TABLE OF CONTENTS
for
APPLICATION PROGRAMS

CURVE FITTING

Suppose we want an equation relating expected physics grade versus years of math study. We could choose a simple linear model such as:

$$y = A + Bx$$

$$\text{physics grade} = A + B(\text{yrs. of math})$$

Such a formula could permit the prediction of physics grade, given years of math by substituting a year's value into the formula.

Data:

x Years of Math	y Physics Grade
1	60
2	65
3	85
4	92

Key sequence:

[VAR #] [2] [x_i; [DATA ENTRY] y_i; [DATA ENTRY]] ← for each data pair

Model 5 calculations:

[BASIC STAT]

N	4.00	= #
x̄	2.50	= #
▲1	1.29	= #
ȳ	75.50	= #
▲2	15.42	= #
R	.97	= #

[LINEAR]

A	46.50	= #
B	11.60	= #
r^2	.94	= #

Although the calculation, BASIC STAT, is not necessary, some of the results may be quite helpful.

Results:

$$y = 46.50 + 11.60\,x$$

expected (physics grade) = 46.50 + 11.60 (yrs. of math)

We can now predict the physics grade expected if a student entered a university after five years of mathematics preparation.

Press:

[5] [ŷ EVAL]

```
            LE
      5.00        *
    104.50        *
```

That's better than a 100% average. Not too bad!

The r^2 value (.94) establishes that the fit is good. But, can the curve fit be improved by attempting a more advanced model? Say, parabolic!

$$y = A + Bx + Cx^2$$

physics grade = A + B(yrs. of math) + C(yrs. of math)^2

Press [PARA]

```
          A       =
      49.00       *

          B       =
       9.10       *

          C       =
        .50       *

r²                =
        .95       *
```

Results:

physics grade = 49 + 9.10(yrs. of math) + .50(yrs. of math)^2

Even with a more complex model, the r^2 value (.95) is not greatly improved.

Once again we can press [5] [ŷ EVAL]

```
  PE
          5.00        *
        107.00        *
```

This time the student has a 107% physics grade after 5 years of math study!!! Since an average greater than 100% is impossible, this example helps to illustrate that statistical results require human interpretation.

HISTOGRAM

The following data set represents the final grades given in a college French class. The instructor wants to determine the distribution of the grades about the mean grade.

Data:

Student	Grade (x)
1	55
2	60
3	71
4	72
5	78
6	81
7	76
8	69
9	51
10	70
11	63
12	61
13	69
14	70
15	68

Student	Grade (x)
16	65
17	64
18	65
19	74
20	78
21	80
22	75
23	93
24	85
25	73
26	76
27	80
28	68
29	70
30	88
31	77

Since the data ranges from 51 to 90, we would choose our histogram offset and cell width values to be:

OFF = 50
W = 5

This key sequence is:

[VAR #] [1]

[OFFSET/WIDTH] [5] [0] [DATA ENTRY]

[5] [DATA ENTRY]

Printout:

```
V1      .......

             OF    *
        50.00      *

             CEL*
         5.00      *
```

Now for each data value (x_i), press: x_i [DATA ENTRY]

To find the average grade and the standard deviation, press [BASIC STAT] and the printout is:

```
     N              *
          31.00     *

   x̄                *
          71.77     *

   σ1               *
           9.16     *
```

Then to calculate the histogram, simply press: [HISTO]

The printout is:

```
        HG
      1.00    #
     50.00    #
      1.00    #
      3.23    #

      2.00    #     ──► cell number
     55.00    #     ──► lower cell boundary
      1.00    #     ──► cell frequency
      3.23    #     ──► cell's relative percent frequency
        o
        o
        o
     10.00    #
     95.00    #
       .00    #
       .00    #
```

To use the plotter to plot the histogram; the X and Y **axes**, the tic intervals, and the X and Y intercepts must be established.

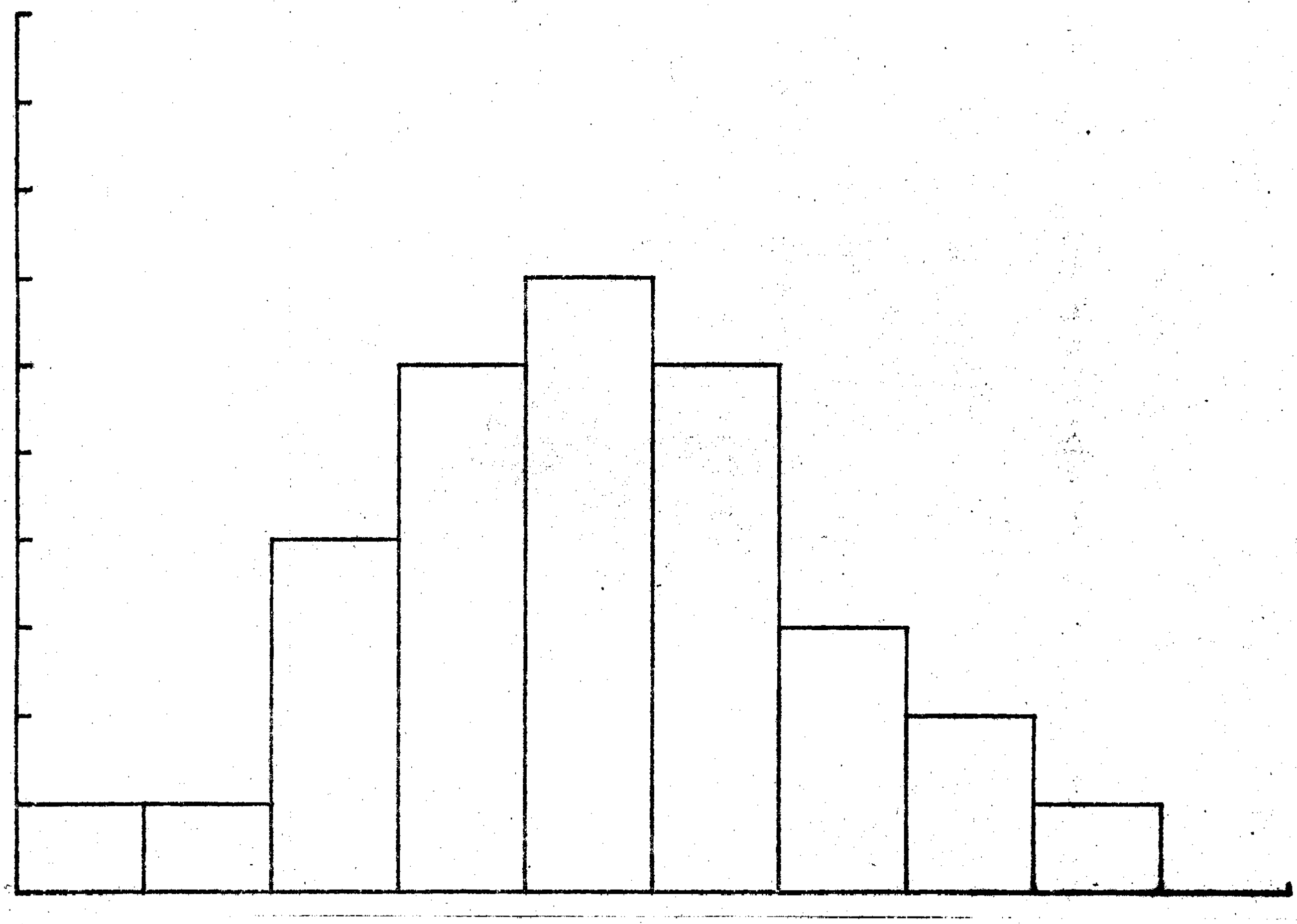
Key Sequence	Printout
[0] [STORE] [1] [0] [AXES] [1]	.00 ← RG1 = .00 # 10.00 #
[0] [STORE] (*) [1] [0] [AXES] [2]	.00 ← RG2 = .00 # 10.00 #
[1] [STORE] [1] (*) [AXES] [3]	1.00 ← T I = 1.00 # 1.00 #
[0] [STORE] [0] (*) [AXES] [4]	.00 ← AI = .00 # .00 #

* These data entries can be omitted since **the previously stored data was identical.**

Note:

We are no longer working with the raw data -- we are working with cells. So when the X axis is set up, you actually establish the minimum and maximum cells to be plotted. Likewise, when the Y axis is set up, you actually establish the minimum and maximum cell frequencies to be plotted.

