ISA R Workshop- Practice

Jonathan Rosenblatt

March 7, 2010

1. Describe in your own words the following data objects:
   (a) matrix
   (b) data frame
   (c) vector
   (d) list
2. Describe in your own words the following vector types ("modes"):
   (a) integer
   (b) numeric
   (c) character
   (d) factor
   (e) logical
3. Which data objects would you need to hold the following information? (More then one option sis possible)
   (a) The heights in your sample.
   (b) The heights and weights in your sample.
   (c) The names, heights and weights in your sample.
4. Create the following data vectors:
   (a) $(1, 5, 4, 6, 8)$. See c( )
   (b) $(11, 11, 11, 11, 11, 11, 11)$. See rep( )
   (c) $(1, 3, 5, 7, 9, 11, 13, 15)$. See seq( )
5. Download the data located at `http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/bone.data`.
   (a) How are the columns separated? Which is the appropriate value for the *sep*= argument of the *read.table* command?

(b) How many observations are in the data?
See the dim( ) command.

(c) How many variables are in the data? What is the class of each?
See the class( ) command.

(d) What is the age of the youngest male? How about the youngest female? What is their respective bone density (spnbmd)?
See subset( ), min( ) and which( ) commands.

(e) Would you say the ages of males and females are similarly distributed in the sample?
Are the means similar?
Are the distributions similar?
See mean( ), tapply( ), plot(dist( )), rug( ), hist( ), boxplot( )

(f) Test the hypothesis of equal location of the ages distribution for different genders.

i. Perform a t-test. Can Normality be assumed? Can Equal variances be assumed?
See t.test( ) and var.test( )

ii. Perform a Wilcokson test.
See wilcox.test( )

(g) Transform the age from years to days and save the transformed data in .csv format.
See getwd( ), setwd( ), write.csv

(h) Visualize the relation between age and bone density. Try three different ways to visualize the gender on this plot. Don't forget to add a title and axis labels.
See plot( ) and it's pch, cex, col parameters.
If you want a legend, have a peek at legend( ) and locator( ).

(i) Fit a simple linear regression to *bone density* using the *age* and gender as predictors (no interaction).
See lm( ), summary( ), coef( ), confint( ), plot( )

i. What are the estimated coefficients?

ii. Construct confidence intervals to coefficient estimates.

iii. Diagnose the fit: Is the variance constant? Do the residuals seem correlated?

6. Simulate the distribution of the correlation coefficient:
See runif( ), ?Control, rpois( ), cor( ), hist( )

(a) Generate 100 observations from a uniform distribution [0,1]. Store these as x.

(b) Multiply each observation by 3 , add 2 and add some Poisson disturbance with rate=1. Store these as Y.

(c) Calculate the correlation between X and Y.

(d) Repeat the process 1000 times and visualize the distribution. Is it bell shaped?

7. Think of one more good practice question and post it to our forum at `http://groups.google.com/group/israel-r-user-group` for other to enjoy.