MYSTAT Workshop NOTES

BY

Dr. Hans Zuuring
School of Forestry
University of Montana
Missoula, MT 59812
USA

AT

PAKISTAN FORESTRY INSTITUTE
PESHAWAR, PAKISTAN

July 1992

Table of Contents

Brief Software Loading Instructions

MYSTAT

For 2 floppy disk drive system:

1. Create a formatted bootable (system tracks present) work diskette (our data sets will be stored here later) containing the following files: COMMAND.COM, CONFIG.SYS, ANSI.SYS and DATA.DEF (consult your DOS manual for instructions on how to format a diskette)
   N.B.: make sure that the following DOS commands appear in your CONFIG.SYS file:

   ```
   files=15
   buffers=20
   device=c:\DIRNAME\ansi.sys
   ```

   where DIRNAME is the name of your DOS subdirectory.

2. Insert the above diskette in drive A and the diskette containing the MYSTAT diskette in drive B. Then close both drive doors.

3. Turn on the microcomputer and its printer if present.

4. When the system prompt appears on the screen (A:\>) type: B:MYSTAT <Enter>.

5. The MYSTAT logo will appear and you are ready to go. Use the HELP command to get information about various commands. Remember that every command must be followed by pressing the <Enter> key.

For HARD DISK system (microcomputer lab):

1. Turn on the microcomputer and its printer if present. (Don't place any diskette in drive A: at this time)

2. When the system prompt appears on the screen (C:\>) do either of the following steps:

   a) if the MYSTAT files have already been loaded onto the hard disk go to step 3.

   b) if the MYSTAT files have not been loaded:
      i) create a subdirectory called MYSTAT using the DOS command => MD MYSTAT <Enter>.

ii) move to this directory with the DOS command => **CD\MYSTAT** <Enter>.
iii) insert the MYSTAT diskette in drive A and copy all the files into the hard disk subdirectory using the DOS command => **COPY A:*.*** <Enter>.
iv) when all the MYSTAT files have been copied to the hard disk, remove the last diskette from drive A:

3. Insert the work diskette in drive A: and type **MYSTAT** <Enter>.

4. The MYSTAT logo will appear on the screen and you are ready to go.

5. To exit from MYSTAT type **QUIT** <Enter> and after listing a summary of all commands issued while in a MYSTAT session the DOS prompt (C:\>) will reappear.

Data File Building

Data files must be created by invoking a slow full screen text editor by typing the MYSTAT command **EDIT** <Enter>. Once the blank spreadsheet appears you can toggle between the body of the spreadsheet and the MYSTAT command line at the bottom of the spreadsheet by pressing the <Esc> key. Then you can type **HELP** <Enter> and a screen of valid EDIT commands appears. Each column in the spreadsheet must be labelled with field name up to 8 letters long. This is done by typing a single quote (') followed by the label name and the <Enter> key. You exit the column heading row by using the left and down arrow keys. Now you can enter your data. When through entering your data press the <Esc> key to exit the spreadsheet and then use the MYSTAT SAVE command to store the data in a MYSTAT file which will have a .SYS extension (type only the file name and not the extension). Now type **QUIT** <Enter> to exit from the MYSTAT editor and return to the main menu.

MYSTAT Commands

MYSTAT recognizes only the first three letters of commands and does nor care whether they are typed in lower or upper case. There are two types of commands: HOT and COLD. A HOT command is executed as soon as the <Enter> key is pressed while a COLD command selects an option which will not be executed until the next HOT command. The order of entry of COLD commands is not important. COLD commands remain in force until they are changed.

