PRINT YOUR NAME: _____

SIGN YOUR NAME: _____

Stat 13 final, Prof. Rick Schoenberg, 3/17/24, 9am-noon, online.

1. Do not turn the page and start the exam until you are told to do so.

2. Enter your name now above.

3. You may use a calculator, a pencil, and any books and notes you want during the exam, but no computers, tablets, phones, or anything that can communicate over the web.

4. Final numerical answers are rounded to 3 significant digits.

5. There are 30 multiple choice questions all worth the same number of points.

6. No partial credit is given for multiple choice questions. Choose ONE answer only.

_____ 1. A researcher is studying whether a drug has an effect on people's blood pressures. She takes a simple random sample of 300 subjects and separates the subjects randomly into 3 groups. The first group takes a placebo, and has a mean blood pressure of 129.2. The second group takes a low dose of the drug, and has a mean blood pressure higher than 129.2, and the third group takes a large dose of he drug, and has a mean blood pressure lower than 129.2. She then does an ANOVA test and the results are below. Which of the following is an appropriate conclusion?

Source	df	SS	MS	F	p-value
Treatment	2	17.28	8.64	2.56	0.0381
Error	297	1001.12	3.37		
Total	299	1018.40			

a. The three groups have mean blood pressures that are not statistically significantly different.
b. The large dose group has a statistically significantly higher blood pressure than the other two groups.
c. The large dose group has a statistically significantly lower blood pressure than the other two groups.
d. At least one of the groups has a statistically significantly different blood pressure than the other two groups.
e. The low dose group has a statistically significantly higher blood pressure than the other two groups.
f. We accept the null hypothesis that, if we had access to the whole population, then the three groups would have the same mean blood pressures.

_____ 2. A researcher takes a simple random sample of 22 subjects and records their incomes as well as whether they prefer cold weather or hot weather. She finds the average income of those 10 subjects who prefer cold weather is $57,000 per year, and the average income of those 12 subjects who prefer hot weather is $56,000 per year. She then does a t-test to see if the differences in incomes is statistically significant. Which of the following is true?
a. The t-test is appropriate here, because she had at least 10 of each weather preference in her sample.
b. The t-test is not appropriate here, because incomes are not normally distributed.
c. The t-test is appropriate here, because simple random samples are always resistant to non-response bias.
d. The t-test is not appropriate here, because simple random samples always have major coverage problems.
e. The t-test is not appropriate here, because experiments are always prone to confounding factors.

_____ 3. For the next two problems, a researcher studies the effect of watching a scary movie on heart rate. She takes a simple random sample of 200 subjects, and randomly divides them into two groups. Group A is shown 10 minutes of a scary movie, and their heart rates are recorded. Group B is shown 10 minutes of a non-scary movie, and their heart rates are recorded. The next week, the subjects return, and Group A is shown 10 minutes of a non-scary movie, Group B is shown 10 minutes of a scary movie and for both groups their heart rates are recorded. The researcher finds the following, for the 200 subjects in groups A and B combined:

Mean heart rate when watching scary movie: 88 bpm.
Mean heart rate when watching non-scary movie: 86 bpm.
SD of heart rates when watching scary movie: 22 bpm.
SD of heart rates when watching non-scary movie: 20 bpm.
SD of differences (scary movie heartrate - non-scary movie heartrate): 8 bpm.

Which of the following is an appropriate Z-statistic for testing whether or not the observed difference in heart rates is statistically significant?
a. 0.572. b. 0.998. c. 1.31. d. 1.77. e. 3.54. f. None of the above.

_____ 4. Continuing the previous problem, suppose you wanted to use simulations to find a null distribution for your standardized statistic. Which of the following would be an appropriate way to conduct the simulations?
a. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the subject's non-scary movie heartrate, and otherwise keep the subject's scores the same.
b. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with another subject's scary movie heartrate, where the other subject is selected at random among the 199 other subjects.
c. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the sample mean heartrate for all 200 subjects.
d. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm.
e. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm +/- 20 bpm.

For the next two problems, a researcher studies whether a certain antidepressant is habit forming. She takes a simple random sample of 212 subjects, and randomly gives half (106 subjects) the antidepressant and the other 106 subjects a placebo. After 6 months, she then tells the subjects to stop taking the antidepressants and asks them whether they have any side effects upon stopping the treatment. She finds that of those who were given the antidepressant, 15% reported side effects upon stopping the treatment, and in the placebo group, 12% reported side effects upon stopping the treatment.

