Math 1235, Test #3, April 26, 2006

Part 1: True/False Questions (4 points each):

Instructions: Circle the appropriate answer.

1. **True**/False  Failing to reject $H_0$ when $H_0$ is false constitutes a Type II error.

2. **True**/False  If an $F$ test rejects $H_0$, then it is appropriate to compare group means using techniques like the least significant difference (LSD) method.

3. **False**  If other factors are unchanged, then a larger sample size will lead to a wider confidence interval.

4. **True**/False  If other factors are unchanged, then a 95% confidence interval will be wider than a 99% confidence interval.

5. **False**  A large $P$-value leads us to reject $H_0$.

Part 2: Other Questions:

Instructions: SHOW YOUR WORK to receive credit.

Problem 6: (16 points) A hospital administrator wants to test whether patients are equally likely to be admitted on each day of the week. She randomly samples 210 patients admitted during the past year, classifying each patient by day of admission. The data are tabulated below. Use a level 0.05 test to determine whether the data are consistent with the hypothesis that each day is equally likely.

<table>
<thead>
<tr>
<th></th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed</strong></td>
<td>36</td>
<td>19</td>
<td>28</td>
<td>24</td>
<td>33</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td><strong>Expected</strong></td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

$$
\chi^2 = \sum_{\text{cells}} \frac{(O-E)^2}{E}
$$

If each day is equally likely, then each day has probability $\frac{1}{7}$. Thus, each expected count is

$$
\text{exp} = 210 \left( \frac{1}{7} \right) = 30.
$$

$$
\begin{align*}
(36-30)^2 &= 6^2 = 36, \\
(19-30)^2 &= 11^2 = 121, \\
(28-30)^2 &= 2^2 = 4, \\
(24-30)^2 &= 6^2 = 36, \\
(33-30)^2 &= 3^2 = 9, \\
(40-30)^2 &= 10^2 = 100, \\
(30-30)^2 &= 0.
\end{align*}
$$

Since there are 7 cells, there are $7 - 1 = 6$ degrees of freedom.

Since $10.20 > 12.592$, we don't reject the hypothesis that each day is equally likely.
Problem 7: (16 points) The table below shows the numbers of wins and losses for the Philadelphia Phillies during the 1980 and 1981 seasons. (The 1981 season was shortened by a strike.) Use a level 0.05 test to decide whether these data provide any evidence that the team’s true winning percentage differed for the two years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Wins</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>91</td>
<td>71</td>
</tr>
<tr>
<td>1981</td>
<td>59</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>119</td>
</tr>
</tbody>
</table>

\[
\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(91-90.33)^2}{90.33} + \frac{(71-71.67)^2}{71.67} + \frac{(59-59.67)^2}{59.67} + \frac{(48-47.33)^2}{47.33} = .028.
\]

There is \((2-1)(2-1) = 1\) degree of freedom, and the cut-off is 3.841. Since 0.028 < 3.841, we do not reject the hypothesis that the winning percentages were the same for the two years.

Problem 8: (8 points) Our textbook recommends making side-by-side boxplots as a means of seeing whether the assumptions needed for ANOVA are violated. Give one example of how side-by-side boxplots might indicate that ANOVA is not appropriate.

Here are some possible answers:

1. The boxplots might show that the different groups have very different spreads. This would make us doubt the equal variance assumption.

2. The boxplots might show that the distributions are not symmetric. This would make us doubt the normality assumption.
Problem 9: (20 points) A biology student wants to study how different fertilizers affect the growth of bean sprouts. He sprouts 12 beans for each of 10 fertilizers, and after one week, he measures the heights of the 120 sprouts in millimeters. Side-by-side boxplots give no indication that the ANOVA assumptions are not met.

Part a: Using the description given above, complete the ANOVA table for the experiment.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fertilizer</td>
<td>9</td>
<td>2073.708</td>
<td>230.412</td>
<td>1.188</td>
</tr>
<tr>
<td>Error</td>
<td>110</td>
<td>21,331.982</td>
<td>193.92</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>23404.791</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here \( k = 10 \) groups,
\[ n = 12 \] beans per group, \( \text{so} \ N = 12(10) = 120 \) total observations.
\[ F = \frac{230.412}{193.92} \approx 1.188 \]

Part b: State the null and alternative hypotheses for the \( F \) test.

\( H_0: \) the mean for each group is the same
\[ (H_0: \mu_1 = \mu_2 = \ldots = \mu_{10}) \]

\( H_a: \) the means are not all the same

Part c: Using your completed ANOVA table from Part a, carry out the \( F \) test at level 0.05. Be sure to state your conclusion in terms of heights and fertilizers.

Since there are 9 and 110 df, the cut-off value is about 1.97. Since 1.188 < 1.97, we retain \( H_0. \) We have no evidence that the bean sprouts grow taller with one fertilizer than with another.
**Problem 10: (20 points)** A economist wants to determine whether there is an association between the size of a baseball team’s payroll and the number of games the team wins. To study this question, he has collected data on payroll (in millions of dollars) and number of wins for the 30 major league baseball teams during the 2005 season. A scatterplot suggests that fitting a linear model is appropriate, and the output given below was obtained from MINITAB. The economist also computed that the average payroll was $72.75$ million dollars.

\[
\begin{array}{llll}
\text{Predictor} & \text{Coeff} & \text{SE Coef} & \text{T} & \text{P} \\
\text{Constant} & 69.378 & 4.200 & 16.52 & 0.000 \\
\text{Payroll} & 0.15975 & 0.05251 & 3.04 & 0.005 \\
\end{array}
\]

\[S = 9.55882\]

**Part a:** Create a 95% confidence interval for the slope of the regression line. What does your interval say about whether there is an association between payroll and number of wins?

Since there are $30 - 2 = 28$ d.f., the critical value is $2.048$.

The CI is

\[b_1 \pm 2.048 \times SE(b_1)\]

\[= 0.15975 \pm 2.048 (0.05251) \]

\[= [0.052, 0.267] \]

Since this interval is entirely on the positive side of 0, we conclude that payroll and number of wins are positively associated.

**Part b:** How many wins would you predict for a team with a payroll of 90 million dollars?

\[\hat{\text{wins}} = b_0 + b_1 \times 90 = 69.378 + 0.15975(90) \]

\[= 83.76\]

**Part c:** Find a 95% confidence interval for the average number of wins for teams with 90 million dollar payrolls.

\[83.76 \pm 2.048 \sqrt{SE(b_1)^2 \left(\frac{x_0 - \bar{x}}{n}\right)^2 + \frac{s^2}{n}}\]

\[= 83.76 \pm 2.048 \sqrt{(0.05251)^2 (90 - 72.75)^2 + \frac{9.55882^2}{30}}\]

\[= 83.76 \pm 4.03 = [79.73, 87.79] \]