

Creativity1.sas: Explanation of code

Goals of code:

- Display the creativity data (histogram, boxplot, dotplot)
 - As one group or for each treatment
- Calculate summary statistics for each group

Information about starting SAS and the structure of a SAS program is in SASstart.pdf.

Reading the data set: proc import;

proc import reads data files in various formats, including comma separated (.csv) and excel spreadsheet (.xlsx), and creates a SAS data set. The datafile= option names the data file to be read. The file extension (.csv) is necessary. The out= option names the SAS data set that will be created.

I generally check the line in the log that says “The data set WORK.CREATIVITY has 47 observations and 2 variables”. I check that SAS read the number of observations I expected and produced the number of variables I expected.

Reading the data set: Import Data wizard

An alternative to code is to use the Import Data wizard. Look in the File menu and click Import Data. The series of dialogs asks you to identify the type of file, then the file name, than some details about the result.

Descriptive statistics and histograms: proc univariate;

PROC UNIVARIATE calculates and reports many different statistics. The VAR statement names the variable to be summarized. You can name multiple variables if you want more than one summary. The optional HISTOGRAM statement requests a histogram in addition to the numeric summaries. The optional title statement provides a title (in quotes).

SAS reports a lot of numbers, including many that we won't talk about. You will probably be interested in:

In the box of results labeled Moments:

N: number of observations

Mean: sample average

Std Deviation: sample standard deviation

In the box of results labeled Basic Statistical Measures:

Mean: a repeat of the sample average

Median: sample median Std Deviation: a repeat of the sample standard deviation

Interquartile Range: the IQR

Coeff Variation: the coefficient of variation

You may be interested in:

In the Quantiles box: 100% Max, 75% Q3, 25% Q1 and 0% Min:
the max, 75'th percentile, 25'th percentile, and minimum.

Some of the other numbers will be used later. The rest have specialized uses.

Descriptive statistics and histograms for groups: `proc univariate; class treatment;`
Adding a class statement to the `proc univariate` code tells SAS to do everything twice, once for each unique value of the treatment. Same format for the numeric output, but the top of the page is labeled `treatment =` to indicate which treatment is being described. The histogram statement with a class statement produces nice side by side histograms.

Note: Remember that the grouping variable goes in the class statement and the result variable (to be described) goes in the var statement. If you reverse these, e.g. `class score; var treatment;`, you get mush or worse.

Boxplots: `proc boxplot;`

`Proc boxplot` produces side-by-side boxplots. One quirk is that it requires the all observations in one group to be together. The easiest way to ensure this is to sort the data by treatment first. The `proc sort; by treatment; run;` does this. The plot statement specifies the response variable and the grouping variable. The response is first, then the group, separated by a *. Again, if you reverse these, the result is either mush or a lot of errors.

Note: If you expected two boxes (for two groups) and got many, the data set is probably not sorted.