NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 101

Midterm 2 Key

Tuesday, April 12

This is a closed book exam. You may use your calculator and a single page of notes.

The classroom may be crowded. You are expected to abide by the Community Standard on this exam.

Report all numerical answers to at least two correct decimal places.

1. **FALSE** True or False: As the sample size increases, the width of a confidence interval for a population mean will increase.

2. **FALSE** True or False: For a given standard error, lower confidence levels produce wider confidence intervals.

3. **FALSE** True or False: A p-value of .08 is more evidence against the null hypothesis than a p-value of .04.

4. **D** In a simple random sample of n units, which of the following is true?

   A: each unit is equally likely to be chosen

   B: sampling is done without replacement

   C: sampling is done with replacement

   D: A and B

   E: A and C

5. A medical research team is investigating the benefits of a new surgical treatment. The researchers know that the recovery time of any particular patient can vary according to a normal distribution with mean 96 hours and variance 100 hours.

   (a) **15.865** What is the chance that a randomly selected patient has a recovery time of less than 86 hours?

   $Z = (86 - 96)/10 = -1$ so $1/2(100 - 68.27) = 15.865$.

   (b) **82.23 OR 69.145 (CLT)** Suppose we take a random sample of 10 patients. What is the chance that at least one of them will have a recovery time of less than 86 hours?

   $1 - P(\text{none}) = 1 - (1 - 0.15865)^{10} = 1 - (0.84135)^{10} = 0.8223$
   or using CLT approx. $Z > (1 - 1.5865)/\sqrt{10 * 0.15865 * 0.84135} = -0.508$ so $38.29 + 1/2(100 - 38.29) = 69.145$

   (c) **btwn 1.22 and 1.07** Suppose we take a random sample of 200 patients. What is the chance that less than 10% of these patients have recovery times less than 86 hours?

   $\hat{p} \approx N(.15865, \sqrt{(.15865)(.84135)/200})$ or $\hat{p} \approx N(.15865, 0.0258)$. $Z = (.10 - .1587)/.0258 = -2.275$ so $1/2(100-(\text{btwn } 97.56 \text{ and } 97.86)) = (\text{btwn } 1.22 \text{ and } 1.07)$

   (d) **15.865** Suppose we take a random sample of 100 patients. What is the chance of getting a sample average recovery time of more than 97 hours?

   $\bar{X} \approx N(96, 10/\sqrt{100})$ or $\bar{X} \approx N(96, 1.0)$. $Z = (97 - 96)/1 = 1$ so $1/2(100 - 68.27) = 15.865$

   (e) **100** How large a random sample of patients would we need to get a standard error for the sample average recovery time of 1 hour?

   see above

6. A study of 1004 randomly selected U.S. adults found that 151 intend to get a better job as a New Year's resolution.

   (a) **NO** Do you need to worry about the FPCF?

   (b) **btwn (0.1284,0.1724) and (0.1278,0.1730)** What is a 95% CI on the true proportion of U.S. adults who intend to get a better job as a New Year's resolution?
   $0.1504 \pm (\text{btwn } 1.95 \text{ and } 2.00) * \sqrt{(0.1504)(0.8496)/1004} = 0.1504 \pm (\text{btwn } 1.95 \text{ and } 2.00) * 0.0113 = 0.1504 \pm (\text{btwn } 0.0220 \text{ and } 0.0226) = \text{btwn } (0.1284, 0.1724) \text{ and } (0.1278, 0.1730).$

   (c) **FALSE** True or False: There is roughly a 5% chance that the true proportion of U.S. adults who intend to get a better job as a New Year's resolution is smaller than the lower limit of the above confidence interval.

7. As part of a N.C. restaurant survey we randomly select 17 restaurants and measure the temperature of the coffee sold at each. The sample average temperature is 162.0^oF with a sample standard deviation of 10.0^oF.

   (a) **158.75** Calculate a 90% **lower** CI for the true average temperature of coffee sold at N.C. restaurants.
   $162- \approx 1.3 * 10/\sqrt{(}16) = 162- \approx 1.3 * 2.5 = 158.75.$

   (b) **FALSE** Suppose that when the restaurant survey was conducted the number of N.C. restaurants was 500,000. However, we discover that an error was made and the number of N.C. restaurants is actually 1,500,000. True or False: Our estimate of average temperature of coffee sold at N.C. restaurants is much less accurate than we thought.

8. You want to determine whether there is a difference in the proportion of female and the proportion of male Internet users who plan to shop online during the next month. You randomly selected 200 adult female and 250 adult male Internet users. 30% of the females and 38% of the males said that they plan to shop online at least once during the next month.

   (a) **FALSE** True of False: The alternative hypothesis is that there is no difference in the proportion of female and the proportion of male Internet users who plan to shop online during the next month.

