

Name: _____

Statistics 111 – Summer 2008

Final

July 3, 2008

Instructions: You have until 12:10pm to finish this exam. You may use a calculator and your prepared “cheat sheet.” **No credit will be given if only a number is provided.** Either show your work or justify your answers using a sentence or two. Partial credit will be given.

Question 1

Does the addition of extra calcium to yogurt affect the taste? Volunteers were divided into two groups of 100 people: group 1 received yogurt with calcium added and group 2 received just regular yogurt. People in both groups scored the “tastiness” of their yogurt on a scale of 1 to 10. The sample mean rating for group 1 was 6.75 with a sample SD of $s_1 = 1.5$, whereas the sample mean rating for group 2 was 7.0 with a sample SD of $s_2 = 2$.

- (a) Assuming the two groups are independent, perform a hypothesis test (at an $\alpha = 0.05$ level) for a significant difference between the population means of the two groups. Give the test statistic and p-value.

(b) Give a 95% confidence interval for the difference between the population mean ratings $\mu_1 - \mu_2$.

(c) Now, you are told that group 1 and group 2 actually consisted of the same people, so the two groups are matched into pairs, and the difference (group 1 – group 2) between the groups for each pair is calculated. The mean of these differences is -0.5 and the s.d. of these differences is 0.75. Perform a hypothesis test (at an $\alpha = 0.05$ level) for a significant mean difference between the two groups.

Question 2

The drug Hismanal is claimed by its manufacturer to be the first non-drowsy allergy medicine. This claim is based on a properly run experiment with random assignment in which 1400 patients were given Hismanal and 900 patients were given a placebo. In the Hismanal group, 176 reported drowsiness, while 90 reported drowsiness in the placebo group.

(a) Give a 95% confidence interval for the population proportion of patients who experience drowsiness after taking Hismanal.

(b) Perform a hypothesis test (at an $\alpha = 0.05$ level) for a significant difference between the proportion of drowsy patients in the Hismanal group and the proportion of drowsy patients in the placebo group. Give the test statistic and the p-value.

(c) Give a 95% confidence interval for the difference between the population proportion of drowsy patients in the Hismanal group and the population proportion of drowsy patients in the placebo group.

(d) Based on the data analysis you performed, does giving Hismanal to a patient decrease his/her reported drowsiness? **Justify your answer or no credit will be given.**

Question 3

A research assistant has conducted a survey of 50 households in a wealthy region of the country and has given a social scientist 95% and 99% confidence intervals for mean household income in the region. The two intervals (in no particular order) are

(\$50069, \$53931) and (\$50530, \$53470)

The research assistant admits that he assumed that the population standard deviation of the household income was known and used a value for the population standard deviation from a previous survey.

(a) What is the sample mean of household income in the assistant's survey?

(b) Which of the two intervals is the 95% confidence interval? How do you know this?

(c) What is the value of the population standard deviation that the research assistant used to construct these confidence intervals?

(d) The research assistant was then asked to design a new study and told that he needed a margin of error of less than \$500 for a 95% confidence interval. How many households must he sample for the new study?

Question 4

Major league baseball is divided into two leagues: the National league (NL) and the American league (AL). A major difference between the leagues is that the AL allows Designated Hitter (DH) to hit instead of the pitcher. We are interested in whether or not the DH rule leads to a difference in the total number of runs scored by teams between the two leagues. We take the 1992 season as a sample from the population of all baseball seasons since the DH rule was introduced.

Assume there were 12 teams in the NL and 12 in the AL in 1992. The AL teams had a mean runs scored of 685.9 runs and a sample standard deviation of 60.6 runs. The NL teams had a mean runs scored of 658.3 runs and a sample standard deviation of 46.6 runs.

- (a) Conduct a hypothesis test to address the question of whether or not the mean runs scored between the two leagues is significantly different (at the $\alpha = 0.05$ level). Make sure you state your null and alternative hypothesis and give a p-value for your test.

- (b) Construct a 95% confidence interval for the difference in mean runs scored between the two leagues. How could you have used this interval to answer part (a)?

Salary is also an important factor in runs scored, so we might get more information if the 1992 teams from the AL and NL were put into pairs that had matched salaries. This pairing was done, and then the difference (AL team – NL team) in runs scored was calculated for each pair. The mean of these differences is 27.7 runs and the standard deviation of these differences is 22.2 runs.

- (c) Conduct a hypothesis test with this new information to address the question of whether or not the mean runs scored between the two leagues is significantly different (at the $\alpha = 0.05$ level). Make sure you state your null and alternative hypothesis and give a p-value for your test.

Question 5

Sleep and Wages. Dr. Hamermesh, an Economist at UT Austin, performed an observational study trying to understand the link between sleeping habits and wages. The surprising conclusion was that people who sleep less tend to have higher average wages. Thus there was a negative association between sleep (the X variable) and wages (the Y variable).

Here's some data:

	Sleep (average hours)	Wages (thousands of \$\$)
Mean	6.5	35.1
Standard Deviation	1.2	15.4
Rsquared	.81	

(a) What is the correlation between sleep and wages?

(b) What is the best fit line given the data above? Be sure to show the formula you used to find the slope, **b**, and for the y-intercept, **a**.

(c) According to your calculations in part b, what is Alex's predicted wage if he sleeps 8hrs per night?

(d) According to your calculations in part b, what is Heathers's predicted wage if she sleeps 5hrs per night?

(e) Test for significance (at the $\alpha = 0.05$ level) for the association between X and Y.
The standard error of **b** is 1.6.

(f) Create a 95% confidence interval for the slope β . The standard error of \mathbf{b} is 1.6.
How can we use this confidence interval to answer the question of a significant association between X and Y?

(g) If Alex decreases his sleep by three hours per night what would you anticipate would happen to his wage?

Question 6 – Tootsie rolls and Lollipops Redux

There is a bowl with only five pieces of candy in it. Three of the candies are tootsie rolls and two are lollipops.

Here's the new twist: Each time you pull out a lollipop a new one is added. This means that every time you reach into the bowl there are two lollipops in the bowl. The number of tootsie rolls may change, but not the number of lollipops.

The bowl is on top of the fridge and you cannot see which candy you will pull out. Assume that you are equally likely to pull out any of the candies in the bowl.

(a) What is the probability of pulling out a tootsie roll on the first draw?

(b) If you pulled out a tootsie roll on the first draw what is the probability of pulling out a tootsie roll on the second draw?

(c) What is the probability of pulling out a tootsie roll on the second draw (if you don't know what you pulled out on the first draw)?

(d) You draw **three** candies from the bowl, one after the other. Let X = the number of tootsie rolls you pull out. Write down the distribution of X .

X	0	1	2	3
P(X)				

(e) What is the mean of X ?