HEWLETT-PACKARD

HP-41C

STAT PAC

NOTICE

INTRODUCTION

The programs in the Stat Pac have been drawn from the fields of general statistics, analysis of variance, regression, test statistics, and distribution functions.

Each program in this pac is represented by one program in the Application Module and a section in this manual. The manual provides a description of the program with relevant equations, a set of instructions for using the program, and one or more example problems, each of which includes a list of the keystrokes required for its solution.

Before plugging in your Application Module, turn your calculator off, and be sure you understand the section "Inserting and Removing Application Modules." Before using a particular program, take a few minutes to read "Format of User Instructions" and "A Word About Program Usage."

You should first familiarize yourself with a program by running it once or twice while following the complete User Instructions in the manual. Thereafter, the program's prompting should provide the necessary instructions, including which variables are to be input, which keys are to be pressed, and which values will be output. A quick-reference card with a brief description of each program's operating instructions has been provided for your convenience.

We hope the Stat Pac will assist you in the solution of numerous problems in your discipline. If you have technical problems with this Pac, refer to your HP-41 owner's handbook for information on Hewlett-Packard "technical support" or "programming assistance."

Note: Application modules are designed to be used in all HP-41 model calculators. The term "HP-41C" is used throughout the rest of this manual, unless otherwise specified, to refer to all HP-41 calculators.

CONTENTS

Test Statistics

Distribution Functions

INSERTING AND REMOVING APPLICATION MODULES

Before you insert an application module for the first time, familiarize yourself with the following information.

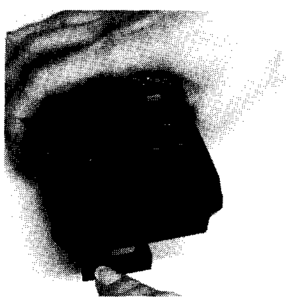
Up to four application modules can be plugged into the ports on the HP-41C. While plugged in, the names of all programs contained in the module can be displayed by pressing ■ CATALOG 2.

> **CAUTION**
>
> Always turn the HP-41C off before inserting or removing any plug-in extensions or accessories. Failure to turn the HP-41C off could damage both the calculator and the accessory.

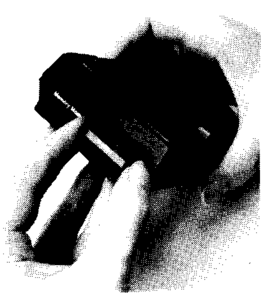
Here is how you should insert application modules:

1. Turn the HP-41C off! Failure to turn the calculator off could damage both the module and the calculator.

2. Remove the port covers. Remember to save the port covers, they should be inserted into the empty ports when no extensions are inserted.



3. With the application module label facing downward as shown, insert the application module into any port **after** the last memory module presently inserted.



4. If you have additional application modules to insert, place them into any port after the last memory module. For example, if you have a memory module inserted in port 1, you can insert application modules in any of ports 2, 3, or 4. **Never insert an application module into a lower numbered port than a memory module.** Be sure to place port covers over unused ports.
5. Turn the calculator on and follow the instructions given in this book for the desired application functions.

To remove application modules:

1. Turn the HP-41C off! Failure to do so could damage both the calculator and the module.
2. Grasp the desired module handle and pull it out as shown.

3. Place a port cap into the empty port.

Mixing Memory Modules and Application Modules

Any time you wish to insert other extensions (such as the HP-82104A Card Reader, or the HP-82143A Printer) the HP-41C has been designed so that the memory modules are in lower numbered ports.

So, when you are using both memory modules and application modules, the memory modules must always be inserted into the lower numbered ports and the application module into any port after the last memory module. When mixing memory and application modules, the HP-41C allows you to leave gaps in the port sequence. For example, you can plug a memory module into port 1 and an application module into port 4, leaving ports 2 and 3 empty.

FORMAT OF USER INSTRUCTIONS

The completed User Instruction Form—which accompanies each program—is your guide to operating the programs in this Pac.

The form is composed of five labeled columns. Reading from left to right, the first column, labeled STEP, gives the instruction step number.

The INSTRUCTIONS column gives instructions and comments concerning the operations to be performed.

The INPUT column specifies the input data, the units of data if applicable, or the appropriate alpha response to a prompted question. Data Input keys consists of 0 to 9 and the decimal point (the numeric keys), [EEX] (enter exponent), and [CHS] (change sign).

The FUNCTION column specifies the keys to be pressed after keying in the corresponding input data.

Whenever a statement in the INPUT or FUNCTION column is printed in gold, the [ALPHA] key must be pressed before the statement can be keyed in. After the statement is keyed in, press [ALPHA] again to return the calculator to its normal operating mode, or to begin program execution. For example, [XEQ]ΣBSTAT means press the following keys: [XEQ] [ALPHA] ΣBSTAT [ALPHA].

The DISPLAY column specifies prompts, intermediate and final answers and their units, where applicable.

