Solutions to Assignment #1

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

$$25, 38, 56, 48, \underline{55}, \underline{64}, \underline{55}, 55, 33, 35,$$

   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.

   > ***Solution:*** Type `mean(c(55,64,55))` to get 58 for the average age of the laid off workers. The average age of the workers who were not laid off in the second round is obtained from
   >
   > ```
   > mean(c(25,38,56,48,55,33,35))
   > ```
   >
   > or 41.   □

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 Wesvaco, 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).

> **Solution:** Run the command
>
> ```
> mean(sample(hourly2,3,replace=F))
> ```
>
> 10 times to obtain the sample means
>
> ```
>         38
>         35
>         48
>         46
>         51
>         45
>         33
>         41
>         56
>         35
> ```
>
> None of the means is bigger than or equal to 58. Therefore the proportion of times that the average age in the laid off worker is 58 or above is 0. □

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.

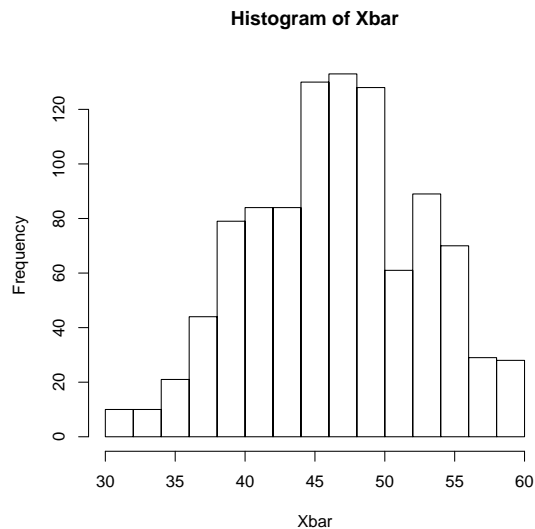   Print the histogram and attach it to your homework solutions.

Histogram of Xbar

Figure 1: Sampling distribution of the mean for 1000 repetitions

Solution: The histogram of `Xbar` is shown in Figure 1.

☐

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.

Solution: To get a better idea of how many values are 58 or above, we need to refine the histogram by introducing more breaks. This is done as follows: Define a "breaks" vectors of values from 30 to 60 in steps of 2:

```
b <- seq(30,60, by=1)
```

and then type

```
hist(Xbar, breaks = b)
```

The refined histogram is shown in Figure 2.

Inspection of the histogram shows that there are about 57 values for `Xbar` which are 58 or higher. Therefore, the proportion of samples
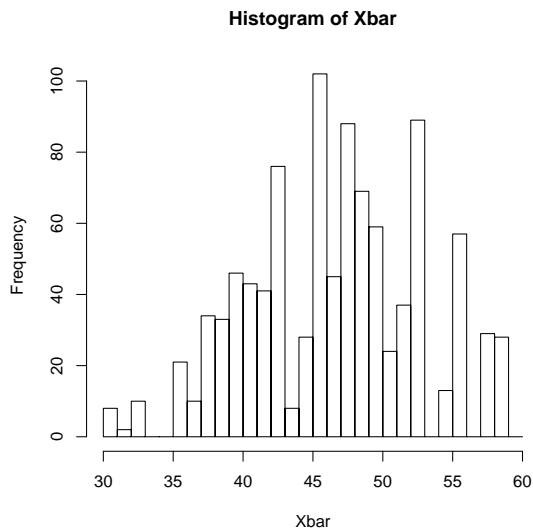
Histogram of Xbar

Figure 2: Sampling distribution of the mean for 1000 repetitions

with a mean of 58 or above is

$$\frac{57}{1000} \quad \text{or} \quad 0.057.$$

□

5. Modify the code in `Simulating2roundHourlyWorkers.R` to run 10,000 simula-
   tions 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.

   Solution: The histogram is shown in Figure 3. To get a better idea of
   how many values are 58 or above, we need to refine the histogram by
   introducing more breaks. This is done as follows: Define a "breaks"
   vectors of values from 30 to 60 in steps of 2:

   ```
   b <- seq(30,60, by=2)
   ```
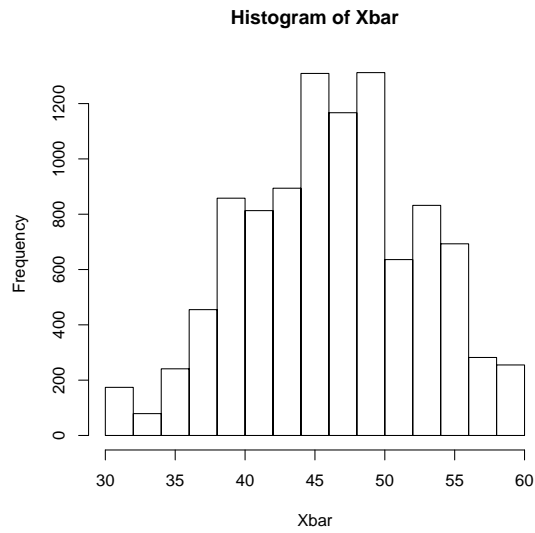
   and then type

   ```
   hist(Xbar, breaks = b)
   ```

Histogram of Xbar

Frequency

Xbar

Figure 3: Sampling distribution of the mean for 10,000 repetitions

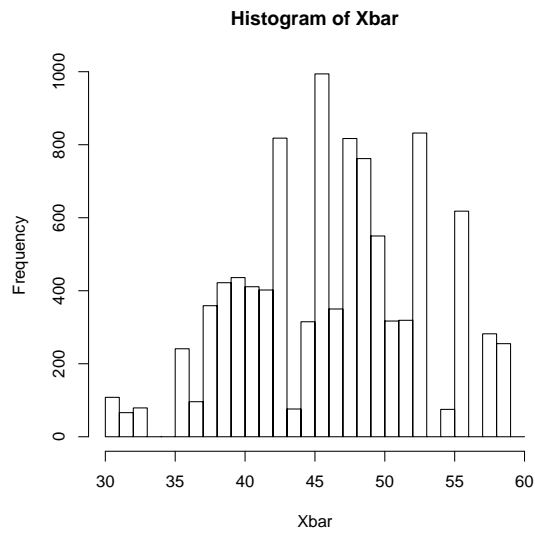**Histogram of Xbar**

Frequency

Xbar

Figure 4: Sampling distribution of the mean for 10,000 repetitions

The refined histogram is shown in Figure 4.

Examining the histogram in Figure 4, it appears that there are about 540 samples with mean of 58 or larger. The proportion of those samples is

$$\frac{540}{1000} \quad \text{or} \quad 0.054.$$

$\square$