(Compile date: Sat Jan 31 10:49:55 2009).

The following homework will help you get acquainted with the R statistical software package. It will seem unfamiliar and awkward at first, but stick with it, soon it will be easy. Don't get frustrated if this HW seems difficult; this is as hard as it gets. I will never expect you to memorize R commands for exams, the quick reference sheet on the website will be provided. During the second half of the course you will find R to be an extremely helpful resource.

If you get stuck with the HW, then take a look at take a look at the R help page: `http://tanbakuchi.com/Resources/R_Statistics/RBasics.html`. If that doesn't help, then send an email to me explaining the problem you are having. Be sure to copy and paste the your R work (and output with errors) into the email.

Some helpful notes:

Implicit multiplication signs Make sure you include all implicit multiplication signs. If you get either of the following errors: `syntax error` or `attempt to apply non-function`, you probably forgot to include the multiplication sign `*`. You will get an error if you type `2a` or `3(4-2)`, you should type `2*a` or `3*(4-2)`.

Order of operations be sure to enter parenthesis when needed. R observes the normal order of operations. Thus $\frac{2+6}{3}$ should be entered as `(2+6)/3`.

Powers in R use the carrot symbol, ie. 2^4 is entered as `2^4`.

Square Root To find the square root in R, use the `sqrt(x)` function, ie. $\sqrt{16}$ is entered as `sqrt(16)`.

Closing parenthesis Make sure you include closing parenthesis and quotations. Typing `sqrt((2+4)*3` won't work since the closing parenthesis for the square root function is missing. The correct expression is `sqrt((2+4)*3)` which has the closing parenthesis. If the R prompt changes from `>` to `+` it indicates you are missing a closing parenthesis or quotation. Type the closing element and hit enter. If you can't get the `>` prompt back, quit and reopen R.

Copy your work into a word document (including any plots). Ensure it is labeled with the question numbers and neat. Only include the correct work, do not include errors.

1. Use R as a calculator to verify that the following statements are true (by evaluating the left hand side to check that it is equal to the right hand side).

   (a) $12 \times 2 - 4.8 = 19.2$

   Solution:

   ```
   > 12 * 2 - 4.8
   [1] 19.2
   ```

   (b)

   $$\frac{8^3 + 2}{4} = 128.5 \tag{1}$$

Solution:

```
> (8^3 + 2)/4
[1] 128.5
```

(c) $\cos(0) = 1$

Solution:

```
> cos(0)
[1] 1
```

(d) $\sqrt{8} = 2.82842712474619$

Solution:

```
> sqrt(8)
[1] 2.828427
```

(e) $\sqrt{\frac{8+43}{5}} = 3.19374388453426$

Solution:

```
> sqrt((8 + 43)/5)
[1] 3.193744
```

2. Define the following variables in R: $a = 5$, $b = 12.3$. Use R to show that the the following statements are true. (If you want to check to see what value is stored in a variable, just type its name and hit enter.) **Don't forget to include implicit multiplication signs.**

Solution:

```
> a = 5
> b = 12.3
```

(a) $3.5a = 17.5$

Solution:

```
> 3.5 * a
[1] 17.5
```

(b) $a - b = -7.3$

Solution:

```
> a - b
[1] -7.3
```

(c) $\frac{12-5}{b} - 5.2^a = -3801.5$

Solution:

```
> (12 - 5)/b - 5.2^a
[1] -3801.471
```

(d) $(b - a)(2a - b) = -16.79$

Solution:

```
> (b - a) * (2 * a - b)
[1] -16.79
```

3. Define the vector (data set) $w = \{-5, 4, 2, 0, 3, 1, -2, 4\}$ in R. Answer the following questions. Type the following commands in R, look at the output and then write one or two *complete* sentences describing what the command did. (Be sure to include your input and output.)

   To create the vector[1] `w` you type: `w=c(-5, 4, 2, 0, 3, 1, -2, 4)`

Solution:

```
> w = c(-5, 4, 2, 0, 3, 1, -2, 4)
```

(a) `w*2`

Solution:

```
> w * 2
[1] -10   8   4   0   6   2  -4   8
```

Squares each value in `w`.

(b) `w[1]`

[1]Throughout this course we will use this method to store a set of data in a variable. Make sure you know how to do this!

Solution:

```
> w[1]
[1] -5
```

Retrieves the first element w_1.

(c) `w[2]`

Solution:

```
> w[2]
[1] 4
```

Retrieves the second element w_2.