Summary of Valid MYSTAT Commands

Command	- - DESCRIPTION - -
DEMO	demonstration of MYSTAT functionality
HELP	information about MYSTAT commands
SYSTAT	information about "mother" statistical analysis package
QUIT	exit MYSTAT
EDIT	create, edit MYSTAT file and compute transformations
USE	open an existing SYSTAT data file
SAVE	save data in MYSTAT file
PUT	save data in ASCII file
SUBMIT	execute group of MYSTAT commands in file
OUTPUT	direct output to screen, disk or printer
MENU	turn MYSTAT menu "on" or "off"
NAMES	display column labels in MYSTAT data file
LIST	display data on screen by case
FORMAT	set # of decimals to print (1 - 9)
NOTE	making comments on output
SORT	sort data in MYSTAT file by user key
RANK	replace observations with ranks
WEIGHT	assign variable weights to observations
PLOT	create a two-way scatter diagram
BOX	create a box and whiskers plot
HISTOGRAM	create a bar chart
STEM	create a stem-and-leaf diagram
TPLOT	create a time series plot
CHARSET	specify a graphics character set
STATS	calculate descriptive statistics
TABULATE	frequency table, chi-square test
TTEST	independent/dependent 2-sample t-tests
PEARSON	calculate correlation coefficients
SIGN	compute a nonparametric sign test
WILCOXON	compute a nonparametric ranks test
FRIEDMAN	perform a nonparametric 2-way ANOVA
MODEL	specify a regression/ANOVA model
CATEGORY	define factors and # of levels/factor for ANOVA
ANOVA	specify dependent variable for ANOVA
COVARIATE	specify covariates for ANOVA
ESTIMATE	estimate the model coefficients

Summary of Valid MYSTAT Editor Commands

Cursor commands (alternative keys) are:

Left arrow (Cntl-S)	Right arrow (Cntl-D)
Down arrow (Cntl-X)	Up arrow (Cntl-E)
Ins (page left, Cntl-A)	Del (page right, Cntl-F)
PgDn (Cntl-C)	PgUp (Cntl-R)
Home (first cell,Cntl-W)	End (last case,Cntl-Z)

EDIT commands are:

Esc (Cntl-Q)	toggles between data window/command line
USE <file name>	fill worksheet with data in MYSTAT file
GET <file name>	fill worksheet with data in ASCII file
SAVE <file name>	saves worksheet values into a MYSTAT file
FIND <expression>	moves cursor to selected case
FORMAT <#>	sets number of decimals displayed
LET <var>=<expression>	transforms or creates variables
IF <expression> THEN LET <var>=<expression>	conditional transformation
DELETE <case>	deletes one or more cases
DROP <variable>	deletes one or more variables
REPEAT <#>	fill worksheet to <#> of cases
HELP <command>	get information about command
NEW	clears the worksheet
QUIT	exit the editor

LET command transforms existing variables or creates new ones.

```
LET <var> = <expression>
```

where <expression> may contain variables and constants with

Numeric operators:	+ - * / ^
Relational operators:	< <= <> == >= >
Logical operators:	AND OR
Functions:	SQR LOG EXP ABS
	SIN COS TAN ASN ACS ATN ATH
	ZCF (cumulative normal)
	ZIF (inverse normal)
Random numbers:	URN ZRN
Case numbers:	CASE

```
LET AVERAGE = (TRIAL1+TRIAL2+TRIAL3)/3
LET SPEED = SQR(SPEED)
LET INCOME = LOG(INCOME)
LET TRENDY = (INCOME>100000) AND (STATE$='California')
```

Sample Problems
and their Solutions

PART I -- Simple Analyses

Q1. Given the following tree volumes in cubic feet per 1/5-acre

450, 410, 370, 405, 395, 400, 360, 440, 405, 420, 295, 340, 330, 380, 365, 325, 300

calculate the mean ($\bar{x}$), standard deviation (s) and standard error of the mean.

A1. First the data must be entered into the computer in a form that can be used by MYSTAT. You have basically <u>two</u> choices:
a) use the MYSTAT spreadsheet editor or
b) construct a DOS text file using any screen editor of your choice (you can even use WordPerfect, a word processor package and save the data file as a DOS text file).
c) extract data files from either Lotus 1-2-3 v2.01 or dBASE III+.

<u>Using the MYSTAT editor</u>:

At the DOS prompt type **MYSTAT** <Enter> to start the MYSTAT program. A copyright notice will appear and follow the instructions at the bottom of the screen, namely press the <Enter> key to display the MYSTAT main menu. This main menu is logically arranged by function and during the course of these exercises you will become familiar with most of them. To invoke the MYSTAT editor type **EDIT** <Enter> and the screen returns after a few moments with a spreadsheet having rows labelled with case numbers and as yet unlabelled columns. You can "toggle" back and forth between the body of the spreadsheet and the command line at the bottom of the screen by pressing the <Esc> key.
The idea is to create labelled columns of numbers. The column labels are constructed by typing unique names using up to eight (8) letters and starting with a single quote ('). Press the <Enter> key after each column name and the highlighted bar will automatically jump to the next column field. When you are finished entering column labels you are ready to enter data values. At this point use the **Arrow** keys to move the highlighted box to the leftmost spreadsheet position and then go <u>down</u> one position. Data values are entered either columnwise or rowwise. To enter data columnwise press the down arrow key after each value. To enter data rowwise press the <Enter> key after each value.
Enter <u>all</u> the above data values in one column labelled VOLUME.