Now to plot this histogram, press: PLOT



Conclusions:

It appears that the final grades are somewhat normally distributed about the average.

DATA PLOTTING

A high school counselor would like to determine what effect mathematics studies have on future student success in various university studies. In particular, does the number of years of math relate to university physics, chemistry, and German? Data was collected for 100 students who received his counseling prior to university study. All 100 students were required to enroll in physics, chemistry, and German. The data table, below, lists the average grades for each group:

Group (yrs. of math)	German	Chemistry	Physics
1	90	65	60
2	85	70	65
3	80	78	85
4	70	85	92

To set up the calculator for data plotting, press: [VAR#] [2]

Now to use the plotter, you must establish the following: the X and Y axes, the tic intervals, and the X and Y intercepts. After referring to our data, we can press:

0 [STORE] 5 [AXES] 1

50 [STORE] 100 [AXES] 2

1 [STORE] 10 [AXES] 3

0 [STORE] 50 [AXES] 4

```
V2      .......
            .00    +

           RG1     =
            .00    *
           5.00    *

          50.00    +

           RG2     =
          50.00    *
         100.00    *

           1.00    +

            T I    =
           1.00    *
          10.00    *

            .00    +

             AI    =
            .00    *
          50.00    *
```

Now that the axes are plotted, individual data points can be plotted too.

The first plot is for German grades. We want the first data point to be unconnected, so we press:

[CHAR#] 0

1 [DATA ENTRY] 90 [DATA ENTRY]

Then to connect this data point to the other points(using the same character code), press:

[CHAR#] 5

Then enter the data:

2 [DATA ENTRY] 85 [DATA ENTRY]

3 [DATA ENTRY] 80 [DATA ENTRY]

4 [DATA ENTRY] 70 [DATA ENTRY]

The chemistry and the physics grades are plotted in the same manner. For chemistry, the unconnected data point is set by CHAR# 2 and the corresponding connected data points are set by CHAR# 7.

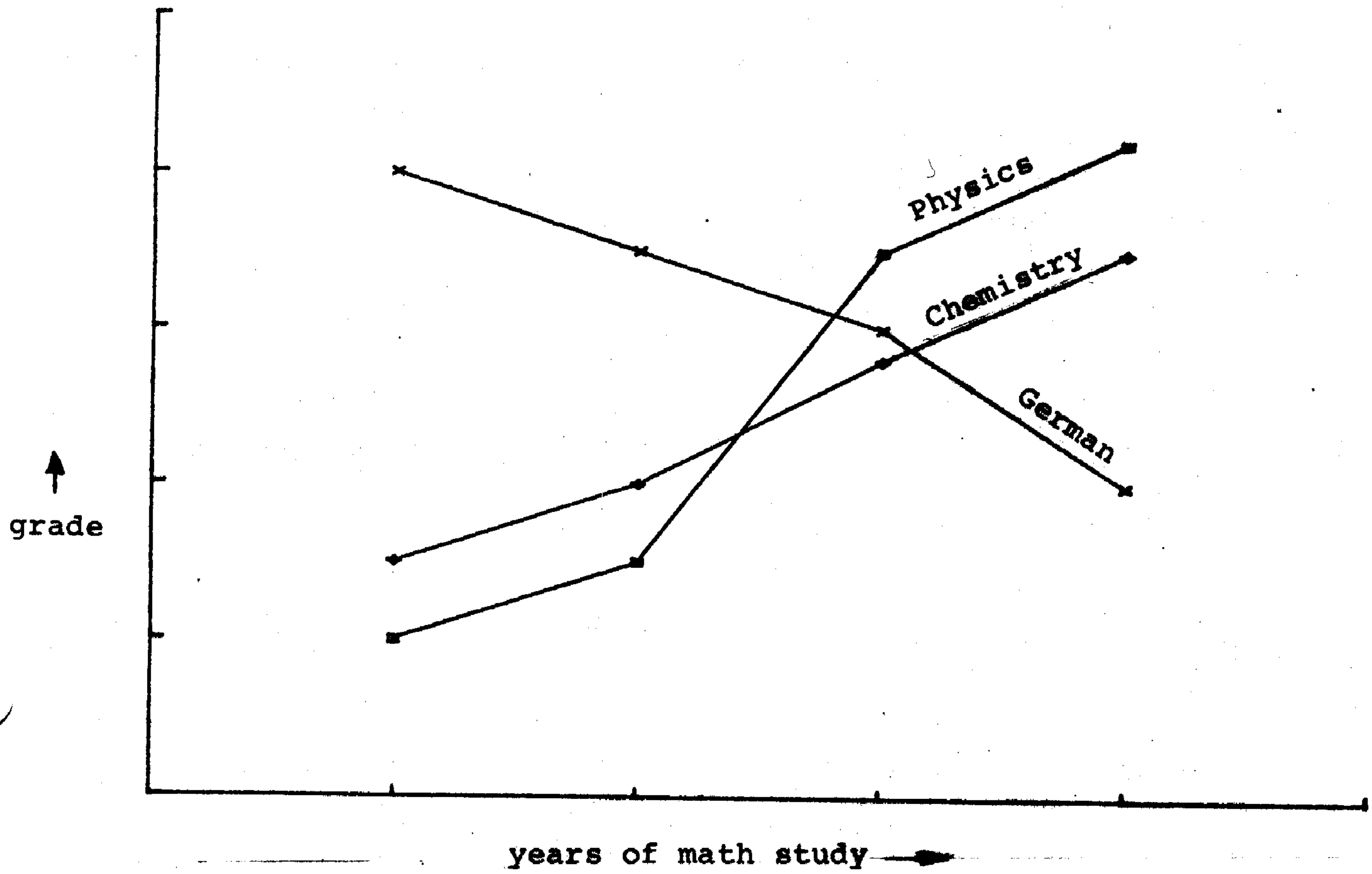
For physics, the unconnected data point is set by CHAR# 3 and the corresponding connected data points are set by CHAR# 8.

(The three plots are shown on the following page.)

Printout

```
      CT0

 1.00      #
90.00      #


      CT5

 2.00      #
85.00      #

 3.00      #
80.00      #

 4.00      #
70.00      #

      CT2

 1.00      #
65.00      #

      CT7

 2.00      #
70.00      #

 3.00      #
78.00      #

 4.00      #
85.00      #

      CT3

 1.00      #
60.00      #

      CT8

 2.00      #
65.00      #

 3.00      #
85.00      #

 4.00      #
92.00      #
```

The final data plots (minus the labels) are as follows.



These plots reveal the following:

a. Mathematics is very useful to university physics; the more, the better.

b. Mathematics is useful to university chemistry; but less than for physics.

c. Mathematics has a negative effect on university German; perhaps the students who concentrated heavily in math studies did so at the sacrifice of language studies.

BASIC STAT

In order to do more forecasting, an insurance company uses statistics to compute the average number of accidents caused by drunken drivers concerning its customers for every month of the preceding year.

Here is the data:

Month	No. of Accidents
1	12
2	7
3	5
4	8
5	5
6	15
7	20
8	6
9	3
10	4
11	10
12	30

The key sequence is:

(VAR #) (1) [each data entry -- (DATA ENTRY)]

← no. of accidents/mo.

Model 5 calculations:

(BASIC STAT)

```
N          =
   12.00   *
x̄          =
   10.42   *
Δ1         =
    7.93   *
```

The monthly mean for accidents ($\bar{x}$) is 10.42 with a standard deviation (Δ1) of 7.93.

Conclusion:

The number of drinking-related accidents varies significantly from month to month. Therefore, using the average number of accidents per month for forecasting techniques is not justified.

PAIRED t

A company producing cigarettes decides to measure the effects of the presentation of the package on the sales figures. During one month in ten cities each records the sales figures in units of one of their brands of cigarettes. The month after, it presents the same brand of cigarettes in a new modern package and records the number of units sold. It compares the results to see if there is any difference due to the change of package.

