Stat 502 Homework 4 Solutions

1.
   a. Several difficulties are present. There is nonconstant variance, the residuals are not normal, and there may be two outliers.

   b. A log transformation works well. It doesn’t matter whether base 10 or the natural log is used. When the response variable is log(y), the residuals have satisfactory properties. Null hypotheses are not rejected for tests of equal variances and normality.

   c. This answer is based on a base 10 logarithm, but you should get the same answer if you used natural log. The first row of the following chart gives the necessary mean and standard error.

   Least Squares Means for logtip
   Message    Sex     Mean   SE Mean
   Happy      Female  1.488  0.03277
   Happy      Male    1.229  0.03205
   None       Female  1.423  0.03205
   None       Male    1.367  0.03622

   Error df = 82 and the exact t-multiplier for the confidence interval is 1.989. We’ll use t = 2 here. In the log scale, the 95% confidence interval is
   \[1.488 \pm (2)(0.03277)\]
   \[1.488 \pm 0.06544\]
   \[1.42256 \text{ to } 1.55344\]

   d. To get the desired confidence interval in the original scale, the antilog of each limit determined in part (c). For base 10, the interval is:
   \[10^{1.42556} \text{ to } 10^{1.55344} \text{ or } 26.64 \text{ to } 35.76\]

2.
   a. The medical school factor is significant, but region and the interaction are not.

   Source     DF   Seq SS    Adj SS   Adj MS   F      P
   Region     3      13.997 5.384 1.795 1.10 0.352
   MedSchl    1       8.587  7.614  7.614  4.67 0.033
   Region*MedSchl 3      7.532  7.532  2.511  1.54 0.209
   Error     105      171.264 171.264 1.631     
   Total     112      201.380  

   It’s possible to get the desired confidence interval by using the Comparisons button of the General Linear Models dialog box. Minitab gives the answer as 0.066 to 1.543.

   Most students probably used a method discussed in class that involves getting least squares means and determining the interval by hand.

   Medical School Mean   SE Mean
   No          4.294  0.1371
   Yes         5.099  0.3463

   The exact multiplier is 1.983, which I’ll use here to match Minitab. It’s okay to use t = 2, and if you do the answer will only be slightly different than this one.
   \[(5.099 - 4.294) \pm 1.983\sqrt{0.1371^2 + 0.3463^2}\]
   \[0.805 \pm (1.983)(.37245)\]
   \[0.805 \pm 0.739, \text{ or } 0.066 \text{ to } 1.544\]
b. Both region and medical school are significant.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>3</td>
<td>103.554</td>
<td>47.019</td>
<td>15.673</td>
<td>5.94</td>
<td>0.001</td>
</tr>
<tr>
<td>MedSchl</td>
<td>1</td>
<td>27.894</td>
<td>26.060</td>
<td>26.060</td>
<td>9.87</td>
<td>0.002</td>
</tr>
<tr>
<td>Region*MedSchl</td>
<td>3</td>
<td>0.617</td>
<td>0.617</td>
<td>0.206</td>
<td>0.08</td>
<td>0.972</td>
</tr>
<tr>
<td>Error</td>
<td>105</td>
<td>277.145</td>
<td>277.145</td>
<td>2.639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>409.210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The confidence interval for the difference (med school vrs not) is approximately 0.55 to 2.43. Using least squares means calculations, the work is:

\[
(10.791 - 9.302) \pm 1.983 \sqrt{0.1744^2 + 0.4405^2}
\]

c. Several difficulties are present. There is nonconstant variance, the residuals are not normal, and there are two outliers.

d. The only significant variable is length of stay.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stay</td>
<td>1</td>
<td>57.305</td>
<td>43.850</td>
<td>43.850</td>
<td>35.79</td>
<td>0.000</td>
</tr>
<tr>
<td>Region</td>
<td>3</td>
<td>8.714</td>
<td>1.630</td>
<td>0.543</td>
<td>0.44</td>
<td>0.722</td>
</tr>
<tr>
<td>MedSchl</td>
<td>1</td>
<td>0.621</td>
<td>0.485</td>
<td>0.485</td>
<td>0.40</td>
<td>0.530</td>
</tr>
<tr>
<td>Region*MedSchl</td>
<td>3</td>
<td>7.325</td>
<td>7.325</td>
<td>2.442</td>
<td>1.99</td>
<td>0.120</td>
</tr>
<tr>
<td>Error</td>
<td>104</td>
<td>127.414</td>
<td>127.414</td>
<td>1.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>201.380</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The confidence interval for the difference (med school vrs not) is approximately −0.475 to 0.882. Using least squares means calculations, the work is:

\[
(4.644 - 4.432) \pm 1.983 \sqrt{0.1210^2 + 0.3096^2}
\]

Notice that taking into account length of stay has made the medical factor become non-significant.

e. Yes and no. It is true that the risk is higher at medical schools (see part a), but the reason is that the length of stay is greater at those hospitals than the others (see part b). In part d, we see that there is no additional risk of infection at medical schools once length of stay is taken into account.