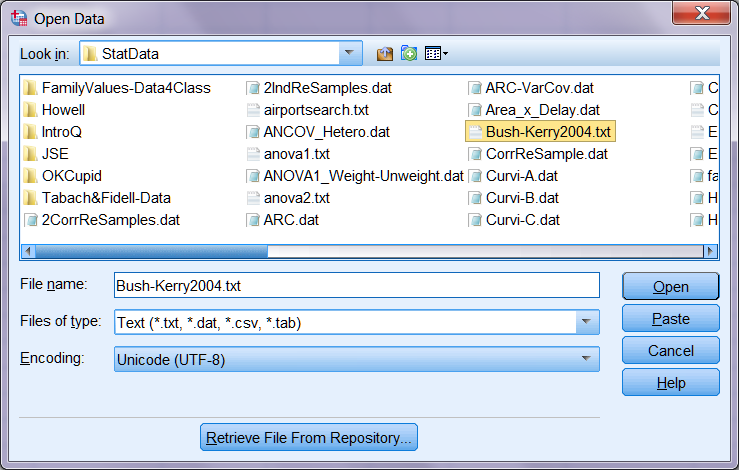
**Reading Plain Text Data Files With SPSS**

I use SAS to simulate data for my students in my statistics classes. The output data are PUT into simple plain text (\*.dat or \*.txt) files, with the blank space used as a delimiter. In my undergraduate classes the students use SPSS. Accordingly, they need to know how to read such a file into SPSS, and this document serves as a lesson on doing just that.

From the page at <http://core.ecu.edu/psyc/wuenschk/StatData/StatData.htm> , download the file Bush-Kerry2004.txt. Then boot up SPSS and click File, Open, Data. Change the file type to Text and Point to the downloaded data file and then click “Open.”



On Step 1 of the Text Import Wizard, verify that “No” is selected and then click Next.

On Step 2, verify that “Delimited” and “No” are selected and then click Next.

On Step 3, verify that the data begin on line 1 (the variable names are not on line 1 in this file), each line represents one case, and you want to import all the cases. Then click next.

On Step 4, verify that the delimiter is a space and there is no text qualifier, then click Next.

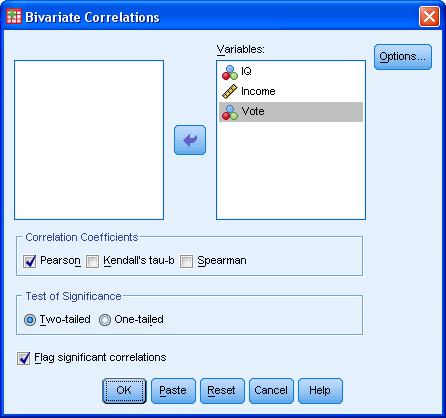
On Step 5 you can name the variables (alternatively you can do this later in the variable view of the datasheet). Under Data Preview highlight the first column. Then, in the pane above, change “V1” to “IQ.” The data format should be “Numeric.” Then highlight the second column. In the pane above, change “V2” to “State.” The data format should be “String.” In a similar fashion, change V3 to “Income” (numeric), V4 to “Vote” (numeric) and V5 to “Candidate” (string). Click Next.

On Step 6 just click Finish.

The “subjects” here are states in the USA (and the District of Columbia) and the variables are

* IQ – Estimated mean IQ of residents of the state
* State – name of the state
* Income – average income of residents of the state
* Vote – the state’s electors voted for which candidate in the 2004 election, where 0 = Kerry and 1 = Bush.
* Candidate – same as vote, but with name of candidate rather than numeric code.

Analyze, Correlate, Bivariate. Scoot all of the numeric variables into the variables box and then click OK.



In the output table each cell gives the value of the Pearson correlation coefficient. Ignore the (Sig) *p* values – they are inappropriate, given that we have the entire population here, not a random sample from a population. Interpret each of the three correlation coefficients, keeping in mind that the vote variable was coded “0” for Kerry and “1” for Bush.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | income | vote |
| iq |  | .264 | -.349 |
| income |  |  | -.635 |

**Interpretation:**

There is a medium-sized correlation between IQ and income. States with higher mean IQ also have higher mean income.

This is a medium-sized negative correlation between IQ and voting for Bush. States that went for Bush had lower mean IQ than did states that went for Kerry.

