Helpful R code and "sesame" data codebook - Homework #1

Note: The following sample code may help if you want to run some of the analyses in R. Please also see the handouts and R information on CoursePlus and discussed in the labs.

```
# Read sesame data in
dta <- read.table("sesame.csv", sep= ",", header=T)
print(names(dta))

# Print out the number of people in the treatment and
# control groups
sum(dta$encour==1)
sum(dta$encour==0)

# Print out some basic information on postlet
mean(dta$postlet)
var(dta$postlet)

# Do a t-test comparing the mean of postlet among the children
# encouraged (encour=1)
# and not encouraged (encour=0)
t.test(dta$postlet[dta$encour==1], dta$postlet[dta$encour==0])

# Do a Wilcoxon rank sum test of postlet among the children
# encouraged (encour=1)
# and not encouraged (encour=0)
wilcox.test(dta$postlet[dta$encour==1], dta$postlet[dta$encour==0])

# Do a t-test comparing the mean values of prelet and postlet
# among the children who were not encouraged (the control group)
t.test(dta$prelet[dta$encour==0], dta$postlet[dta$encour==0])

# Run a linear regression model of postlet as a function of
# encour, and then print out the results
lm1 <- lm(postlet ~ encour, data=dta)
print(summary(lm1))

# Create a new variable, delta, which is the change in letter
# scores from pre to post
dta$delta <- dta$postlet - dta$prelet

# Some code that may be useful for Problem 3, such as how to
# create a vector of numbers
# (Note that this is not crucial; you can answer Problem 3
# without using this specific code)
village <- c(1,2,3,4,5,6,7,8)
tx <- c(1,1,1,1,0,0,0,0)
baseline <- c(52,55,48,58,55,49,53,60)
oneyear <- c(44,47,44,50,53,47,51,58)
# bind vectors into matrix called "data1", then coerce "data1"
```

```
# into data frame
data1 <- cbind(village,tx,baseline,oneyear)
data1 <- as.data.frame(data1)
```

Code book with variable names for Sesame Street data

id : subject identification number

site : 1 =Three to five year old disadvantaged children from inner city areas in various parts of the country.
2 = Four year old advantaged suburban children.
3 = Advantaged rural children.
4 = Disadvantaged rural children.
5 = Disadvantaged Spanish speaking children.

sex male=1, female=2

age age in months

viewcat frequency of viewing
1=rarely watched the show
2=once or twice a week
3=three to five times a week
4=watched the show on average more than 5 times a week

setting: setting in which Sesame Street was viewed, 1=home 2=school

viewenc : treatment condition 1=child encouraged to watch, 2=child not encouraged to watch

encour: treatment condition 0=child not encouraged to watch, 1=child encouraged to watch

regular: frequency of viewing: 0=rarely watched the show, 1= watched once/week or greater

prebody : pretest on knowledge of body parts (scores range from 0-32)

prelet : pretest on letters (scores range from 0-58)

preform : pretest on forms (scores range from 0-20)

prenumb : pretest on numbers (scores range from 0-54)

prerelat : pretest on relational terms (scores range from 0-17)

preclasf : pretest on classification skills

postbody : posttest on knowledge of body parts (0-32)

postlet : posttest on letters (0-58)

postform : posttest on forms (0-20)

postnumb : posttest on numbers (0-54)

postrelat : posttest on relational terms (0-17)

postclasf: posttest on classification skills

peabody: mental age score obtained from administration of the Peabody Picture Vocabulary test as a pretest measure of vocabulary maturity