

hamburger.sas: Explanation of code

Goals of code:

- Read an excel sheet (.csv or .xlsx file)
- Wilcoxon rank sum test

Reading an excel sheet: `proc import datafile=;`

SAS provides two different ways to read excel spread sheets. One is to store a specific spreadsheet as a comma delimited file (.csv format). The other is to read the Excel workbook (.xlsx format file). Both use the data importing procedure, `proc import`. We've already seen how to read a .csv file. Here's how to read a .xls or .xlsx file.

Importing a .xlsx file: 2nd `proc import`

Same arguments as if you were importing a .csv file. You can also read .xls (pre-2010) format file. An Excel workbook can have multiple spreadsheets. By default, SAS will read the one called Sheet1. You can add arguments to specify exactly which sheet if you have more than one and you have named them.

Using the Import Wizard:

There is also an interactive Import Wizard that allows you to mouse select the file to be read. From the main menu, choose File / Import Data. You can choose the type of file and then use the mouse to select a file. On the screen that says "Choose the SAS destination", leave library WORK alone and enter a data set name in the box labeled Member. That is the name of the data set into which the contents of the file will be stored. Don't worry about saving the SAS code (but you can if you want to).

Using `proc import` or the import wizard with text files:

You may be asking why I have you write a data step to read text files. There are `proc import` options and wizard options to read .txt files. The reason is that a csv file or an excel sheet have a specific separator character between fields. There is only one separator between each column. Two adjacent separator characters indicate a missing value in a column. Text files may have multiple spaces between fields. That format is read correctly by a data step but not by the import stuff (`proc` or wizard).

Wilcoxon rank sum test: `proc npar1way wilcoxon;`

Note: The `proc` name has the number one in it, not an "el". Shorthand for nonparametric, 1 way.

The syntax is just like for a two sample t-test using `proc ttest`. The `class` statement names the grouping variable; the `var` statement names the response variable. The `exact` statement requests an exact p-value. This can require substantial computing time if there are lots of observations per group. The `/maxtime=` specifies the maximum time SAS is to spend on the computation. The example code specifies 10 seconds. The `/mc=10000` option requests a monte-carlo approximation to

the exact p-value using the specified number of randomizations. I usually use 10000, but more will give you a more precise estimate of the p-value. Without the `/mc` option, SAS will try to enumerate all permutations of treatments. The keyword `wilcoxon` on the `proc npar1way` statement tells SAS to only give you the wilcoxon score analysis. If you omit this, you get many different types of analysis. Wilcoxon is the most frequently used.