

Stat 13 midterm, Prof. Rick Paik Schoenberg,

1. You may use a calculator, a pencil, and any books and notes you want during the exam, but do not use computers, tablets, phones, or anything that can communicate over the web. All students must solve all the problems completely by themselves.
2. Final numerical answers are rounded to 3 significant digits.
3. There are 20 multiple choice questions worth 5 points each.
4. No partial credit is given for multiple choice questions. Choose ONE answer only.

Consider the following for the next 5 problems. Adults with a Body Mass Index (BMI) above 29.0 are sometimes considered *obese*. There are many serious known problems with BMI, but for the purpose of these problems, we will use the term obese for an adult with BMI above 29.0. About 10.0% of adults are known to suffer from knee pain. Miranda et al. (2002) obtained a simple random sample of several thousand adults, and investigated whether those who were obese were more likely to develop knee pain. In their sample, 255 of the adults were obese, and after surveying these subjects, the researchers found that 46 of the 255 obese adults had knee pain.

1. Which of the following is true?

- a. A **parameter** of interest is the fraction of obese adults in the sample who have knee pain, and it is used to estimate the overall fraction of adults who have knee pain, which is a **random variable**.
- b. A **parameter** of interest is the overall fraction of adults with knee pain who are obese, and the **random variable** used to estimate it is the fraction of adults in the sample who are obese.
- c. A **parameter** of interest is the overall fraction of obese adults who have knee pain, and the **random variable** used to estimate it is the fraction of obese adults in the sample who have knee pain.
- d. A **parameter** of interest is the central limit theorem, and the **random variable** used to estimate it is the conditional proportion of the t-test confidence interval.
- e. None of the above.

2. In order to test whether the risk of knee pain depends on obesity, what would the **null hypothesis** be?

- a. 10.0% of obese adults have knee pain.
- b. Of all adults with knee pain, the percentage who are obese is not 10.0%.
- c. Of all adults with knee pain, the percentage who are obese is 10.0%.
- d. Of all obese adults with knee pain, 10% made it into the sample.
- e. None of the above.

3. Based only on the sample of 255 obese adults, find a 95% CI for the percentage of obese adults who have knee pain.

- a. 18.0% +/- 3.29%.
- b. 18.0% +/- 3.65%.
- c. 18.0% +/- 4.17%.
- d. 18.0% +/- 4.72%.
- e. None of the above.

4. Under this null hypothesis, what is the **standardized Z** statistic, summarizing the difference between the sample percentage of obese people who have knee pain and the population percentage of obese people who have knee pain? Hint: use the value of π under the null hypothesis to compute the SE. Also, be careful not to round your answers until you get the final answer.

- a. 3.58.
- b. 4.28.
- c. 7.28.
- d. 7.78.
- e. None of the above.

5. Which of the following can be concluded?

- a. The difference between the percentage with knee pain in the sample and the percentage with knee pain in the population under the null hypothesis is statistically significant.
- b. The percentage of obese people who have knee pain is 10.0%.
- c. Knee pain does not appear to be associated with obesity.
- d. The percentage of obese people who have knee pain is less than 10.0%.
- e. None of the above.

6. In the study on clofibrate, researchers found that 25% of patients in the clofibrate group who stopped taking clofibrate died within 5 years, compared to just 15% of those in the clofibrate group who continued to take it. Does this prove that clofibrate works?

- a. No. Adherers were healthier than non-adherers, even in the placebo group.
- b. Yes. For middle aged men suffering from heart problems, taking clofibrate causes them to increase their chance of living 5 years by 10%.
- c. No. The clofibrate group was younger than the placebo group.
- d. No. The study was not double-blind, and this likely influenced the results.

7. Suppose researchers take a simple random sample of 400 American adults and find a mean systolic blood pressure of 120 mm Hg and an SD of 50 mm Hg. A 95% CI for the mean systolic blood pressure of all American adults is 120 +/- ____ mm Hg.

- a. 0.594.
- b. 0.785.
- c. 1.47.
- d. 4.90.
- e. None of the above.

8. A survey of 1000 Americans selected at random by random digit dialing finds that those who brush their teeth more than 14 times per week have 30% lower rates of heart disease than those who brush their teeth fewer than 3 times per week. The study concluded that frequent brushing of the teeth may be beneficial for the heart. What is the main problem with this conclusion?

- a. These 1000 Americans are probably wealthier than the overall population.
- b. Those who rarely brush their teeth are probably less health conscious in general.
- c. The explanatory confidence interval is greater than the significance level.
- d. Genes linked to heart disease might make people less likely to brush their teeth.
- e. None of the above.

9. A scientist surveys patients at an exclusive research hospital in Beverly Hills, and records information on the respondents. Which of the following is true?

- a. The subjects are likely to be taller and more talkative than the overall population.
- b. This is an experiment because the data are recorded by researchers.
- c. The subjects are likely wealthier and better educated than the overall population.
- d. This study is double-blind with paired data on each respondent.
- e. None of the above.

For the next three problems, use the following information. In order to investigate if squid or anchovies have higher mercury levels, scientists took a simple random sample of 100 squid and found their mean mercury concentration was 22.0 parts per billion (ppb), with an SD of 7 ppb. They also took a simple random sample of 144 anchovies and found their mean mercury concentration was 15.0 ppb and the SD was 11 ppb.

