

Instructor: Mary Parker

Math 1342

Review for Test 4 - updated Dec. 5, 2007 (Includes two sample tests)

Prepare your notes for the test. Use those. DO NOT use the examples in the book or any other assistance.

Use only the normal table and the t table in the back cover of your book. Those are the only tables you will have available for the test.

On the actual test, you will not have to compute any sample standard deviations of quantitative data from the data. (As you did in Chapters 18 and 19.) If some are needed, I will give you a value to use. On this practice test, if you need a sample standard deviation like this, do it as you were doing in the homework. Then do the rest of the problem using only the same calculator you'll be able to use on the test.

On each sample test (as on the real test) for every confidence interval, interpret it in terms of the question in the problem. For every hypothesis test, show the hypotheses, define the parameter, sketch the sampling distribution of the test statistic (not absolutely required for problems involving two means or two proportions,) find the value of the test statistic from the data, find the p-value, and interpret the result in terms of the question in the problem.

How to do each sample test:

Very short version for very well-prepared students: Answer the first three questions as they appear. For the rest, identify which type of technique is needed. (Example: confidence interval, two proportions. Second example: probability calculation in the sampling distribution of a sample proportion.)

Main method: Work the problems fully using pencil and paper. You will have to prove to me that you have worked on each of them in order to be allowed to see the answer key. **If you will have to submit your proof by email, please look at the end of page 3 here now** to see how you will submit that. But even if you will submit your proof by email, work the problems using pencil and paper, showing your work clearly, so that it will be easy for you to grade when you see the answer key.

First sample test:

1. Using the same set of data you compute a 95% confidence interval and a 99% confidence interval. Which of the following statements is correct?
 - a. The intervals have the same width.
 - b. The 99% interval is wider.
 - c. The 95% interval is wider.
 - d. You cannot determine which interval is wider unless you know n and s .

In each of 3 - 5, the situation requires a test of hypotheses about a population mean μ . Define that population mean and state the appropriate null hypothesis H_0 and alternative hypothesis H_a .

3. A national travel agency is interested in whether the average length of stay in Europe and the Mediterranean by travelers has changed since 1990. According to the U.S. Bureau of Economic Analysis, in 1990, the average stay was 20.6 days. They commission a poll to gather data for 1999.

5. A medical researcher is interested in whether a new cold remedy is "better" than the standard remedy in terms of the average number of days the cold symptoms persist. He will do a random sample of cold sufferers, give each the new remedy, and measure how many days their symptoms persist. With the old cold remedy, symptoms persist an average of 2.7 days. (Hint: Think carefully about what sorts of scores mean "better" here.)

7. The article "What Kinds of People Do Not Use Seat Belts?" (*Amer. J. of Public Health* (1977): 1043-1049) reported on a survey with the objective of studying characteristics of drivers and seat belt usage. Suppose that the investigators wished to estimate the proportion of all drivers with seat belts who use them and their estimate should be correct to within 0.02 with 95% confidence.

- At the outset of the study, how large a sample would be needed? Discuss why it is so large.
- After looking at how large the sample they would need, the investigators decided to take a preliminary study of 50 drivers with seat belts and found that 45 of them used their seat belts. Using the results of this preliminary study, how large a sample would be needed to form an estimate within 0.02 of the correct answer with 95% confidence?

9. A city planner working on bikeways needs information about the experiences and opinions of local bicycle commuters. She designs a questionnaire.

- If we are going to use the methods of this course to analyze the results, how should she choose the people to whom she gives the questionnaire?

One of the questions asks how many minutes it takes the rider to pedal from home to his or her destination. The data from the survey of 22 local bicycle commuters yields times with $\bar{x} = 25.82$ minutes and $s = 7.71$ minutes.

- Find and interpret a 90% confidence interval for the population mean commuting time, μ , for local bicycle commuters.
- Discuss what conditions must be made about the times in order for the procedure you used in part a. to be valid.

11. The Health Insurance Association of America reports that, in 1983, the average daily room charge for a semi-private room in U.S. hospitals was \$195. In the same year, a random sample of 30 Massachusetts hospitals yielded a sample mean semi-private room charge of $\bar{x} = \$202.68$ with a sample standard deviation of $s = \$12.77$. Let μ be the mean charge for a semi-private room in a Massachusetts hospital that year.

- Test the hypotheses $H_0: \mu = 195$ versus $H_a: \mu > 195$ at the 5% significance level, and state your conclusion about room charges.
- Discuss what conditions must be made about these sampling method in order for the procedure you used in part a. to be valid.
- Discuss what conditions must be made about the distribution of the population in order for the procedure you used in part a. to be valid.
- Is there any reason to doubt whether the conditions listed in either (b) or (c) were met? Discuss and address both the conditions in part (b) and those in part (c).

