An Introduction to SAS for Windows

Prepared by Gigi Luk

Exercise set two.

One thing to remember for this exercise -- When you use the INFILE statement, remember to specify where the file is in YOUR computer. For example, if I save the datasets into My Documents in my home computer, my INFILE statement should be:

```
infile 'c:\Documents and Settings\My Documents\DATASET.txt';
```

So, you have to pay attention to the destinations of the file that you save the .txt files. If you don't have SAS at home, you can simply set it as `'f:\DATASET.txt'` in the lab computers and check your answer next class.

1) Read in **data1.txt** using the INFILE statement. The data set should be attached along with this exercise. The first observation starts at the 5th line and you don't have to count the position of the data point – simply separate each variable name by a space will do. The variables are ID, order, ageinyrs, gender, hand and RT. Except for ageinyrs and RT, all the variables contain character information.

2) Now read in **data1.txt** again without the variable order. (Hint: The first digit of the variable ageinyrs is at column 7). This problem is for you to practice how to read in a dataset using both list and column formats.

3) Read in **data3.txt** and **data4.txt** into your computer. For datasets this small, you can either copy and paste them to SAS or use an INFILE statement. If you use an INFILE statement, remember to specify where you store the data files in your computer. The variables in **data3.txt** and **data4.txt** are the same. Except for ID and GENDER, all the other variables are numeric. Here are the variable names:

   ID GENDER AGEINMOS EPPVTSTD EREAD ERAVEN

4) *This is not a problem, just a step you need to take before moving onto the next problem.*
   Data in these two datasets were collected over a 2-year period. There were 30 children participated in the study (20 in the first year and 10 in the second year). Combine these two datasets into a single file, name the combined dataset as **complete** by using the following statements:

```
data complete;
set data3 data4;
run;
```

5) Sort the data by GENDER and complete the following:
   a. Get descriptive statistics on the data separated by GENDER;
   b. Run a correlation for variables AGEINMOS, EPPVTSTD, EREAD, and ERAVEN;
   c. Run a correlation again but by GENDER this time.
   d. Compare the performance on EREAD by GENDER. (Hint: You can use t-test that we've looked at from the previous exercise). What is your conclusion?

Bonus question

6) This problem involves notes that are on page 23. If you have time, try to read the notes and work on the example.

Below is an annotated printout of the dataset **data2.txt** which is sent to you as an attachment in the email. The data set contains 48 observations and is about diamond size and price. The two variables are carat and price, both of them are numeric. For more information about this dataset, check: http://exploringdata.cqu.edu.au/dia_asn.htm

Read in **data2.txt** either by the INFILE statement or simply copy and paste the data from the text file. In the INPUT statement, use INFORMATS to obtain an output like the following: Remember to open up the data and see what the data is like before setting your INFORMATS.

Obs	carat	price
1	0.17	355
2	0.16	328
3	0.17	350
4	0.18	325
5	0.25	642
6	0.16	342
7	0.15	322
8	0.19	485
9	0.21	483
10	0.15	323