_____ 5. What would the Z-statistic be, for summarizing whether or not the observed difference in the percentage reporting side effects upon stopping treatment is statistically significant?
a. 1.25. b. 1.49. c. 1.62. d. 1.70. e. 1.83. f. None of the above.

_____ 6. Continuing the previous problem, the researcher finds the p-value is 10.6%, for a 1-sided test. What is the appropriate conclusion from this study?
a. Those in the treatment group did not have statistically significantly higher rates of side effects than those in the control group, so this study provides strong evidence that the drug is not habit forming.
b. The drug exhibits statistically significant habit forming tendencies in this sample.
c. Those in the treatment group had higher rates of side effects than those in the control group, but the difference was not statistically significant, so a larger study may be needed to determine whether the drug is habit forming.
d. Those in the treatment group had statistically significantly higher rates of side effects than those in the control group, so this study provides strong evidence that the drug is habit forming.
e. This study provides strong evidence that the drug is not habit forming, but since observational studies are always prone to confounding factors, the results may be somewhat biased especially because of differences in incomes or genetics between the treatment and control groups.

For the next 8 problems, suppose a researcher wants to see if sugar consumption influences blood pressure. She take a simple random sample of 100 subjects, and carefully monitors their sugar intake over a 24 hour period, and also records each subject's systolic blood pressure. She finds the subjects have a mean sugar consumption of 42g with a sample SD of 23g. The 100 subjects have a mean systolic blood pressure of 131 mmHG, with a sample SD of 22 mmHG. The sample correlation, r, between sugar consumption and systolic blood pressure is 0.40.

_____7. What is the slope of the regression line, in mmHg/g, for predicting systolic blood pressure based on sugar consumption?
a. 0.302. b. 0.383. c. 0.422. d. 0.458. e. 0.490. f. None of the above.

_____8. Using the regression line, what is the predicted systolic blood pressure for a patient whose daily intake of sugar is 0g?
a. 115 mmHg. b. 120 mmHG. c. 125 mmHG. d. 130 mmHG. e. None of the above.

_____9. Using the regression line, what is the predicted systolic blood pressure for a patient whose daily intake of sugar is 55g?
a. 131 mmHG. b. 132 mmHG. c. 133 mmHG. d. 134. e. 136 mmHG. f. None of the above.

_____10. How much would the predicted systolic blood pressure for a patient whose daily intake of sugar is 144g typically be off by?
a. 20.2 mmHG. b. 21.1 mmHG. c. 21.6 mmHG. d. 22.3 mmHG. e. None of the above.

_____11. What is the standardized t-statistic, for summarizing whether the correlation is statistically significantly different from zero?
a. 3.29. b. 4.32. c. 5.09. d. 5.85. e. 6.14. f. None of the above.

_____12. Continuing the previous problem, suppose the p-value corresponding to this t-statistic is 0.00187%. Which conclusion is appropriate?
a. Consuming sugar makes your blood pressure increase by a statistically significant amount.
b. Increasing your blood pressure makes your sugar consumption decrease by a statistically significant amount.
c. Consuming more sugar makes your predicted blood pressure increase by an amount that is statistically significantly different from zero.
d. Increasing your blood pressure makes your predicted sugar consumption decrease by a statistically significant amount.
e. This is a randomized, controlled experiment with placebo double-blind t-test confidence intervals based on the normal distribution from the central limit theorem.

_____13. What can we conclude, based on the fact that the observed correlation $r = 0.40$, and $r^2 = 0.16$?
a. 16% of the variability in people's systolic blood pressures is explained by their sugar consumption.
b. Each additional gram of sugar one consumes corresponds to an increase in their predicted systolic blood pressure by 0.40 mmHG.
c. Each additional mmHG of increase in one's systolic blood pressure corresponds to an increase in their predicted sugar consumption of 0.40g.
d. 16% of subjects have sugar consumptions and systolic blood pressures within the 95% confidence interval.
e. We have 16% confidence that 95% of subjects have sugar consumptions and systolic blood pressures within the range of the observations in the sample.

_____ 14. What would be the *main* problem with using the regression line to predict the systolic blood pressure for a subject whose daily sugar consumption was 400g?
a. Extrapolation. b. Non-response bias. c. Coverage. d. Heteroskedasticity. e. Non-normality.

For the next two problems, suppose a researcher named Dr. Lee is studying 5 factors (hair color, shoe size, alcohol consumption, grade point average, and hygiene level) to see if they might be related to the risk of depression. Dr. Lee samples 100 subjects at random from the population, records these 5 explanatory variables for each subject, and also evaluates whether the subjects suffer from depression. Dr. Lee will then do a Z-test to see if each of the factors is significantly related to depression rate. Suppose that, in the overall population, we happen to know that these 5 variables are actually unrelated to depression.