There is a large negative correlation between mean income and voting for Bush. States that went for Bush had lower mean income than did states that went for Kerry

**Column (Fixed Field) Input**

One of my graduate students downloaded a plain text data file from <http://www.stat.ufl.edu/~winner/datasets.html> . The coding sheet is copied below: Dataset: trouthead.dat

Source: C. McC. Mottley (1941). "The covariance method of comparing the head-lengths of trout from different environment," *Copeia*, *1941*(3), 154-159

Description: Head-Lengths and Standard-Lengths of 50 male and 50 female trout from Kootenay Lake and 10 male and 10 female trout from Lake Wilson.

Variables/Columns

Sex 8 /\* 1=M, 2=F \*/

Lake 16 /\* 1=KL, 2=LW \*/

Head-Length (mm) 22-24

Standard-Length (mm) 30-32

I changed the word “Gender” to “Sex,” as I have a hard imagining that trout have gender identity distinct from their biological sex. The sex variable is coded 1 male, 2 female, and the scores are in column 8. The lake information is in column 16. Head lengths are in columns 22 to 24, and standard lengths in columns 30-32. To read these data into SAS one would use code like this:

options pageno=min nodate formdlim='-' FORMCHAR="|----|+|---+=|-/\<>\*";

**Proc** **Format**; Value mf **1**='male' **2**='female'; Value lk **1**='Kootenay' **2**='Wilson';

**Data** Trout; Infile 'C:\Users\Vati\Desktop\PDS\_2017\trouthead.dat';

Input sex **8** lake **16** head **22**-**24** standard **30**-**32**;

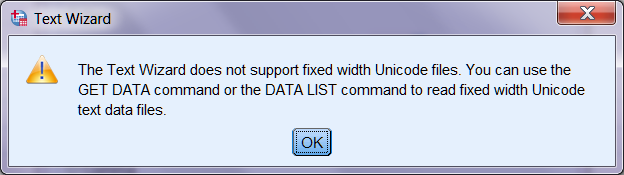
Ratio=head/standard;

Format sex mf. lake lk. ;

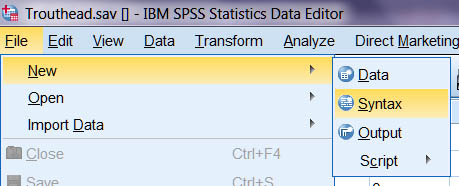
**Proc** **Ttest**; Class lake; Var ratio; **run**;

Note the use of [Proc Format](http://core.ecu.edu/psyc/wuenschk/SAS/Help/Format-SAS.htm) and the Format statement to create and apply value labels to the codes for sex and lake. Also note that a created a new variable, ratio, the ratio between head length and standard length.

I also tried to read these data into SPSS using the import wizard, but got this error message:



I had to use Google to find out how to use the SPSS DATA LIST command. Here is what you need to do. First, open a new syntax window – File, New, Syntax



In the syntax window I entered this code:

data list fixed file='C:\Users\Vati\Desktop\PDS\_2017\trouthead.dat'

/ gender 8 lake 16 head 22-24 standard 30-32.

list

Then I clicked Run, All. The data were successfully imported. I used Transform, Compute to create the new ratio variable and I used the values column in the datasheet variable view to provide value labels for sex and lake.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Statistics** | | | | | | |
|  | | sex | lake | head | standard | ratio |
| N | Valid | 120 | 120 | 120 | 120 | 120 |
| Missing | 0 | 0 | 0 | 0 | 0 |
| Skewness | | .000 | 1.812 | 1.434 | 1.223 | .188 |
| Kurtosis | | -2.034 | 1.303 | 1.052 | .555 | -.192 |

I downloaded the article referred to in the coding sheet. The researcher had applied a log transformation to both length variables. As you can see, they exhibit positive skewness. The researcher then used Analysis of Covariance to compare the two lakes on head length, adjusted for standard length, within sex. For trout of both sexes, the head lengths were longer at Lake Wilson than at Kootenay Lake, after adjusting for standard length.

I decided to take a different approach, using the ratio variable. As you can see above, the ratio data are approximately normally distributed, no transformation needed. A *t* test revealed that the ratio of head to standard length was significantly greater (*p* < .001) at Lake Wilson than at Kootenay Lake.

I did also conduct an ANCOV on the untransformed data ignoring sex, using standard length as a covariate. The mean adjusted head length was significantly greater (*p* < .001) at Lake Wilson (74.245 mm) than at Kootenay Lake (70.001).

[Karl L. Wuensch](http://core.ecu.edu/psyc/WuenschK/KLW.htm), August, 2017

[Return to Wuensch’s SPSS Lessons Page](http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Lessons.htm)