   (b) **-1.7937** Find the test statistic for determining if there is a difference in the proportion of female and the proportion of male Internet users who plan to shop online during the next month.
   $ts = (0.30 - 0.38 - 0)/\sqrt{\frac{(0.30)(0.70)}{200} + \frac{(0.38)(0.62)}{250}} = -0.08/0.0446 = -1.7937$

   (c) **YES** Can we reject the null hypothesis if we require a significance probability less than 10% to make this decision?
   $p - value = (100 - 92.81) = 7.19$ so we cannot reject

9. You want to buy a vacuum cleaner and a salesperson tells you the repair costs for the 'Big-Eats' model are larger than the repair costs for the 'No-Goo' model. You decide to test this claim. You research the repair costs of 34 randomly selected 'Big-Eats' vacuum cleaners and 46 randomly selected 'No-Goo' vacuum cleaners. For the 'Big-Eats' model, the sample average repair cost is \$60 with a sample standard deviation of \$10; for the 'No-Goo' model, the sample average repair cost is \$50 with a sample standard deviation of \$18.

   (a) State the null and alternative hypotheses.
   $H_0 : \mu_{be} - \mu_{ng} \leq 0$ or the mean repair costs of the 'Big-Eats' model are less than or equal to the mean repair costs of the 'No-Goo' model.
   $H_1 : \mu_{be} - \mu_{ng} > 0$ or the the mean repair costs of the 'Big-Eats' model are greater than the mean repair costs of the 'No-Goo' model.

   (b) **3.165** Find the test statistic for determining if the average repair cost for the 'Big-Eats' model is larger than the average repair cost for the 'No-Goo' model.
   $ts = (60 - 50 - 0)/\sqrt{\frac{10^2}{34} + \frac{18^2}{46}} = 10/3.1599 = 3.165$

   (c) **0.0815** Find the p-value for your test statistic.
   $p - value = 1/2(100 - 99.837) = 0.0815$

10. **C** A 95% confidence interval obtained from a random sample of 1000 people has

    A: a better chance of containing the population percentage than a 95% confidence interval obtained from a random sample of 500 people.

    B: less of a chance of containing the population percentage than a 95% confidence interval obtained from a random sample of 500 people.

    C: the same chance of containing the population percentage than a 95% confidence interval obtained from a random sample of 500 people.

11. The Bureau of Labor Statistics administers the **Current Population** Survey, which is performed by the Census Bureau. The purpose of this survey is to estimate the unemployment rate.

12. **FALSE** True or False: To reduce the width of a confidence interval for a population proportion by a factor of two (i.e., in half), you have to double the sample size.

13. **NO** Yes or No: A researcher who tried to learn statistics without taking a formal course does a hypothesis test and gets a p-value of .024. He says "there is a 98.6% chance that the alternative hypothesis is false, so the null hypothesis is true." Is the researcher's statement correct?

14. **FALSE** True or False: The statement, "the 95% confidence interval for the population mean is (350,400)", is equivalent to the statement, "there is a 95% probability that the population mean is between 350 and 400".

15. A statistician draws a card from a deck 100 times and gets 60 red cards. His null hypothesis says that the deck is fair; the alternative, that the deck is biased - the probability of drawing a red card is more than 50%. He conducts a hypothesis test and finds that the significance probability of his test statistic is 3%. True or False:

    (a) **TRUE** If the deck is fair, the chance of getting 60 or more red cards is about 3%.

    (b) **FALSE** Given that 60 red cards were drawn, there is only about a 3% chance for the deck to be fair.

    (c) **FALSE** Given that 60 red cards were drawn, there is about a 97% chance for the deck to be biased.

16. Results from a survey five years ago asking where coffee drinkers typically drink their first cup of coffee showed that 50% drink their first cup at home, 25% drink their first cup at work, and 25% drink their first cup while commuting. To determine whether this distribution has changed, you randomly select 581 coffee drinkers and ask each where they typically drink their first cup of coffee. The results:

Location	Frequency
At home	389
At work	110
While commuting	82

(a) **C** Which test would you conduct in order to determine whether the distribution of where coffee drinkers typically drink their first cup of coffee has changed?

A: the one-sample z-test

B: the two-sample z-test

C: the Chi-Square test for goodness-of-fit

D: the Chi-Square test for independence

(b) **33.3985** Calculate the contribution of the first cell (At home) to the test statistic.
$E_1 = 581 * 0.50 = 290.5$ so $contrib = ((389 - 290.5)^2)/290.5 = 33.3985$

(c) **2** What are the degrees of freedom?
$k = (3 - 1) = 2$

(d) **YES** The value of the test statistic is found to be 69.496. What do you conclude? Has the distribution changed?
$p - value < 0.01$ so conclude has changed