36-711 HW01: Due Friday September 6, 2002
Treating Depression: A Randomized Clinical Trial

The data set `depression.dat`, available in the data area of the 711 website, contains data from a multicenter clinical trial to study treatment of depression by various drugs. We will use this data set to explore working with Splus/R.

The study design was as follows:

- There were 5 medical clinics across the country that participated in this trial. Using many clinics enables the investigators to enroll many more patients into the study and allows for a diversity of patients to participate.
- There were 3 treatment groups. Patients received either Imipramine (Imip), Lithium (Li), or a Placebo (Pl). Imip and Li are active drugs. Patients were randomized to one of the 3 treatment groups, using a random device (like rolling a 3 sided die).
- Patients were followed from 2-4 years to see whether or not they had a recurrence of depression. If they did not have a recurrence within this time frame, then their treatment was considered a Success. If they did have a recurrence, it was considered a Failure. The study was double-blinded.
- A number of additional background variables were measured for each patient.

The variables in the data set are:

HOSPT	Which hospital: 1, 2, 3, 5 or 6.
TREAT	0=*Lithium*; 1=*Imipramine*; 2=*Placebo*.
OUTCOME	0=*Success* 1=*Failure* (recurrence of depression)
TIME	number of weeks until a recurrence (if OUTCOME=1) or until study ended (if OUTCOME=0)
ACUTET	How many days the patient was depressed before the start of the study
AGE	Age in years
GENDER	1=*Female* 2=*Male*.

Read the data into Splus using a command like

```
> depr <- read.table("depression.dat",header=T)
```

and then do the exercises starting on the next page of this hw. (Remember to use `help()` and `help.start()` often. Also, remember that your data and functions are saved from one session to the next in Splus.)

Note: *For this assignment, please turn in a paper copy of your Splus work, annotated to indicate what problem you are working on, answer any questions in text that need answering, etc. Attach paper copies of graphs at the end of your hw. Example of "Annotated Splus output":*

```
Brian Junker
36-711 Statistical Computing
HW 1 -- Due Sep 6 2002

READING IN THE DATA

> depr <- read.table("depression.dat",header=T)

PROBLEM #1:

> names(depr)
[1] "Hospt"  "Treat"  "Outcome"  "Time"  "AcuteT"  "Age"  "Gender"
> dim(depr)
[1] 109   7
The result of the ''dim'' function shows that my ''depr'' data frame
has 109 rows (observations) and 7 columns (variables), as expected.
The ''names'' command shows the names of the columns (variables) and
they are what I expected.

PROBLEM #2:

...
```

You can prepare this file by cutting and pasting from your Splus window into emacs, and adding annotations in emacs. Use "enscript" to print it out.

1. **[Not to be graded.]** Use the `names()`. `dimnames()` and `dim()` functions in Splus to verify the following:
   - There are 109 observations and 7 variables.
   - The variable names are "Hospt", "Treat", "Outcome", "Time", "AcuteT", "Age", and "Gender".
2. Use `summary()` to produce a summary of each variable in the data set.
   - You can also use `sqrt(var())` to produce standard deviations if you wish.
   - You can do each summary by hand, e.g. `summary(depr$Hospt)`, or `summary(depr$Treat)`, etc., or you can experiment with a "for" loop:

     ```
     for (i in names(depr)) {
        print(summary(depr[,i]))
     }
     ```

3. For "Time", "AcuteT", and "Age" the summaries you just produced make sense. For the other variables they do not, really.

   (a) Why not?

   (b) For the other five variables try `table()` or `summary(as.factor())`. What do you learn from these summaries?

4. Use `search()` to find out what databases you have access to in your Splus session. Use `attach()` to attach the depression data.frame as a data base. Use `search()` again to see where the depression data is now.

5. Produce stem and leaf plots of the three continuous variables, using `stem()`. Since you have attached the depression data, you won't have to type so much.

6. Use `motif()` (or in R use `x11()`) to open a graphics window. Use `par(mfrow=c(2,2))` to divide the window into four subgraphs. Make histograms of each of the continuous variables, using `hist()` and `title()` once for each graph (use the title to give a short description of each variable, e.g. "Time" is "Time to recurrence or study end", etc.). Print the result (three histograms on one page) out, to be turned in [use the Graph...Print menu in the graphics window; you may have to set print options with the Options menu there first].

7. How are treatments distributed across hospitals? use `table(Hospt, Treat)` to find out. Now try `table(Hospt,Outcome,Treat)`. Do some treatments work better at some hospitals?

   This may require some side calculation. A quick way to do the side calculation is

   ```
   > bozo _ table(Hospt,Outcome,Treat)
   > for (i in 1:3) bozo[,,i] _ bozo[,,i]/apply(bozo[,,i],1,sum)
   > round(bozo,2)
   ```

   Play around with the pieces of this (and consult `help()`*, etc.) and describe what each piece of this calculation does.*

8. Are the times to recurrence of depression the same across the hospitals? Use `par(mfrow=c(1,1))` to use the whole graphics window for a single plot, and then make parallel boxplots of `Time` using either

   - `plot(as.factor(Hospt),Time,xlab="Hospital",ylab="Time")` or
   - `boxplot(split(Time,Hospt),xlab="Hospital",ylab="Time")`

   Print the graph and turn it in. *What other relations would be interesting to explore this way?*

9. It might be interesting to explore the relationship between `AcuteT` and `Time` at each hospital. For example, `plot(AcuteT[Hospt==1],Time[Hospt==1])` would show this for Hospital 1. Make a single plot with six subplots showing this relationship for each of the (five!) hospitals in the data set, print it out, and turn it in. *What other relations would be interesting to explore this way?*