Computing Mean and Median Using R

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. **Bimodal Examples**. Many of the histograms we meet will look like the normal curve (bell curve). Some appear to be mixtures of two (or more) normal curves.

   a) Run the following in R. Explain what the data set is and comment on the shape of the histogram.

   ```
   > hist(faithful$eruptions, breaks=15, probability=T)
   > ?faithful
   ```

   b) How big is the data set?

   ```
   > length(faithful$eruptions)
   > dim(faithful)
   ```

   c) Calculate the mean and median of the Old Faithful eruption durations.

   d) Repeat a) – c) using `quakes$depth`.

4. **Sampling from a uniform distribution**. If a data set is uniformly distributed between -1 and 1, its mean and median are zero.

a) Take a small random sample from a uniform distribution on $[-1, 1]$. Why is the histogram not quite uniform? Why are the mean and median not exactly 0?

```
> x <- runif(n=10, min=-1, max=1)

> hist(x, breaks=10, probability=T)

> mean(x)

> median(x)
```

Repeat a) a few times to see how the sample changes.

Repeat a) using $n = 100$ and $n = 1000$. What happens?

5. **Loading a Spreadsheet File**.

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

$$\texttt{http://www.adjoint-functors.net/su/web/314/goldData.csv}$$

b) Load the file into R using the following command (you may need to specify the complete file path).

```
> g <- read.csv("http://www.adjoint-functors.net/su/web/314/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")
```

f) Compute the mean return. Comment on its sign and magnitude.

g) Compute the median return. Comment on its relationship to the mean.