When all the values have been entered press the <Esc> key followed by the command **SAVE filename** <Enter> and a message appears telling you that the records have been saved on the computer's hard disk. Now type **QUIT** <Enter> to exit from the MYSTAT editor and return to the MYSTAT main menu.

Using a screen text editor:

Prior to starting the MYSTAT program we could use any screen editor such as WordPerfect provided we remember three things:

a) save the data file as a DOS text file with a .DAT extension
b) separate multiple column data values with spaces or commas entering them in rowwise fashion, i.e.

1,25,6		1 25 6
2,12,14	OR	2 12 14
3,5,3		3 5 3

c) no column labels are typed.

Note: MYSTAT will accept up to a maximum of 50 fields (variables or columns) and 32,000 records (rows)

At the DOS prompt type **MYSTAT** <Enter> to start the MYSTAT program. A copyright notice will appear and follow the instructions at the bottom of the screen, namely press the <Enter> key to display the MYSTAT main menu. Type **EDIT** <Enter> to invoke the MYSTAT editor and when the spreadsheet appears use the up arrow key to position the highlighted bar in the column label fields and enter the same number of labels as there are data values in each record of your external DOS text file. When completed press the <Esc> key to return to the command line and type: **GET filename** <Enter> where filename = DOS text file name. In a few minutes these data values will be "read into" the MYSTAT spreadsheet and you can then proceed to save this MYSTAT data file by typing **SAVE filename** <Enter> followed by **QUIT** <Enter> to exit the MYSTAT editor.

Extracted data files:

Data can be extracted from Lotus 1-2-3 v2.01 by translating the file into a dBASE III+ file provided there is only one row designated for field (column) labels. Once the data is in a file with the .DBF extension you can start the dBASE III+ program, select the previously created database file, press the <Esc> key to exit the ASSIST mode and invoke the dot (.) prompt. Now type **COPY TO filename.DAT SDF** <Enter> to create an undelimited DOS text (ASCII) file. Press the <F2> key to return to the ASSIST mode and quit dBASE.

From here on the same instructions with respect to MYSTAT apply as written under the previous section.

Before any analyses can be performed by MYSTAT you must indicate which data file is to be "used". This is done by typing **USE filename** <Enter> where filename = a DOS filename up to eight characters long with no embedded blanks or special symbols or characters. After typing the USE command the screen goes blank and returns with a listing of the existing column label names and the user is prompted to press the <Enter> key when ready to return to the MYSTAT main menu. There is extensive on-line help which can be obtained by typing **HELP** <Enter> and a list of valid MYSTAT commands is presented. Type **HELP STATS** <Enter> to obtain help on the syntax of the STATS command. Press the <Enter> key to return to the MYSTAT main menu. Type **STATS VOLUME/ MEAN,SD,SEM** <Enter> and a table of descriptive statistics appears on the screen in a few seconds.

Note: the / is a qualifier which modifies the STATS command so that only the requested summary statistics are computed and not the standard ones -- minimum, maximum, mean, standard deviation. You can also calculate these statistics by some grouping variable and typing the phrase **BY field name** before pressing the <Enter> key in the above command string.

Q2. Calculate the same summary statistics as Q1 given that the first ten (10) observations belong to site 1 and the remainder to site 2.

