The Basics of Statistical Computing in Your UNIX Account

Introduction

The computing network in Biostatistics is primarily composed of Sun workstations running the UNIX operating system. When a Biostat account is created for you, e.g., with login ID (user-name) `pubh7430a`, directory space is available for you to store your files on the UNIX system. The directory is called, e.g., `/home/muskie/ph7430a/` where `muskie` is the name of one of the Sun workstations. You can save and access files stored there from any UNIX terminal or from home. Statistical software is available on the Sun workstations. Microsoft editing and other such software is NOT available on the Sun workstations.

Logging on

Log on: X-term room (Mayo A-316). Your Biostat course instructor can give you the keypad pass-code to get into the room with the UNIX terminals (called *X-terms*). If your X-term is dark, move the mouse to wake it up and a login window should appear. First, type your user-name after `login`. Then press the *Return* or *Enter* key, and type your preliminary password. You must type it exactly as it appears, distinguishing between upper and lower case. Press *Return* again. Click and hold the left mouse button, move the cursor down to Xterm Window, and then move to the right for a list of machines; choose merganser. Windows directly to a workstation like merganser are called *terminal windows*, in contrast to windows for editors which we will discuss below.

Log on: from your home. First, make sure your computer has an internet connection to the U. Otherwise, you may need the Internet Kit provided by Shepherd Labs (East Bank), 50 Coffey Hall (St. Paul) or 93 Blegen Hall (West Bank) so that you can log in from home. Once you connect to the U, you can use any secure telnet connection to log into your Biostat account. Standard telnet software is part of the Internet Kit or may be on your PC/Mac already. However, standard telnet connections to the Biostat machines are no longer allowed. Instead, you will need to use SSH (Secure Shell - a secure, encrypted replacement for telnet) by downloading the PuTTY Telnet and SSH terminal software to access our machines from home.

PuTTY allows Windows users to connect to remote systems over the Internet via Telnet and SSH. It is available for free use by download at `www.chiark.greenend.org.uk/ sgtatham/putty/`. Click on "Download" at the top of the webpage and then on the active link "putty.exe" under the boldfaced title "For Windows 95, 98, ME, NT, 2000, XP and Vista on Intel x86." Your computer will ask you if you want to save the file to your computer or execute it. You should save it to your computer.

You must then install PuTTY by running the file you downloaded. (For PCs running Windows, use Run in the Start menu or find the program's .exe file in the Windows Explorer and double click on it.) Accept all the defaults when you run the install. The desktop shortcut will be named PuTTY. Once installed, to connect to the Biostat computer named merganser:

1. Connect your computer to the University network through the modem pool, or however you usually connect. Log in with your X500 University account and password as you normally would.

2. Start 'PuTTY' and type in the following setting:

Host Name: `merganser.biostat.umn.edu`

Leave the defaults as they are and click on "Open".

A window will pop up asking you to make sure the host is the correct one to connect when you first log on `merganser` by PuTTY. Click "OK".

3. Then a dark window appears on the screen.

login as: `your biostat user-name`

4. At the password prompt, enter the password for your Biostat account.

If you have a cable modem, `merganser` may refuse your connection because Comcast doesn't assign you a permanent IP address. In this case, connect first to `muskie.biostat.umn.edu` and then, once connected, type `ssh merganser` in the `muskie` window to connect to `merganser`.

Choose a new password. After you log in for the first time, the system will ask you to change your preliminary password and choose your new password. It will ask you to type your current password and then your new password twice. Your new password should be 6–8 characters long and should contain numbers and an internal upper case letter, but no more than 2 upper case letters, and no punctuation (–, ;). You might use part of the home address you memorized when you first went to school as a child.

Moving files from Biostat to a home PC. PuTTY can also be used to move files from your Biostat account to your home PC, or vice versa. While at home, click on Start, then Programs, then the PuTTY folder. You should see PSFTP. Click on this, and then you can type in "Open merganser.biostat.umn.edu". Login by your user-name and password.

1. Move files from the server to your home PC. Enter `get myfile.dat newname.dat`. This will fetch the file called myfile.dat on the server and store it as newname.dat in your home PC. You may specify the path for the newname.dat. For example, if you want to save myfile.dat as newname.dat under D:\mycourses, then you need to enter `get myfile.dat D:\mycourses\newname.dat`.

2. Upload files from your home PC to the server. Enter `put myfile.dat newname.dat`. This will send myfile.dat from your home PC to the server and name it as newname.dat.