Above the DISPLAY column is a box which specifies the minimum number of registers necessary to execute the program. Refer to pages 73 and 117 in the Owner's Handbook for a complete description of how to size calculator memory.

A WORD ABOUT PROGRAM USAGE

Catalog

When an Application Module is plugged into a port of the HP-41C, the contents of the Module can be reviewed by pressing ■ [CATALOG] 2 (the Extension Catalog). Executing the [CATALOG] function lists the name of each global label in the module, as well as functions of any other extensions which might be plugged in. Remember that the catalog function lists the extension in port 1 first, followed by the extensions in ports 2-4.

ALPHA and USER Mode Notation

This manual uses a special notation to signify ALPHA mode. Whenever a statement on the User Instruction Form is printed in gold, the [ALPHA] key must be pressed before the statement can be keyed in. After the statement is input, press [ALPHA] again to return the calculator to its normal operating mode, or to begin program execution. For example, [XEQ]ΣBSTAT means press the following keys: [XEQ] [ALPHA] ΣBSTAT [ALPHA].

In USER mode, when referring to the top two rows of keys (the keys have been re-defined), this manual will use the symbols [A], [C], [E], ■ [A] and [R/S] on the User Instruction Form and in the keystroke solutions to sample problems.

Using Optional Printer

When the optional printer is plugged into the HP-41C along with this Applications Module, all results will be printed automatically. You may also want to keep a permanent record of the values input to a certain program. A convenient way to do this is to set the Print Mode switch to NORMAL before running the program. In this mode, all input values and the corresponding keystrokes will be listed on the printer, thus providing a record of the entire operation of the program.

Downloading Module Programs

If you wish to trace execution, to modify, to record on magnetic cards, or to print a program in this Application Module, it must first be copied into the HP-41C's program memory. For information concerning the HP-41C COPY function, see the Owner's Handbook. It is *not* necessary to copy a program in order to run it.

Program Interruption

These programs have been designed to operate properly when run from beginning to end, without turning the calculator off (remember, the calculator may turn itself off). If the HP-41C is turned off, it may be necessary to set flag 21 (SF 21) to continue proper execution.

Use of Labels

You should generally avoid writing programs into the calculator memory that use program labels identical to those in your Application Module. In case of a label conflict, the label within program memory has priority over the label within the Application Pac program. All program labels used in this Pac are listed in appendix B, "Program Labels."

Key Assignments

If you have customized your keyboard with the ASN function, those reassignments will take precedence over the local labels A, C, and E used in this Pac.

Flag 03

If flag 03 is set when a Stat Pac program is executed, the statistical registers may not be cleared and incorrect results may occur.

BASIC STATISTICS FOR TWO VARIABLES

This program calculates means, standard deviations, covariance, correlation coefficient, coefficients of variation, sums of data points, sum of multiplication of data points, and sums of squares of data points derived from a set of ungrouped data points $\{(x_i, y_i), i = 1, 2, \ldots, n\}$, or grouped data points $\{(x_i, y_i, f_i), i = 1, 2, \ldots, n\}$. f_i denotes the frequency of repetition of (x_i, y_i).

$$\text{means } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\text{standard deviations } s_x = \sqrt{\frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}} \quad \text{(of sample)}$$

$$\text{or } s_x{}' = \sqrt{\frac{\Sigma x_i^2 - n\bar{x}^2}{n}} \quad \text{(of population)}$$

$$s_y = \sqrt{\frac{\Sigma y_i^2 - n\bar{y}^2}{n-1}} \quad \text{(of sample)}$$

$$\text{or } s_y{}' = \sqrt{\frac{\Sigma y_i^2 - n\bar{y}^2}{n}} \quad \text{(of population)}$$

$$\text{covariance} \quad s_{xy} = \frac{1}{n-1}\left(\Sigma x_i y_i - \frac{1}{n}\Sigma x_i \Sigma y_i\right) \quad \text{(of sample)}$$

$$\text{or } s_{xy}{}' = \frac{1}{n}\left[\Sigma x_i y_i - \frac{1}{n}\Sigma x_i \Sigma y_i\right] \quad \text{(of population)}$$

$$\text{correlation coefficient } \gamma_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$\text{Coefficients of variation } V_x = \frac{s_x}{\bar{x}} \cdot 100, \qquad V_y = \frac{s_y}{\bar{y}} \cdot 100$$

Note n is a positive integer and $n > 1$.

