MEMORIAL UNIVERSITY OF NEWFOUNDLAND
DEPARTMENT OF MATHEMATICS AND STATISTICS
MIDTERM EXAM - STATISTICS 3521 - WINTER 2007, SECTION 01

Instructor: A. Oyet Date: February 22, 2007

Name(Surname First):________________________Student Number ________________

Time Allowed: 75mins

INSTRUCTIONS

1. Answer all questions.

2. Make sure that your examination paper has 3 QUESTIONS.

3. Important formula(s) are provided at the end of each question.

4. No books or notes are allowed. You may use your calculator.

5. Clearly outline your answers. Show your work and answer each question carefully.

6. Statistical tables can be found on the last page.

<table>
<thead>
<tr>
<th>Question #</th>
<th>Mark</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
Problem 1 (10 points)

In developing empirically a cost function from observed data on a complex chemical experiment, an analyst employed the normal error regression model

\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n. \]

The parameter \( \beta_0 \) was interpreted here as the cost of setting up the experiment. The analyst hypothesized that this cost should be \$7.5\) thousand and wished to test the hypothesis by means of a general linear test.

(a) Indicate the alternative conclusions for the test.
(b) Specify the full and reduced models.
(c) Specify the degrees of freedom for both the reduced model and the full model.
Problem 2 (20 points)

In a study on the relationship between the dollar cost of correcting typographical errors \((y)\) and the number of galleys for a manuscript, a random sample of recent orders handled by a firm specializing in technical manuscripts is shown in the table below.

<table>
<thead>
<tr>
<th>(i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_i)</td>
<td>7</td>
<td>12</td>
<td>4</td>
<td>14</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>(y_i)</td>
<td>128</td>
<td>213</td>
<td>75</td>
<td>250</td>
<td>446</td>
<td>540</td>
</tr>
</tbody>
</table>

Assume that the regression model

\[
y_i = \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, 6,
\]

is appropriate, with normally distributed independent error terms.

(a) Using the least squares method, write down and solve the least squares normal equation for the regression model.

(b) Given that

\[
\begin{align*}
\sum_{i=1}^{6} x_i &= 92; \\
\sum_{i=1}^{6} x_i^2 &= 1930; \\
\sum_{i=1}^{6} y_i &= 1652; \\
\sum_{i=1}^{6} y_i^2 &= 620394; \\
\sum_{i=1}^{6} x_iy_i &= 34602,
\end{align*}
\]

compute an estimate for \(\beta_1\). Write down your estimated regression function.

(c) Suppose that a company wishes to estimate the cost of correcting typographical errors in 5 galleys handled by a St. John’s firm. Use a 95% interval to estimate the cost assuming the estimated function in part (b) is applicable.

**Formula:** 

\[
S_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2}} \quad \text{where} \quad \hat{\sigma}^2 = MS_E = \frac{SS_E}{n-1}.
\]
Problem 3 (20 points): Points may not be awarded for vague and very general answers. You must be specific in your answers.

(a) A pharmacologist employed linear regression model to study the relation between the concentration of a drug in plasma \((y)\) and the log-dose of the drug \((x)\). Figure 1(a) is a residual plot from the analysis. What does the plot suggest?

(b) An economist studying the relation between household electricity consumption \((y)\) and number of rooms in the home \((x)\) employed linear regression and obtained the residual plot in Figure 1(b). What problem appear to be present in the plot?

(c) A chemist studied the concentration of a solution \((y)\) over time \((x)\). Figure 1(c) shows a residual plot from a regression analysis of \(y\) versus \(x\). What does the plot suggest?

(d) A marketing researcher studied annual sales of a product that had been introduced ten (10) years ago by means of regression analysis. The residual plot from the analysis is shown in Figure 1(d). What does the plot suggest?