Math 17: Intro Stats
Final Exam
December 18, 2009

Directions: Before you leave, you must turn in both this exam sheet and any statistical tables. If not, you will receive a significant grade reduction. You are allowed to use a calculator and a two-sided sheet of notes for this exam. All cell phones, PDAs, iPods, laptops, etc, should be turned off and put out of sight. You may not discuss the exam with anyone but me. In total, this exam is worth 200 points. You have the entire period to complete this exam.

Part I – Multiple Choice: There is only ONE correct response per question. Each question is worth 5 points. There are a total of 10 questions for a combined total of 50 points. Clearly circle or write your answer in front of each question.

<table>
<thead>
<tr>
<th>Part I</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td></td>
</tr>
<tr>
<td>Possible Points</td>
<td>50</td>
</tr>
</tbody>
</table>

Part II – Free Response: You must show all work in order to receive full credit. Each question is worth a different amount of points and this value is noted in the table below. There are a total of 8 questions with multiple parts for a combined total of 150 points.

<table>
<thead>
<tr>
<th>Part II</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible Points</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>20</td>
<td>15</td>
<td>150</td>
</tr>
</tbody>
</table>

Here is my suggestion:
Read all questions before beginning and try to complete the ones you know best first.

GOOD LUCK!!!
PART I: MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1. Olivia wants to learn a foreign language. To get an idea of how satisfied other students were after taking a foreign language course, she decides to take a random sample of 20 students. If she randomly selects 5 students from French, 5 from German, 5 from Spanish, and 5 from Chinese, then what sampling method did she use?
   A) simple random sampling
   B) stratified random sampling
   C) cluster sampling
   D) sampling with replacement
   E) systematic sampling

2. Which of the following correlation values indicates the strongest linear relationship between two quantitative variables?
   A) r = -0.65
   B) r = -0.30
   C) r = 0.00
   D) r = 0.11
   E) r = 0.60

3. A short quiz has two true/false questions and one multiple choice question with four choices. A student guesses at each question. Assuming the choices are all equally likely, what is the probability that the student gets all three correct (assume independence)?
   A) 1/3
   B) 1/4
   C) 1/8
   D) 1/16
   E) 1/32

4. What statement is true about both \( \hat{p} \) and \( \mu \)?
   A) They are both parameters.
   B) They are both statistics.
   C) They are both symbols pertaining to proportions.
   D) \( \hat{p} \) is a statistic and \( \mu \) is a parameter.
   E) \( \hat{p} \) is a parameter and \( \mu \) is a statistic.

5. Suppose I know that SAT scores are normally distributed with a mean of 1000 and a standard deviation of 200. Roughly what percent of all SAT scores are above 1200?
   A) 10%
   B) 16%
   C) 64%
   D) 84%
   E) 90%
6. A University administrator says that in the past, the proportion of students who binge drink was 0.45. But, he also believes that due to various university ad campaigns, the proportion is now less than 0.45. To resolve the issue, a random sample of 400 students is obtained and it is found that 156 (or 39%) admit that they have engaged in binge drinking.

<table>
<thead>
<tr>
<th>Test of $p = .45$ versus $p &lt; .45$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
</tr>
<tr>
<td>156</td>
</tr>
</tbody>
</table>

Which of the following statements describes the p-value that was calculated in this situation?
A) It is the probability that $z$ is greater than $-2.41$.
B) It is the probability that $z$ is less than $-2.41$.
C) It is the probability that $z$ does not equal $-2.41$.
D) It is the probability that $z$ equals $-2.41$.
E) None of the above.

7. Suppose a random variable $X$ has a mean of 26.74 and a standard deviation of 1.06. What is the standardized score (z-score) for a value of $x = 26.74$?
A) 0.00
B) 1.06
C) 1.36
D) 7.89
E) 25.68

Questions 8-9: A child is observing squirrels in the park and notices that some are black and some are gray. For the next five squirrels she sees, she counts $X$ = the number of black squirrels. Suppose $X$ is a binomial random variable with $n = 5$ and $p = 0.50$.