Town	x Units Sold Old Package	y Units Sold Modern Package
1	250	290
2	270	280
3	280	300
4	250	310
5	300	320
6	275	335
7	360	380
8	280	375
9	365	390
10	295	400

To enter the data, the key sequence is:

(VAR #) (2) [x_i (DATA ENTRY) y_i (DATA ENTRY)] ← for each data pair

Hypothesis Testing:

$H_0 : \bar{x} = \bar{y}$

if $|t| \leq t_{table}$ -- accept H_0

if $|t| > t_{table}$ -- reject H_0

Now press: (BASIC STAT)

and the following items are output:

- sample size (N)
- mean of x ($\bar{x}$)
- standard deviation of x ($\Delta 1$)
- mean of y ($\bar{y}$)
- standard deviation of y ($\Delta 2$)
- **simple correlation coefficient (R)**

```
N          =
     10.00 *
x̄          =
    292.50 *
Δ1         =
     40.29 *
ȳ          =
    338.00 *
Δ2         =
     44.61 *
R          =
       .69 *
```

Now to test our hypothesis, press: (t)

```
      TP              =
-           4.30      #

            D         =
            9.00      #
```

Now we compare the absolute value of our calculated t value (4.30) with the t value from the table. But to use the t table, we need to know two things:

The degrees of freedom -- which we find to be 9.00 from the previous printout.

The confidence level --- which we determine ourselves; in this case, let's use 95%.

With these two factors we see that the t table value is 2.26 .

Since

$|-4.30| > 2.26$ -- we reject the hypothesis that there is no difference between the types of packaging.

Furthermore, by comparing the BASIC STAT results for $\bar{x}$ and $\bar{y}$, we see that the average y is considerably greater than the average x. So we conclude that the new package presentations account for the significant increase in sales figures.

DATA PLOTTING

The data table shown describes the profits of an automobile company over a six-month period.

Month (x)	Total Profit (y)
1	10.42
2	9.35
3	8.65
4	8.52
5	8.58
6	9.1

This data can be easily plotted after you set up the axes:

But first press [VAR#] [2] to erase previous data and to set up the calculator for data plotting.

```
V2     .......
```

Then referring to the data, press:

0 [STORE] 7 [AXES] 1 — sets up min. and max. values for X axis

```
      .00   →
     RG1    =
      .00   #
     7.00   #
```

0 [STORE] 11 [AXES] 2 — sets up min. and max. values for Y axis

```
      .00   →
     RG2    =
      .00   #
    11.00   #
```

1 [STORE] 1 [AXES] 3 — sets up tic marks for unit intervals of 1

```
     1.00   →
      T I   =
     1.00   #
     1.00   #
```

0 [STORE] 0 [AXES] 4 — sets up X and Y intercepts of 0,0

```
      .00   →
       AI   =
      .00   #
      .00   #
```

Now to plot the first data point, use a CHAR# from 0 through 4. This ensures the point will be plotted without a connecting line. Press:

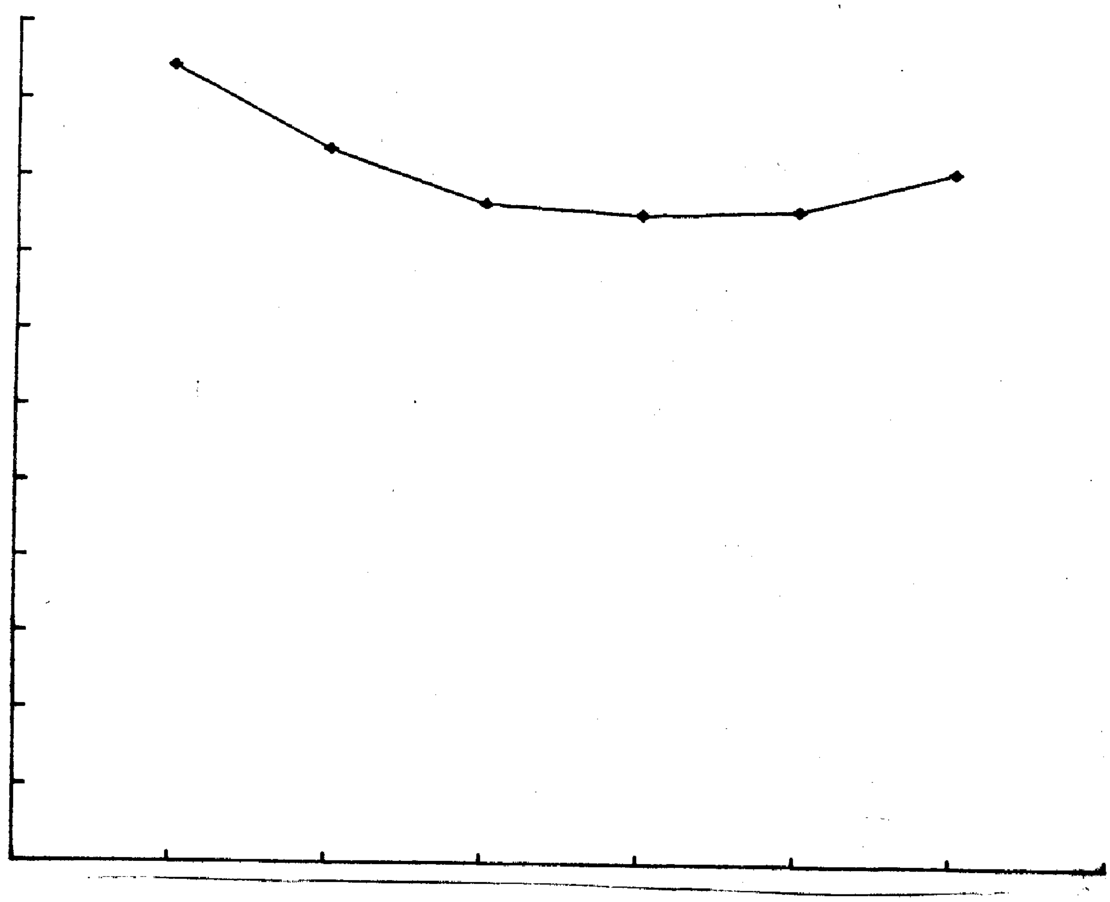
[CHAR#] [2] then 1 [DATA ENTRY] 10.42 [DATA ENTRY]

Then the rest of the data can be plotted with lines connecting the points by using a CHAR# from 5 through 9. Furthermore, by using a CHAR# five greater than the previous one, the same symbol is used for the plotted points. Press:

[CHAR#] [7] then [x_i [DATA ENTRY] y_i [DATA ENTRY]] ← for each data pair

CT2	
1.00	#
10.42	#
CT7	
2.00	#
9.35	#
3.00	#
8.65	#
4.00	#
8.52	#
5.00	#
8.58	#
6.00	#
9.10	#

The plot looks like this:



The plot above reveals a slight drop in total profits and a definite negative trend over the six months for the automobile company. On the following page, the company manager expands the statistical study to further explain this trend.

The company manager decides to investigate the individual components of the total profits: new cars, used cars, service.

The accumulated data is as follows:

Month	New	Used	Service	Total
1	3.12	2.3	5	10.42
2	3.25	2.6	3.5	9.35
3	3.35	2.8	2.5	8.65
4	3.42	3.1	2	8.52
5	3.48	3.6	1.5	8.58
6	3.60	4.0	1.5	9.1

Once again to erase previous data and to set up the calculator for data plotting, press: [VAR #] [2]

Then key in the axes information in the same manner as shown previously.

