Shepherd UNIVERSITY

Math 314: Statistics

More R Exercises

1. **Mean (Average) and Median**. You can use R to calculate the mean and median of a list of data.

   a) Enter the following R commands. Why are the mean and median equal?

   ```
   > x <- c(1,2,3)
   > mean(x)
   > median(x)
   ```

   b) Enter the following R commands. Why are the mean and median different?

   ```
   > x <- c(1,2,4)
   > mean(x)
   > median(x)
   ```

   b) Enter the following R commands a few times. What do the commands do? Why does the mean change?

   ```
   > x <- sample(1:6, 10, replace=TRUE)
   > x
   > mean(x)
   ```

2. **Histogram Example**.

   a) Run the following three commands a few times. How and why does the histogram change?

   ```
   > x <- sample(1:6, 10, replace=TRUE)
   > x
   > hist(x, probability=TRUE, breaks=seq(0.5,6.5,1), main=''Die Tosses'', xlab=''Number of Dots'', ylab=''Percent'')
   ```

   b) Increase the number of rolls from 10 to 100. How and why does the histogram change?

3. **Standard Deviation**. You can use R to calculate the standard deviation of a list of numbers.

   a) Enter the following commands then calculate the mean and standard deviation by hand.

   ```
   > x <- c(9,9,10,10,10,12)
   > mean(x)
   > sd(x)
   > sd(x)*sqrt(5/6)
   ```

   b) The built-in R command `sd` calculates what our book calls $\mathtt{SD}^+$. It is $\mathtt{SD}^+ = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \text{mean})^2}$.

The bult-in R function sd has $n-1$ in the denominator rather than n. Type the following to load a short program that loads an implementation of SD.

```
> source("http://www.adjoint-functors.net/SD.R")
> SD(x)
> x <- c(1, 3, 4, 5, 7)
> SD(x)
```

4. **Loading a Spreadsheet File**.

a) Open a web browser and use it to download the following file to your desktop.

`http://www.adjoint-functors.net/su/web/314/goldData.csv`

b) Load the file into `R` using the following command.

```
> g <- read.csv(‘‘goldData.csv’’)
```

c) Now take a look at the first few rows of the first five columns of the data. Explain what the data contains.

```
> head(g[1:5])
```

d) Take a look at the last few rows of the first five columns of the data. What does this tell you.

```
> tail(g[1:5])
```

e) Enter the following commands in `R`. What are the commands doing? What does the histogram tell you?

```
> gdollar <- g$US.dollar
> g1 <- gdollar[1:(length(gdollar)-1)]
> goldReturns <- diff(gdollar)/g1
> hist(goldReturns, probability=TRUE, main=’’Daily Gold Returns’’, ylab=’’Percent’’)
```