

Name: \_\_\_\_\_

## Lab Problems 9-12 (worth 20 pts.)

Statistics 1000

Dr. Nancy Pfenning

9. A student survey was completed by 446 students in introductory statistics courses at a large university in the fall of 2003. Students were asked to pick their favorite color from black, blue, green, orange, pink, purple, red, yellow.
- (a) If colors were equally popular, what proportion (to three decimal places) of students would choose each color? \_\_\_\_\_
  - (b) Pick a color that you suspect will be *less* popular than others. \_\_\_\_\_ Using software to access the survey data, report the sample proportion who preferred the color you chose. \_\_\_\_\_ Is it in fact lower than the proportion you calculated in (a)? \_\_\_\_\_
  - (c) Use software to produce a 95% confidence interval for the proportion of all students who would choose that color. \_\_\_\_\_
  - (d) Does your confidence interval (i) contain the proportion you calculated in (a), or is it (ii) strictly above, or (iii) strictly below?
  - (e) Use software to carry out a hypothesis test to see if the sample proportion choosing your color was low enough to assert that, overall, students picked that color less than if they were choosing at random from eight colors. Report the standardized sample proportion  $z$  \_\_\_\_\_ and the p-value \_\_\_\_\_, and state your conclusions, using 0.05 as the cut-off for small p-values:

10. Several hundred students enrolled in introductory statistics courses at a large university were surveyed on various days of the week, and asked to report the number of minutes they'd spent exercising the day before.
- (a) Use software to access the survey data and construct a 95% confidence interval for the mean number of minutes of daily exercise for the population of students.
- 
- (b) According to the American Time Use Survey published by the Bureau of Labor Statistics, the mean daily hours of exercise for Americans aged 15 to 24 at that time (2003) was about 31 minutes. Use your confidence interval to choose from the following:
- 31 is a plausible value for mean daily amount of exercise for the population of students
  - 31 is not a plausible value for mean daily amount of exercise for the population of students: apparently they exercise more
  - 31 is not a plausible value for mean daily amount of exercise for the population of students: apparently they exercise less
- (c) Use mathematical notation to write null and alternative hypotheses to test if the population mean time for students could be 31 minutes:

11. Does whether or not a person has ears pierced tell us something about whether the person wears contacts, glasses, or neither?
- (a) Use software to access the student survey data, and carry out a chi-square test for a relationship between the two variables: report the chi-square statistic \_\_\_\_\_ and the p-value \_\_\_\_\_, which provides evidence that the variables *are* related.
  - (b) Which sampled group has a higher proportion wearing glasses: those who (i) do or those who (ii) do not have ears pierced? Explain why gender is a possible confounding variable in the relationship between pierced ears and eyewear.
  - (c) Use software to separate data on ears pierced and eyewear for males and females. Report the p-value for females. \_\_\_\_\_ Is there evidence that for females, having ears pierced or not is related to eyewear? \_\_\_\_\_
  - (d) Report the p-value for males. \_\_\_\_\_ Is there evidence that for males, having ears pierced or not is related to eyewear? \_\_\_\_\_
  - (e) What is the name for the phenomenon whereby the nature of a relationship changes when groups are combined? \_\_\_\_\_

12. Several hundred students enrolled in introductory statistics courses at a large university were surveyed, and asked to pick a whole number at random from 1 to 20. Since the mean of the numbers from 1 to 20 is 10.5, if selections are truly random, they should average 10.5 in the long run.
- (a) Tell whether we would opt for a (i)  $z$  or (ii)  $t$  procedure if population standard deviation were unknown. Tell whether we would opt for a (i)  $z$  or (ii)  $t$  procedure if we take into account that the standard deviation of the numbers 1 through 20 is 5.766.
- (b) Use software to access the data and, with 5.766 as population standard standard deviation, construct a 95% confidence interval for mean selection by all students.
- \_\_\_\_\_
- (c) With 5.766 as population standard deviation, carry out a test to see if the students' random number selections were consistent with random selections from a population whose mean is 10.5: report the sample mean \_\_\_\_\_ and p-value \_\_\_\_\_, and say whether or not selections could have been truly random: \_\_\_\_\_
- (d) Would the null hypothesis have been rejected against the one-sided alternative  $H_a : \mu > 10.5$ ? \_\_\_\_\_ Explain:
- (e) Would the null hypothesis have been rejected against the one-sided alternative  $H_a : \mu < 10.5$ ? \_\_\_\_\_ Explain:
- (f) Do people apparently perceive larger or smaller numbers to be more random? Explain:
- (g) Note that the sample standard deviation  $s = 5.283$  is smaller than the assumed population standard deviation  $\sigma = 5.766$ . [This is partly due to the phenomenon that students tend to avoid the extremes 1 and 20 when making a "random" selection.] If  $t = \frac{\bar{x} - \mu_0}{5.283/\sqrt{n}}$  had been used instead of  $z = \frac{\bar{x} - \mu_0}{5.766/\sqrt{n}}$ , would  $t$  have been (i) larger than  $z$  or (ii) smaller than  $z$  or (iii) the same size as  $z$ ?
- (h) If  $t = \frac{\bar{x} - \mu_0}{5.283/\sqrt{n}}$  had been used instead of  $z = \frac{\bar{x} - \mu_0}{5.766/\sqrt{n}}$ , would the p-value have been (i) larger than or (ii) smaller than or (iii) the same size as the one obtained using  $z$ ?