MAT 222 Final
Fall 2002

Name: 

Student Number: 

Instructor’s Name: 

READ THIS BEFORE YOU BEGIN

This examination contains 5 problems for a total of 100 points. There are 9 pages in this booklet. It is your responsibility to make sure that they are all present. Your solutions must be written legibly and contain all the necessary steps which enabled you to arrive at your answer in order to receive full credit. Unsupported answers will receive little credit.

1. 
2. 
3. 
4. 
5. 
TOTAL _____
1. (15 pts) Independent random samples of 2,065 males and 2,235 females were obtained in order to compare the percentage of males who smoked to the percentage of females who smoked. Of the males sampled, 580 were cigarette smokers; of the females sampled, 525 were cigarette smokers. Do the data provide sufficient evidence to conclude that the percentage of males who smoke cigarettes exceeds the percentage of females who smoke cigarettes?

(a) State the null and alternative hypotheses

(b) Calculate the test statistic.

(c) Is the test significant at the 1% level?
2. (20 pts) Radon levels in houses vary greatly by location. Samples are taken from houses in Syracuse and in Rochester with the following results: In 10 randomly selected houses in Syracuse the average radon level was $\bar{x} = 10.2$ with standard deviation $s = 3.4$. In 8 randomly selected houses in Rochester the average was $\bar{x} = 13.1$ with standard deviation $s = 4.1$. We wish to test whether the radon levels in Rochester are higher than those in Syracuse.

(a) State the suitable null and alternative hypotheses.

(b) Is there significant evidence of unequal population standard deviations at the 1% level?
(c) Is there significant evidence at the 1% level that radon levels are higher in Rochester than in Syracuse?

(d) Construct a 95% confidence interval for the difference in the radon levels between Rochester and Syracuse.
3. (15 pts) The table below gives the results of a study of 452 individuals enrolled in a methadone-treatment program. The number of positive HIV tests and negative HIV tests are provided separately for people with different durations of intravenous drug use.

<table>
<thead>
<tr>
<th>Durations of Drug Use</th>
<th>HIV positive</th>
<th>HIV negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>23</td>
<td>53</td>
</tr>
<tr>
<td>1-4 years</td>
<td>77</td>
<td>141</td>
</tr>
<tr>
<td>&gt; 4 years</td>
<td>78</td>
<td>80</td>
</tr>
</tbody>
</table>

(a) You are to conduct a chi-squared test on the table. State the null hypothesis that this statistic is used to test.

(b) Calculate the chi-squared statistic and find its P-value.

(c) Perform any further calculations that are necessary in order to state your conclusion. State your conclusion.
4. (25 pts) Do tall parents have tall children? Sir Francis Galton, a cousin of Charles Darwin, undertook a detailed study of human characteristics. One of his interest was in determining whether a relationship exists between the heights of parents and the heights of their sons and he believed that there is a very close relationship between the heights of parents and the heights of their sons. To test this claim, a researcher took a random sample of 10 men (sons) and recorded their heights with their fathers’ and mothers’ heights, all measured in inches. The multiple linear regression model

\[ \text{Son’s height} = \beta_0 + \beta_1 \cdot \text{father’s height} + \beta_2 \cdot \text{mother’s height} + \text{error} \]

was used, where the errors are assumed to be independent and normally distributed with mean 0 and standard deviation \( \sigma \). The model was fit with MINITAB and the following output was obtained.

**Regression Analysis: Son versus Father, Mother**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-32.38</td>
<td>25.65</td>
<td>-1.26</td>
<td>0.247</td>
</tr>
<tr>
<td>Father</td>
<td>1.0037</td>
<td>0.3507</td>
<td>2.86</td>
<td>0.024</td>
</tr>
<tr>
<td>Mother</td>
<td>0.5161</td>
<td>0.1363</td>
<td>3.79</td>
<td>0.007</td>
</tr>
</tbody>
</table>

\( S = 2.030 \quad R^2 = 77.3\% \quad R^2(\text{adj}) = 70.8\% \)

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>98.059</td>
<td>49.029</td>
<td>11.90</td>
<td>0.006</td>
</tr>
<tr>
<td>Residual Error</td>
<td>7</td>
<td>28.841</td>
<td>4.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>126.900</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father</td>
<td>1</td>
<td>39.001</td>
</tr>
<tr>
<td>Mother</td>
<td>1</td>
<td>59.058</td>
</tr>
</tbody>
</table>

(a) State the null and alternative hypotheses tested by the ANOVA F-statistic for this problem. What conclusion do you draw from the F-test?

(b) What height does this model predict for a son whose father is 70 inches tall and mother is 66 inches tall?
(c) If the father's height is held constant, how much does the son's height increase for each additional inch in the mother's height, according to the fitted model?

(d) Construct a 90% confidence interval for $\beta_2$, the coefficient of mother's height.

(e) It has been argued that the father's height is not significant in determining the son's height once the mother's height has been taken into account. Conduct a test of this hypothesis.
5. (25 pts) The U.S. Energy Information Administration gathers data on residential energy consumption and publishes its findings in *Residential Energy Consumption Survey*. A researcher wants to know if there is a difference in the energy consumption in the four geographic regions of the U.S., namely the Northeast, Midwest, South and West. Data regarding the energy consumption in a random sample of homes in each of the regions was collected. The Minitab analysis of this data is given below.

### One-way ANOVA: Northeast, Midwest, South, West

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>3</td>
<td>97.50</td>
<td>32.50</td>
<td>6.32</td>
<td>0.005</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>82.30</td>
<td>5.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>179.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Individual 95% CIs For Mean Based on Pooled StDev**

- **Northeast**: Mean = 13.000, StDev = 1.871
  - Lower CI = 10.000
  - Upper CI = 16.000
- **Midwest**: Mean = 14.500, StDev = 2.588
  - Lower CI = 11.000
  - Upper CI = 18.000
- **South**: Mean = 10.000, StDev = 2.582
  - Lower CI = 7.000
  - Upper CI = 13.000
- **West**: Mean = 9.200, StDev = 1.924
  - Lower CI = 6.000
  - Upper CI = 12.000

**Pooled StDev = 2.268**

(a) Explain in words what ANOVA tests in this setting. What is the conclusion of the test.

(b) Is the assumption of equal population standard deviations reasonable here? Explain
(c) Calculate $R^2$.

(d) What is the contrast that compares energy consumption in the west to the average of the other three regions.

(e) Test at the 5% level, the hypothesis that the west uses less energy than the average of the other three regions.