M358K Sp 06 (Smith)

REVIEW PROBLEMS FOR SECOND EXAM

Please Note:

- Exam will cover Chapters 4 – 8.
- This review sheet is *not* intended to tell you what will or what will not be on the exam. However, many of these problems have appeared on or are very similar to problems that have appeared on previous exams in this course. Some others might not appear on an exam, but are designed to help you learn things that might be on the exam.
- In all problems involving a hypothesis test, you are expected to i) state your null and alternate hypotheses clearly, ii) calculate a test statistic, iii) give a P-value as precise as your tables allow, iv) specify degrees of freedom (if appropriate), and v) state your conclusion in terms of the context of the problem.

Remember that on the exam you will be expected to give reasons and explain your work clearly.

Textbook problems: 5.59, 5.63, 6.97, 6.109, 6.111, 7.121, 8.21, 8.23, 8.59, 8.63

Other problems:

1. What is a sampling distribution? Why are sampling distributions important? Illustrate with examples.

2. What kind of analysis (hypothesis test, confidence interval, correlation, regression; one, matched pair, or two sample procedure; z or t statistic; none of the above) should be used in each of the following situations? Be as precise as possible (stating all of the above options that apply – for example, "two-sample hypothesis test using a t-statistic") and *give reasons for your choices*.

a. The level of phosphorus in a person's blood tends to vary normally over time. The level of phosphorus in the blood of a kidney dialysis patient was measured on six consecutive clinic visits. The measurements (in milligrams of phosphorus per deciliter of blood) were

5.6 5.1 4.6 4.8 5.7 6.4.

These can be considered an SRS of the patient's blood phosphorus level. Find a 90% confidence interval for the patient's mean phosphorus level.

b. A water quality inspector wants to compare the average amounts of a certain chemical in the local water supply before and after a new chemical processing plant starts operating. She takes water samples at several randomly chosen locations before the plant opens and samples at the same locations after the plant has been in operation for a month.

c. A basketball player suspects that she makes shots better from the left side of the court than from the right. She decides to conduct an experiment to test this out. She decides to make 40 shots and randomly selects twenty of theses to be made from the left side of the court; the rest are made from the right side. All shots from the left are made from the same spot, and all shots from the right side are made from the symmetric spot on the right side of the court.

d.A researcher wants to test the effectiveness of a new weight loss program. He advertises the program and recruits 40 people to try it. He plans to compare the average weight loss of these 40 people with the average weight loss of 40 people randomly selected from those following an existing weight loss program.

e. You wish to find out how much UT students study per week, on average. To obtain your sample, you ask all the students in the TV lounge of your dorm on Wednesday night how many hours per week they study.

f. A researcher wished to compare the average amount of time spent in extracurricular activities by high school students in a suburban school district with that in a school district of a large city. The researcher will use an SRS from the suburban district and another SRS from the urban district.

g. Twenty years ago, the average number of years an employees had worked for the postal service was 7.5 years. You wish to determine whether this is still the case. You take an SRS of 100 postal employees and find that the average time these employees have worked for the postal service is 7 years with (sample) standard deviation 2. (You may assume that the distribution of the time the population of employees have worked for the postal service is approximately normal.)

h. You wish to find the average number of hours per week the 40 students living in a certain co-op spend studying .

i. A teacher compares the scores of students in her class on a standardized test with the national average score.

j. A political analyst speculated that younger adults were becoming more conservative. To test this, he decided to see if the mean age of registered Republicans was lower than that of registered Democrats. He selected an SRS of 128 registered Republicans form a list of registered Republicans and determined the mean age to be 39 years with a standard deviation of 8 years. He also selected an independent SRS of 200 registered Democrats from a list of registered Democrats and determined the mean age to be 40 years with a standard deviation of 10 years.

k. You want to decide what fraction of the variation in college GPA's is accounted for by SAT scores.

m. A sample of 24 couples obtaining marriage licenses in Cumberland County, Pennsylvania in 1993 had ages as shown on the next page. Do these data support the hypothesis that women marry men older than they are?

Couple number	Husband's age	Wife's age
1	25	22
2	25	32
3	51	50
4	25	25
5	38	33
6	30	27
7	60	45
8	54	47
9	31	30
10	54	44
11	23	23
12	43	39
13	25	24
14	23	22
15	19	16
16	71	73
17	26	27
18	31	36
19	26	24
20	62	60
21	29	26
22	31	23
23	29	28
24	35	36

2. For each of the above examples that involves a hypothesis test, state the null and alternate hypotheses.

3. You are studying a random variable X whose distribution is not normal, but not too far off from normal. You do calculations using normal random variables to approximate each of the following:

a. The probability of obtaining a value of X that is ≤ 4.
b. The probability that the mean x-bar from a random sample of 10 from X is ≤ 4.
c. The probability that the mean x-bar from a random sample of 100 from X is ≤ 4.

Which of your three calculations is most accurate? Which is least accurate? Why?