_____ 15. If we want to be at least 95% sure Dr. Lee will not find any of the 5 explanatory variables to be a statistically significant predictor of depression, what could we recommend Dr. Lee do?
a. Increase the sample size to 500 instead of 100.
b. Decrease the significance level for the Z-tests to 0.01 instead of 0.05.
c. Increase the number of factors to 25, instead of 5.
d. Increase the number of response variables to 5, instead of 1.
e. Increase the significance level for the Z-tests to 0.25, instead of 0.05.

_____ 16. Suppose Dr. Lee finds a statistically significant relationship in his sample between shoe size and depression. What would be the explanation?
a. This is a Type I error.
b. This is a Type II error.
c. Dr. Lee's sampling method is biased.
d. Non-response bias.
e. Heteroskedasticity.

_____ 17. Suppose a researcher wants to investigate the possible risk factors for a certain disease. She intends to reach out to a simple random sample of 500 Americans with the disease and give them a survey in which she asks them each 1000 questions about their various daily activities, foods and drinks they consume, and clothes they wear, in order to try to determine if any of the behaviors are different among this population compared to known behaviors in the overall population. For example, if she finds that 5% of the subjects in her study are vegetarian, she would compare that to the percentage of all Americans who are vegetarian, which was estimated in previous research. Which of the following is likely to be a huge problem with this study?
a. Non-normality. b. Non-response. c. Heteroskedasticity. d. Standard error.
e. Sampling variability.

For the next two problems, suppose a researcher takes a simple random sample of subjects to study the relationship between smoking and cavities. They find that the 100 smokers have on average 2.3 cavities with a standard deviation of 1.2 cavities, and the 200 nonsmokers average 2.0 cavities with a standard deviation of 1.1 cavities.

_____ 18. Find a 95%-CI for the mean additional number of cavities per person the smokers have.
a. 0.3 +/- 0.212. b. 0.3 +/- 0.237. c. 0.3 +/- 0.259. d. 0.3 +/- 0.280. e. None of the above.

_____ 19. What can we conclude based on the confidence interval in the previous problem?
a. The difference in the number of cavities between the smokers and nonsmokers is not statistically significant.
b. The smokers have statistically significantly more cavities than the nonsmokers.
c. We fail to reject the null hypothesis that the smokers and nonsmokers have the same number of cavities on average.
d. We fail to reject the alternative hypothesis that the smokers and nonsmokers have the same number of cavities on average.
e. None of the above.

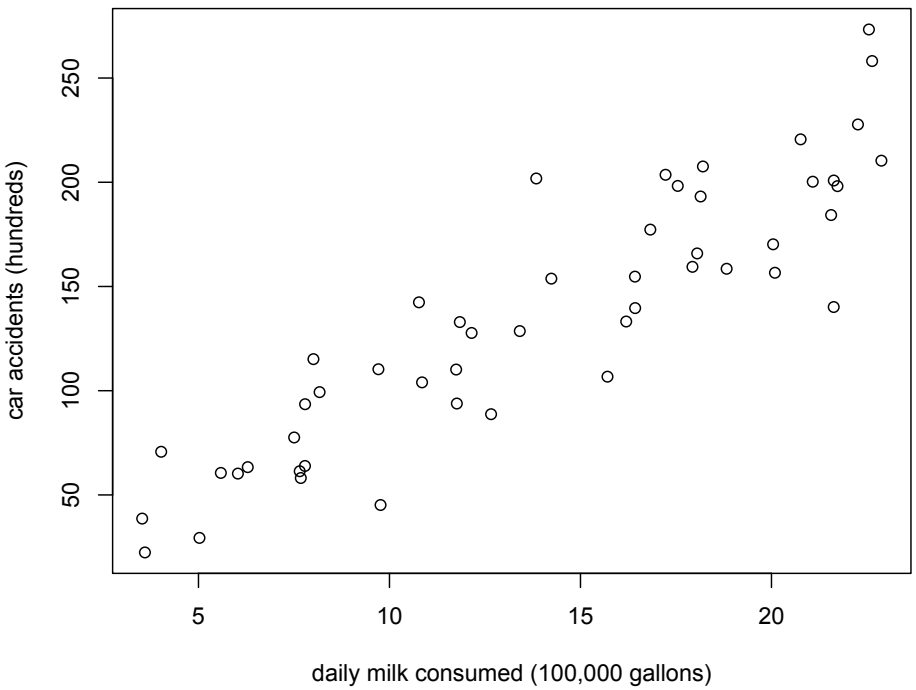
For the next 3 problems, suppose in a simple random sample of 1000 New Yorkers, the number of cavities they have had in their life has the following properties.
Mean = 4.0 cavities.
SD = 2.0 cavities.
Median = 4 cavities.
25th percentile = 1 cavity.
75th percentile = 8 cavities.