				SIZE: 012
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
	Ungrouped Data			
1.	Initialize the program.		[XEQ] ΣBSTAT	ΣBSTAT
2.	Repeat step 2~3 for i=1,2,...,n. Input: x_i y_i	 x_i y_i	 [ENTER↑] [A]	 (i)
3.	If you made a mistake in inputting x_k and y_k, then correct by	x_k y_k	[ENTER↑] [C]	 (k−1)
4.	Go to step 8 for basic statistic calculations.			
	Grouped Data			
5.	Initialize the program.		[XEQ] ΣBSTG	ΣBSTG
6.	Repeat step 6~7 for i=1,2,...,n. Input: x_i y_i f_i	 x_i y_i f_i	 [ENTER↑] [ENTER↑] [A]	 (Σf_i)
7.	If you made a mistake in inputting x_k, y_k, and f_k, then correct by	x_k y_k f_k	[ENTER↑] [ENTER↑] [C]	 $(\Sigma f_i - f_k)$
8.	To calculate basic statistics: $\bar{x}$ $\bar{y}$ s_x s'_x s_y s'_y v_x* v_y* s_{xy} s'_{xy} γ_{xy} Σx_i Σy_i $\Sigma x_i y_i$ Σx_i^2 Σy_i^2		 [E] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S]	 XBAR=$(\bar{x})$ YBAR=$(\bar{y})$ SX=(s_x) SX.=(s'_x) SY=(s_y) SY.=(s'_y) VX=(v_x) VY=(v_y) SXY=(s_{xy}) SXY.=(s'_{xy}) GXY=(γ_{xy}) ΣX=(Σx_i) ΣY=(Σy_i) ΣXY=$(\Sigma x_i y_i)$ ΣX2=(Σx_i^2) ΣY2=(Σy_i^2)
9.	Repeat step 8 if you want the results again.			
10.	To use the same program for another set of data, initialize the program by → then go to step 2 or step 6.		[■] [A]	ΣBSTAT or ΣBSTG
11.	To use the other program, go to step 1 or step 5.			

NOTE: "DATA ERROR" will be displayed if $\bar{x}$ or $\bar{y}$ is zero. Press [R/S] and proceed.

Example 1:

For the following set of data, find the means, standard deviations, covariance, correlation coefficient, coefficients of variation, and the sums.

x_i	26	30	44	50	62	68	74
y_i	92	85	78	81	54	51	40

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 012	
[XEQ] [ALPHA] ΣBSTAT [ALPHA]	***ΣBSTAT***
26 [ENTER↑] 92 [A]	
100 [ENTER↑] 100 [A]	
100 [ENTER↑] 100 [C]	
30 [ENTER↑] 85 [A]	
44 [ENTER↑] 78 [A]	
50 [ENTER↑] 81 [A]	
62 [ENTER↑] 54 [A]	
68 [ENTER↑] 51 [A]	
74 [ENTER↑] 40 [A]	***7.00***
[E]	***XBAR=50.57***
[R/S]	***YBAR=68.71***
[R/S]	***SX=18.50***
[R/S]	***SX.=17.13***
[R/S]	***SY=20.00***
[R/S]	***SY.=18.51***
[R/S]	***VX=36.58***
[R/S]	***VY=29.10***
[R/S]	***SXY=-354.14***
[R/S]	***SXY.=-303.55***
[R/S]	***GXY=-0.96***
[R/S]	***ΣX=354.00***
[R/S]	***ΣY=481.00***
[R/S]	***ΣXY=22200.00***
[R/S]	***ΣX2=19956.00***
[R/S]	***ΣY2=35451.00***

Example 2:

Apply the program to the following set of grouped data.

x_i	4.8	5.2	3.8	4.4	4.1
y_i	15.1	11.5	14.3	13.6	12.8
f_i	1	3	1	6	2

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 012	
[XEQ] [ALPHA] ΣBSTG [ALPHA]	*ΣBSTG*
4.8 [ENTER↑] 15.1 [ENTER↑] 1 [A]	
5.2 [ENTER↑] 11.5 [ENTER↑] 3 [A]	
3.8 [ENTER↑] 14.3 [ENTER↑] 1 [A]	
4.4 [ENTER↑] 13.6 [ENTER↑] 6 [A]	
4.1 [ENTER↑] 12.8 [ENTER↑] 2 [A]	***13.00***
[E]	***XBAR=4.52***
[R/S]	***YBAR=13.16***
[R/S]	***SX=0.45***
[R/S]	***SX.=0.43***
[R/S]	***SY=1.11***
[R/S]	***SY.=1.07***
[R/S]	***VX=9.93***
[R/S]	***VY=8.42***
[R/S]	***SXY=-0.31***
[R/S]	***SXY.=-0.28***
[R/S]	***GXY=-0.62***
[R/S]	***ΣX=58.80***
[R/S]	***ΣY=171.10***
[R/S]	***ΣXY=770.22***
[R/S]	***ΣX2=268.38***
[R/S]	***ΣY2=2266.69***

MOMENTS, SKEWNESS AND KURTOSIS (FOR GROUPED OR UNGROUPED DATA)

For grouped or ungrouped data, moments are used to describe sets of data, skewness is used to measure the lack of symmetry in a distribution, and kurtosis is the relative peakness or flatness of a distribution. For a given set of data

$$\{x_1, x_2, \ldots, x_n\}:$$

1st moment $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

2nd moment $m_2 = \frac{1}{n}\Sigma x_i^2 - \bar{x}^2$

3rd moment $m_3 = \frac{1}{n}\Sigma x_i^3 - \frac{3}{n}\bar{x}\,\Sigma x_i^2 + 2\bar{x}^3$