(d) `w==4`

Solution:

```
> w == 4
[1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
```

Shows the elements in w that are equal to 4.

(e) `w>2`

Solution:

```
> w > 2
[1] FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE
```

Shows the elements in w that are greater than 2.

(f) `w[w>2]`

Solution:

```
> w[w > 2]
[1] 4 3 4
```

Retrieves all the elements in w that are greater than 2.

(g) What would you type in R to find all the values in `w` that are less than 0?

Solution:

```
> w[w < 0]
[1] -5 -2
```

4. Define the following vectors in R just as you did for `w` in the previous question:

$$y = \{65, 22, 14, 19, 20\}$$
$$z = \{8, 3, 2, 5, 7, 8\}$$

Solution:

```
> y = c(65, 22, 14, 19, 20)
> z = c(8, 3, 2, 5, 7, 8)
```

(a) To sum up all the numbers in a vector `x`, you can use the function `sum(x)`. Thus, to find the sum of all the values in `y` you would type:

```
> sum(y)
[1] 140
```

Use R to find the sum of all the values in `z`.

Solution:

```
> sum(z)
[1] 33
```

(b) The function `max(x)` returns the maximum value in a vector. Thus, to find the maximum value in `z` you would type:

```
> max(z)
[1] 8
```

Use R to find the maximum value in `y`.

Solution:

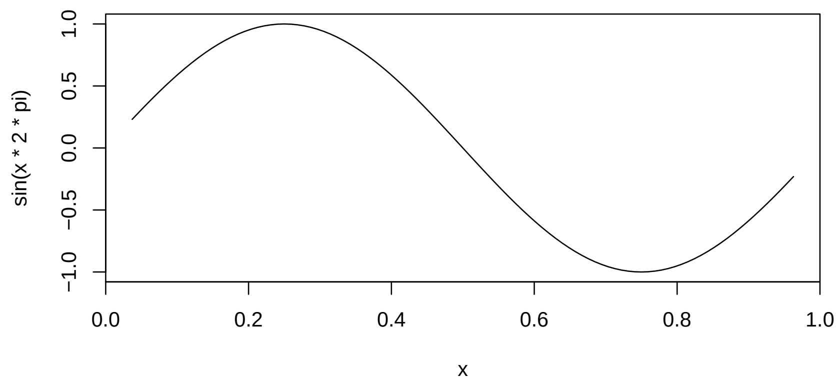
```
> max(y)
[1] 65
```

5. R is capable of making many types of graphs. We can use R's `curve` function to plot polynomials.

(a) Type in the following command: `curve(sin(x*2*pi))`
What function did R plot?

Solution:

*> curve(sin(x * 2 * pi))*



R plotted the sine function. $f(x) = \sin(x2\pi)$

(b) What is the range of x values plotted for the previous graph you made?

Solution: $(0, 1)$

(c) Now type in: `curve(sin(x*2*pi), xlim=c(-2, 2))`
We now have added an optional argument to the function which changes the default behavior. What is the new range of x values plotted on the graph?

Solution: $(-2, 2)$

(d) What is the default range of x values plotted for the curve function?

Solution: $(0, 1)$

(e) What does the optional argument `xlim` do?

Solution: Changes the range of x values plotted.

(f) What would you type into R to make the the above graph have a x range of $(0, 5)$?

Solution: `curve(sin(x*2*pi), xlim=c(0, 5))`

(g) Type the following command: `curve(x^3, xlim=c(-10, 10), main="Polynomial")`
This time we are graphing $f(x) = x^3$. What does the optional argument `main` do?

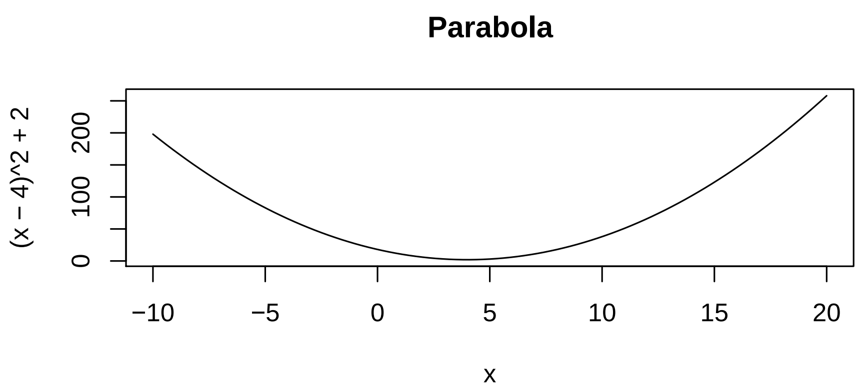
Solution: Sets the main title of the graph.