A2. Let us add another column to our data file and label it SITE. Type **EDIT** <Enter> and the spreadsheet returns after the data from the currently specified data file has been read. Use the arrow keys to move the highlighted box one position up and one to the right and type **the label 'SITE** <Enter>. Use the arrow keys again to move the highlighted box below this label and type **1** followed by the down arrow key. Do this ten (10) times. Continue for the remaining seven (7) rows using the number **2**. Press the <Esc> key and type **SAVE old filename** <Enter> and a message appears asking you whether you want to overwrite the existing data file (Y or N). Type **Y** and the old MYSTAT data file is updated. Type **QUIT** <Enter> to exit from the MYSTAT editor and return to the MYSTAT main menu. Type **USE filename** <Enter> to access the updated data file followed by the <Enter> key to return to the main menu. Type **STATS VOLUME/MEAN,SD,SEM BY SITE** <Enter> and a table of descriptive statistics appears on the screen for each site.

Q3. Use hypothesis testing procedures to determine if the mean volumes are different due to site with 95% confidence.

A3. Type **HELP TTEST** <Enter> to determine the proper syntax of this command. The correct form is example three. Do you know what needs to be typed in this case? Press the <Enter> key to return to the MYSTAT main menu and type:
TTEST VOLUME*SITE <Enter> and the screen displays the results in a few seconds.
We observe that the calculated t-value is larger than that expected by random chance since the calculated PROB value is less than 0.05, the specified alpha (α) level. Therefore, we reject H_0 and accept H_1. Do you know what the hypotheses are?

Q4. Draw Box plots displaying the median and hinges of VOLUME by SITE and interpret the results.

A4. Type **HELP BOX** <Enter> to determine the proper syntax of this command. Do you know what needs to be typed in this case? Press the <Enter> key to return to the MYSTAT main menu and type:
BOX VOLUME*SITE <Enter> and the screen displays the results in a few seconds.

PART II -- More Complicated Analyses

Q5. A wildlife biologist wants to investigate the effects of sex and diet on the mean calcium concentration of blood taken from birds of a single species population. Ten (10) female and ten (10) male birds were captured and five of each sex were given a plain diet while the remaining birds were given a diet containing a growth hormone. After a suitable time period blood samples were taken from all twenty birds and the plasma calcium concentration was recorded in mg/100 ml. Given the data below:

Regular Diet		Diet with hormone	
Female	Male	Female	Male
16.5	14.5	39.1	32.0
18.4	11.0	26.2	23.8
12.7	10.8	21.3	28.8
14.0	14.3	35.8	25.0
12.8	10.0	40.2	29.3
74.4	60.6	162.6	138.9

Construct an analysis of variance table with all the proper headings and labels (What are the underlying assumptions associated with any ANOVA?).

A5. Start MYSTAT and invoke the editor. Create <u>three</u> column labels as follows: **'Calcium** <Enter> **'Diet** <Enter> **'Sex** <Enter> and then use the arrow keys to position the highlighted box at case 1 under the CALCIUM label. Enter the above data values columnwise using the down arrow key after each entry. Next press the <Esc> key and type the following commands:

IF CASE<11 THEN LET DIET=1 <Enter> and wait for MYSTAT to perform the transformations.
IF CASE>10 THEN LET DIET=2 <Enter> and wait for MYSTAT to perform the transformations.
LET SEX=1 <Enter> and wait for MYSTAT to perform the task.
IF (CASE>5 AND CASE<11) OR CASE>15 THEN LET SEX=2 <Enter> and wait for MYSTAT to perform the transformations.

<u>Note</u>: you are creating two classification (dummy) variables representing the various levels of the factors Diet and Sex. They are necessary to perform an ANOVA. Diet represents the leftmost (slowest moving) subscript while Sex represents the rightmost (fastest moving) subscript of y_{ijk}, the calcium response for the i<u>th</u> diet, j<u>th</u> sex and k<u>th</u> replicate.

Save the data values by typing: **SAVE BIRDS** <Enter>
and exit the MYSTAT editor by typing **QUIT** <Enter>. Again inform the MYSTAT processor which data file you wish to "use" by typing **USE BIRDS** <Enter>. Type **HELP ANOVA** <Enter> to determine the proper syntax of this command and the others that are associated with an analysis of variance (ANOVA). Do you know what needs to be typed in this case? Press the <Enter> key to return to the MYSTAT main menu and type the following three commands:

CATEGORY DIET=2,SEX=2 <Enter>
ANOVA CALCIUM <Enter>
ESTIMATE <Enter>

Note: most MYSTAT commands can be abbreviated to the first two or three letters, i.e. CAT and EST.
In a few seconds the screen displays an ANOVA table with the various F-values and associated probabilities.