4th moment $m_4 = \frac{1}{n}\Sigma x_i^4 - \frac{4}{n}\bar{x}\,\Sigma x_i^3 + \frac{6}{n}\bar{x}^2\,\Sigma x_i^2 - 3\bar{x}^4$

Moment coefficient of skewness

$$\gamma_1 = \frac{m_3}{m_2^{3/2}}$$

Moment coefficient of kurtosis

$$\gamma_2 = \frac{m_4}{m_2^2}$$

This program also provides the option for calculating those statistics for grouped data (using similar formulas as for ungrouped data):

data	x_1	x_2	...	x_m
frequency	f_1	f_2	...	f_m

Note that for this case, 1st moment

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i\, x_i}{\sum_{i=1}^{m} f_i}$$

Reference:

Theory and Problems of Statistics, M.R. Spiegel, Schaum's Outline, McGraw-Hill, 1961

				SIZE: 012
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
	Ungrouped Data			
1.	Initialize the program.		XEQ ΣMMTUG	ΣMMTUG
2.	Repeat step 2~3 for i=1,2,...,n. Input x_i.	x_i	A	(i)
3.	If you made a mistake in inputting x_k, then correct by	x_k	C	(k−1)
4.	Go to step 8 for moments calculations.			
	Grouped Data			
5.	Initialize the program.		XEQ ΣMMTGD	ΣMMTGD
6.	Repeat step 6~7 for j=1,2,...,m Input: x_j f_j	x_j f_j	ENTER↑ A	 (j)
7.	If you made a mistake in inputting x_h and f_h, then correct by	x_h f_h	ENTER↑ C	 (h−1)
8.	Calculate moments etc.: $\bar{x}$ m_2 m_3 m_4 γ_1 γ_2		 E R/S R/S R/S R/S R/S	 XBAR=($\bar{x}$) M2=(m_2) M3=(m_3) M4=(m_4) GM1=(γ_1) GM2=(γ_2)
9.	Repeat step 8 if you want the results again.			
10.	To use the same program for another set of data, initialize the program by → then go to step 2 or step 6.		[illegible] A	ΣMMTUG or ΣMMTGD
11.	To use the other program, go to step 1 or step 5.			

Examples:

1. Ungrouped data

i	1	2	3	4	5	6	7	8	9
x_i	2.1	3.5	4.2	6.5	4.1	3.6	5.3	3.7	4.9

$$\bar{x} = 4.21,\ m_2 = 1.39,\ m_3 = 0.39,\ m_4 = 5.49$$

$$\gamma_1 = 0.24,\ \gamma_2 = 2.84$$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 012	
[XEQ] [ALPHA] ΣMMTUG [ALPHA]	***ΣMMTUG***
2.1 [A] 3.5 [A] 4.0 [A] 4.0 [C]	
4.2 [A] 6.5 [A] 4.1 [A] 3.6 [A]	
5.3 [A] 3.7 [A] 4.9 [A]	***9.00***
[E]	***XBAR=4.21***
[R/S]	***M2=1.39***
[R/S]	***M3=0.39***
[R/S]	***M4=5.49***
[R/S]	***GM1=0.24***
[R/S]	***GM2=2.84***

2. Grouped data

i	1	2	3	4	5
x_i	3	2	4	6	1
f_i	4	5	3	2	1

$$\bar{x} = 3.13,\ m_2 = 1.98,\ m_3 = 2.14,\ m_4 = 11.05$$

$$\gamma_1 = 0.77,\ \gamma_2 = 2.81$$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 012	
[XEQ] [ALPHA] ΣMMTGD [ALPHA]	*ΣMMTGD*
3 [ENTER↑] 4 [A] 2 [ENTER↑] 5 [A]	
4 [ENTER↑] 4 [A] 4 [ENTER↑] 4 [C]	
4 [ENTER↑] 3 [A] 6 [ENTER↑] 2 [A]	
1 [ENTER↑] 1 [A]	*5.00*
[E]	*XBAR=3.13*
[R/S]	*M2=1.98*
[R/S]	*M3=2.14*
[R/S]	*M4=11.05*
[R/S]	*GM1=0.77*
[R/S]	*GM2=2.81*

ANALYSIS OF VARIANCE (ONE WAY)

The one-way analysis of variance is used to test if observed differences among k sample means can be attributed to chance or whether they are indicative of actual differences among the corresponding population means. Suppose the i^{th} sample has n_i observations (samples may have equal or unequal number of observations). The null hypothesis we want to test is that the k population means are all equal. This program generates the complete ANOVA table.