In setting up the axes, you can use the data shown on the printout tc the right.

```
V2      .......
           .00     →

           RG1     =
           .00     #
          7.00     #

           .00     →

           RG2     =
           .00     #
          7.00     #

          1.00     →

           T I     =
          1.00     #
          1.00     #

           .00     →

            AI     =
           .00     #
           .00     #
```

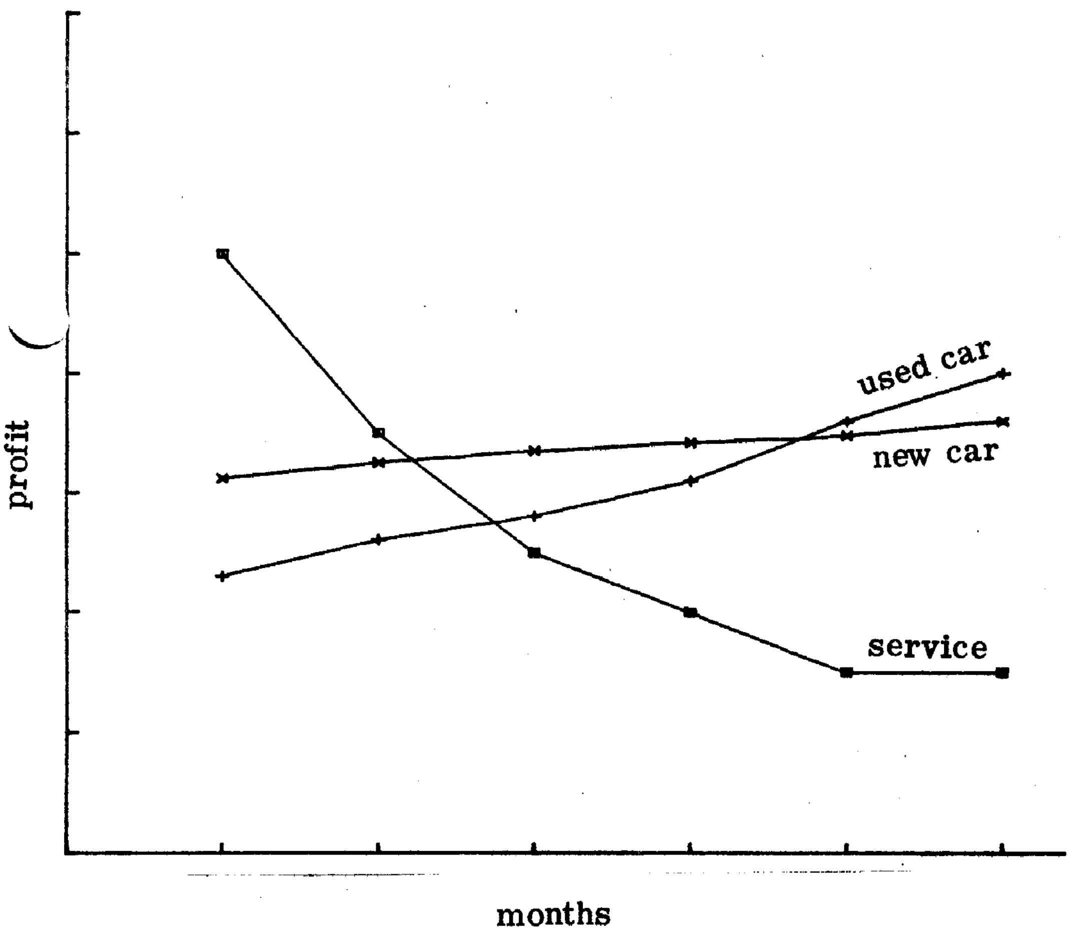
Now the new car profits, the used car profits, and the service profits can each be graphed by month.

Just use the same type of procedure as used for the previous plot.

Remember:
For each graph, when you mark the first data point(for month 1), a CHAR# from 0 through 4 should be used.

Use the data shown on the printout to the right:

The graph, minus the labels, is as follows:



The composite plot reveals several business trends:

a. **New car profit is growing -- but slowly.**

b. **Used car profit is growing -- more rapidly than new cars.**

c. **Service profit is falling -- but almost leveling off.**

```
  CT0
1.00      #
3.12      #

  CT5
2.00      #
3.25      #

3.00      #
3.35      #

4.00      #
3.42      #

5.00      #
3.48      #

6.00      #
3.60      #

  CT1
1.00      #
2.30      #

  CT6
2.00      #
2.60      #

3.00      #
2.80      #

4.00      #
3.10      #

5.00      #
3.60      #

6.00      #
4.00      #

  CT3
1.00      #
5.00      #

  CT8
2.00      #
3.50      #

3.00      #
2.50      #

4.00      #
2.00      #

5.00      #
1.50      #

6.00      #
1.50      -
```

TWO-SAMPLE t

The data sets below represent the hours of production lost due to sickness in two departments of a factory during 2-1/2 yrs. Before analyzing the results to re-adjust the production cost, the accounting department needs to confirm a significant difference between the two departments concerning sick leave.

Month	x 1st department	y 2nd department
1	10.10	↑ no data available ↓
2	18.20	
3	11.30	
4	17.00	
5	12.10	
6	16.10	
7	13.60	14.10
8	16.20	18.10
9	14.20	17.20
10	15.00	17.70
11	15.70	16.10
12	15.60	17.00
13	15.10	16.40
14	14.10	14.60
15	14.40	17.10
16	12.60	16.30
17	13.00	14.00
18	13.00	16.20
19	12.00	15.00
20	14.20	15.00
21	11.10	15.70
22	14.60	16.60
23	13.70	13.10
24	11.30	17.70
25	14.00	16.00
26	10.10	15.10
27	15.70	18.10
28	13.60	19.00
29	16.40	15.90
30	17.80	18.20

To set up the two-sample t problem, press:

[VAR #] [0]

```
2T .......
```

Then enter all the data from the first sample(x):

[x_i [DATA ENTRY]]

```
DATA 1
10.10    #
18.20    #
```

Then enter all the data from the second sample(y):

[CHANGE SAMPLE] [y_i [DATA ENTRY]]

```
DATA 2
14.10    #
18.10    #
```

Now press: [BASIC STAT]

and the following items are output:

sample size of 1st set (N_1)

mean of first set ($\bar{x}$)

standard deviation of 1st set (∆1)

sample size of 2nd set (N_2)

mean of 2nd set ($\bar{y}$)

standard deviation of 2nd set (∆2)

```
N1          =
      30.00 #
x           =
      14.06 #
∆1          =
       2.14 #

N2          =
      24.00 #
y           =
      16.26 #
∆2          =
       1.50 #
```

Hypothesis test:

H_0: $\bar{x} = \bar{y}$

if $|t| \leq t_{table}$ -- accept H_0

if $|t| > t_{table}$ -- reject H_0

assume a confidence level of, say, 95%.

Now to test our hypothesis, press: [t]

```
      T          =
 -    4.26       #
      D          =
     52.00       #
```

With D = 52 and a confidence level of 95%, the t_{table} value is: approximately 2.00.

Since $|-4.26| > 2.00$ -- we reject the hypothesis that there is no difference between the two departments.

Since $\bar{x}$ is considerably less than $\bar{y}$ as seen in the BASIC STAT printout, the production costs should be re-adjusted accordingly.

PAIRED t

The first group of data below represents the drying time of a traditional paint on ten different types of surfaces. The second set of data represents the drying time on the same ten surfaces, but with a new plastic paint. Before buying this new, more expensive paint, a contractor wants to know if there is a real difference between the drying times of the two paints.

Surface	x Traditional Paint	y New Plastic Paint
1	2.4	1.5
2	2.1	1.3
3	3.1	2.6
4	4.2	3.7
5	3.2	2.8
6	4.4	5.2
7	6.2	4.4
8	5.3	3.6
9	5.5	5.5
10	6.2	4.5

To enter the data, the key sequence is:

[VAR #] [2] [x_i [DATA ENTRY] y_i [DATA ENTRY]]

↑ for each data pair

Now press: [BASIC STAT]

and the following items are output:

- sample size (N)
- mean of x ($\bar{x}$)
- standard deviation of x (s1)
- mean of y ($\bar{y}$)
- standard deviation of y (s2)
- simple correlation coefficient (R)

N		=
	10.00	#
$\bar{x}$		=
	4.26	#
s1		=
	1.52	#
$\bar{y}$		=
	3.51	#
s2		=
	1.45	#
R		=
	.35	#

Hypothesis Testing:

$H_0 : \bar{x} = \bar{y}$ (that is, there is no significant difference between the drying times of the two paints)

if $|t| \leq t_{table}$ -- accept H_0

if $|t| > t_{table}$ -- reject H_0

assume a confidence level = 95%

Now to test our hypothesis, press: [t]

```
TP              =
      2.87      #

      D         =
      9.00      #
```

With D = 9 (from the printout) and with a confidence level = 95%, the t_{table} value = 2.26, while the computed t value = 2.87 .