13. A pediatrician began measuring the blood cholesterol levels of her young patients. She was surprised to find that many of them had levels over 200, indicating increased risk of artery disease. Ten such patients were randomly selected to take part in a nutritional program designed to lower blood cholesterol.. Two months following the beginning of the program, the pediatrician measured the blood cholesterol levels of the 10 patients again. The results are

| | | | | | | | | | | |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Patient | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Before program | 210 | 217 | 208 | 215 | 202 | 209 | 207 | 210 | 221 | 218 |
| After program | 212 | 210 | 210 | 213 | 200 | 208 | 203 | 199 | 218 | 214 |

- Test whether the program lowers cholesterol, on the average. Write hypotheses, define the parameter(s) used, do the test, report the P value, and explain what this hypothesis test says about the benefits of the program.
- Discuss what conditions must be made about these data in order for the procedure you used in part a. to be valid.
- Discuss how you would check whether these conditions are met. Are they met for this data?

15. The state of Georgia's HOPE scholarship program guarantees fully paid tuition to Georgia public universities for Georgia high school seniors who have a B average in academic requirements as long as they maintain a B average in college. It was reported that 53.3% of a random sample of 137 students entering Ivan Allen College at Georgia Tech (social science and humanities) with a HOPE scholarship lost the scholarship at the end of the first year because they had a GPA of less and 3.0. It was also reported that 72 of a random sample of 111 students entering the College of Computing with a B average had lost their HOPE scholarship by the end of the first year.

- Is there significant evidence that the proportion who lose HOPE scholarships is different for the College of Computing than for the Ivan Allen College?
- What conditions are necessary in order for the technique you used in part a to be appropriate?
- Do you think they are reasonable in this case? Discuss and make sure that you address all the conditions .

17. A psychological study was conducted to compare the effect of caffeine on reaction times (in seconds) of people. Assume that one hundred people are available for this study. The two treatments are: (1) no caffeine for the past 48 hours and (2) caffeinated drinks total of 200 mg of caffeine per day for the last two days (48 hours.)

- Describe in words and a diagram an appropriate completely randomized experimental design for this experiment.
- Describe in words an appropriate matched-pairs experimental design for this experiment. Suppose the experiment you described in part (a) was performed. The data for Treatment “caffeine” are summarized by: $\bar{x} = 3.6$ and $s = 0.424$. The data for Treatment “no caffeine” are summarized by: $\bar{x} = 3.8$ and $s = 0.374$.
- Do the data present sufficient evidence to suggest a difference between true mean reaction times for the two groups? Use a 5% significance level.
- What conditions are necessary to use the technique you used?

19. Consider the data about the reaction times in problem 17.

- Find and interpret a 95% confidence interval for the difference of the two means.
- Based on this confidence interval, does there seem to be evidence that the mean reaction times for the two groups are different?

21. According to government data, 22% of American children under the age of six live in households with incomes lower than the official poverty level. A random sample of 200 children is selected for a study of learning in early childhood. What is the probability that at least 50 of the children in the sample live in poverty?

23. Consider the same data as in problem 15.

- Find a 95% confidence interval for the proportion of students from the College of Computing who lost their HOPE scholarship by the end of the first year.
- What conditions are necessary for the technique you used in part (a) to be appropriate?
- Are those conditions met in this case? Discuss.
- Would you have done anything differently in part (a) if there had been only 3 of the 111 students who had lost their HOPE scholarship by the end of the first year? Explain.

This is the end of the first sample test. If you are submitting your answers by email, and you did full solutions, a one- or two-line answer in the email message will be adequate if it shows enough details. (But be prepared to show me your handwritten work when asked.) In the email message, give me the numerical answer you found (for the interval, or the t-score and p-value, or the probability) and some part of the answer to the questions about conditions or interpretations.

Do not put more than is necessary in the email message. That is a waste of time and you need to get on to learning rather than just typing solutions.

Use the same instructions as the first sample test

2. You have measured the systolic blood pressure of a random sample of 25 employees of a company located near you. A 95% confidence interval for the mean systolic blood pressure for the employees of this company is (122,138). Which of the following statements gives a valid interpretation of this interval?
- 95% of the sample of employees have a systolic blood pressure between 122 and 138.
 - 95% of the population of employees have a systolic blood pressure between 122 and 138.
 - If the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population mean systolic blood pressure.
 - The probability that the sample mean blood pressure is between 122 and 138 is .95.

In each of 3 - 5, the situation requires a test of hypotheses about a population mean μ . Define that population mean and state the appropriate null hypothesis H_0 and alternative hypothesis H_a .