1. Mean of observations in the i^{th} sample ($i = 1, 2, \ldots, k$)

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

2. Standard deviation of observations in the i^{th} sample

$$s_i = \left[\left(\sum_{j=1}^{n_i} x_{ij}^2 - n_i\bar{x}_i^2\right)\Big/(n_i - 1)\right]^{1/2}$$

3. Sum of observations in the i^{th} sample

$$\text{Sum}_i = \sum_{j=1}^{n_i} x_{ij}$$

4. Total sum of squares

$$\text{TSS} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum_{i=1}^{k} n_i}$$

5. Treatment sum of squares

$$\text{TrSS} = \sum_{i=1}^{k}\frac{\left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i} - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum_{i=1}^{k} n_i}$$

6. Error sum of squares

$$ESS = TSS - TrSS$$

7. Treatment degrees of freedom

$$df_1 = k - 1$$

8. Error degrees of freedom

$$df_2 = \sum_{i=1}^{k} n_i - k$$

9. Total degrees of freedom

$$df_3 = df_1 + df_2 = \sum_{i=1}^{k} n_i - 1$$

10. Treatment mean square

$$TrMS = \frac{TrSS}{df_1}$$

11. Error mean square

$$EMS = \frac{ESS}{df_2}$$

12. The F ratio

$$F = \frac{TrMS}{EMS} \text{(with degrees of freedom } df_1, df_2\text{)}$$

Reference:

J.E. Freund, *Mathematical Statistics*, Prentice Hall, 1962.

STEP	INSTRUCTIONS	INPUT	FUNCTION	SIZE: 020 DISPLAY
1.	Initialize the program.		[XEQ] ΣAOVONE	ΣAOVONE
2.	Repeat step 2~5 for i=1,2,...,k.			
3.	Repeat step 3~4 for j=1,2,...,n_i. Input x_{ij}.	x_{ij}	[A]	(j)
4.	If you made a mistake in inputting x_{im}, then correct by	x_{im}	[C]	(m−1)
5.	Calculate: mean $\bar{x}_i$ standard deviation s_i sum Sum_i		[R/S] [R/S] [R/S]	XBAR=($\bar{x}_i$) S=(s_i) SUM=(Sum_i)
6.	To calculate ANOVA Table: TSS TrSS ESS df_i df_2 df_3 TrMS EMS F		 [E] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S]	 TSS=(TSS) TRSS=(TrSS) ESS=(ESS) DF1=(df_1) DF2=(df_2) DF3=(df_3) TRMS=(TrMS) EMS=(EMS) F=(F)
7.	Repeat step 6 if you want the results again.			
8.	For another set of data, initialize the program by → then go to step 2.		■ [A]	ΣAOVONE

Example:

The following random samples of achievement test scores were obtained from students at four different schools:

i \ j	1	2	3	4	5	6	7
School 1	88	99	96	68	85		
School 2	78	62	98	83	61	88	
School 3	80	61	74	92	78	54	77
School 4	71	65	90	46			

Calculate the ANOVA table and test the null hypothesis that the differences among the sample means can be attributed to chance. Use significance level $\alpha = 0.01$.

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 020	
[XEQ] [ALPHA] ΣAOVONE [ALPHA]	ΣAOVONE
88 [A] 99 [A] 96 [A] 68 [A]	
85 [A]	5.00
[R/S]	XBAR=87.20
[R/S]	S=12.15
[R/S]	SUM=436.00
78 [A] 62 [A] 98 [A] 83 [A]	
61 [A] 88 [A]	6.00
[R/S]	XBAR=78.33
[R/S]	S=14.62
[R/S]	SUM=470.00
80 [A] 61 [A] 74 [A] 92 [A]	
78 [A] 54 [A] 77 [A]	7.00
[R/S]	XBAR=73.71
[R/S]	S=12.61
[R/S]	SUM=516.00
71 [A] 66 [A] 66 [C] 65 [A]	
90 [A] 46 [A]	4.00
[R/S]	XBAR=68.00
[R/S]	S=18.13
[R/S]	SUM=272.00
[E]	TSS=4530.00
[R/S]	TRSS=930.44
[R/S]	ESS=3599.56
[R/S]	DF1=3.00
[R/S]	DF2=18.00
[R/S]	DF3=21.00
[R/S]	TRMS=310.15
[R/S]	EMS=199.98
[R/S]	F=1.55

ANOVA Table

	SS	df	MS	F
Treatments	930.44	3	310.15	1.55
Error	3599.56	18	199.98	
Total	4530.00	21		

Since F = 1.55 does not exceed $F_{.01,3,18}$ = 5.09, the null hypothesis can not be rejected. Thus we have no evidence to conclude that the means of the scores for the four schools are significantly different.

ANALYSIS OF VARIANCE (TWO WAY, NO REPLICATIONS)

The analysis of variance is the analysis of the total variability of a set of data (measured by their total sum of squares) into components which can be attributed to different sources of variation.