Q6. Use hypothesis testing procedures to determine whether bird sex, diet, and their interaction affect mean blood plasma calcium concentrations with 95% confidence.

A6. Hypotheses:

(1) $H_0 : \mu_A = \mu_B$ versus $H_1 : \mu_A$ not $= \mu_B$

(2) $H_0 : \mu_M = \mu_F$ versus $H_1 : \mu_M$ not $= \mu_F$

(3) $H_0 : \mu_{DxS} = 0$ versus $H_1 : \mu_{DxS}$ not $= 0$

ANOVA table

Source	SS	DF	MS	F-ratio	
Treat	1,461.33	3	487.11	21.27	
Diet	1,386.11	1	1,386.11	60.53	(1)
Sex	70.31	1	70.31	3.07	(2)
D x S	4.91	1	4.91	0.21	(3)
Error	366.37	16	22.90		
Total	1,827.70	19			

Rejection Rules:

(1) Reject H_0 if $F_c = 60.53 > F_{1,16,0.05} = 4.49$

(2) Reject H_0 if $F_c = 3.07 > F_{1,16,0.05} = 4.49$

(3) Reject H_0 if $F_c = 0.21 > F_{1,16,0.05} = 4.49$

Conclusions:

(1) An examination of the PROB value (= .000) reveals that H_0 should be rejected and H_1 should be accepted and we conclude that there is a difference in mean calcium concentration of the blood due to diet with 95% confidence, and

(2) An examination of this PROB value (= .099) indicates that we fail to reject H_0 and conclude there is no evidence that there is any difference in mean calcium concentration of the blood due to sex with 95% confidence.

(3) An examination of this PROB value (= .650) indicates that we fail to reject H_0 and conclude there is no evidence that there is any difference in mean calcium concentration of the blood due to a Diet by Sex interaction with 95% confidence.

Q7. A survey was conducted among villagers who lived at two different locations regarding their attitudes about range land management activities taking place on local communal areas. A number of questions were asked about the management of these lands, three of which are presented here, as well as some socioeconomic data such as family size and income (in thousands of Rs). Given the data below:

LOCATION	FAMSIZE	INCOME	Q1	Q2	Q3
1	13	35	1	0	3
1	6	120	1	1	1
1	19	60	1	0	2
1	11	100	1	1	2
1	8	45	1	0	3
1	8	200	1	1	3
1	6	135	0	1	3
1	8	75	1	0	2
1	15	110	0	1	2
2	8	85	0	1	2
2	11	130	1	0	3
2	14	165	0	0	3
2	18	70	0	1	2
2	8	55	0	1	1
2	8	105	0	1	1
2	6	140	1	0	1
2	15	95	0	0	1
2	9	175	0	0	2

a) Is there any relationship (dependency) between location and the responses to Q1, Q2 and Q3 at the 0.05 level of significance?
b) Is there any relationship (dependency) between family size and the responses to Q1, Q2 and Q3 at the 0.05 level of significance?
c) Is there any relationship (dependency) between income and the responses to Q1, Q2 and Q3 at the 0.05 significance level?

A7. Standard approach:
Start MYSTAT and invoke the editor by typing **EDIT** <Enter>. Create six column labels as follows: **'Location** <Enter> **'Famsize** <Enter> **'Income** <Enter> **'Q1** <Enter> **'Q2** <Enter> **'Q3** <Enter> and then use the arrow keys to position the highlighted box at case 1 under the LOCATION label. Enter the above data values columnwise using the down arrow key after each entry. Repeat this same process for the other five columns. After all the data has been entered press the <Esc> key to return to the command line.
[Skip the next paragraph if you are NOT using that method]

Alternative approach:
I have created a DOS text file containing the above data using WordPerfect. The name of this file is **SURVEY.DAT** which I can copy into your hard disk MYSTAT subdirectory. Invoke the MYSTAT editor as above and also create the same labels. Then press the <Esc> key to return to the command line and type **GET SURVEY** <Enter> to load the data from the DOS text file into the MYSTAT spreadsheet. This process may take several minutes depending on the size of the file.