Helpful UNIX commands

In any X-term window, the ↑ key will bring back the last command, which you can then re-enter or edit. You can use the mouse to cut and paste: Highlight text with the mouse by left clicking and dragging. Release and then left click on the location where you want to paste. Click the middle button on the mouse to do the actual pasting. Here are some commonly used commands:

```
ls -l                                    lists current directories/files
emacs &                                  starts the emacs editor
gv &                                     starts the ghostview postscript viewer
man command                              displays help file about command

cp file1 file2                           makes copy of file1 named file2
cp /home/muskie/ph7430a/file1 file1      copies file1 from course account to current directory
lpr filename                             prints the file to current printer
reduce filename                          prints the file to current printer, two pages per page
rm filename                              deletes filename
mv file1 file2                           renames file1 as file2
mv file1 directoryname/.                 moves file1 into directory
more filename                            scrolls a file on the screen, Ctrl-C or q exits the scroll
mkdir directoryname                      makes a new directory
cd ~                                     moves you into your course account main directory
cd directoryname                         moves you into the specified subdirectory
diff file1 file2                         finds differences between file1 and file2
chmod 700 directoryname                  sets protection to make this directory private
```

For example, data sets for PubH 7430 are stored in the directory `/home/merganser/course_data/correlated.data/`. If you want to look at the description that goes with a particular data set, for example, you would type

```
more /home/merganser/course_data/correlated.data/timetrial.readme
```

and then press space bar to scroll one page at a time or *Return* to scroll one line at a time.

Text editors: vi, pico, emacs

If you are working from home over a modem, you should probably use vi or pico to edit programs. The Biostat internal website (see below) has sources for help on these editors, and there are books in the computing reference section at the Coffman Union bookstore.

On the X-terms in Mayo, you can use vi or pico, but Emacs is the easiest and most flexible editor to use. You can use the mouse to move the cursor and for cutting and pasting between windows exactly as described above for cutting and pasting UNIX commands. The *GNU Emacs Manual* is in Mayo A-316 and the Biostat Reading Room.

Use the command emacs & in a terminal window to open an Emacs window. Left click on File at the top, and in the menu choose New Frame. This gives you a second Emacs window. You can move these windows around with a left click and drag on the gray bar at the top of the window. This is also how you bring a window to the front.

Left click on File at the top of one window, and in the menu choose Open File. The cursor will move to the bottom of the window, where you type in the name of the file to open. Many menu commands can also be done at the keyboard by pressing two keys simultaneously. Here are some commonly used features of emacs:

```
M-w                       Copy highlighted text
C-w                       Cut highlighted text
C-y                       Paste highlighted text at cursor
                          or center click to paste copied text
C-x C-f                   "find" (open) file
C-x C-s                   save file
M-x replace-string        global search and replace from cursor through end of file
```

M stands for the *meta*-key, which on the UNIX keyboards is the *Alt* key on the lower left side of the keyboard. C stands for the *Ctrl* key on the far left side of the keyboard.

Using SAS, S-Plus and R on our UNIX system

From the Biostat X-term room:

1. SAS

   Open an Emacs window to edit your SAS code by typing `emacs filename.sas &`. You should always give your SAS code file the extension .sas in order to have the system recognize it as SAS code. Save the file in the same directory in which you want SAS to run.

   Use one X-term window (for either `sashimi` or `merganser`) to run SAS. To run the SAS program, simply type `sas9 filename`. (Typing `sas filename` will run SAS version 6.12 on some machines.) After the program has run, SAS makes a file filename.log containing a record of the steps the computer took in running your file. This file is useful for diagnosing programming errors. If the program ran successfully, then SAS also makes filename.lst, containing the output from the program.

   Use another X-term window to look at the log and output files. You can open them in Emacs, or type `more filename.log` or `less filename.log` to scroll through the file. Hit *Return* to scroll one line at a time or the space bar to scroll one page at time.

   If you like a windowing interface, you can also run Window SAS by typing `sas9 &` from a `merganser` or `sashimi` window. Three new windows will pop up: Program Editor, Log, and Output. You can edit your SAS code in the Program Editor and then click Locals and select Submit to run your program. You can view and save the log file in the Log window and the output in the Output window.

   For SAS 9, there is on-line documentation at `support.sas.com/documentation/onlinedoc/sas9doc.html`.

2. S-Plus and R:

   Open an Emacs window to edit your S-Plus (or R) source code by typing `emacs filename.s &` (or `emacs filename.r &`). Do this from the same directory in which you want S-Plus or R to run. Type `Splus6` to initiate S-Plus 6 or `R` to initiate R. Then enter the following command to run the code:

   <div align="center">

   `source(''filename'')`

   </div>

   If the source code runs successfully, S-Plus or R will return with any output followed by a new prompt (>). Otherwise, you will see some error messages; in this case, edit the code to try to correct the errors, save your edits, and source the file again.

Use a second Emacs window as a scrapbook to store output you want to print, or use the `sink` command at the beginning of your program code to write all your output to a file. In the X-term window where your S-Plus or R code ran, click and drag the left mouse button to highlight the output you want to save, then move the cursor to this Emacs window and click the middle mouse button to paste the output. It can also be used in reverse to copy and paste each line of the S-Plus or R code from the source file into the X-term window. This is particularly useful when trying to locate in which line of code the errors are occurring. When you have finished, save your output file to e.g. hw1.out and print it from the X-term window using `reduce hw1.out` or `lpr hw1.out`, or from the Emacs Print Buffer command under the File menu.