The two way analysis of variance tests the row effects and the column effects independently. This program will generate the ANOVA table for the case such that (1) each cell only has one observation and (2) the row and column effects do not interact.

Equations:

1. Sums

$$\text{Row } RS_i = \sum_j x_{ij} \qquad i = 1, 2, \ldots, r$$

$$\text{Column } CS_j = \sum_i x_{ij} \qquad j = 1, 2, \ldots, c$$

2. Sums of squares

$$\text{Total } TSS = \Sigma\Sigma x_{ij}^2 - (\Sigma\Sigma x_{ij})^2/rc$$

$$\text{Row } RSS = \sum_i \left(\sum_j x_{ij}\right)^2 /c - (\Sigma\Sigma x_{ij})^2/rc$$

$$\text{Column } CSS = \sum_j \left(\sum_i x_{ij}\right)^2 /r - (\Sigma\Sigma x_{ij})^2/rc$$

$$\text{Error } ESS = TSS - RSS - CSS$$

3. Degrees of freedom

$$\text{Row } df_1 = r - 1$$

$$\text{Column } df_2 = c - 1$$

$$\text{Error } df_3 = (r - 1)(c - 1)$$

4. F ratios

$$\text{Row } F_1 = \frac{RSS}{df_1} \Big/ \frac{ESS}{df_3}$$

$$\text{Column } F_2 = \frac{CSS}{df_2} \Big/ \frac{ESS}{df_3}$$

Reference:

Dixon and Massey, *Introduction to Statistical Analysis,* McGraw-Hill, 1969.

SIZE: 018

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
1.	Initialize the program.		[XEQ] ΣAOVTWO	ΣAOVTWO
	Row-Wise			
2.	Repeat step 2~5 for i=1,2,...,r.			
3.	Repeat step 3~4 for j=1,2,...,c. Input x_{ij}	x_{ij}	[A]	(j)
4.	If you made a mistake in inputting x_{im}, then correct by	x_{im}	[C]	(m−1)
5.	Calculate row sum and initialize the program for the next row.		[R/S]	SUM=(RS_i)
6.	After completion of the last row, (row r) initialize the program for column-wise data entry.		[R/S]	COLUMN-WISE
	Column-Wise			
7.	Repeat step 7~10 for j=1,2,...,c.			
8.	Repeat step 8~9 for i=1,2,...,r. Input x_{ij}.	x_{ij}	[A]	(i)
9.	If you made a mistake in inputting x_{hj}, then correct by	x_{hj}	[C]	(h−1)
10.	Calculate column sum and initialize the program for the next column.		[R/S]	SUM=(CS_j)
11.	Calculate ANOVA Table: RSS CSS TSS ESS df_1 df_2 df_3 F_1 F_2		 [E] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S] [R/S]	 RSS=(RSS) CSS=(CSS) TSS=(TSS) ESS=(ESS) DF1=(df_1) DF2=(df_2) DF3=(df_3) F1=(F_1) F2=(F_2)
12.	Repeat step 11 if you want the results again			
13.	For another set of data, initialize the program by → then go to step 2		■ [A]	ΣAOVTWO

Example:

Apply this program to analyze the following set of data.

i \ j		Column 1	2	3	4
Row	1	7	6	8	7
	2	2	4	4	4
	3	4	6	5	3

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 018	
[XEQ] [ALPHA] ΣAOVTWO [ALPHA]	***ΣAOVTWO***
7 [A] 6 [A] 8 [A] 7 [A]	***4.00***
[R/S]	***SUM=28.00***
2 [A] 4 [A] 4 [A] 4 [A]	***4.00***
[R/S]	***SUM=14.00***
4 [A] 7 [A] 7 [C] 6 [A] 5 [A]	
3 [A]	***4.00***
[R/S]	***SUM=18.00***
[R/S]	***COLUMN-WISE***
7 [A] 2 [A] 4 [A]	***3.00***
[R/S]	***SUM=13.00***
6 [A] 4 [A] 6 [A]	***3.00***
[R/S]	***SUM=16.00***
8 [A] 4 [A] 5 [A]	***3.00***
[R/S]	***SUM=17.00***
7 [A] 4 [A] 3 [A]	***3.00***
[R/S]	***SUM=14.00***
[E]	***RSS=26.00***
[R/S]	***CSS=3.33***
[R/S]	***TSS=36.00***
[R/S]	***ESS=6.67***
[R/S]	***DF1=2.00***
[R/S]	***DF2=3.00***
[R/S]	***DF3=6.00***
[R/S]	***F1=11.70***
[R/S]	***F2=1.00***

ANOVA

	SS	df	F ratio
Row	26.00	2	11.70
Column	3.33	3	1.00
Error	6.67	6	
Total	36.00		

ANALYSIS OF COVARIANCE (ONE WAY)

The one way analysis of covariance program tests the effect of one variable separately from the effect of a second variable, if the second variable represents an actual measurement for each individual (rather than a category).

