

**Assignment #1****Due on Wednesday September 10, 2008**

**Background Information.** In the problems in this assignment you are asked to use R to perform calculations and analysis on the data provided in the first activity: *An age discrimination case?* The data are contained in the MS Excel file *Westvaco.xls*, which may be downloaded from the course webpage at

<http://pages.pomona.edu/~ajr04747>

R is an integrated suite of programs that are useful in analyzing data and producing graphical displays. It is free and may be downloaded from

<http://cran.stat.ucla.edu/>

Follow the instructions to download according to the computer platform that you have. It is available for Windows, MacOS X and Linux.

**Do** the following problems

1. In the Westvaco data for Activity #1 (*An Age Discrimination Case?*), consider the hourly payed workers involved in the second round of layoffs. The ages of the 10 workers are

25, 38, 56, 48, 55, 64, 55, 55, 33, 35,

the underlined values being the ages of the 3 workers that were laid off in that round. These values may be entered into a vector called `hourly2` by using the R command

```
hourly2 <- c(25,38,56,48,55,64,55,55,33,35)
```

The R function `c()` concatenates values in parentheses to form a vector or one-dimensional array.

Use the `mean()` function in R to compute the average age of those that were laid off in the second round and that of those that remained.

2. The R function `sample()` can be used to select random samples from a vector. For instance, the command

```
sample(hourly2,3,replace=F)
```

selects a random sample of size 3 without replacement from the values stored in the vector `hourly2`.

Selecting a random sample of size 3 simulates the decision process Westvaco went through in the second round of layoffs assuming the the process was done purely at random; in particular, assuming that the company did not consider age as a criterion for the firings.

Simulate the second round of layoffs at Westvaco, under the assumptions that the hourly workers were selected at random for layoffs, 10 times. Compute the average age of the laid off workers in each run and compare with the average age of the workers that actually were laid off. Compute the proportion of the times that average came out to be larger than or equal to 58 (the average age of the hourly workers that were actually laid off).

3. The R code contained in the file `Simulating2roundHourlyWorkers.R`, which may be downloaded from the course website, generates 1000 simulations of the firings in the 2 round of layoffs at Westvaco, assuming selections were done at random, and puts them in a vector called `Xbar`.

Use the `source()` function in R to run the code. This may be done from the “Source R code” option in the file menu or by typing

```
source('Simulating2roundHourlyWorkers.R')
```

Use the `hist(Xbar)` command in R to generate a histogram of all the average ages of the laid off workers in the simulations.

Print the histogram and attach it to your homework solutions.

4. Use the histogram obtained in the previous problem to estimate the proportion of samples that yield an average age bigger than or equal to 58 for the laid off hourly workers.
5. Modify the code in `Simulating2roundHourlyWorkers.R` to run 10,000 simulations of the lay offs under the assumption that the selection was done at random. Plot a histogram of the average ages of the 3 workers that are laid off in each run and use the the histogram to estimate the proportion of samples that yield an average age bigger than or equal to 58 for the laid off hourly workers.