Since $|2.87| > 2.26$ -- we reject the hypothesis that there is no difference between the two types of paint.

Furthermore, since $\bar{y}$ is significantly less than $\bar{x}$ as seen in the BASIC STAT printout, we conclude that the drying time of the new plastic paint is indeed much faster than is the drying time of the traditional paint.

BASIC STAT

The following set of data represents the weights of twelve different cars having the same length. The question is, "Do cars having the same length vary significantly in their weight?"

Car	Weight
1	38
2	40
3	41
4	39
5	38.5
6	43
7	45
8	40.5
9	41
10	37.5
11	42
12	42

The key sequence is:

[VAR #] [1] [each data entry -- [DATA ENTRY]]
← weight

Model 5 calculations:

[BASIC STAT]

```
     N               =
          12.00      ¤
     x               =
          40.63      ¤
     ▵1              =
           2.20      ¤
```

The mean of the weight of the twelve cars ($\bar{x}$) is **40.63 with a standard** deviation (▵1) of 2.20.

Conclusion:

For cars that have the same length, the weight does not differ very much between the twelve cars selected.

DATA PLOTTING

The data given in the table represents the average weights of 27 cans of tuna, each hour, over an eight hour period. Averages are given for two tuna packing machines.

Hour	Machine 1	Machine 2
1	113	104
2	110	106
3	113	107
4	112	110
5	114	109
6	110	112
7	112	113
8	111	112

To set up the calculator for data plotting, press: [VAR#] [2]

To set up the axes for the plotter, press:

0 [STORE] 10 [AXES] 1

← setting up X min. and max. (hours)

100 [STORE] 120 [AXES] 2

← setting up Y min. and max. (ave. wgt.)

1 [STORE] 2.5 [AXES] 3

← setting up X and Y tic intervals

0 [STORE] 100 [AXES] 4

← setting up the X and Y intercepts

```
V2     .......
           .00    →

          RG1     =
           .00    #
         10.00    #

        100.00    →

          RG2     =
        100.00    #
        120.00    #

          1.00    →

           T I    =
          1.00    #
          2.50    #

           .00    →

           A I    =
           .00    #
        100.00    #
```

Now that the axes are plotted, individual data points can be plotted too.

The first plot is for machine 1. We want the first data point to be unconnected, so we press:

[CHAR#] 0

1 [DATA ENTRY] 113 [DATA ENTRY]

Then to connect this data point to the other points (using the same character code), press:

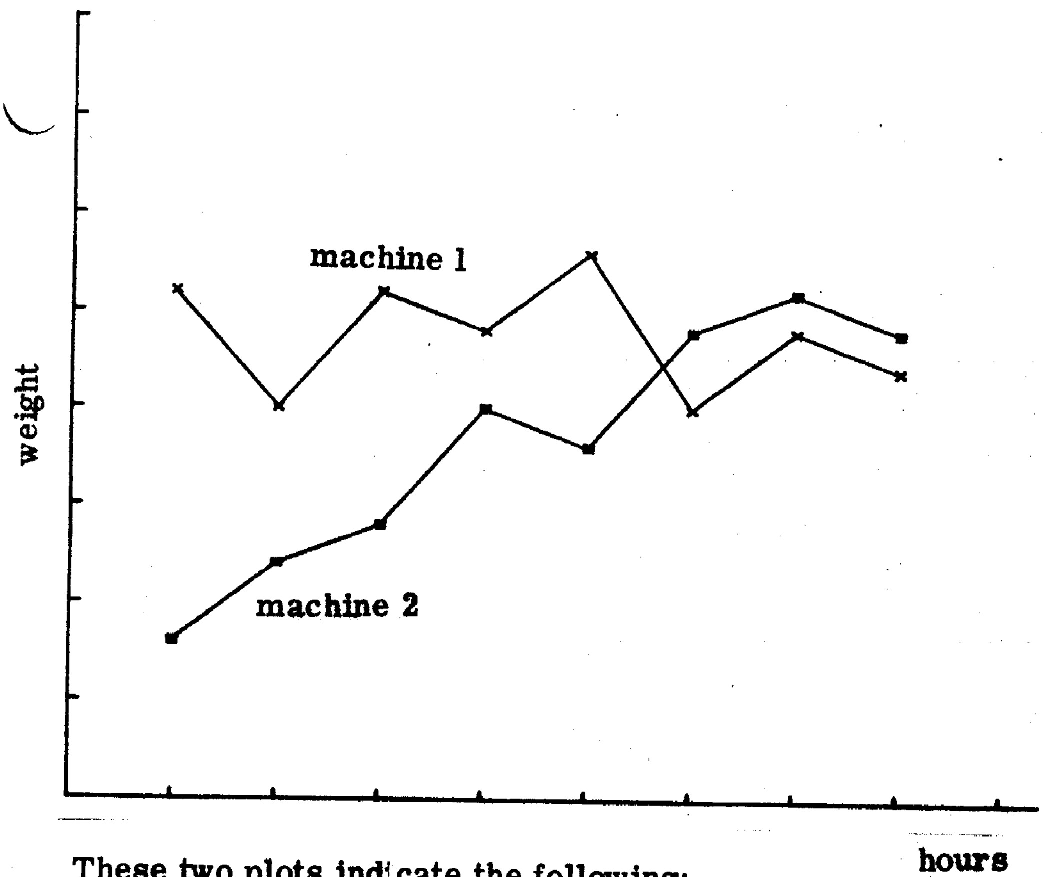
[CHAR#] 5

Then enter the data in pairs:

2 [DATA ENTRY] 110 [DATA ENTRY] etc. for each data pair

For machine 2, use the same technique. Use CHAR#3 for the unconnected data point, and use CHAR#8 for the connected data points.

Here are the plots.



These two plots indicate the following:

a. Machine 1 appears to have a constant (trend free) output.

b. Machine 2 appears to have an upward trend(increased wgts.).

```
        CT0
  1.00      #
113.00      #
        CT5
  2.00      #
110.00      #
  3.00      #
113.00      #
  4.00      #
112.00      #
  5.00      #
114.00      #
  6.00      #
110.00      #
  7.00      #
112.00      #
  8.00      #
111.00      #
        CT3
  1.00      #
104.00      #
        CT8
  2.00      #
106.00      #
  3.00      #
107.00      #
  4.00      #
110.00      #
  5.00      #
109.00      #
  6.00      #
112.00      #
  7.00      #
113.00      #
  8.00      #
112.00      #
```

CURVE FITTING

Suppose we want to determine a formula relating the time required to cut down a tree (felling time) to the tree's diameter. We could choose a simple linear model, such as:

$$y = A + Bx$$

$$\text{felling time} = A + B\ (\text{tree diameter})$$

Such a formula could permit the prediction of felling time, given the tree's diameter, by substituting the tree's diameter into the formula.