Suppose (x_{ij}, y_{ij}) represents the j^{th} observation from the i^{th} population ($i = 1,2, \ldots, k, j = 1,2, \ldots, n_i$). Note that samples may have equal or unequal number of observations. The analysis of covariance tests for a difference in means of residuals. The residuals are the differences of the observations and a regression quantity based on the associated second variable. The analysis of covariance procedure is based on the separations of the sums of squares and the sums of products into several portions. This program will generate the complete ANOCOV table.

Equations:

1. Sums and sums of squares

$$Sx_i = \sum_j x_{ij} \ (i = 1,2,\ldots,k)$$

$$TSSx = \Sigma\Sigma x_{ij}^2 - \frac{(\Sigma\Sigma x_{ij})^2}{\sum_i n_i}$$

$$ASSx = \sum_i \frac{\left(\sum_j x_{ij}\right)^2}{n_i} - \frac{(\Sigma\Sigma x_{ij})^2}{\sum_i n_i}$$

$$WSSx = TSSx - ASSx$$

2. Degrees of freedom

$$df_1 = k - 1$$

$$df_2 = \sum_i n_i - k$$

3. Mean squares and F statistic

$$AMSx = \frac{ASSx}{df_1}$$

$$WMSx = \frac{WSSx}{df_2}$$

$$F_x = \frac{AMSx}{WMSx} \text{ with degrees of freedom } df_1, df_2$$

By changing x_{ij} to y_{ij}, similar formulas for y_{ij} can be obtained.

4. Sums of products

$$TSP = \Sigma\Sigma x_{ij}\, y_{ij} - \frac{(\Sigma\Sigma x_{ij})\,(\Sigma\Sigma y_{ij})}{\sum_i n_i}$$

$$ASP = \sum_i \frac{\left(\sum_j x_{ij}\right)\left(\sum_j y_{ij}\right)}{n_i} - \frac{(\Sigma\Sigma x_{ij})\,(\Sigma\Sigma y_{ij})}{\sum_i n_i}$$

$$WSP = TSP - ASP$$

5. Residual sums of squares

$$TSS\hat{y} = TSSy - \frac{(TSP)^2}{TSSx}$$

$$WSS\hat{y} = WSSy - \frac{(WSP)^2}{WSSx}$$

$$ASS\hat{y} = TSS\hat{y} - WSS\hat{y}$$

6. Residual degrees of freedom

$$df_3 = k - 1$$

$$df_4 = \sum_i n_i - k - 1$$

7. Residual mean squares and F statistic

$$AMS\hat{y} = \frac{ASS\hat{y}}{df_3}$$

$$WMS\hat{y} = \frac{WSS\hat{y}}{df_4}$$

$$F = \frac{AMS\hat{y}}{WMS\hat{y}} \text{ with degrees of freedom } df_3, df_4$$

ANOCOV Table

	degrees of freedom	SSx	SP	SSy	Residuals degrees of freedom	Residuals SSŷ	Residuals MSŷ	Residuals F statistic
Among means	df_1	ASSx	ASP	ASSy	df_3	ASSŷ	AMSŷ	F
Within groups	df_2	WSSx	WSP	WSSy	df_4	WSSŷ	WMSŷ	
Total		TSSx	TSP	TSSy		TSSŷ		

Remarks:

- F_x can be used to test if the X means are equal (ANOVA for X).
- F_y can be used to test if the Y means (not making use of the X values) are equal (ANOVA for unadjusted Y).

Reference:

Dixon and Massey, *Introduction to Statistical Analysis*, McGraw-Hill, 1969.

				SIZE: 026
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
1.	Initialize the program.		[XEQ] ΣANOCOV	ΣANOCOV (pse) NEW I=1.00
2.	Repeat step 2~6 for i=1,2,...,k.			
3.	Repeat step 3~4 for $j=1,2,\ldots,n_j$. Input x_{ij} and y_{ij}.	 x_{ij} y_{ij}	 [ENTER↑] [A]	 (j)
4.	If you made a mistake in inputting x_{im} and y_{im}, then correct by	 x_{im} y_{im}	 [ENTER↑] [C]	 (m−1)
5.	Calculate the i^{th} sums: Sx_i Sy_i		[R/S] [R/S]	$SX=(Sx_i)$ $SY=(Sy_i)$
6.	Initialize for new i.		[R/S]	NEW I=(i)
7.	To calculate ANOCOV Table: TSSx ASSx WSSx TSSy ASSy WSSy df_1 df_2 Fx Fy TSP ASP WSP TSSŷ WSSŷ ASSŷ df_3 df_4 AMSŷ WMSŷ F		 [E] [R/S]	 TSSX=(TSSx) ASSX=(ASSx) WSSX=(WSSx) TSSY=(TSSy) ASSY=(ASSy) WSSY=(WSSy) $DF1=(df_1)$ $DF2=(df_2)$ FX=(Fx) FY=(Fy) TSP=(TSP) ASP=(ASP) WSP=(WSP) TSSY.=(TSSŷ) WSSY.=(WSSŷ) ASSY.=(ASSŷ) $DF3=(df_3)$ $DF4=(df_4)$ AMSY.=(AMSŷ) WMSY.=(WMSŷ) F=(F)
8.	Repeat step 7 if you want the results again.			
9.	For another set of data, initialize the program by → then go to step 2.		■ [A]	ΣANOCOV (pse) NEW I=1.00

