Statistics 101: Practice Exam A for Final Exam

This sheet contains practice problems for the final exam. It is longer than the actual final so that you have extra problems. Other material in the text and from lectures may appear on the final exam.

Questions 1–19 refer to a random sample of 500 heads of households. The data are sampled from the households collected in the March 2000 Current Population Survey. For this problem, we’ll assume that the data are a simple random sample of 500 households from the entire U.S. population.

People who own their homes have to pay taxes on their property. Below is a histogram that shows property taxes for all 500 households in the sample.

![Histogram of Property Taxes](image)

1. True or False: More than 45% of these households have property taxes greater than the mean.

2. Choose the value that you think is closest to the standard deviation of these property taxes: 10, 100, 1000, 10000.

3. True or False: If we remove all the houses that have property taxes equal to zero, the average of the remaining taxes would be larger than $1,050.

4. True or False: A normal curve can be used to determine the percentage of houses paying above $500 with very good accuracy.

5. Estimate the percentage of households that pay more than $2000 in property taxes. _____

Below is a box plot of property tax by marital status. The marital status 1 is for a married householder; 5 is for a divorced householder; and, 7 is for a single householder. There are 219 married people, 84 divorced people, and 108 single people. The other people in the sample who have other marital statuses are not displayed in this graph.
6. Order the three groups by median property tax, going from largest to smallest. ________

7. True or false: The standard deviation of property taxes for these married people is closer in value to the standard deviation for these single people than it is to the standard deviation for these divorced people.

8. True or false: A larger percentage of these married people have property taxes below 500 than do these divorced people.

9. Which of these three groups has the largest percentage of people not paying any property tax?

10. True or False: The standard error of the sample average property tax for divorced people is larger than the standard error for the sample average property tax for married people.

Below is a box plot of property taxes for male and female household heads. Men are coded with a 1, and women are coded with a 2.
11. Are the assumptions for using confidence intervals or hypothesis tests involving sample average of property taxes likely to hold? Explain your reasoning.

12. Give a 95% confidence interval for the difference in average property taxes for male and female household heads.

13. Based on the interval, do you think there is overwhelming evidence that the average property tax amounts differ?

14. Check all of the following that are true:

___ There is a 95% chance that the population difference in average property taxes is between the two values you determined in 14.

___ If we pick two households at random so that one is headed by a male and the other by a female, the difference in their property taxes will fall within the upper and lower limits 95% of the time.

___ If we took another random sample of 500, then another, then another, and so on, we’d expect 95% of the formed confidence intervals to contain the population difference in average property taxes.

15. Test the null hypothesis that there is no difference in average property taxes between male and female household heads. State your null and alternative hypotheses, the test statistic, the p-value, and your conclusions. Consider a p-value near 0.05 to be small.
16. Check all of the following that are true.

____ The probability that the null hypothesis is true equals the p-value from the previous part.

____ It may be the case that the results are due to chance, and our conclusion from the hypothesis test is wrong.

____ The chance of getting a value of the test statistic as or more extreme than what was observed, assuming the null hypothesis is true, equals the p-value.

Below are scatter plots of property tax, household income, number people in the household, and age of the household head.

Questions 17 - 19 refer to the plot above.
17. Order income, number of people, and age of household head in terms of correlation with property tax, going from largest to smallest.

18. The value of the correlation between household income and age of household head is closest to which of the following values: - .5, -.25, 0, .25, .5.

19. If you fit a regression between property tax (outcome) and income (predictor), which of the following statements would be true. You can choose more than one.

___ The slope of the line would be positive.
___ The intercept of the line would be greater than 1000.
___ The root MSE of the regression would be larger than the SD of property tax.

20. Does taking additional vitamin C help prevent the common cold? Nobel Laureate Linus Pauling (1901 - 1994) performed a randomized experiment to address this question and reported his results in the Proceedings of the National Academy of Sciences. Pauling randomly assigned 279 French skiers to be in one of two groups: a group that took vitamin C supplements or a group that took a placebo (a sugar pill). The numbers of people for each category are summarized below:

<table>
<thead>
<tr>
<th></th>
<th>Got a cold</th>
<th>Did not get a cold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin C</td>
<td>17</td>
<td>122</td>
</tr>
<tr>
<td>Placebo</td>
<td>31</td>
<td>109</td>
</tr>
</tbody>
</table>

Pretend that you are the consulting statistician for Linus Pauling (a lofty honor indeed!)