4. A consumer advocacy group thinks that Wheat Flakes brand cereal contains, on the average, less than the advertised weight of 15 ounces per box. They do a random sample of size 64 to gather data to test this claim.

6. A quality control engineer in a bakery goods plant needs to estimate the mean weight, μ , of bags of potato chips that are packed by machine. She knows from experience that $\sigma = 0.1$ oz for this machine. She wants to find a 99% confidence interval with width less than .02 oz. How large a sample must she take?

8. White remains the most popular car color in the United States, but its popularity appears to be slipping relative to other colors. According to an annual survey by DuPont, white was the color of 20% of the vehicles purchased during 1993, a decline of 4% from the previous year. (According to a DuPont Spokesperson, white represents "innocence, purity, honesty, and cleanliness.") A random sample of 400 cars purchased during this period in a certain large metropolitan area resulted in 100 that were white.

- Does the proportion of all cars purchased in this area that were white appear to be higher than the national percentage?
- What conditions are necessary to use the technique you used?
- Discuss whether this study meets each of those conditions.

10. A major oil company has developed a new gasoline additive that is supposed to increase mileage. In the experiment, 10 cars are randomly selected, and each is driven with and without the additive. The data, d , are the mpg with the additive minus the mpg without the additive. Summary statistics are $\bar{d} = 0.66$ mpg and $s_d = 0.44$ mpg.

- Find and interpret a 95% confidence interval for μ , the average difference in mpg attributable to the additive in the entire population.
- Discuss what conditions must be made about these data in order for the procedure you used in part a. to be valid.

12. The U.S. Federal Highway Administration reports that the average passenger vehicle was driven 8.9 thousand miles in 1982. A random sample of $n = 500$ passenger vehicles gave an average of $\bar{x} = 8.7$ thousand miles driven for last year, with a standard deviation of $s = 6.0$ thousand miles. Let μ denote the mean distance driven per passenger vehicle for last year for the entire population.

- Test the hypotheses $H_0: \mu = 8.9$ thousand miles versus $H_a: \mu \neq 8.9$ thousand miles. Report the P value and state your conclusion about car mileage.
- Discuss what conditions must be made about these data in order for the procedure you used in part a. to be valid, in order for this procedure to be valid.

14. A person released from prison before completing the original sentence is placed under the supervision of a parole board. If that person violates specified conditions of good behavior during the parole period, the board can order a return to prison. The article "Impulsive and Premeditated Homicide: An Analysis of the Subsequent Parole Risk of the Murderer" (*J. of Criminal Law and Criminology* (1978): 108-114) reported the accompanying data on parole behavior. One random sample of individuals had served time in prison for impulsive murder and the other random sample had served time for premeditated murder.

| | Impulsive | Premeditated |
|--------------------------|-----------|--------------|
| Sample size | 42 | 40 |
| Number with no violation | 13 | 22 |

- Construct a 98% confidence interval for the difference in the proportion who successfully completed parole for impulsive and premeditated murderers.
- What conditions are necessary in order for the technique you used to be appropriate?
- Do you think they are reasonable in this case? Discuss.
- Without further computation**, is there a significant difference between the two proportions at a significance level of 2%? Discuss.

16. A study was conducted by the Florida Fish and Game Commission to assess the amounts of chemical residues found in the brain tissue of brown pelicans. In a test for DDT, random samples of 10 juveniles and 13 nestlings gave the following results. The data were measured in parts per million. Juveniles: $\bar{x} = 0.041$ and $s = 0.017$. Nestlings: $\bar{x} = 0.026$ and $s = 0.006$.

- Test the hypothesis that there is no difference between mean amounts of DDT found in juveniles and nestlings versus the alternative that the juveniles have a larger mean. Use a 5% significance level. (This test has important implications regarding the accumulation of DDT over time.)
- What conditions must you make about these data to use the technique that you used?
- How could you check these conditions ?

18. In a study of the effects of acid rain, a random sample of 100 trees from a particular forest are examined. Forty percent of these show some sort of damage. Which of the following statements is correct? (Circle one.)

- 40% is a parameter.
- 40% is a statistic.
- Forty percent of all trees in this forest show some sort of damage.
- More than forty percent of all trees in this forest show some sort of damage.
- Less than forty percent of all trees in this forest show some sort of damage.

20. Consider the data in problem about DDT and pelicans in problem 16. Find and interpret a 95% confidence interval for the difference in the mean accumulation of DDT in juvenile and nestling brown pelicans.

22. In a country of 40,250 households, the mean annual income per household is \$11,387 and the standard deviation is \$3419.

- What is the distribution of the mean income for samples of size 900?
- Find the probability that, in a sample of 900 households, the mean annual income is between \$10,500 and \$11,500.

Submit your answers the same way as for the first sample test.