

Statistics 480, Homework 4, Spring 2003

File Transfers: Part One

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 A data set containing crime statistics for each of the 50 states exists as a SAS-data-set on VINCENT. Your goal is to perform a regression analysis using SAS and then use JMP to obtain residual plots. It is easier to get presentation quality graphics from JMP than from SAS (unless you are an expert at SAS/GRAPH PROCs). This exercise assumes you have created a `/stat480/` directory in your home directory on VINCENT.

Part I The data set is called `crime.sas7bdat` and is located in a "public" or "anonymous" FTP directory on VINCENT. You will need to login to this directory and get the file. SAS-data-sets should be transferred as binary files rather than ascii/text files. You will also want to get the file called `crimedata.sas` (a text file). This file performs the required analysis of the crime SAS-data-set.

1. Go to a VINCENT machine and login.
2. Execute the FTP commands from your home directory. Your FTP session will require you to enter commands (in order, one at a time). The commands appear in bold. Prompts from the computer are not in bold.

```
vincent% ftp likelihood.stat.iastate.edu
Name (likelihood.stat.iastate.edu:your_user_name): anonymous.wrstephe
Password: your_email_address
ftp> ls
incoming
crimedata.sas
crime.sas7bdat
ftp> binary
200 Type set to I.
ftp> get crime.sas7bdat
ftp> ascii
200 Type set to A.
ftp> get crimedata.sas
ftp> quit
```

After you quit the FTP session, type `ls` at the VINCENT prompt and verify that you have both files in your home directory now.

3. The file `crime.sas7bdat` needs to be in your SASUSER directory before you can use the file `crimedata.sas` to analyze it. To find your SASUSER directory, from the VINCENT prompt in your home directory, enter the command `ls sasuser*`. You should find a directory `sasuser.800`. If so,

```
mv crime.sas7bdat sasuser.800/
```

to move the data set to your SASUSER directory. (Your SASUSER directory might be named something other than `sasuser.800` since this name changes from time to time with different versions of SAS.)

Note: Your SASUSER directory may not be in your home directory; instead, it may be located inside your stat480 directory—in this case, you'll need to execute the command

```
mv crime.sas7bdat stat480/sasuser.800/.
```

You will also want to execute

```
mv crimedata.sas stat480/ and cd stat480 before the next step.
```

4. Now, you are ready to analyze the data with SAS. From the VINCENT prompt in your stat480 directory, execute

```
add sas
sas crimedata
```

This should result in two new files being created: **crimedata.log** and **crimedata.lst**. Print the file **crimedata.lst**.

5. Logout of VINCENT

Part II In Part I, you created a file called **crimedata.lst** in your VINCENT home (or stat480) directory. This file contains a PROC REG printout for the crime data and (near the bottom of the file) a listing of all the data along with the residuals and predicted values for each observation. In this step, you want to use JMP to generate a histogram of the residuals, a normal quantile plot of the residuals, and a scatterplot matrix of the residuals, predicted values, and x_j 's (rape, robbery, assault, burglary, larceny, auto).

1. Go to either a PC/Windows or Macintosh computer and open the FTP program. Fetch is the name of the Macintosh FTP program and WIN_FTP is the name of the Windows FTP program.
2. FTP on Mac and Windows is a more graphical task than on VINCENT (Unix). You will still have to establish a connection to a VINCENT machine to begin the FTP session. You should login as yourself (not anonymous.wrstephe). You want to get the file **crimedata.lst** from your VINCENT home (or stat480) directory onto the Mac or PC you are using. Pay close attention to where on the Mac or PC hard drive the file is transferred because you'll need to be able to open it in a text editor and in JMP.
3. Once the file has been transferred to your Mac or PC, you may quit the FTP program. The file **crimedata.lst** contains more than just the data, so we need to open it in a text editor and delete all the lines except the last 50 lines containing the data (with residuals and predicted values appended). On the Mac, you can use SimpleText, and, on the PC, you can use Notepad under Accessories. Either of these will work for deleting the unwanted lines from **crimedata.lst**. If you use a word processor (like Microsoft Word) to do this step, make sure you save the file as text/ascii and not in the default Word format. JMP can read text files but not Word files. Save the edited version of the file as **crimedata.txt**.
4. Open JMP and select Open... from the File menu. In this dialog box, you will have to tell JMP to look for Text files and to open the file as Text with a Preview. Ask JMP to open the file **crimedata.txt**. When opening a Text file with a Preview, JMP provides several options and you can see how the first few rows of data will look after being imported using the options you select (be sure to click Apply Settings if you change the importing options to see how the changed options will affect your data). If the worksheet you obtain in JMP is incorrect (e.g. the first row of data might be used as the column names in JMP rather than as the first row of data!), just choose Open from the File menu and try adjusting some of the options.
5. With the data set successfully opened in JMP, obtain the residual plots described above. The menu items Analyze→Distribution and Analyze→Multivariate will be useful. You may also want to name the columns for easy reference. Print each of these plots.

Part III Describe any difficulties you had in completing any of the above steps—turn this and your printouts in next week.