Here is the 'raw' data:

x Tree Diameter	y Felling Time
10	1.1
15	1.6
20	2.3
25	3.2
30	4.3
35	5.6

Let's plot this data as we enter it. To do so requires that we set up our axes. But first press:

[VAR#] [2]

Now set up the axes:

0 [STORE] 50 [AXES] 1

0 [STORE] 10 [AXES] 2

10 [STORE] 1 [AXES] 3

0 [STORE] 0 [AXES] 4

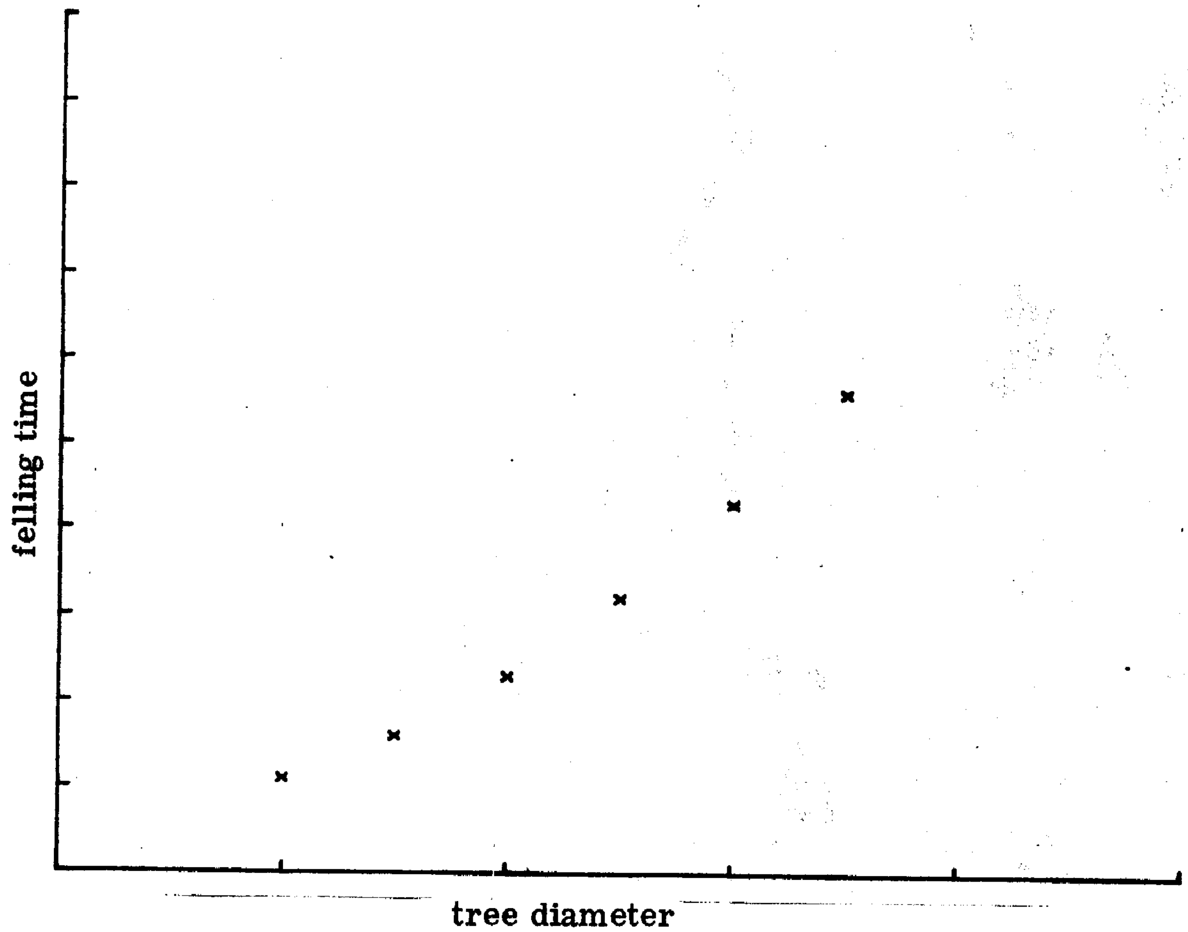
```
V2     .......
          .00    →

          RG1    =
          .00    #
        50.00    #

          .00    →

          RG2    =
          .00    #
        10.00    #

        10.00    →

          T I    =
        10.00    #
         1.00    #

          .00    →

          A I    =
          .00    #
          .00    #
```

Now as we enter the data, it will be plotted too. Enter the data as follows:

[x_i DATA ENTRY y_i DATA ENTRY]

↑ for each data pair

The plot of the data points is shown below:



Now BASIC STAT and the linear fit can be determined. Press:

BASIC STAT

```
N         =
    6.00  #
x̄         =
   22.50  #
s1        =
    9.35  #
ȳ         =
    3.02  #
s2        =
    1.71  #
R         =
     .99  #
```

LINEAR

```
        A   =
-  1.03     #
        B   =
    .18     #
r²          =
    .97     #
```

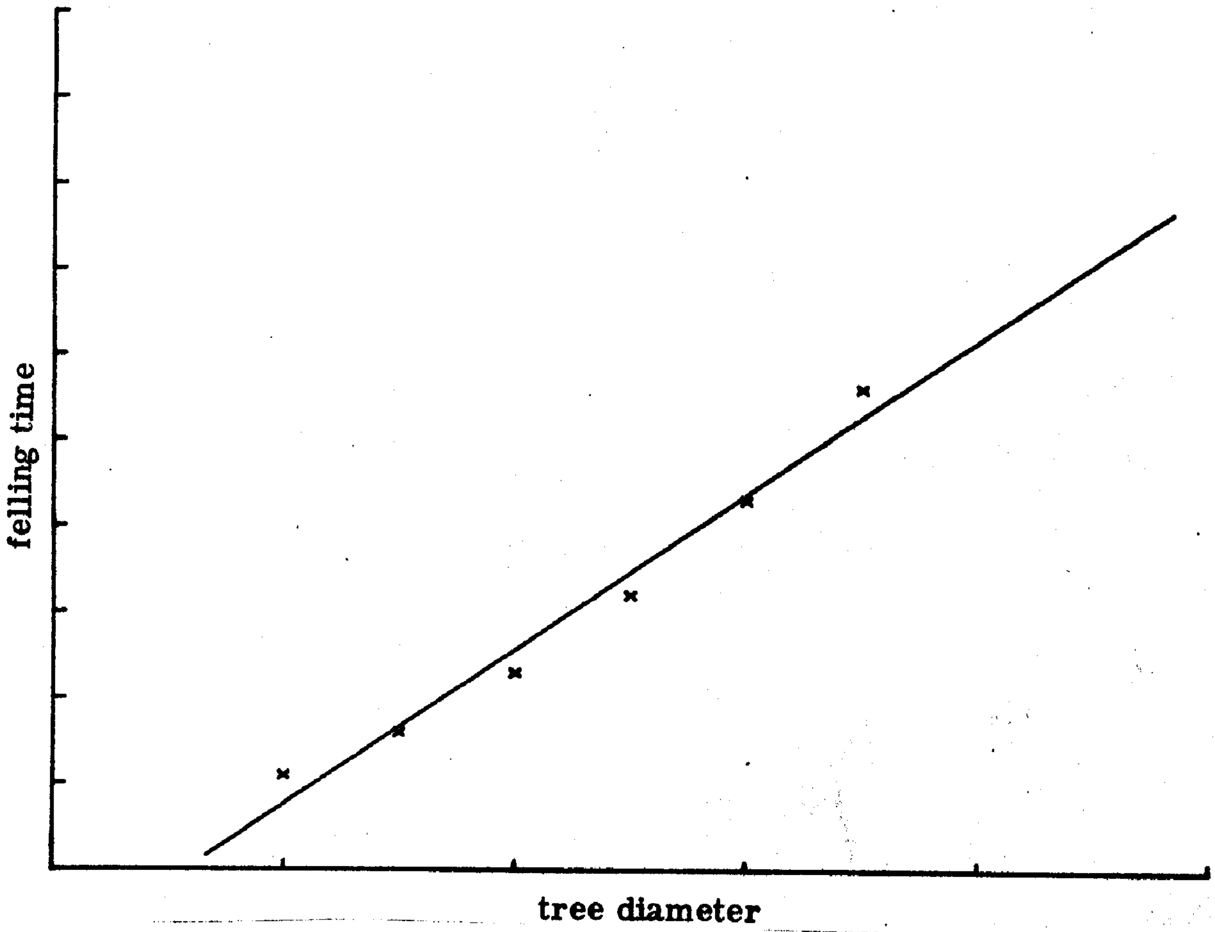
Results:

$$y = -1.03 + 0.18x$$

$$\text{felling time} = -1.03 + 0.18\,(\text{tree diameter})$$

Notice $r^2 = .97$, which indicates a reasonable fit although not the best possible.

Let's plot the data using this linear fit. Press: [PLOT]



We readily notice that the felling time is not linear. This should not come as a surprise because tree area increases as the square of the tree diameter. Thus, a parabolic model would probably yield a better fit.