Now save the data in a MYSTAT file by typing: **SAVE SURVEY** <Enter>. Exit the editor by typing: **QUIT** <Enter> and you return to the main menu. Again type **USE SURVEY** <Enter> to open this MYSTAT data file.
Note: the data associated with Q1, Q2 and Q3 are coded data where Q1 and Q2 have Yes/No responses (coded as a 1 or 0 respectively) and Q3 has a Low/Medium/High response (coded as 1, 2 or 3 respectively).

To answer part a) we can proceed directly to the analysis provided that we recognize that a cross tabulation or contingency table with a Chi-square test are needed. Type **HELP TABULATE** <Enter> to determine the proper syntax of this command. Upon return to the MYSTAT main menu type **TABULATE LOCATION*Q1/PERCENT** <Enter> and in a few seconds

several screens of results appear. Pay attention to the Pearson Chi-square test statistic value and its associated PROB value as well as the magnitude of CRAMER's V, the only measure of association bounded by 0 and +1. Repeat this same process for Q2 and Q3. What are your conclusions?

Note: You can save all the results displayed on the screen by typing the command: **OUTPUT RESULT7A** <Enter> before typing the **TABULATE** command. A DOS text file named RESULT7A.DAT is created which can later be retrieved by WordPerfect.

To answer part b) and c) requires additional work. Why? Because the data associated with FAMSIZE and INCOME is interval scale data (the former discrete and the latter continuous). We need to create two new coded variables that represent classes of FAMSIZE and INCOME but the question is: How do we construct these classes, i.e. how many and the width of each? To help us make some decisions we will look at stem plots. Type **HELP STEM** <Enter> to determine the proper syntax of this command. Upon return to the main menu type **STEM FAMSIZE,INCOME** <Enter> and use the information provided here to decide on the number of classes and their widths for FAMSIZE and INCOME. I suggest the following classes and associated codes:

FAMSIZE class	FAMCLASS code	INCOME class	INCCLASS code
4 - 7	1	30 - 90	1
8 - 10	2	91 - 150	2
11 - 14	3	151 - 210	3
15+	4		

Invoke the MYSTAT editor again and add two new column labels named FAMCLASS and INCCLASS. Press the <Esc> key to return to the command line and type the following commands:

IF FAMSIZE<8 THEN LET FAMCLASS=1 <Enter> and wait for MYSTAT to perform the transformations.
IF FAMSIZE<11 AND FAMSIZE>7 THEN LET FAMCLASS=2 <Enter> and wait for MYSTAT to perform the transformations.
IF FAMSIZE<15 AND FAMSIZE>10 THEN LET FAMCLASS=3 <Enter> and wait for MYSTAT to perform the transformations.
IF FAMSIZE>14 THEN LET FAMCLASS=4 <Enter> and wait for MYSTAT to perform the transformations.

IF INCOME<91 THEN LET INCCLASS=1 <Enter> and wait for MYSTAT to perform the transformations.

IF INCOME<151 AND INCOME>90 THEN LET INCCLASS=2 <Enter> and wait for MYSTAT to perform the transformations.
IF INCOME>150 THEN LET INCCLASS=3 <Enter> and wait for MYSTAT to perform the transformations.

Now save the transformations by typing **SAVE SURVEY** <Enter> and type **Y** to the MYSTAT prompt asking if the existing data file is to be overwritten. Finally, type **QUIT** <Enter> to exit the editor and return to the MYSTAT main menu.
Note: All the above MYSTAT commands could be placed in a DOS text file named **SURVEY.CMD** and from the MYSTAT prompt at the main menu type **SUBMIT SURVEY** <Enter> to execute the commands in that file.

We are now ready to perform the necessary Chi-square analyses as follows:

For part b):

TABULATE FAMCLASS*Q1/PERCENT <Enter> and in a few seconds several screens of results appear. Pay attention to the Pearson Chi-square test statistic value and its associated PROB value as well as the magnitude of CRAMER's V, the only measure of association bounded by 0 and +1. Repeat this same process for Q2 and Q3. What are your conclusions?

For part c):

TABULATE INCCLASS*Q1/PERCENT <Enter> and repeat this command with Q2 and Q3. Again, what are your conclusions?
Note: You can save all the results displayed on the screen by typing the command: **OUTPUT RESULT7B** <Enter> before typing the **TABULATE** command. A DOS text file named RESULT7B.DAT is created which can later be retrieved by WordPerfect.