10. Find a 95% CI for how many fewer ppb of mercury are in anchovies than in squid.

- a. 7.0 +/- 2.04.
- b. 7.0 +/- 2.26.
- c. 7.0 +/- 3.72.
- d. 7.0 +/- 3.98.
- e. None of the above.

11. Is the difference between the mean mercury concentration in anchovy and the mean mercury concentration in squid statistically significant?

- a. No.
- b. Yes.
- c. It cannot be determined because this is an observational study.

12. If scientists took another simple random sample of 100 squid and 144 anchovies, computed the difference between the sample mean mercury concentration for squid minus the sample mean mercury concentration for anchovies, and used this to estimate the difference between the two population means, how much would the estimate typically be off by?

- a. About 0.350 ppb.
- b. About 0.794 ppb.
- c. About 0.963 ppb.
- d. About 1.15 ppb.
- e. None of the above.

13. In general, the p-value of a test of significance is:

- a. the probability, assuming the null hypothesis is true, that the test statistic will take a value at least as extreme as that actually observed.
- b. the probability the null hypothesis is true.
- c. the probability the null hypothesis is false.
- d. the probability, assuming the alternative hypothesis is true, that the test statistic will take a value at least as extreme as that actually observed.
- e. None of the above.

14. Which of the following is **not** an advantage of having a slightly larger sample size, assuming everything else is held fixed.

- a. The resulting significance test will have greater power.
- b. The sample will be more representative of the population.
- c. The resulting confidence interval will be narrower.
- d. The resulting margin of error will be smaller.

15. Suppose a random sample of 100 Californians is asked if they regularly listen to music. A 95% confidence interval for the proportion of all Californians regularly listening to music is found to be 0.70 to 0.90. If a 90% confidence interval was calculated instead, how would it differ from the 95% confidence interval?

- a. The 90% confidence interval would have the same width as the 95% confidence interval.
- b. The 90% confidence interval would be narrower.
- c. The 90% confidence interval would be wider.
- d. More information is needed in order to determine if the 90% CI would be wider or narrower than the 95% CI.
- e. None of the above.

16. For the next two problems, suppose you take a simple random sample of 15 residents of a certain city and find their mean income is \$62,000, the median income is \$35,000, and the SD is \$10,000. Which of the following is true?

- a. The incomes appear to be right skewed.
- b. The incomes appear to be normally distributed.
- c. The incomes appear to be left skewed.
- d. None of the above.

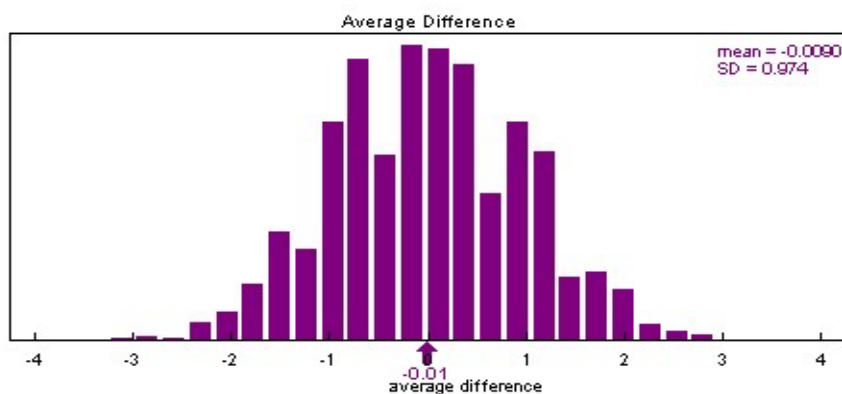
17. You want to find a 95% CI for the mean income of residents in this city, and are considering using $\$62,000 \pm 1.96 \times \$10,000/\sqrt{15}$. Which of the following is true?

- a. You should use this formula because of the central limit theorem, since the sample size is large.
- b. You should use this formula because incomes are normally distributed.
- c. You should not use this formula because the study is an observational study rather than an experiment.
- d. You should not use this formula because the sample size is small, the population SD is unknown and incomes are probably not normally distributed.

18. Suppose echinacea truly does decrease the risk of getting sick, but in a certain study researchers take a simple random sample of 100 people, randomly give half echinacea and the other half placebo, and find the difference in number of sick days between the echinacea group and the placebo group is not statistically significant. Which of the following is true?

- a. The study has too much statistical power.
- b. The sample size of the observational study has a statistically significant adherer bias.
- c. The study is making a Type II error.
- d. The experiment is prone to nonresponse bias.
- e. None of the above.

19. Researchers investigated the number of words 500 subjects could memorize when listening to music or when not listening to music. They recorded each subject twice, once with music and once without music, in random order for each student. Assuming the null hypothesis of no difference as a result of listening to music, the researchers used simulations to generate the following null distribution of the difference between the number of words memorized with music and the number memorized without music.



Based on these simulations, if the researchers want to construct a 95% CI for the mean difference in the number of words memorized, what would be the margin of error?

- a. 1.25 words.
- b. 1.48 words.
- c. 1.91 words.
- d. 2.53 words.
- e. None of the above.

20. In the portacaval shunt example, why did the studies with historical controls find that the portacaval shunt seemed to be associated with lower death rates?

- a. Those getting the shunt smoked more.
- b. Those getting the shunt were healthier.
- c. Those getting the shunt were genetically predisposed to die younger.
- d. The explanatory variable is a confounding factor t-test with 95% central limit theorem.
- e. None of the above.