

Class Exercise 7

This exercise is based upon Chapter 2 of Delwiche and Slaughter’s “The Little SAS book”. It is a short demonstration of an evolution in reading difficult data sets.

Type the following 5-variable data set into an Excel file; the worksheet should have 3 rows and 5 columns when finished. Note that the data contains many difficult formats: dollar amounts with commas, clock time, text with embedded spaces, and numbers with embedded commas.

Dollars	Date	Time	Numbers with Commas	Text with Space
\$1,523	10/28/59	10:12	1,324	Good morning
\$1,329,200.23	10/14/03	23:14:12	10	Good
\$0.46	10/6/98	5:12.6	1,200,000.45	Some random text

As you type in the data set, does Excel seem to be formatting it automatically? What sort of format is it using for each column?

Now import the data set into SAS using the PROC IMPORT step below (substitute your class directory).

```
proc import out=tough
datafile="/home/grego1/STAT 540/ToughSpreadsheet.xlsx"
dbms=xlsx replace; *Using excel instead of xlsx may work better here;
sheet="Sheet1";
run;
```

Now go to the WORK library in Libraries and open Tough. *Note: Alternatively, you may also select the **Output Data** tab then select Tough.* How did SAS do in reading the data set? Were all of Excel’s formats preserved, or have they been modified? You can generally answer these questions simply by inspecting the worksheet, but it may also help to click on each variable in turn, and look at the data formats.

The main purpose of this portion of the exercise is to show you that most informats, even complicated ones are handled readily in Excel without specifying formats yourself. These formats are generally preserved when importing Excel spreadsheets into SAS.

Now download the comma-delimited file listed on the webpage under **Exported comma-delimited file for cumulative daily COVID-19 confirmed cases**. Upload it to SAS Studio, and use a PROC IMPORT statement to save it as a SAS dataset. Just as we did in R with cumulative daily COVID-19 deaths, we will eventually transform this “long” dataset into a “tall” dataset. For now, inspect the dataset and discuss any changes you think will be needed (which variables to remove, any formatting issues, etc).