Lab 1
Stat 133

Joel Hanson
397 Evans
jhanson@stat.berkeley.edu

September 2, 2005

The primary purpose of this lab is to establish an SCF class account so that you will be able to start exploring the R computing language, which you will need to use immenently on your first homework assignment. For those of you who would like to access the SCF computers from your home computers, some limitted instruction will be provided.

1 Activating Your Class Accounts

Before explaining the mechanics of establishing the SCF account, I want to emphasize that security concerns are taken very seriously at the SCF. You can and will lose the privelege of having an SCF account if you violate any SCF regulation—which could well jeapordize your successful completion of this course. **Take all precautions to prevent any misuse or unauthorized use of your SCF account.** Carefully read and comply with all the security requirements in the assignment form that you sign.

The assignment form that you have provided is self-explanatory, but a few comments and some warnings may be helpful:

- The login name and password are case-sensitive—so make sure you type them in the login window exactly as written. Be sure to check the `Caps Lock` key if you have any problems.
- Once you manage to login, you will be prompted for a new password. Your new password must contain 6-8 characters with at least one letter and one number. *Do not write your new password down or share it with anyone else—But be sure to remember it.*
- After changing your password you will need to log out in order for your account to be fully activated. After several hours your account should be accessible. *The failure to log out of your account when finished working is an SCF Rules Violation—Do not make that careless error.*

- Once your account is fully accessible, type `help` for a list of SCF help topics. In particular, type `help rules` and carefully read and comprehend all the SCF rules. *Failure to comply with any rule therein can and will result in the loss of your account priveleges, as well as exposing you to other disciplinary actions.* Additional help files that you should read include: `accounts login quotas diskuse background & bigjobs.`

2 Exploring R

R is a freely available GNU statistical computing language that is explained in great detail at `http://cran.r-project.org/`. In order to get started using R, I recommend that you go to CRAN and click on `Manuals` under `Documentation` and then download **An Introduction to R**. In particular, the sample session in Appendix A will show you how to get started with R. For your convenience, I am including an extended excerpt:

```
Login, start your windowing system.
$ R
Start R as appropriate for your platform.
The R program begins, with a banner.

(Within R, the prompt on the left hand side will not be shown to avoid confusion.)

help.start()
Start the HTML interface to on-line help (using a web browser available at your machine)
Iconify the help window and move on to the next part.

x = rnorm(50)
y = rnorm(x)
Generate two pseudo-random normal vectors of x- and y-coordinates.
plot(x, y)
Plot the points in the plane. A graphics window will appear automatically.
ls()
See which R objects are now in the R workspace.
rm(x, y)
Remove objects no longer needed. (Clean up).
x = 1:20
Make x = (1, 2, ..., 20).
w = 1 + sqrt(x)/2
A 'weight' vector of standard deviations.
dummy = data.frame(x=x, y= x + rnorm(x)*w)
dummy
Make a data frame of two columns, x and y, and look at it.
fm = lm(y ~ x, data=dummy)
summary(fm)
Fit a simple linear regression of y on x and look at the analysis.
```

```
fm1 = lm(y ~ x, data=dummy, weight=1/w^2)
summary(fm1)
Since we know the standard deviations, we can do a weighted regression.
attach(dummy)
Make the columns in the data frame visible as variables.
lrf = lowess(x, y)
Make a nonparametric local regression function.
plot(x, y)
Standard point plot.
lines(x, lrf$y)
Add in the local regression.
abline(0, 1, lty=3)
The true regression line: (intercept 0, slope 1).
abline(coef(fm))
Unweighted regression line.
abline(coef(fm1), col = "red")
Weighted regression line.
detach()
Remove data frame from the search path.
plot(fitted(fm), resid(fm),
     xlab="Fitted values",
     ylab="Residuals",
     main="Residuals vs Fitted")
A standard regression diagnostic plot to check for heteroscedasticity.
Can you see it?
qqnorm(resid(fm), main="Residuals Rankit Plot")
A normal scores plot to check for skewness, kurtosis and outliers.
(Not very useful here.)
rm(fm, fm1, lrf, x, dummy)
Clean up again.
```

Professor Nolan recommends the two following tutorials that are also avaible at the CRAN site by clicking on `Contributed` under `Documentation`:

- **Using R for Data Analysis and Graphics - Introduction, Examples and Commentary** by John Maindonald.
- **R for Beginners** by Emmanuel Paradis.

Be sure to utilize the help system within R. Start by typing `help.start()` after opening R, and whenever you have a question about an R object type `help(`*object*`)`.

Finally, to quit R type `q()`.

3 Configuring Your Home PC For SCF Access

Begin by downloading SSH and Exceed from `http://software.berkeley.edu` and installing them. Then take the following steps:

1. `Click on Start -> All programs -> Hummingbird Connectivity -> Exceed -> Exceed`
2. `Double click on SSH icon and click on Edit -> Settings -> Profile Settings -> Tunneling -> Tunnel X11 Connections -> OK`
3. `In SSH, click on Quick Connect and then login into yout account.`

Acceptable hosts include one of the SCF machines scf-ugXX.berkeley.edu, where XX is from 01 to 27, or strider.berkeley.edu or treebeard.berkeley.edu (for big jobs only).