Stat 13 midterm, Prof. Rick Schoenberg, 7/16/25, 11am-12:50pm.

1. You may use a calculator, a pencil, and any books and notes you want during the exam, but no internet searching or communicating or use of the computer other than to read your notes or submit your answers.

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.

5. Use the exam that matches the FIRST LETTER of your FIRST NAME!!!!!!!!!
If the FIRST LETTER OF YOUR FIRST NAME is A through K, then use the exam called midA.pdf .
If the FIRST LETTER OF YOUR FIRST NAME is L through S, then use the exam called midL.pdf .
If the FIRST LETTER OF YOUR FIRST NAME is T through Z, then use the exam called midT.pdf .
If there is ambiguity about which of your names is first, use whatever comes first on your "from" when you send an email.

6. You have from 11am to 12:50pm. By 12:50pm you must EMAIL me your answers, to frederic@stat.UCLA.edu . Your email should just contain your name and your answers, like ADDBC CDAAB BBCCD DAABC. You do not need to show work. Make sure your answers are exactly in the correct order!

7. You must zoom in to the usual zoom while taking the exam. If you have a question during the exam, ask it using chat.

Consider the following for the next 13 problems. Researchers take a simple random sample of 1000 Californians and find that 52 of them have done yoga today. The researchers also take a simple random sample of 500 Nevadans and find that 17 of them have done yoga today.
Of the 69 subjects who have done yoga today in California and Nevada combined, the researchers find their average systolic blood pressure to be 115 mm HG, with a SD of 15 mm HG, and of the 1431 who have not done yoga today, their average systolic blood pressure is 120 mm HG, with a SD of 16 mm HG.

_____ 1. Find a 95%-CI for the percentage of Californians who have done yoga today.
a. 5.2% +/- 0.723%. b. 5.2% +/- 1.14%. c. 5.2% +/- 1.38%. d. 5.2% +/- 2.52%.
e. 5.2% +/- 2.81%. f. None of the above.

_____ 2. If one were to test whether the percentage of all Californians who have done yoga today is 4%, what would the size of the Z-statistic be?
a. 1.68. b. 1.94. c. 2.45. d. 3.23. e. 3.96.

_____ 3. Find a 95%-CI for the difference between the percentage of Californians who have done yoga today and the percentage of Nevadans who have done yoga today.
a. 1.30% +/- 2.72%. b. 1.30% +/- 2.92%. c. 1.80% +/- 1.95%.
d. 1.80% +/- 2.10%. e. None of the above.

_____ 4. If one were to test whether the percentage of Californians who have done yoga today is statistically significantly different from the percentage of Nevadans who have done yoga today, what would the Z-statistic be? (Hint: use the pooled percentage who have done yoga in computing your standard error.)
a. 0.808. b. 1.57. c. 1.71. d. 1.83. e. 1.90. f. None of the above.

_____ 5. Which of the following conclusions is appropriate, based on either the 95%-CI or Z-statistic found in the previous two problems?

a. Statistically significantly more Californians did yoga today than Nevadans.

b. We reject the null hypothesis that the same fraction of Californians as Nevadans did yoga today.

c. The null hypothesis of zero non-response bias is rejected, due to the power of the low sample size of the confidence level for the confounding factors.

d. The difference between the fraction of Californians who did yoga today and the fraction of Nevadans who did yoga today is not statistically significant.

e. None of the above.

_____ 6. Find a 95%-CI for the mean systolic blood pressure, in mm HG, among those who did yoga today, in California and Nevada combined.

a. 115 +/- 1.72. b. 115 +/- 2.04. c. 115 +/- 2.53. d. 115 +/- 3.54.

e. 115 +/- 3.71. f. None of the above.

_____ 7. If we were to compute a 99% CI instead of 95%, how would the interval change?

a. The 99% CI would be wider.

b. The 99% CI would be narrower.

c. The 99% CI would be the same.

_____ 8. If one were to test whether the mean systolic blood pressure among those in both California and Nevada who did yoga today is statistically significantly different from the mean systolic blood pressure among those in California and Nevada who did not do yoga today, what would the size of the Z-statistic be?

a. 2.70. b. 4.40. c. 4.81. d. 8.77. e. 9.62. f. None of the above.

_____ 9. Find a 95%-CI for the mean systolic blood pressure among those who did not do yoga today minus the mean systolic blood pressure among those who did yoga today, in mm HG.

a. 5 +/- 2.02. b. 5 +/- 2.59. c. 5 +/- 3.12. d. 5 +/- 3.64.

e. 5 +/- 4.14. f. None of the above.

_____ 10. Which of the following conclusions is appropriate?

a. The difference between the mean systolic blood pressure among those who did yoga today and the mean systolic blood pressure among those who did not do yoga today is not statistically significant.

b. We accept the null hypothesis that there is no difference in blood pressure between those who did yoga today and those who did not.

c. The sample size is a statistically significant confidence interval with power corresponding to Type II errors due to adherer bias.

d. Those who did yoga today had statistically significantly lower average systolic blood pressures.

e. None of the above.