Listing of SURVEY.DAT file

```
1  13   35  1  0  3
1   6  120  1  1  1
1  19   60  1  0  2
1  11  100  1  1  2
1   4   45  1  0  3
1  12  200  1  1  3
1   6  135  0  1  3
1   8   75  1  0  2
1  15  110  0  1  2
2   8   85  0  1  2
2  11  130  1  0  3
2  14  165  0  0  3
2  18   70  0  1  2
2   8   55  0  1  1
2   9  105  0  1  1
2   6  140  1  0  1
2  15   95  0  0  1
2   9  175  0  0  2
```

Listing of SURVEY.CMD file

```
use survey
edit
E
F
D
'FAMCLASS
'INCCLASS
Q
format 1
if FAMSIZE<8 then let FAMCLASS=1
if FAMSIZE<11 and FAMSIZE>7 then let FAMCLASS=2
if FAMSIZE<15 and FAMSIZE>10 then let FAMCLASS=3
if FAMSIZE>14 then let FAMCLASS=4
if INCOME<91 then let INCCLASS=1
if INCOME<151 and INCOME>90 then let INCCLASS=2
if INCOME>150 then let INCCLASS=3
save survey1
quit
```

PART III -- Model Building

Q8. A forester wants to construct a volume equation for blue pine growing in a specific region of Pakistan. For this purpose **68** trees were selected at random and they were cut down. The DBH (in inches) and HT (in feet) was recorded for each tree as well as its VOLUME (in cubic feet). The volumes were obtained by conducting a stem analysis, i.e. each tree was cut into ten sections and the section volumes (computed via Smalian's formula) were summed. The raw data was stored in a DOS text file named **CHAMPION.DAT**

a) Compute the following three new independent variables: DBH*DBH, DBH*HT, DBH*DBH*HT

b) Plot VOLUME versus each of the five independent variables, i.e. DBH, HT, and the three from part a). Do you observe any linear trends?

c) Calculate the simple correlation coefficients between VOLUME and the above five independent variables. Which variables exhibit the strongest linear association?

A8. Start MYSTAT and invoke the editor. Create three column labels as follows: **'VOLUME** <Enter> **'DBH** <Enter> **'HT** <Enter> and press the <Esc> key to return to the command line. Type **GET CHAMPION** <Enter> to load the data from the DOS text file into the MYSTAT spreadsheet. This process may take several minutes depending on the size of the file. When the cursor returns press the <Esc> key again to return to the body of the spreadsheet and create three more column labels as follows: **'DBH2** <Enter> **'DBHHT** <Enter> **'DBH2HT** <Enter> . Press the <Esc> key again to return to the command line and perform the following transformations:

```
LET DBH2=DBH*DBH <Enter>          wait for MYSTAT to finish!
LET DBHHT=DBH*HT <Enter>          wait for MYSTAT to finish!
LET DBH2HT=DBH2*HT <Enter>        wait for MYSTAT to finish!
```

Now type **SAVE CHAMPION** <Enter> to store the newly created MYSTAT data file followed by **QUIT** <Enter> to exit the editor and return to the main menu.

b) Type **USE CHAMPION** <Enter> to open the MYSTAT data file and then type **PLOT VOLUME*DBH** <Enter>, **PLOT VOLUME*HT** <Enter>, **PLOT VOLUME*DBH2** <Enter>, **PLOT VOLUME*DBHHT** <Enter>, and **PLOT VOLUME*DBH2HT** <Enter> to generate five scatterplots.
Note: if you want to save these results in a DOS text file type **OUTPUT RESULTQ8** <Enter> before the PLOT commands.

c) Type **PEARSON** <Enter> to compute all the simple correlation coefficients between the variables taken two at a time.

Q9. Fit the following <u>five</u> linear models to the data:

(1) VOLUME = B_0 + B_1 (DBH2)
(2) VOLUME = B_0 + B_1 (DBH2HT)
(3) VOLUME = B_0 + B_1 (DBH) + B_2 (HT)
(4) VOLUME = B_0 + B_1 (DBH) + B_2 (DBHHT)
(5) VOLUME = B_0 + B_1 (DBH) + B_2 (DBH2HT)

and select the "best" volume equation. Which statistics are important and should be considered when choosing your final model?