6. Use the `curve()` function in R to plot the following function over the domain $(-10, 20)$. Set the title of the plot to "Parabola". (Be sure to copy and paste your plot into the HW.)

$$f(x) = (x - 4)^2 + 20 \tag{2}$$

Solution:

```
> curve((x - 4)^2 + 2, xlim = c(-10, 20), main = "Parabola")
```



7. Load the book data into R (download the .RData file on the website under the R resources and double click on it). This will load a bunch of data tables.

   (a) One of the data tables is named `MM`. This table contains information on the weights and colors of M&M's observed in a study. Type `MM` and hit enter. This will display the data in the table. What are the column names (you may have to scroll up)?

   Solution:

   ```
   > names(MM)

   [1] "WEIGHT" "COLOR"
   ```

   (b) An easier way to determine the names of the columns is to use the `names()` function. Now type: `names(MM)`. What did this do?

   Solution: Listed the names of the columns in the `MM` table.

   (c) Type `MM$WEIGHT`. What did this do?

Solution: Retrieved the weight of the M&M data.

(d) Now find the mean weight of the M&M's using the above statement and the same method we used previously to find the mean of a vector.

Solution:

```
> mean(MM$WEIGHT)

[1] 0.85649
```

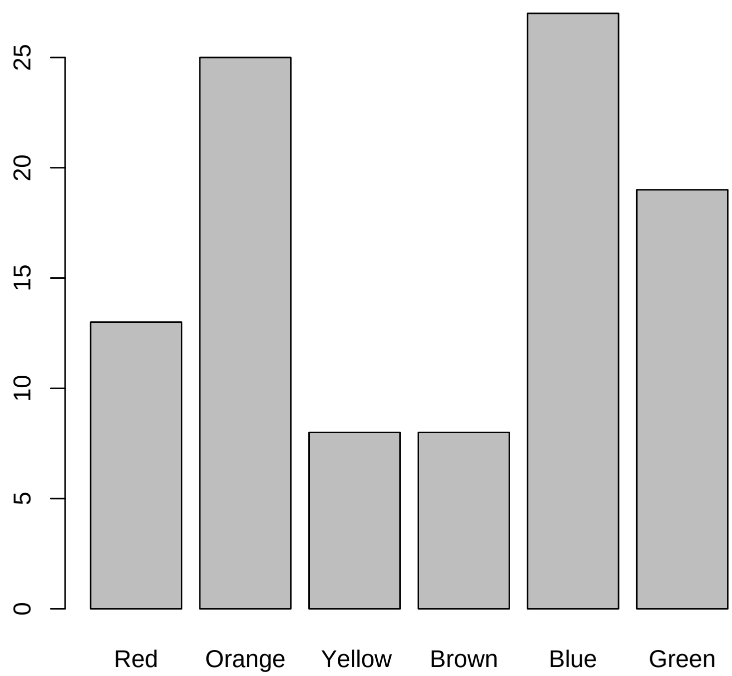
(e) Now make a histogram plot of the M&M weights by typing: `hist(MM$WEIGHT)`
Hopefully now you can see how R is able to do allot of work with just a little typing. Yes, the trivial calculations can seem tedious, but more complex calculations and plots are made easily!

```
> hist(MM$WEIGHT)
```

(f) Type the following: `plot(MM$COLOR)`
Which color of M&M were observed the most in the study?

Solution:

```
> plot(MM$COLOR)
```



The blue colored M&M's were observed most in the study.

(g) Type the following: `summary(MM)`
What does the above command do?

Solution: It summarizes each column.

```
> summary(MM)

     WEIGHT          COLOR
 Min.   :0.6960   Red   :13
 1st Qu.:0.8287   Orange:25
 Median :0.8580   Yellow: 8
 Mean   :0.8565   Brown : 8
 3rd Qu.:0.8810   Blue  :27
 Max.   :1.0150   Green :19
```

(h) Now find the mean weight of the blue M&M's by typing

```
blue=MM$WEIGHT[MM$COLOR=="Blue"]
mean(blue)
```

Solution:

```
> blue = MM$WEIGHT[MM$COLOR == "Blue"]
> mean(blue)

[1] 0.856037
```

(i) Next find the mean weight of the green M&M's by modifying what you did in the previous problem.

Solution:

```
> green = MM$WEIGHT[MM$COLOR == "Green"]
> mean(green)

[1] 0.8635263
```