i) Pauling seeks to know if there is evidence that the population incidence rate of colds for people who take Vitamin C is less than the population incidence rate of colds for people who take the placebo. What do you tell him? State clearly and justify your null and alternative hypotheses, the test statistic, the p-value, and conclusions.

ii) Asking people to take the sugar pills is expensive because you have to buy sugar pills and distribute them to the skiers. Pauling requests that the next experiment--Nobel Laureates always try to replicate results--avoid the sugar pills to save resources. Instead, he suggests the control group be randomly assigned to take nothing. Are you willing to comply with Pauling's request? Explain why or why not.

iii) Suppose you could replicate this experiment with 1000 skiers—500 in each treatment group. Approximate the standard error that you’d use in a 99% confidence interval.

iv) Discuss the types of conclusions that you can draw from these data. That is, what do the results suggest about the ability of Vitamin C to prevent colds?
21. True or False

For each statement, if you think the statement is always true, just say it is true. If you think the statement is always false or sometimes false, say it is false and explain why or when it is false in two or less sentences.

i) Your research colleague calculates by hand a value of the correlation of -1.24. He says this shows that there is a strong, negative linear association between the two variables.

ii) When data are randomly sampled from the same population, a 95% confidence interval constructed from a sample of 100 units should be narrower than a 95% confidence interval constructed from a sample of 200 units.

iii) In a regression analysis, a non-random pattern in the plot of residuals versus fitted values is consistent with the assumptions of the regression model.

iv) You perform a hypothesis test with a sample size of four units, and you do not reject the null hypothesis. This statistical test provides conclusive evidence against the alternative hypothesis.

v) A group of teachers attend a summer program designed to improve their foreign language skills. The teachers take a foreign language test at the start of the summer before the program begins. After the program ends, the teachers take another language test of similar difficulty. Based on a matched pairs hypothesis test, the average increase in scores is significantly greater than zero (p-value = .002). These data demonstrate that the summer training program improved the foreign language skills of the teachers.

vi) A professor is considering which of two exams to give. Scores on the first exam follow a normal distribution with mean of 75 and standard deviation of 5. Scores on the second exam follow a normal distribution with mean of 75 and standard deviation of 10. She wants to pick the exam likely to result in a relatively small number of people scoring below 60. She should pick the second exam.

vii) A certain company employs 10,000 people. Fifty percent of these employees are women. One hundred of these employees are in management positions. Of these 100 managers, 35 are women. In a court case against the company, a defense attorney argues that his client does not discriminate on the basis of sex when hiring managers. As evidence, he says, "The chance that a randomly selected employee is in management and a woman equals Pr(is a woman) * Pr(is in management) = .50 * .01 = .005. The chance that a randomly selected employee is in management and a man equals Pr(is a man) * Pr(is in management) = .50 * .01 = .005. These are equal probabilities, so that there is no evidence of discrimination." The prosecutor's calculations are valid.

22. A random sample of 100 is taken from a population with 10% minority and 90% non-minority members.

i) True or False: The number of minorities in the sample will be around 10 give or take 3 or so.

ii) True or False: There is about a 68% chance that the number of minorities in the sample will be between 9% and 11%.

iii) True or false: The population has about 10% minority members give or take .3% or so.

iv) True or false: In a particular sample of 100, it would be nearly impossible to see more than 16 minority members.
23. A standard deck of cards has 52 cards. Each card has on it one of the following: a number from 1 to 10, a Jack, a Queen, a King, or an Ace. There are four cards of each type. Each card is worth points equal to its number. The 10, Jack, Queen, and King are each worth 10 points. The player can choose to make an Ace worth 11 points or one point.

i) You are dealt an Ace. What is the chance that you will get a sum of 21 if you get only one more card?

ii) You are dealt two cards face down. What is the chance that the sum of their values will equal 21?

iii) You are dealt a six and an eight. What is the chance that you go over 21 by taking one more card?

iv) You are dealt a six, a five, and an 8. What is the chance that you get 21 if you can take two cards? If you get 21 after taking the first card, you don’t take a second card.

24. You have to decide whether or not to study hard for your Stats final. The professor tells you that, in the past, 75% of the As belong to people who study hard, and 20% of the non-As belong to people who study hard. Furthermore, experience shows that about 40% of people get As on the final.

i) What is the probability of getting an A, given that you study hard?

ii) What is the probability of getting an A, given that you do not study hard?