio.

   a. Given that radio is a news source, what is the probability that television is also a news source? Justify your answer.

   b. Given that TV is a news source, what is the probability that radio is also a news source? Justify your answer.

   c. Given that both newspaper and radio are news sources, what is the probability that TV is not a news source? Justify your answer.

3. The accuracy of a medical diagnostic test, in which a positive result indicates the presence of a disease, is often stated in terms of its **sensitivity**, the proportion of diseased people that test positive or P(+|Disease), and its **specificity**, the proportion of people without the disease who test negative or P(-|No Disease). The proportion of the population that has the disease is called the **prevalence** rate. A diagnostic test for the disease has 97% sensitivity and 96% specificity.

   a. Suppose that the prevalence rate of the disease is 10% and that a person's test result is positive. What is the probability that the person actually has the disease? Justify your answer

   b. Suppose that the prevalence rate of the disease is 0. 1 % and that a person's test result is positive. What is the probability that the person actually has the disease? Justify your answer.

   c. Do you think 20-year-old males should be tested for prostrate cancer? Do you think 20-year-old females should be tested for breast cancer? Justify your answer.

4. An experiment measures the number of particle emissions from a radioactive substance. The number of emissions has a Poisson distribution with rate $\lambda = 6$ particles per week.

    a. What is the probability of at least one emission in a week?

    b. What is the probability of at least one emission in a year?

5. Let X = the time to failure of a light bulb. Assume that X is exponentially distributed

    a. If the mean time to failure is 4,000 hours, what is the median time to failure?

    b. What is the probability that the bulb will last at least 2000 hours?

6. Let X be a normal random variable with mean 20 and standard deviation 6.

    a. What is the probability of X being between 12 and 26?

    b. What is its 90 % percentile?

7. A cohort study examined a group of elderly patients who showed no signs of dementia. After five years it was found that those that engaged in mental activity such as reading, playing chess, or doing crossword puzzles had a significantly lower rate of dementia. Does this prove that mental activity helps prevent dementia? Explain.

8. Explain why the regression fallacy is a problem in pretest posttest situations.

9. A researcher compares the precision of two assay methods, a standard method S and a new method N. A solution is prepared and divided into 40 samples. The variance of the concentration readings from the two methods is compared.

    a. Diagram a completely randomized design for this experiment.

    b. Four technicians perform the assays that have slightly individualized techniques in running the assays. Diagram a randomized block design using the technicians as blocks.

10. Hospitals are graded based on their success rates in treating different categories of patients. We want to compare two hospitals – A, which is a university affiliated research hospital, and B, which is a general community hospital – with respect to success rates for a certain complicated surgery. The data classified by low risk patients and high risk patients are shown in the following table.

	Low Risk				High Risk		
	Success	Failure	Total		Success	Failure	Total
Hospital A	400	100	500	Hospital A	160	640	800
Hospital B	300	200	500	Hospital B	30	180	210
Total	700	300	1000	Total	190	820	1010

a. Calculate the success rates for each category of patients for both hospitals. Which hospital is better?

b. Aggregate the data over the two categories of patients and calculate the overall success rates for both hospitals. Now which hospital is better?

c. Explain the discrepancy between the results obtained in (a) and (b).

d. What are the adjusted (standardized) success rates for each hospital?

e. Do you favor publishing success rates of hospitals? Under what circumstances would you favor publishing these success rates.