Statistics 480, Homework 5, Spring 2003

Murphy's Law for Data Sets #8: *The data set you need will always exist in a different format than you need it in for analysis and on a different type of computer system than you use for analysis.*

This exercise is designed to familiarize you with some of the basics of file transfer. While it is impossible to expose you to all possible variations of this task called file transfer, the hope is that you become aware of some of the issues as well as some of the solutions.

Exercise

Overview Minitab (available in 307 Snedecor) comes with a sample data set called **pancake.mtw**. This data comes from an experiment involving the quality of pancake mix for different supplements and amounts of whey in the mix. Your goal is to move this data from the PC in Minitab format to VINCENT in several text/ascii formats and analyze the data with SAS. The steps involve opening the data set in Minitab, saving it as a text file (**pancake.txt**) and saving it as an Excel file (**pancake.xls**), then opening the Excel file (in Excel) and saving it as "Text (Tab Delimited)" (name it **pancake_tab.txt**) and saving it as "Formatted Text (Space Delimited)" (name it **pancake_sp.txt**). Finally, open **pancake.txt** in JMP, and save it as a text data file from JMP naming it **pancake_jmp.txt**. You will then FTP these four text files (SAS has no hope of reading the **pancake.xls** file) to your VINCENT account, where you will edit them (with pico or emacs and then the SAS Program Editor) to add the required SAS commands to the top and bottom of the files before submitting them to SAS.

Part I At a PC in 307 Snedecor, you will prepare four text versions of the pancake data for transfer to VINCENT. **Note: There are only 5 PCs and 4 Macs in 307 Snedecor and these are the only machines that have Minitab**.

1. Open Minitab and open the **pancake.mtw** file (**.mtw** stands for Minitab Worksheet) which should be in Minitab's Data folder.

2. You will want to use the Save Current Worksheet As command in the File menu of Minitab to save the data first as a text file (name it **pancake.txt**) and second as an Excel file (name it **pancake.xls**).

3. Open Excel and open the file **pancake.xls**. Use Excel's Save As command to save the data first as a tab delimited file (name it **pancake_tab.txt**) and second as a space delimited file (name it **pancake_sp.txt**).

4. Open JMP and open the file **pancake.txt**. You may want to open this as a Text file with Preview so that you can be sure the data is imported correctly. Save the file as a Text Export file from JMP (name it **pancake_jmp.txt**).

5. Now, use the FTP software to "put" the four text files into your VINCENT account.

Part II From a VINCENT machine, you will now edit the four text files putting SAS commands before and after the data to obtain SAS output. You will do this twice for each file—once with the pico editor (or emacs editor) and then once with the SAS Program Editor. The order here is important—use pico (or emacs) and **then** use the SAS Program Editor for the second round of edits and submits. When using the pico (or emacs) editor, you will edit each file, then submit them to sas in "batch mode" (i.e. type **sas filename** at the VINCENT prompt). When you use the SAS Program Editor to edit the files, you will submit them via the Run→Submit menu command in SAS.

1. Login at a VINCENT machine.

2. Type **ls** to make sure you are in the same directory as the pancake text files (most likely your home directory).

3. Using pico (or emacs), edit each text file. You will need to open each file and insert the following lines around the data:

```
data pancake;
     input supplmnt whey quality;
cards;

<data goes here>
;
proc print;
proc glm;
     class supplmnt whey;
     model quality=supplmnt whey/p;
run;
```

You may find it convenient to edit and submit one file at a time to SAS, or you may prefer to edit all three files and then submit them one at a time. Print each of the files after you've inserted the SAS code and note which VINCENT text editor you used (pico or emacs).

4. The next step is a little open-ended and vague because it depends on what you get as output from SAS, or if you don't get any output, what the SAS log says. Your goal is to try and get output from SAS from all four pancake files **without** editing the actual numbers themselves! You can make changes to the SAS code around the data, but you should not edit the actual data itself (imagine if you had thousands of observations—you would not edit them by hand). If you don't get output, read the log and see if it sounds like something you can fix without editing the data. If you believe the problem cannot be fixed without actually editing the numbers, then print the log and explain what you believe is the problem with the data. If you get output, then print the output.

5. Repeat steps 3 and 4 using the SAS Program Editor instead of pico or emacs.

Part III Describe any difficulties you had in completing any of the above steps. Also, indicate the differences between the four text files you obtained and include their printouts. Turn in the printouts of your SAS logs and SAS outputs too.