Data Analysis Using SAS for Windows

Prepared by Gigi Luk

Exercise set three.

Read in **bkground.sas7bdat** and **reading.sas7bdat.** These datasets are sent to you as SAS database files. Save the files to your computer and read them in by creating a library that includes all the datasets in that drive. Before moving on, create a copy for each dataset to the work library for further analysis.

1) These data from the two files are collected with the same group of students in two different batches of variables. So, you will see the same ID in the two files. How can we properly merge together the data in the two files? (Call the combined dataset as **Final**)

2) Set up a dataset that only includes female children. (Name the dataset as **Girls**).

3) Rename the variable ID to SUBJECT in **Girls**.

4) The researcher found no difference in the performance from the two genders, s/he wants to drop the variable gender from the **Final** dataset. How can s/he do that?

5) Create four datasets:
   a. HiERead → include observations that have English reading scores greater than 60.
   b. LoERead → include observations that have English reading scores less than or equal to 20.
   c. HiCRead → include observations that have Chinese reading scores greater than 17.
   d. LoCRead → include observations that have Chinese reading scores less than 10.

6) Instead of putting the datasets into 4 subsets, we can also create a variable call "level" to indicate the levels of reading skills for these children. Use the dataset names indicated in 5) and create two new variables called **Level_E** and **Level_C** which indicates the level of English and Chinese reading skills. Use the same cutoff in 5). Also, use the words "high" and "low" instead of the short forms.

7) Find the average between the two reading performance variables, **Eread** and **Cread**. Create an appropriate variable name.

Bonus questions

The following bonus questions require you to look at two Procs that we do not cover in this course. But they are very simple statistical techniques which I think it would be handy to have them in your vocabulary. A hint is to look at the list of procedures covered in the help menu and see which one is good to use. These questions also require you to have a little background on statistics to do.

A) Now the researcher wants to compare whether the children performed the same in the two reading tasks. What test can she use? (*Hint:* The two variables were measured <u>repeatedly</u> in the same child)

B) Can you find out how many males and females students there are in the dataset? (*Note:* Don't count them manually!! Use SAS to figure this out.)