_____ 20. What is the IQR of the number of cavities the New Yorkers have had?
a. 4. b. 5. c. 6. d. 7. e. 8. f. None of the above.

_____ 21. Based on this sample, a typical New Yorker has had a number of cavities that deviates from 4.0 by about how many?
a. 0.02. b. 0.04. c. 0.196. d. 2.0. e. 4.0. f. None of the above.

_____ 22. Based on this sample, find a 95%-CI for the mean number of cavities a typical New Yorker has had.
a. 4.0 +/- 0.124. b. 4.0 +/- 0.279. c. 4.0 +/- 0.490. d. 4.0 +/- 0.972. e. None of the above.

_____ 23. Researchers collect data for each of the 50 United States on the total amount of milk consumed daily in the state and the average number of car accidents in the state. The data are shown below, and the correlation is 0.90. Each circle on the plot represents a state. If someone were to propose banning milk in order to decrease the number of car accidents, which of the following would be the most persuasive confounding factor?
a. Genetics. b. Gender. c. Healthcare quality. d. Population. e. Education.



For the next two problems, suppose a simple random sample of 400 Minnesotans is obtained, and they are asked whether or not they trust their doctor. In the sample, 300 say they trust their doctor, and 100 say they do not.

_____ 24. Find a 95%-CI for the percentage of Minnesotans who trust their doctor.
a. 75% +/- 3.02%. b. 75% +/- 3.53%. c. 75% +/- 4.24%. d. 75% +/- 4.87%.
e. None of the above.

_____ 25. Suppose we know the percentage of Americans who trust their doctor is 65%, and we would like to test whether the percentage of Minnesotans who trust their doctor differs significantly from the national percentage. What would the Z-statistic be?
(Hint: use the national percentage to compute the standard error.)
a. 3.78 b. 4.19 c. 5.02 d. 6.17 e. 6.82 f. None of the above.

_____ 26. A study found 25% of patients experiencing post-operative pain claimed they felt "significant, prompt pain relief" after receiving a pill they were told contained pain medication, even though the pill just contained salt and water. What is the best explanation for this result?
a. These 25% of patients were on the verge of recovery and were about to feel better either way, whether they received the pill or not.
b. The idea that the patients were given a healing pill was effective at instantly healing their post-operative wounds via the mind-body connection.
c. People are not very accurate at measuring their own pain level and are easily influenced by the suggestion that their symptoms should be reduced.
d. Salt and water is a highly effective treatment for post-operative wounds.
e. These 25% of patients may have instantly healed by chance alone.

_____ 27. Suppose a researcher decides to do a test at a significance level of 1%, instead of the usual 5%. She computes a p-value of 3%. Which of the following is the appropriate conclusion?
a. The null hypothesis would not be rejected.
b. The null hypothesis would be rejected.
c. The alternative hypothesis would be rejected.
d. The alternative hypothesis would be respected, regretted and ejected.

_____ 28. Which of the following does not directly influence the p-value for a 1 sample Z-test, assuming everything else remains the same?
a. The sample size.
b. The population size.
c. The effect size.
d. The standard deviation of the values in the sample.
e. The decision to do a 1-sided test versus a 2-sided test.

For the next two questions, a researcher is studying how much higher the risk of irritable bowel syndrome (IBS) is among non-vegetarians compared to vegetarians. She takes a simple random sample of 100 vegetarians and 150 non-vegetarians, and finds that 14 of the vegetarians have IBS, and 29 of the non-vegetarians have IBS. She would like to find a 95%-CI for the decrease in risk of IBS among vegetarians.

_____ 29. What is the standard error for the difference between the two percentages? Use the unpooled SE you would use in constructing a 95%-CI.
a. 2.23%. b. 2.78%. c. 3.55%. d. 4.74%. e. 5.02%. f. None of the above.

_____ 30. What is a 95%-CI for the percentage with IBS among non-vegetarians minus the the percentage with IBS among vegetarians?
a. 5.33% +/- 1.91%. b. 5.33% +/- 3.70%. c. 5.33% +/- 5.35%.
d. 5.33% +/- 7.37%. e. 5.33% +/- 9.28%. f. None of the above.