Press [PARA] and the printout is:

```
          A  *
       .70   *
          B  *
       .00   *
          C  *
       .00   *
             *
      1.00   *
```

Results:

$$y = A + Bx + Cx^2$$

$$\text{Felling Time} = A + B(\text{Tree Diameter}) + C(\text{Tree Diameter})^2$$

$$\text{Felling Time} = 0.70 + 0.0(\text{Tree Diameter}) + .004(\text{Tree Diameter})^2$$

The r^2 value (1.00) indicates a perfect fit. We can now use the formula to predict the felling times for any size trees.

To determine the felling time for a tree with a diameter = 5, press:

[5] [ŷ EVAL]

and the printout is:

```
PE
        5.00    *
         .80    *
```

So for a tree diameter of 5, the felling time = .8 .

To determine the felling time for a tree with a diameter = 22.5, press:

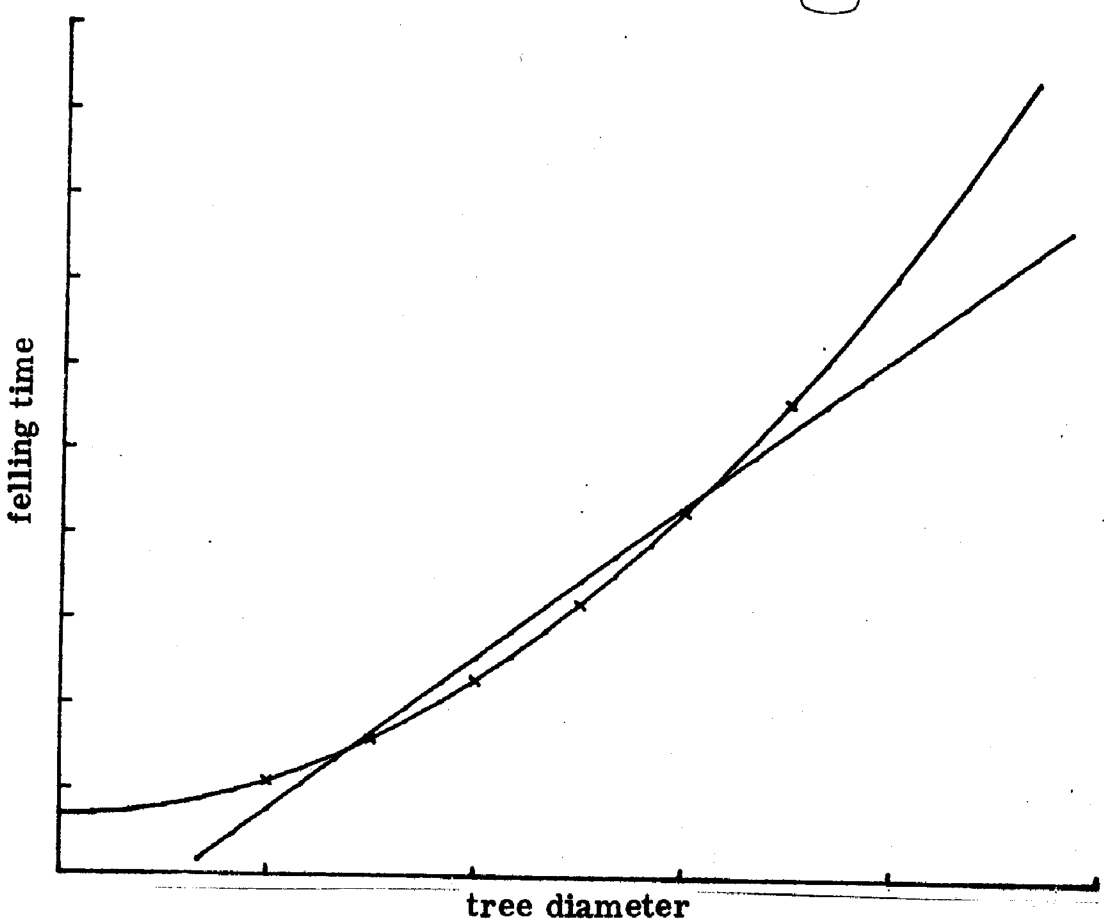
[2] [2] [.] [5] [ŷ EVAL]

and the printout is:

```
PE
        22.50    *
         2.73    *
```

So for a tree diameter of 22.5, the felling time = 2.73 .

Now to plot the data using this parabolic fit, press: [PLOT]



Notice the parabolic curve goes through all the data points, whereas the linear curve fit does not.

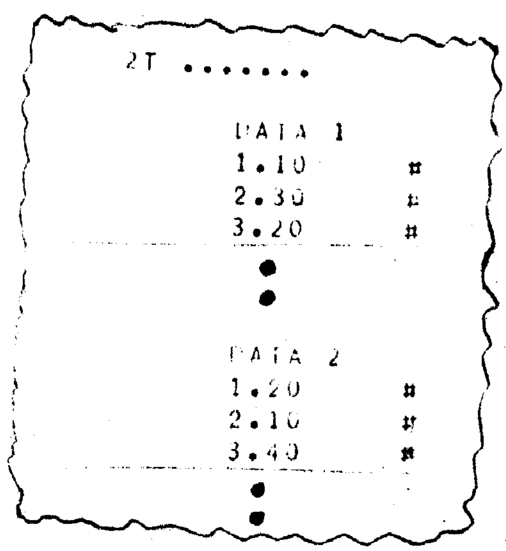
TWO-SAMPLE t

In a hospital, two groups of premature babies **that were born in the same month** have been set up. Group A is composed of baby boys; group B is composed of baby girls. Before trying a new **treatment** on them, the doctors need to know if there is a difference in weight between the two groups. The following data sets represent the weights of these babies:

Group A	Group B
1.1	1.2
2.3	2.1
3.2	3.4
4.4	4.3
5.6	3.1
4.5	2.2
3.3	1.3
2.1	
1.2	

To set up the two-sample t problem, press:

[VAR #] [0]



Then enter all the data from the first sample(x):

$[x_i$ [DATA ENTRY]$]$

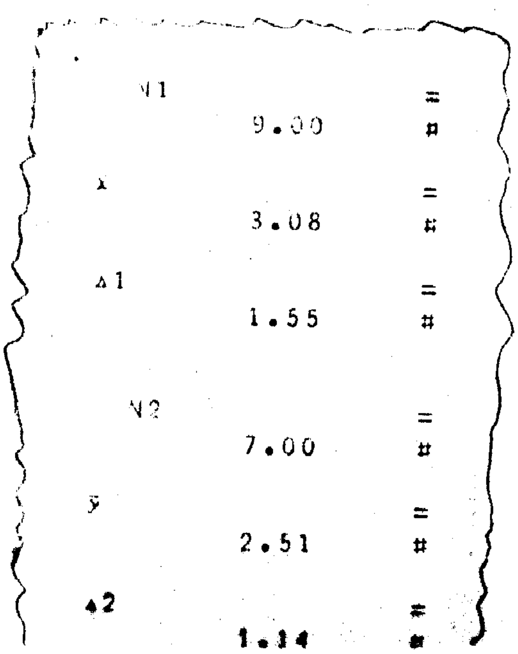
Then enter all the data from the second sample(y):

[CHANGE SAMPLE] $[y_i$ [DATA ENTRY]$]$

Now press: [BASIC STAT]

and the following items are output:

- sample size of 1st set (N_1)
- mean of 1st set ($\bar{x}$)
- std. deviation of 1st set ($\Delta 1$)
- sample size of 2nd set (N_2)
- mean of 2nd set ($\bar{y}$)
- std. deviation of 2nd set ($\Delta 2$)



Hypothesis Test:

$H_0 : \bar{x} = \bar{y}$ (that is, there is no statistical difference between the two groups)

if $|t| \leq t_{table}$ -- accept H_0

if $|t| > t_{table}$ -- reject H_0

assume a confidence level = 95%

Now to test our hypothesis, press: t

```
   T               =
             .81   ¤

            D      =
          14.00    ¤
```

With D = 14 and a confidence level of 95%, the t_{table} value = 2.14 ; and the calculated t = .81 .

Since $|.81| < 2.14$ -- we accept the hypothesis that there is no difference between the two groups; that is, they belong to the same population.

So the doctors conclude that there is no real difference in weight between male and female premature infants. Therefore, the same treatment can be used on both groups.