You can also run S-Plus in batch mode. Type the following command in the X-term window:
```
Splus6 < filename > & filename.out
```
The printed output and error messages will be stored in the file filename.out.

To run R in batch mode, type the following command in the X-term window:
```
R --no-save < filename.r > filename.out
```
The printed output and error messages will be stored in the file filename.out. The `--no-save` option tells R it does not need to save all the variables created by your program.

Before quitting S-Plus, delete the variables you no longer need. You can use the command
```
remove(''var1'')
```
to remove a single variable, or
```
remove( c(''var1'',''var2'',''var3'', ...))
```
to remove a list of several variables. Note that the amount of storage space on your account is limited, and not removing unneeded S-Plus variables will use up your space FAST!

The command to quit S-Plus or R is `q()`. R will ask whether or not it should save all the variables you created. In general, you should type `n` for "no."

3. Ghostview:

You can work with graphics in S-Plus or R by writing the graph to a postscript (ps) file and then converting it to a pdf file by typing `ps2pdf graph.ps`. You can then print out the graph using Acrobat (open the file in Acrobat with `acroread graph.pdf`), or import it into other documents.

From home:

Once you connect to the Biostat UNIX system through PuTTY, you can submit batch jobs from home. To submit a SAS job, type the following command at the UNIX prompt:
```
sas8.2 filename.sas &
```
To submit an S-Plus job, type the following command at the UNIX prompt:
```
Splus6 < filename > & filename.out
```
To submit an R job, type the following command at the UNIX prompt:
```
R --no-save < filename.r > filename.out
```
Note that PuTTY does not allow a graphics connection. This means that SAS, R, and S-Plus programs can be run from the UNIX command line, but no graphics created by those programs can be viewed

instantly. Graphics must be created in the program and then downloaded to your home computer and viewed separately.

Software does exist to enable graphics to be created and viewed as they are being created in SAS, R, or S-Plus under UNIX. For example, X-Win32 is available from `www.starnet.com`. The demonstration version is available for free, but only works for two hours at a time. After two hours, you must quit the program, but then you can restart it right away. Once this program is installed on your PC, here is how it works:

1. Dial up and connect to one of our machines via the modem pool or however you usually connect to the U.

2. In the PuTTY window for our machine, type `finger` and look in the Where column to see what port the modem pool has assigned to you, e.g., pub-14-0000.dialup.umn.edu.

3. In the PuTTY window, type `setenv DISPLAY pub-14-0000.dialup.umn.edu:0.0`. This tells the UNIX machine where to send the output from the X-term.

4. In the PuTTY window, type `xterm`. Then an X-term will pop up on your home computer desktop.

This is still a little slow, even with a DSL/cable modem. However, now everything works as if you were sitting in front of an X-term in Mayo.

How to get S-Plus or R for a home PC:

A student version of S-Plus for Windows is supposedly available for students for free by download from `elms03.e-academy.com/splus/`. R is available for free for any platform (including Linux) by download from `www.r-project.org`.

How to Get Help

The Biostat internal website:

Much documentation is available from the Biostat internal home page, `web.ccbr.umn.edu`, which is different from our public home page, `www.biostat.umn.edu`. To access the internal web page:

1. From the X-term room: After logging into the UNIX system, open a terminal window and type `firefox &` to initiate Firefox. Then click on Home.

2. From home: Open any web browser and type in `https://web.ccbr.umn.edu`.

   Your browser may ask you to verify that you want to go to that site. Keep clicking on Next or Continue. It will also ask you for your Biostat user-name and password. If your password is not accepted, then send an email to `kien@ccbr.umn.edu` and ask him to check on this for you.

Documentation for UNIX:

If you have questions regarding UNIX commands, an on-line help manual can be found at `unixhelp.ed.ac.uk`. There is also an introduction to UNIX book in the Biostat Reading Room (Mayo A-460). The Biostat internal website also has a UNIX FAQ section.

Documentation for statistical software:

Rudimentary documentation for R is available at `www.r-project.org`. On-line tutorials for R are available at `www.faculty.uaf.edu/ffnt/teaching/popgen/R-tutorial/R-tutorial.html` and `www.math.ilstu.edu/dhkim/Rstuff/Rtutor.html`. On-line documentation for S-Plus is available at `www.insightful.com/support/documentation.asp`, but you must have an up-to-date browser. (If you try to go to this site, and you get re-directed to `www.redesign.com`, then your browser is not up-to-date.)

Documentation books for SAS and S-Plus are available in the Biostat Reading Room (Mayo A-460).

Written by: Lynn Eberly, Will Thomas, Xin (Eric) Zhi, Yanni Zhu, and Chelsey Schmeichel.