Lab Assignment 4: Correlation

Investigating the Relationships between Variables.
Open the data set Crime.JMP from JMP IN Data.

1. Let’s look at the correlations between all of the continuous variables in the dataset. Select Analyze → Multivariate then highlight all continuous variables, select Y, Columns → OK. The red circles/ellipses encloses approximately 95% of the points.

   • How does the shape of the red circle/ellipse correspond to the correlation (ie. What correlation corresponds to a circle versus an ellipse)?
     A circle corresponds to a weak correlation, whereas an ellipse corresponds to a stronger association.

   • Which two variables have the strongest positive correlation?
     larceny and burglary, r=.7921

   • Which two variables have the strongest negative correlation?
     There are no pairs of variables with negative correlations.

   • Which two variables have the weakest correlation?
     larceny and murder, r=.1019

2. We can keep track of the output analysis by saving it in a journal. This will be an extremely valuable tool when working on your project. To open a journal window, click on Edit → Journal, the correlation output should appear in the new window.

3. Plot ”Robbery” (X), by ”Rape” (Y), Analyze → Fit Y by X. Is there a relationship between ”Robbery” and ”Rape”?
   Yes...an increasing one.
   Do you see any potential outlying points?
   Yes, corresponding to New York and Alaska

   • To overlay the correlation ellipse, click on the red arrow next to the variable name, select Density Ellipse → 0.95. Are the points you chose to be outlying values indeed exceptional values, ie. do they fall outside of the 95% density ellipse? To what state do these values correspond?
     Yes...Nevada, as well!

   • What happens if the density ellipse is altered to include only 90%?
     California is also an outlier now.

   • Which state has the lowest value for ”Rape”? North Dakota Highest? Alaska
   • Repeat for ”Robbery”? North Dakota and New York, respectively
4. Fit a linear regression line to the plot of "Rape" vs. "Robbery" by clicking on the red arrow and selecting **Fit Line**.

- Do cities with more robberies have more rapes?
  **Yes, for the most part.**

- Does this mean that more robberies cause more rapes?
  **No: association is NOT the same as causation! There could be confounding variables.**

- What are the slope and intercept of the line? (The equation for the line can be found directly below **Linear Fit.**)
  **slope=0.072 and intercept=16.79**

- Click on **Edit → Journal** to add this latest analysis to the journal.

- Double click at the end of the journal window to add text.

- To save the journal, go to **Window → Close all Journals** then click "Yes" in the pop-up menu, and then another menu will appear, verifying the file name and location of the newly created file. **Important:** the file extension must be .rtf (rich text format).

- Close or Minimize **JMP-IN**, and open Microsoft Word or Word Perfect. Open the .rtf file that was saved in the above instructions. Once opened, graphs can be moved using cut and paste and additional text can be added.

**Importing Data into JMP-IN**  
Use the **File → Open** command. Under **Files of type** choose **Text Import Preview**—this will allow you to import data from other packages. You can choose the format in which your data is listed (ie, tab-, comma-, etc-delimited).

Some things to watch for:

1. Before saving the file, eliminate all excess text. This can confuse JMP-IN and make it (and you) very unhappy! Note that when importing files, JMP-IN has an option that specifies that the first row is the column labels.

2. Make sure you have saved the document as a plain (tab delimited) text file (.txt) before importing it. Tabs will tell JMP-IN where a new column begins.