HISTOGRAM

The following data sets represent blood cell diameters of 54 patients undergoing medical treatment. The patients have been grouped into two categories based upon diet. One group contains 30 patients, the other 24.

Diet A	Diet B
10.1	14.1
18.2	18.1
11.3	17.2
17.0	17.7
12.1	16.1
16.1	17.0
13.6	16.4
16.2	14.6
14.2	17.1
15.0	16.3
15.7	14.0
15.6	16.2
15.1	15.0
14.1	15.0
14.4	15.7
12.6	16.6
13.0	13.1
13.0	
12.0	17.7
14.2	
11.1	16.0
14.6	
13.7	15.1
11.3	
14.0	18.1
10.1	
15.7	19.0
13.6	
16.4	15.9
17.8	18.2

Since the blood cell diameters from diet A range from 10.1 to 18.2 , for our first histogram we would choose the offset and cell width values to be:

OF = 10
CEL = 1

This key sequence is:

[VAR #] [1]

[OFFSET WIDTH] [1] [0] [DATA ENTRY]

[1] [DATA ENTRY]

Printout:

```
V1      .......

            OF      =
        10.00       #

            C E L=
          1.00      #
```

Now for each diet A data value(x_i), press:

x_i [DATA ENTRY]

To find the average blood cell diameter and the standard deviation, press [BASIC STAT] and the printout is:

```
N            =
     30.00   #
x̄            =
     14.06   #
Δ1           =
      2.14   #
```

Then to calculate the histogram, simply press: [HISTO]

The printout is:

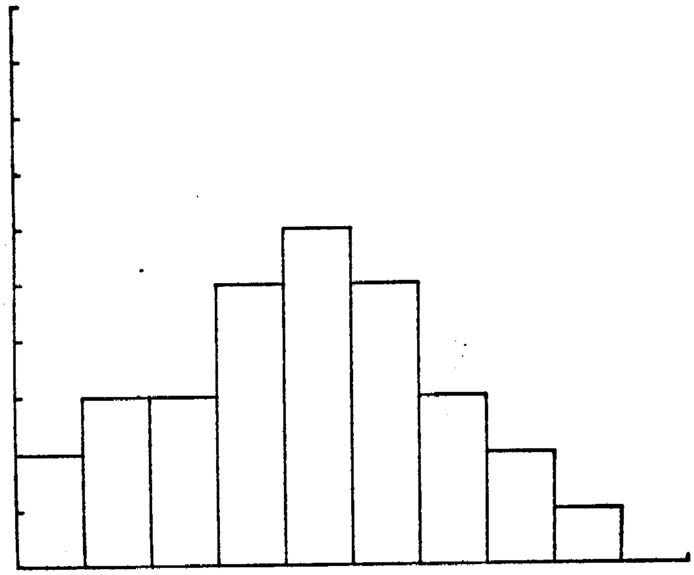
```
     HG
   1.00   #   → cell number
  10.00   #   → lower cell boundary
   2.00   #   → cell frequency
   6.67   #   → cell's relative percent frequency

   2.00   #
  11.00   #
   3.00   #
  10.00   #
     •
     •
```

To plot this histogram, you must first set up the axes:

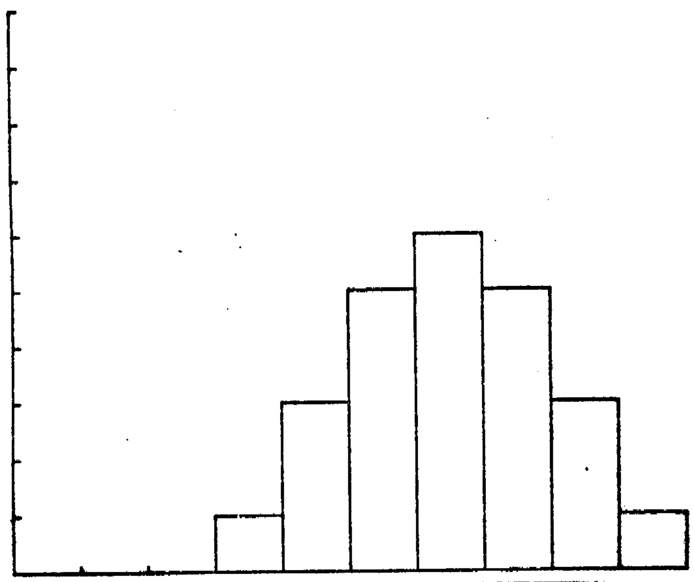
0 [STORE] 10 [AXES] 1	sets up cell min. and max. for X axis
0 [STORE] 10 [AXES] 2	sets up min. and max. cell frequency for Y axis
1 [STORE] 1 [AXES] 3	sets tic mark intervals for X and Y axes
0 [STORE] 0 [AXES] 4	sets up intercepts for X and Y axes

```
      .00    ◆
   RG1       =
      .00    #
    10.00    #

      .00    ◆
   RG2       =
      .00    #
    10.00    #

     1.00    ◆
   T I       =
     1.00    #
     1.00    #

      .00    ◆
   AI        =
      .00    #
      .00    #
```

Now to plot this histogram, press: (PLOT)



For our second histogram, we can use the diet B data. Press (VAR #) (1) to erase the old data; then follow the same procedure that was used for the diet A data. You can use the same parameters for the OFFSET/WIDTH and for the X and Y axes that were used previously. The plot of the diet B data looks like this:



Discussion:

a. Both data sets (blood cell diameters) are approximately normal.

b. There may be a slight difference in average blood cell diameter.

0	0	50	62	100	144	150	226	200	310
1	1	51	63	101	145	151	227	201	311
2	2	52	64	102	146	152	230	202	312
3	3	53	65	103	147	153	231	203	313
4	4	54	66	104	150	154	232	204	314
5	5	55	67	105	151	155	233	205	315
6	6	56	70	106	152	156	234	206	316
7	7	57	71	107	153	157	235	207	317
8	10	58	72	108	154	158	236	208	320
9	11	59	73	109	155	159	237	209	321
10	12	60	74	110	156	160	240	210	322
11	13	61	75	111	157	161	241	211	323
12	14	62	76	112	160	162	242	212	324
13	15	63	77	113	161	163	243	213	325
14	16	64	100	114	162	164	244	214	326
15	17	65	101	115	163	165	245	215	327
16	20	66	102	116	164	166	246	216	330
17	21	67	103	117	165	167	247	217	331
18	22	68	104	118	166	168	250	218	332
19	23	69	105	119	167	169	251	219	333
20	24	70	106	120	170	170	252	220	334
21	25	71	107	121	171	171	253	221	335
22	26	72	110	122	172	172	254	222	336
23	27	73	111	123	173	173	255	223	337
24	30	74	112	124	174	174	256	224	340
25	31	75	113	125	175	175	257	225	341
26	32	76	114	126	176	176	260	226	342
27	33	77	115	127	177	177	261	227	343
28	34	78	116	128	200	178	262	228	344
29	35	79	117	129	201	179	263	229	345
30	36	80	120	130	202	180	264	230	346
31	37	81	121	131	203	181	265	231	347
32	40	82	122	132	204	182	266	232	350
33	41	83	123	133	205	183	267	233	351
34	42	84	124	134	206	184	270	234	352
35	43	85	125	135	207	185	271	235	353
36	44	86	126	136	210	186	272	236	354
37	45	87	127	137	211	187	273	237	355
38	46	88	130	138	212	188	274	238	356
39	47	89	131	139	213	189	275	239	357
40	50	90	132	140	214	190	276	240	360
41	51	91	133	141	215	191	277	241	361
42	52	92	134	142	216	192	300	242	362
43	53	93	135	143	217	193	301	243	363
44	54	94	136	144	220	194	302	244	364
45	55	95	137	145	221	195	303	245	365
46	56	96	140	146	222	196	304	246	366
47	57	97	141	147	223	197	305	247	367
48	60	98	142	148	224	198	306	248	370
49	61	99	143	149	225	199	307	249	371
								250	372
								251	373
								252	374
								253	375
								254	376
								255	377
								256	400

Ø Ø-255
4ØØ 256-511 (SUBTRACT 256)
1ØØØ 512-767 (SUBTRACT 512)
14ØØ 768-1023 (SUBTRACT 768)