1. A sociologist is interested in the relationship between the homicide rate, number of homicides per 100,000 population (Y) and a city’s population (X1) in thousands of people, the percentage of families with yearly incomes less than $10,000 (X2), and the rate of unemployment (X3). Data are provided for a random sample of 20 cities.

<table>
<thead>
<tr>
<th>City</th>
<th>Y</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>City</th>
<th>Y</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.2</td>
<td>587</td>
<td>16.5</td>
<td>6.2</td>
<td>11</td>
<td>14.5</td>
<td>7895</td>
<td>18.1</td>
<td>6.0</td>
</tr>
<tr>
<td>2</td>
<td>13.4</td>
<td>643</td>
<td>20.5</td>
<td>6.4</td>
<td>12</td>
<td>26.9</td>
<td>762</td>
<td>23.1</td>
<td>7.4</td>
</tr>
<tr>
<td>3</td>
<td>40.7</td>
<td>635</td>
<td>26.3</td>
<td>9.3</td>
<td>13</td>
<td>15.7</td>
<td>2793</td>
<td>19.1</td>
<td>5.8</td>
</tr>
<tr>
<td>4</td>
<td>5.3</td>
<td>692</td>
<td>16.5</td>
<td>5.3</td>
<td>14</td>
<td>36.2</td>
<td>741</td>
<td>24.7</td>
<td>8.6</td>
</tr>
<tr>
<td>5</td>
<td>24.8</td>
<td>1248</td>
<td>19.2</td>
<td>7.3</td>
<td>15</td>
<td>18.1</td>
<td>625</td>
<td>18.6</td>
<td>6.5</td>
</tr>
<tr>
<td>6</td>
<td>12.7</td>
<td>643</td>
<td>16.5</td>
<td>5.9</td>
<td>16</td>
<td>28.9</td>
<td>854</td>
<td>24.9</td>
<td>8.3</td>
</tr>
<tr>
<td>7</td>
<td>20.9</td>
<td>1964</td>
<td>20.2</td>
<td>6.4</td>
<td>17</td>
<td>14.9</td>
<td>716</td>
<td>17.9</td>
<td>6.7</td>
</tr>
<tr>
<td>8</td>
<td>35.7</td>
<td>1531</td>
<td>21.3</td>
<td>7.6</td>
<td>18</td>
<td>25.8</td>
<td>921</td>
<td>22.4</td>
<td>8.6</td>
</tr>
<tr>
<td>9</td>
<td>8.7</td>
<td>713</td>
<td>17.2</td>
<td>4.9</td>
<td>19</td>
<td>21.7</td>
<td>595</td>
<td>20.2</td>
<td>8.4</td>
</tr>
<tr>
<td>10</td>
<td>9.6</td>
<td>749</td>
<td>14.3</td>
<td>6.4</td>
<td>20</td>
<td>25.7</td>
<td>3353</td>
<td>16.9</td>
<td>6.7</td>
</tr>
</tbody>
</table>

The data are on the class web page

http://www.public.iastate.edu/~wrstephe/stat401.html

(a) Use JMP to fit a multiple regression model with homicide rate as the response variable and population, low income and unemployment as the three explanatory variables. Turn in the JMP output with your answers to the following questions.

(b) Give the multiple regression equation.

(c) Give the value of the estimate of the error standard deviation, \( \sigma \).

(d) Give the value and an interpretation of \( R^2 \).

(e) What is the value of the adjusted \( R^2 \)?

(f) Test the hypothesis \( H_0: \beta_1 = \beta_2 = \beta_3 = 0 \) versus the alternative that at least one slope parameter is not zero. What does this test tell you about the relationship between the homicide rate and the three explanatory variables?

(g) Look at the plot of residuals versus unemployment rate. What does this plot tell you about the linear model?

(h) Describe what you see in the analysis of residuals; histogram, box plot and normal quantile plot. What does this indicate about the conditions necessary for multiple regression analysis?

2. Use JMP to fit the models necessary to answer the following questions. Be sure to support your answers with information from the appropriate JMP output. Turn in the JMP output with your answers.

(a) Is there a statistically significant relationship between homicide rate and the percentage of families with yearly incomes less than $10,000?
(b) Does unemployment rate add significantly to the model that already contains population?
(c) Does population add significantly to the model that contains the two variables percentage of families with yearly incomes less than $10,000 and unemployment rate?
(d) There are three models that satisfy the first two criteria for being the “best” model, i.e. the model is useful and all variables in the model add significantly. Summarize the models that are both statistically useful and have all explanatory variables adding significantly. In your summaries give:
   - The explanatory variables in the model.
   - The test statistic and P-value for the test of model utility.
   - The test statistic and P-value for each of the explanatory variables that add significantly to the model.
   - $R^2$, adjusted $R^2$, and the Root Mean Square Error.

(e) Of the three models summarized in d), which is the “best” model? Explain briefly.
(f) For the “best” model look at the plot of residuals versus predicted values and the analysis of the distribution of residuals. Indicate what you see in the plots and what this tells you about the conditions necessary for conducting the statistical analysis.

3. Use JMP Fit Y by X to fit a simple linear model with Homicide rate as the response and Population as the explanatory variable. Turn in the JMP output with your answers.

(a) How much of the variation in homicide rate is explained by the linear relationship with population?
(b) Is this model useful? Support your answer with an appropriate test of hypothesis.
(c) Looking at the plot of homicide rate versus population, and the plot of residuals versus population, there is one point that seems to influence where the regression line goes. Which point is this? Give city number, population, percentage of low income families and unemployment rate.
(d) Click on this point, go to Rows and Exclude the point. Go to the red triangle pull down next to Bivariate Fit in the output window and chose Fit Line. You will get a second line on your graph (this one does not include the point you clicked on). How is this line different from the original line (indicate how the intercept and slope have changed)?
(e) How much of the variation in homicide rate is explained by the linear relationship with population once the one point is excluded?
(f) Is this model useful? Support your answer with an appropriate test of hypothesis.
(g) If we fit a linear model with number of homicides, rather than the rate of homicides, as the response variable and population as the explanatory variable, would you expect to see a statistically significant linear relationship? Explain your answer. You do not need to actually fit this model to answer the question.