8. What is the expected number of black squirrels she will see, $E[X]$?
A) 2.0
B) 2.5
C) 3.0
D) 3.5
E) 4.0

9. What is the variance for the number of black squirrels she will see?
A) 1.00
B) 1.12
C) 1.25
D) 2.50
E) 2.75

10. True or False: In hypothesis testing, a result is statistically significant if it is unlikely that chance alone can explain the observed results.
A) True
B) False
PART II: FREE RESPONSE. Write the word or phrase that best completes each statement or answers the question.

President Marx, the president of University AAA of more than 40,000 students, has formed a committee of twenty six to analyze a survey recently done in a huge intro stats class. This sample consists of responses from \( n = 452 \) students and is believed to be representative of the student population in this school. The main mission of this committee is to discover some interesting findings about college students.

The Independence Assumption is satisfied, so do not worry about checking the randomization condition and the 10% condition in Q1 – Q8. (You do need to check other conditions when doing inference below.)

1. Seven committee members, Dan, Emma, John, Kate, Ben Sam and Mike, are concern about the use of marijuana on campus and notice that among 452 students in the survey, 195 admitted that they have been smoked marijuana in the past six months.

   A. Based on the survey results, construct and interpret a 95% confidence interval for the proportion of students who have smoked marijuana in the past six months.

   B. Explain the meaning of “95% confidence” in part A.
Another eight committee members, Meron, Maikha, Allison, Asia, Humza, Jiaqi, Angelique, and Lehua, are more interested in students’ grade-average point. In particular, they want to know if gender plays an important role on students’ academic performance.

A. Indicate what inference procedure you would use to see if there is a significant difference in GPA between male and female students. Why?

B. Write appropriate hypotheses.

C. Assume the assumptions for the test are met. Partial R output is given below. Use the output to complete the test procedure at a .01 significance level.

```
data:  GPA by Sex
t = 4.4794, df = 389.805, p-value = ----- 
alternative hypothesis: true difference in means is not equal to 0
sample estimates:  mean in group Female  mean in group Male
                    3.255498          3.052261
```

P-value:

Conclusion:

D. Given your conclusion above, which type of error (Type I or Type II) could be made? Explain this type of error in context.
3. A group of three, Thebe, Edward, and Caca, are more curious about students’ social life. They wish to determine whether or not students’ feelings about religious importance and their opinions about uncommitted sex are related.

A. What is the appropriate analysis to perform (be specific) and state appropriate hypotheses.

Analysis:

Null:

Alternative:

B. Expected counts for this test are given below with some missing. Compute and fill in the missing expected counts in the table. Observed (Expected) is the table setup.

<table>
<thead>
<tr>
<th>ReligImp \ SexUncom</th>
<th>YES</th>
<th>NO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairly Important</td>
<td>133 (132.61)</td>
<td>89 (89.39)</td>
<td>222</td>
</tr>
<tr>
<td>Not Important</td>
<td>99 (     )</td>
<td>29 (51.54)</td>
<td>128</td>
</tr>
<tr>
<td>Very Important</td>
<td>38 (     )</td>
<td>64 (41.07)</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>270</td>
<td>182</td>
<td>452</td>
</tr>
</tbody>
</table>

C. How many degrees of freedom?

D. Assume the assumptions for the test are met. The test statistic works out to be 37.9346 with a P-value less than .0001. State your complete conclusion in context.

E. Compute the chi-square component of the cell \{Very Important\ YES\}. Does this cell arouse your suspicion? What additional information do you get from this cell?
4. Committee members, Philip, Katherine, Elisabeth, and Madeleine, are also interested in student's religious views. However, they are more interested in knowing whether students' religious views affect the number of days per month they drink alcohol. Perform an ANOVA to look for differences in the mean number of days per month students drink alcohol for the three religion groups. Use a .01 significance level.

A. State appropriate hypotheses.

B. What assumptions need to hold in order for the ANOVA to be valid?

C. Assuming the assumptions hold, use the partial output from R below.

```
> AnovaModel.1 <- aov(AlcoholDays ~ ReligImp, data=Project)
> summary(AnovaModel.1)

                 Df Sum Sq Mean Sq F value Pr(>F)
ReligImp          2  1268.7   634.3  ???    1.513e-07 ***
Residuals       449 17510.1    39.0
```

What is the missing value of the test statistic?

