36-401 Homework #1a (Due Sept. 6, 2001) REVISED Solutions

1. **[Exercise 1, "Introduction to Splus" p.8] Check the arguments for t.test(). Check the help for t.test(). Perform a 2 sample t-test for d1_c(1, 3, 4, 5, 4, 3, 1), d2_2:8. Repeat, dropping the assumption that the variances are equal.**

   The arguments for t.test() are x, y, alternative, mu, paired, var.equal, conf.level, VALUE, statistic, parameters, p.value, conf.int, estimate, null.value, alternative, method, and data.name (see help on t.test for details).

   A two-sample t-test for d1 and d2 (assuming equal variances) results:
   > t.test(d1, d2, alternative="two.sided", var.equal=T)

   Standard Two-Sample t-Test

   data: d1 and d2
   t = -2, df = 12, p-value = 0.0687
   alternative hypothesis: true difference in means is not equal to 0
   95 percent confidence interval:
   -4.1788128 0.1788128
   sample estimates:

mean of x	mean of y
3	5

   Without the assumption of equal variances, we get:
   > t.test(d1, d2, alternative="two.sided", var.equal=F)

   Welch Modified Two-Sample t-Test

   data: d1 and d2
   t = -2, df = 10.8, p-value = 0.0713
   alternative hypothesis: true difference in means is not equal to 0
   95 percent confidence interval:
   -4.2059682 0.2059682
   sample estimates:

mean of x	mean of y
3	5

2. **[Exercise 3, "Introduction to Splus" p.8-9] Make a plot of x=1,2,3,4,5 vs. y=1.2, 2.3, 3.1, 4.8, 6.2 using plot(x,y). Pull up the searchable help system using help.start(). Redo the plot with a title, and some x and y labels. Use lines() to overlay connecting lines**

between the symbols. Add your name as a subtitle.

> x_1:5
> y_c(1.2, 2.3, 3.1, 4.8, 6.2)
> plot(x,y)

Figure 1 appears as a result.

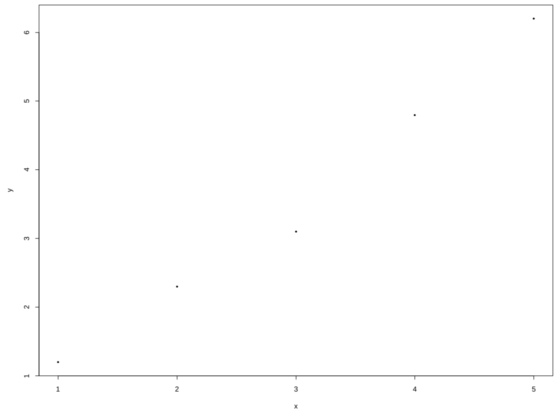

Figure 1: Initial Scatterplot

To add a title, subtitle, and x- and y- axis labels, type

> title(main="Main Title", sub="Kimberly Sellers", xlab="x-axis label", ylab="y-axis label", axes=T)
Figure 2 displays the new graph.

Finally, to connect the points, type:
> lines(x,y,type="l")
The result is Figure 3.

3. **[Exercise 4, "Introduction to Splus" p.9] Generate 100 random numbers from a Gaussian distribution with mean = 10 and sd=2 using "rn_rnorm(100,10,2)". Plot a histogram of them. Try changing the number of bars in the histogram. Use par(mfrow=c(3,2), oma=c(1,1,3,1)) to divide your plot area up into 3 rows of 2 pots each and save space for an overall title on top. Plot 6 different sets of 100 random N(10,4) points using symplot() to get a feel for how far true normal data of size 100 varies from completely**

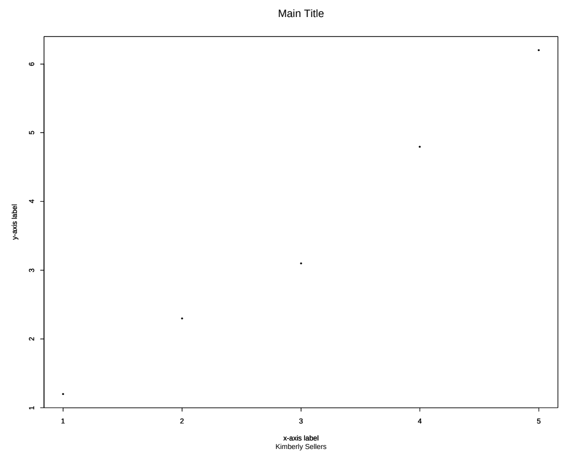

Figure 2: Scatterplot including titles and axes

symmetrical. Use "mtext("The overall title", cex=1.2, outer=T)" to get an overall title.

```
> rn1_rnorm(100,10,2)
> rn2_rnorm(100,10,2)
> rn3_rnorm(100,10,2)
> rn4_rnorm(100,10,2)
> rn5_rnorm(100,10,2)
> rn6_rnorm(100,10,2)
> rn_data.frame(rn1, rn2, rn3, rn4, rn5, rn6)
> motif()
> par(mfrow=c(3,2), oma=c(1,1,3,1))
> for (i in names(rn)) symplot(get(i))
> mtext("The overall title", cex=1.2, outer=T)
> source("ps.q")
> ps("symplots")
Generated postscript file "symplots.ps".
```

The output is Figure 4.