_____ 11. Can we conclude that doing yoga lowers your systolic blood pressure?

a. No, because this was an experiment and is therefore prone to confounding factors.

b. No, because this was an observational study, and stress is a confounder since those who had time to do yoga were probably less stressed out to begin with and thus had lower blood pressures.

c. Yes, because this was an experiment and experiments are not as prone to confounding factors.

d. No, because this was an observational study, and genetics is a plausible confounding factor because it is likely that a gene that makes you like yoga also lowers your blood pressure.

e. None of the above.

_____ 12. Of the 69 subjects who have done yoga today, a **typical** one of them has a systolic blood pressure of about

a. 115 +/- about 2 mm HG. b. 115 +/- about 5 mm HG.

c. 115 +/- about 7 mm HG. d. 115 +/- about 15 mm HG.

_____ 13. Would it be appropriate to do a t-test to determine if the percentage of Nevadans who did yoga today is significantly different from the percentage of Californians who did yoga today?

a. Yes, because the sample size is small.

b. Yes, because the variable being studied is approximately normally distributed.

c. No, because the variable being studied is binary and therefore not normally distributed.

d. Yes, because the variable being studied is binary and the sample size is small.

_____ 14. Which is an appropriate interpretation of the results of the study on Echinacea by Oneil et al. (2008)?

a. The study showed that Echinacea fails to curb the common cold.

b. The study showed that Echinacea won't help much in curing the common cold.

c. Those who took Echinacea had 5 fewer sick days on average than those in the placebo group, but the difference was not statistically significant due to the study's small sample size.

d. The study found that those who took Echinacea had nearly the same number of sick days on average as those in the placebo group, but even the small observed difference was found to be statistically significant due to the large sample size.

e. The study was an observational study and thus prone to confounding factors, such as age, since those in the Echinacea group were significantly younger than those in the placebo group, which could explain why those in the Echinacea group had fewer sick days on average.

_____ 15. Since later studies have shown that Echinacea is effective at reducing the risk of sickness, assuming those later studies are correct, what can we conclude about the study by Oneil et al. (2008)?

a. The study by Oneil et al. (2008) made a Type I error.

b. The study by Oneil et al. (2008) made a Type II error.

c. The study by Oneil et al. (2008) had very high power.

d. The study by Oneil et al. (2008) was an observational study.

e. The study by Oneil et al. (2008) had a large sample size.

_____ 16. Suppose the income of the average American is known to be $50,000 per year. A simple random sample of 17 Nevadans is obtained, and their sample mean income is $47,000 per year, with a sample standard deviation of $10,000 per year. A researcher does a t-test, to see if the incomes of Nevadans are significantly different from Americans as a whole. The researcher calculates the test statistic t = ($47,000 - $50,000) / ($10,000 / √17) = -1.24, and finds the corresponding p-value, for a 2-sided test, is 23.4%. Why is this test inappropriate?

a. Incomes are known not to be normally distributed, so the t-test in inappropriate here.

b. The sample size is too large for a t-test to be appropriate.

c. There are too many confounding factors for a t-test to be appropriate.

d. Whether someone lives in Nevada or not is a binary variable, and the t-test should not be used for binary variables.

e. None of the above.

_____ 17. Researchers take a simple random sample of 400 healthy adult Americans, and randomly assign 200 subjects to take a drug designed to reduce the risk of cancer, and the other 200 to take a placebo. They find that 13 of those taking the drug get cancer during the observation period, compared with 15 in the placebo group, and they find the difference is not statistically significant, with a p-value of 58%. Which of the following conclusions is most appropriate?

a. The data are compatible with the null hypothesis that the drug is ineffective.

b. The data are incompatible with the null hypothesis that the drug is ineffective.

c. The data are incompatible with the null hypothesis that the drug is effective.

d. The data prove that the drug is effective.

e. The data prove that the drug is ineffective.

For the next 3 problems, suppose a researcher takes a simple random sample of 100 high school students, and finds that the number of books read per year per student has a mean of 7.2, a median of 7.0, a 25th percentile of 2.0, and a 75th percentile of 11.0. She finds a 95% CI for the mean number of books read per year per student among high school students in general is 7.2 +/- 2.0.

_____ 18. Find s, the sample standard deviation of the number of books read per year among the 100 students in the sample.

a. 5.40. b. 7.23. c. 7.90. d. 8.24. e. 10.2. f. None of the above.

_____ 19. What is the IQR of the number of books read per year per student?

a. 7.0 b. 8.0. c. 9.0 d. 10.0 e. 11.0. f. None of the above.

____ 20. Student X read 17 books per year, student Y read 20, and student Z read 25. Which ones are formally considered outliers?

a. X, Y and Z.

b. Only Y and Z.

c. Only Z.

d. None of them.