What is the P-value of the hypothesis test?

D. What is your conclusion?

E. Would it be appropriate to run multiple comparisons based on your conclusion?

F. Name a nonparametric alternative to this test procedure.
5. Speaking of Alcohol Days, four committee members, Max, Nicholas, Steven and Alina, have done some analysis with it. In addition, they also have a closer look at the variable Party Days, the number of days per month students attend parties. One interesting study would be to investigate the association between these two variables. Suppose the committee would like to build a model that will predict students’ average alcohol days based on their party days. The plots attached below are the scatterplot, the residuals plot, and a histogram of the residuals (in order), with the regression analysis for the data. Use this information to analyze the association between party days and alcohol days of college students.

A. Provided and interpret the value of the squared correlation ($R^2$) for this regression.

B. What is the equation of the least squares regression line for predict average Alcohol days on Party days.

C. For a student who answered 10 party days with 12 alcohol days, what was his/her residual?
D. Is there an association between Party Days and Alcohol Days? Write appropriate hypotheses, check and explain if the assumptions for regression satisfied, provide the corresponding test statistic and P-value, and then state your conclusion about the association.

E. Create a 95% confidence interval for the true slope and explain in context what your interval means.
6. In this survey, each student was asked their actual height and their desired height. Some committee members want to know if college men would like to be taller than what they are.

A. Explain why this data is an example of paired data.

B. Write appropriate hypotheses (in words and in symbols).

C. What are the assumptions that much be satisfied to perform this test? Assume they are met for the parts below.

D. Partial R output for the paired t test is provided. Use the output to complete the test procedure. Make sure to state your conclusion in context.

<table>
<thead>
<tr>
<th>data: ProjectMale$IdealHt and ProjectMale$Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>t = 12.633, df = 196, p-value &lt; 2.2e-16</td>
</tr>
<tr>
<td>alternative hypothesis: true difference in means is greater than 0</td>
</tr>
</tbody>
</table>

E. Suppose that these committee members in fact have the same curiosity about college women, and would like to know if this desire is the same for both genders. If they want to determine whether this difference between student’s actual height and ideal height is the same for males and females, what inference procedure should they consider?
7. One committee member noticed that, in this survey, the variable LongFing gives responses for which finger is longer — index, ring, or same. This reminds him of some studies about differences in the finger length patterns of males and females. He then creates a two-way table shown below.

<table>
<thead>
<tr>
<th>Longer finger</th>
<th>Index</th>
<th>Ring</th>
<th>Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>91</td>
<td>124</td>
<td>38</td>
</tr>
<tr>
<td>Males</td>
<td>38</td>
<td>140</td>
<td>21</td>
</tr>
</tbody>
</table>

A. What is the conditional distribution (in %) of finger length pattern for females?

Index ________  Ring ________  Same ________.

For males? Index ________  Ring ________  Same ________.

B. Write a short paragraph to describe differences between men and women with regard to finger length pattern.

C. Now, focus only on the proportions of females and males with a longer index finger. Based on the results, construct and interpret a 95% confidence interval for the difference in the proportions of females and males with a longer index finger.
8. The following R output gives the descriptive statistics for the variable, *BookCost*, the amount students expect to spend on textbooks per semester.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>0%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BookCost</td>
<td>315.1956</td>
<td>115.4370</td>
<td>0</td>
<td>250</td>
<td>300</td>
<td>400</td>
<td>600</td>
</tr>
</tbody>
</table>

A. Based on the output above, answer the following questions:

IQR = ___________; Range = ___________; 

About 1/4 of the sample said they spent more than ___________.

The rising textbook costs start drawing the committee's attention. So, some members decide to investigate the textbook costs at local bookstores. Suppose that a committee found a study showing that the new textbooks at local bookstores cost an average of 110 dollars and a standard deviation of 12 dollars.

B. A random sample of 60 new textbooks from a local bookstore is examined. What is the distribution of the sample mean for these 60 new textbooks? (Provide all features of the distribution).

C. What is the probability that the average price for the 60 new textbooks selected will be greater than 118 dollars?