An Introduction to Bivariate Correlation Analysis in SPSS

IQ, Income, and Voting

We shall use the data set “Bush-Kerry2004.sav,” which is described at http://core.ecu.edu/psyc/wuenschk/StatData/Bush-Kerry2004.doc and is available at http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Data.htm. Our research units are the states (and the District of Columbia) of the United States. Download the data and bring them into SPSS.

Variable “iq” is the estimated IQ of the residents of each state. Variable “income” is the estimated personal income of residents of each state. We want to determine whether or not there is a relationship between state intelligence and state income.

Click Analyze, Correlate, Bivariate. Slide IQ, Income, and Vote into the Variables box.

The output will show you that the correlation between intelligence and income falls just short of statistical significance.

Let us look at a scatter plot of the data. Click Graphs, Legacy Dialogs, Scatter/Dot. Select “Simple Scatter” and click Define.
Scoot "income" into the Y Axis box and "iq" into the X Axis box. Click OK.

The plot shows the general trend for income to increase with IQ. Let us edit the graph a bit. Double-click on the graph to open the chart editor.
Click “Elements,” “Fit Line at Total.” Click “Close” on the resulting Properties window and then close the chart editor.

As you can see, SPSS has added the “best-fitting” line that describes the relationship between state IQ and state income.

Direct your attention to the upper left corner of the plot. There is a case that clearly does not fit the general pattern – a case with relatively low IQ but high income. I just had to know what case that is, so I went back to the data file.

The culprit is identified – it is the District of Columbia. I should have known that, what with all the elected officials and their staff located there. 😊
Let us see what happens if we delete that one unusual case. Highlight that case and hit the “Delete” key. You now have the data for the fifty states, no DC. Re-compute the correlation coefficients.

**Correlations**

<table>
<thead>
<tr>
<th></th>
<th>iq</th>
<th>income</th>
<th>vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>iq</td>
<td>1</td>
<td>.502**</td>
<td>-.426**</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>income</td>
<td>.502**</td>
<td>1</td>
<td>-.643**</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>vote</td>
<td>-.426**</td>
<td>-.643**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Notice that we now have a much stronger, and statistically significant, relationship between state IQ and state income.

**Point Biserial R**

Look at the relationship between IQ and “vote.” “Vote” indicates which candidate won the state in the presidential election, the Democrat (0) or the Republican (1). The significant negative correlation indicates that Republican states had residents who were less intelligent than those in the Democratic states. The p value here (.002) comes from a t statistic which SPSS has not reported to us, but we can compute it easily.

\[ t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.426\sqrt{48}}{\sqrt{1-.426^2}} = 3.262. \]
Let us use the more common method of comparing one group mean with another, the independent samples $t$ test.

Notice that the pooled $t$ test is identical to the correlation analysis. Independent samples $t$ tests are just a special case of a correlation analysis. Think about that the next time some fool tells you that you can infer causality from the results of a $t$ test but not from the results of a correlation analysis. What is important, with respect to inferring causality, is how the data were gathered (experimentally or not), not how they were analyzed.

**Overall GPA and Grades in PSYC 2101 (Psychological Statistics)**

Download and bring into SPSS the data at http://core.ecu.edu/psyc/wuenschk/SPSS/GPA_Overall-2101_PSYC-majors.sav. For each PSYC major taking my PSYC 2101 (one semester in 2015) we have two scores: Their overall GPA and their final grade in PSYC 2101. Overall GPA is on a scale that runs from 0 to 4.0. GPA2101 is on a scale that runs from 0 to 4.3 (4.3 being an A+, a grade the ECU Registrar does not recognize).

Compute the correlation between Overall GPA and GPA2101 and prepare a scatter plot showing the relationship. Plot GPA2101 on the ordinate and Overall GPA on the abscissa. Write an APA style report in which you describe the results of your analysis. Then check my report versus the one you wrote.
More Examples

- Correlation between homework grades and exam grades
- Correlates of infant mortality in African nations

Return to Wuensch’s SPSS Lessons Page.

Karl L. Wuensch, March, 2016