4. The weights of extra large eggs have a normal distribution with a mean of one ounce and a standard deviation of 0.1 ounces.

a. What is the probability that a single extra large egg weighs less than 0.9 ounces?

b. What is the probability that the average weight of the twelve eggs in a randomly selected carton of a dozen extra large eggs is less than 0.9 ounces?

5. The Gallup Poll once found that about 15% of American adults jogged regularly. Suppose that in fact the proportion of the American adult population who jogged regularly at that time was p = 0.15. You want to find the probability that a simple random sample of 100 American adults would contain at least 20 regular joggers.

a. What are the mean and standard deviation of the distribution of p-hat, the proportion of joggers in a SRS of 100 American adults at that time?

b. Explain carefully why, in this situation, you can use the normal approximation to the sampling distribution of p-hat.

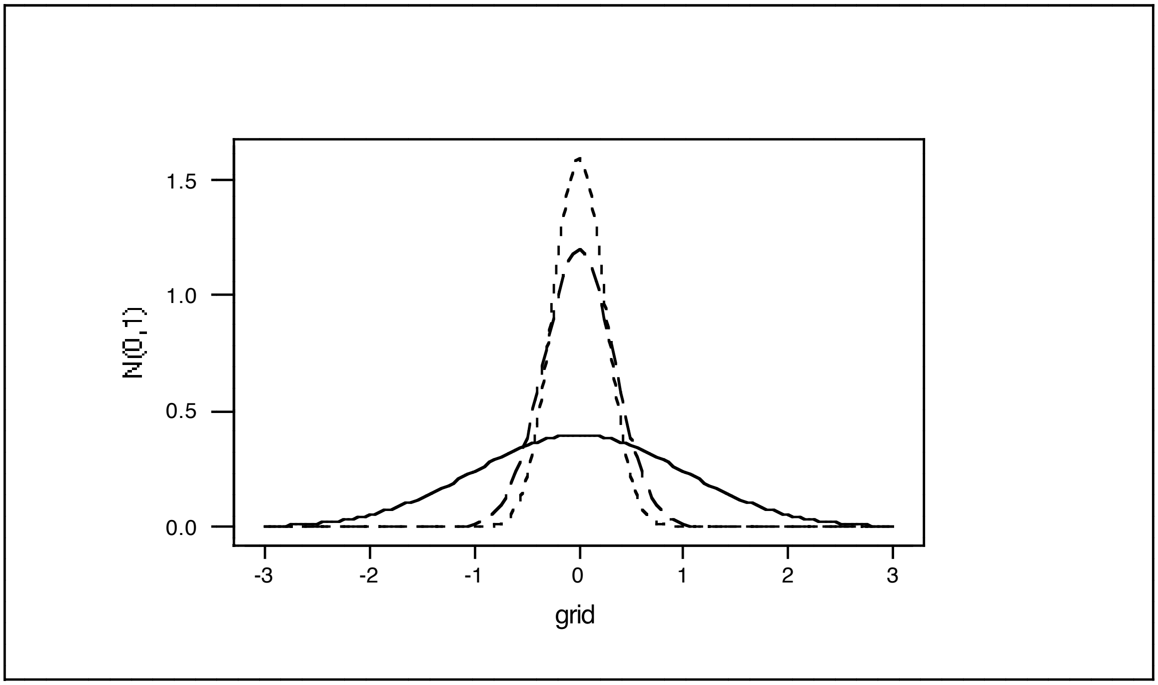
c. Use the normal approximation to find the approximate probability in question.

6. Sometimes people describe a 95% confidence interval for a mean by saying, “90% of the time, the true mean is in the confidence interval,” or ,”The probability that the population mean is in the confidence interval is .95.” The trouble with these descriptions is that they are ambiguous: with one possible interpretation they are correct, but with another possible interpretation they are incorrect. Explain what these two possible interpretations are and which one is correct.

7. The scores of individual students on the American College Testing (ACT) Program composite college entrance examination have a normal distribution with mean 18.6 and standard deviation 6.0. A simple random sample of 36 students take the test. What is the distribution of the average (sample mean) score for the 36 students? What would the distribution be if we were sampling 49 students?

8. The weights of medium oranges packaged by an orchard are normally distributed with a mean of 14 ounces and a standard deviation of 2 ounces. The weights of large oranges packaged by this orchard are normally distributed with a mean of 18 ounces and a standard deviation of 3 ounces. Find the probability that a medium packaged orange selected at random weighs more than a large packaged orange selected at random.

9. The diagram below shows three distributions: One is the distribution of a certain random variable x, one is the sampling distribution of the sample means x-bar for samples of size 9, and one is the sampling distribution of the sample means x-bar for samples of size 16. Which is which? How do you know?



10. A survey asks a simple random sample of 1500 adults in Ohio if they support an increase in the state sales tax from 5.5 to 6%, with the additional revenue going to education. Let p-hat denote the proportion in the sample that say they support the increase. Suppose that 40% of all adults in Ohio support the increase.

a. What are the mean and standard deviation of p-hat?

b. How large would the sample need to be to guarantee that the standard deviation of p-hat is no more than 0.01?

c. What is the probability that p-hat is more than 0.50?