A9. Type **HELP MODEL** <Enter> to determine the proper syntax of this command. Upon return to the main menu type **OUTPUT RESULTQ9** <Enter> to save your regression results. Then type the five sets of commands listed below:

```
For model (1):      MODEL VOLUME=CONSTANT+DBH2 <Enter>
    [Optional] ==> SAVE RESIDS1 <Enter>
                   EST <Enter>
                   <Enter>
For model (2):      MODEL VOLUME=CONSTANT+DBH2HT <Enter>
    [Optional] ==> SAVE RESIDS2 <Enter>
                   EST <Enter>
                   <Enter>
For model (3):      MODEL VOLUME=CONSTANT+DBH+HT <Enter>
    [Optional] ==> SAVE RESIDS3 <Enter>
                   EST <Enter>
                   <Enter>
For model (4):      MODEL VOLUME=CONSTANT+DBH+DBHHT <Enter>
    [Optional] ==> SAVE RESIDS4 <Enter>
                   EST <Enter>
                   <Enter>
For model (5):      MODEL VOLUME=CONSTANT+DBH+DBH2HT <Enter>
    [Optional] ==> SAVE RESIDS5 <Enter>
                   EST <Enter>
                   <Enter>
```

<u>Note</u>: if you want to examine the residuals from each model they can be saved by typing **SAVE RESIDSx** <Enter> <u>after</u> each MODEL statement and before the EST command where **x** takes on the values 1 to 5.

To select the "best" model look at the adjusted R-squared value, standard error of estimate, tolerances, t-values associated with the regression coefficients and the ANOVA F-value. The number of independent variables in the model and the contribution made by each towards "explaining" the variation in Y (VOLUME) is also considered.

Listing of CHAMPION.DAT file

VOL DBH HT <= this record is not in data file

```
 0.12,  1.6,  12.
 0.18,  1.9,  15.
 0.22,  2.1,  17.
 0.24,  2.2,  18.
 0.35,  2.3,  26.
 0.37,  2.4,  19.
 0.44,  2.6,  20.
 0.58,  2.8,  27.
 0.62,  3.0,  23.
 0.73,  3.1,  27.
 0.85,  3.3,  29.
 1.20,  3.4,  36.
 1.34,  3.6,  37.
 1.30,  3.9,  33.
 1.46,  4.1,  30.
 1.47,  4.3,  34.
 2.39,  4.5,  41.
 2.28,  4.7,  42.
 2.53,  4.8,  40.
 3.45,  5.4,  44.
 2.70,  5.4,  39.
 2.75,  5.9,  35.
 4.92,  6.0,  52.
 4.44,  6.4,  46.
 6.05,  6.7,  54.
 5.77,  6.9,  53.
 5.70,  7.0,  48.
 6.91,  7.3,  53.
 8.47,  7.8,  62.
 7.88,  8.0,  55.
 8.51,  8.4,  58.
 8.27,  8.6,  58.
 6.34,  8.8,  42.
11.52,  9.2,  62.
17.90,  9.6,  70.
12.01, 10.0,  56.
11.87, 10.3,  68.
12.98, 11.0,  61.
16.00, 11.4,  53.
17.15, 11.5,  56.
19.69, 11.7,  74.
13.27, 11.9,  51.
21.76, 12.0,  66.
27.84, 12.3,  76.
```

VOL	DBH	HT	<= this record is not in data file
23.81,	12.4,	75.	
24.30,	12.7,	67.	
28.60,	13.3,	83.	
29.28,	13.6,	77.	
33.33,	14.1,	77.	
26.94,	14.4,	61.	
34.77,	14.7,	80.	
34.05,	14.9,	77.	
25.39,	15.3,	66.	
41.19,	15.7,	78.	
36.85,	16.0,	72.	
48.81,	16.7,	83.	
48.09,	16.8,	84.	
34.47,	17.0,	63.	
53.32,	17.2,	86.	
40.78,	17.4,	77.	
46.48,	18.0,	78.	
55.35,	18.3,	76.	
56.05,	18.8,	79.	
51.39,	19.1,	74.	
60.70,	19.6,	78.	
69.47,	20.0,	82.	
83.62,	21.2,	98.	
114.11,	25.1,	105.	