4. **[Exercise 6, "Introduction to Splus" p.9] Write your own function to describe a vector. It should print the mean, median, sd, min, and max. You can use commands**

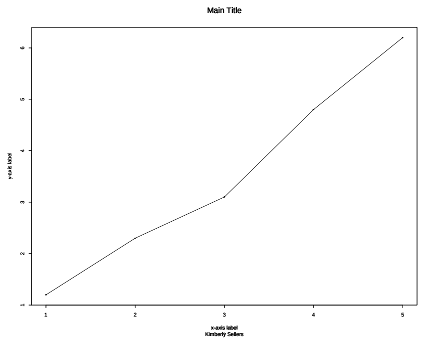

Figure 3: Line segments connecting scatter points

of the form: cat("The mean is", mean(x), "\n"). Add an optional second argument "plotit=F", which will make a histogram if you use the form "describe(conc$educat, T)". Use your function to describe conc$educat.

My function, titled "createfunc.q"

```
output_function(x, plotit)

{ cat("The mean is ", mean(x), "\n",
"The median is ", median(x),"\n",
"The standard deviation is ", sd(x), "\n",
"The minimum is ", min(x), "\n",
"The maximum is ", max(x), "\n")
if (plotit==T) hist(x)
}
```

In S-plus:

```
> motif()
> source("createfunc2.q")
> conc_read.table("concord1.dat", header=T)
> output(conc$educat, plotit=F)
The mean is 14.0040322580645
The median is 13
The standard deviation is 3.09054992834016
The minimum is 6
The maximum is 20
NULL
```

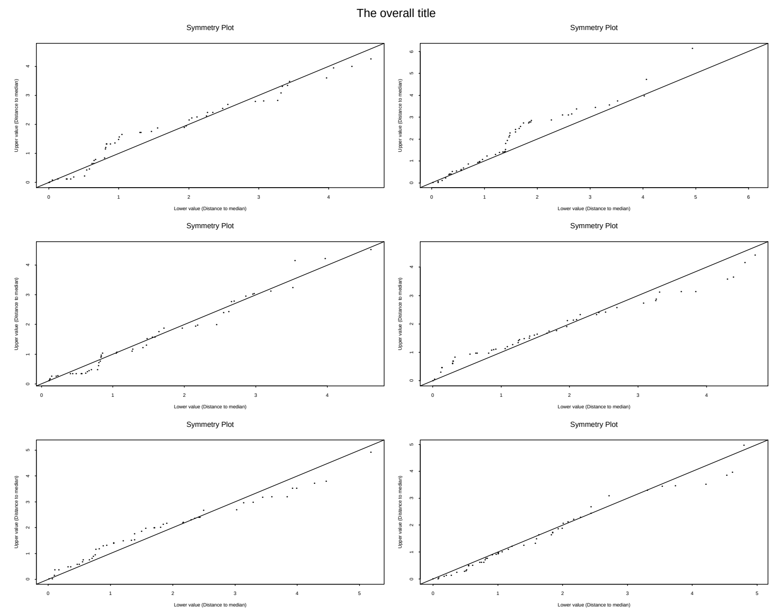

Figure 4: Symmetry plots for six randomly generated number sets, each containing 100 numbers

```
> output(conc$educat, plotit=T)
The mean is 14.0040322580645
The median is 13
The standard deviation is 3.09054992834016
The minimum is 6
The maximum is 20
```

```
[1]   5   7   9   11   13   15   17   19
```

Figure 5 displays the histogram.

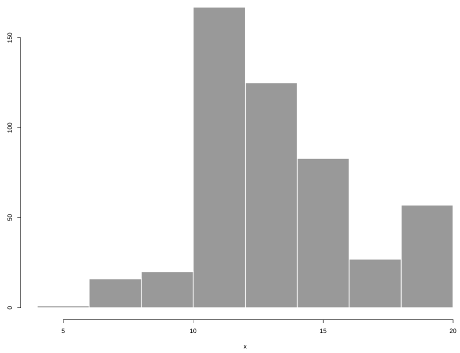

Figure 5: Histogram of education from Concord Water Study dataset