11. A 95% confidence interval for the mean mu of a population is computed from a random sample and found to be 9±3. Decide which of the statements below are true, false, or ambiguous. Try to identify where any confusions have arisen.

a. There is a 95% probability that mu is between 6 and 12.

b. There is a 95% probability that the true mean is 9 and there is a 95% chance that the true margin of error is 3.

c. If we took many, many additional random samples and from each computed a 95% confidence interval for mu, approximately 95% of these intervals would contain mu.

d. There is a 95% chance that the mean of a sample collected from this population is between 6 and 12.

e. 95% of all test statistics will fall in the range 6 to 12.

f. 95% of the sample data will fall between 6 and 12.

12. In a study comparing ninth graders in junior high schools and ninth graders in senior high schools, researchers considered the grade point averages of a simple random sample from each group. The results are shown below.

	Number	Mean GPA	Sample Standard Deviation
Ninth graders in Jr. High School	771	2.59	0.89
Ninth graders in Sr. High School	825	2.24	1.11

a. Find a 99% confidence interval for the mean GPA of the ninth graders in senior high school.

b. The researchers believed that there would be no significant difference in mean GPA between the two groups. Carry out an appropriate hypothesis test on the data. Be sure to do all of the following that are appropriate: state hypotheses clearly, find a test statistic, state what degrees of freedom you are using, find a p-value as precisely as possible from the table available, and state your conclusion in the context of the study.

13. For each statement below decide which category best fits: true, false, incomplete, or ambiguous. Explain, and also try to identify any confusions.

a. The p-value is the probability, assuming the null hypothesis is false, that the test statistic will take a value at least as extreme as that actually observed.

b. The p-value is the probability, assuming the null hypothesis is true, that the test statistic will take a value at least as extreme as that actually observed.

c. The p-value is the probability, assuming the alternate hypothesis is false, that the test statistic will take a value at least as extreme as that actually observed.

d. The p-value is the probability, assuming the alternate hypothesis is true, that the test statistic will take a value at least as extreme as that actually observed.

e. The p-value is the probability that the null hypothesis is false.

f. The p-value is the probability that the null hypothesis is true.

g. In a hypothesis test for a mean, with alternate hypothesis mu > 1, if p = .99,
then the sample average is almost always greater than 1.

h. In a hypothesis test with null hypothesis mu =0, if p = .001, then the chances that mu = 0 are very small.

i. In a hypothesis test for a mean, with alternate hypothesis mu > 1, if p = .99,
then there is more than a 99% chance that mu > 1.

j. If you obtain a p-value of 0.01, you have absolutely disproved the null hypothesis.

k. If you obtain a p-value of 0.01, you have absolutely proved the alternate hypothesis.

l. From a hypothesis test, you can deduce the probability that the alternate hypothesis is true.

m. The p-value is the probability that the alternate hypothesis is true.

n. If the p-value in a matched pairs test is .05, then the two measurements are equal on average 5 times out of every 100 samples.

o. The p-value is the probability that you are wrong if you reject the null hypothesis.

p. If you obtain a p-value of .01 from an experiment, then you know that if the experiment were repeated a large number of times, you would obtain a significant result on 99% of these occasions.

14. Suppose that the population of the scores of all high school seniors who took the SAT-M (SAT-math) test this year follows a normal distribution with mean mu and standard deviation 100. You read a report that says, “On the basis of a simple random sample of 100 high school seniors who took the SAT-M this year, a confidence interval for mu is 512.00 ± 25.76.” What is the confidence level for this interval?

15. To assess the accuracy of a laboratory scale, a standard weight which is known to weigh 1 gram is repeatedly weighed a total of n times and the mean x-bar of the weighings is computed. Suppose the scale readings are normally distributed with unknown mean mu and standard deviation 0.01. How large should n be so that a 95% confidence interval for mu has a margin of error of .0001?

16. X and Y are independent random variables with respective means μ_X and μ_Y and respective standard deviations σ_X and σ_Y.

a. What is the mean of X – Y? Explain why. Does your answer use the fact that X and Y are independent? If so, where?

b. What is the variance of X – Y? Explain why. Does your answer use the fact that X and Y are independent? If so, where?

c. Expand on the above to explain why the standard error for a two sample t-test is taken to be $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$. Explain how your explanation shows why the samples in a two-sample t-test must be independently chosen.

17. A technician tests whether or not a certain part is defective by taking a random sample of measurements of its output and performing a hypothesis test. The alternate hypothesis of the test is equivalent to "the part is defective." The p-value of the test is 0.04, so the technician rejects the null hypothesis in favor of the alternate. Which of the following is true:

a. The part is definitely defective.

b. The technician has decided that the part is defective, but it might not be.

c. Neither (a) nor (b) is true.