Example:

		j: 1	2	3	4
i = 1	x	3	2	1	2
	y	10	8	8	11
i = 2	x	4	3	3	5
	y	12	12	10	13
i = 3	x	1	2	3	1
	y	6	5	8	7

$(k = 3, n_1 = n_2 = n_3 = 4)$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 026	
[XEQ] [ALPHA] ΣANOCOV [ALPHA]	*ΣANOCOV (Pse)*
	NEW I=1.00
3 [ENTER↑] 10 [A] 2 [ENTER↑] 8 [A]	
5 [ENTER↑] 5 [A] 5 [ENTER↑] 5 [C]	
1 [ENTER↑] 8 [A] 2 [ENTER↑] 11 [A]	*4.00*
[R/S]	*SX=8.00*
[R/S]	*SY=37.00*
[R/S]	*NEW I=2.00*
4 [ENTER↑] 12 [A] 3 [ENTER↑] 12 [A]	
3 [ENTER↑] 10 [A] 5 [ENTER↑] 13 [A]	*4.00*
[R/S]	*SX=15.00*
[R/S]	*SY=47.00*
[R/S]	*NEW I=3.00*
1 [ENTER↑] 6 [A] 2 [ENTER↑] 5 [A]	
3 [ENTER↑] 8 [A] 1 [ENTER↑] 7 [A]	*4.00*
[R/S]	*SX=7.00*
[R/S]	*SY=26.00*
[R/S]	*NEW I=4.00*
[E]	*TSSX=17.00*
[R/S]	*ASSX=9.50*
[R/S]	*WSSX=7.50*
[R/S]	*TSSY=71.67*
[R/S]	*ASSY=55.17*
[R/S]	*WSSY=16.50*
[R/S]	*DF1=2.00*
[R/S]	*DF2=9.00*
[R/S]	*FX=5.70*

[R/S]	***FY=15.05***
[R/S]	***TSP=27.00***
[R/S]	***ASP=20.75***
[R/S]	***WSP=6.25***
[R/S]	***TSSY.=28.78***
[R/S]	***WSSY.=11.29***
[R/S]	***ASSY.=17.49***
[R/S]	***DF3=2.00***
[R/S]	***DF4=8.00***
[R/S]	***AMSY.=8.75***
[R/S]	***WMSY.=1.41***
[R/S]	***F=6.20***

ANOCOV Table

					Residuals			
	df	SSx	SP	SSy	df	SSŷ	MSŷ	F
Among means	2	9.50	20.75	55.17	2	17.49	8.75	6.20
Within groups	9	7.50	6.25	16.50	8	11.29	1.41	
Total		17.00	27.00	71.67		28.78		

CURVE FITTING

For a set of data points (x_i, y_i), $i = 1, 2, \ldots, n$, this program can be used to fit the data to any of the following curves:

1. Straight line (linear regression); $y = a + bx$.
2. Exponential curve; $y = ae^{bx}$ $(a > 0)$.
3. Logarithmic curve; $y = a + b \ln x$.
4. Power curve; $y = ax^b$ $(a > 0)$.

The regression coefficients a and b are found from solving the following system of linear equations:

$$\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix} \begin{bmatrix} A \\ b \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma Y_i X_i \end{bmatrix}$$

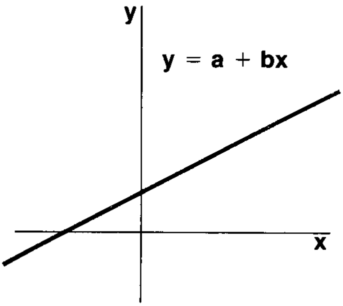
where the variables are defined as follows:

Regression	**A**	$\mathbf{X_i}$	$\mathbf{Y_i}$
Linear	a	x_i	y_i
Exponential	ln a	x_i	$\ln y_i$
Logarithmic	a	$\ln x_i$	y_i
Power	ln a	$\ln x_i$	$\ln y_i$

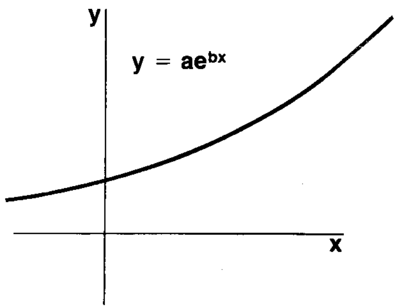
The coefficient of determination is:

$$R^2 = \frac{A\Sigma Y_i + b\Sigma X_i\, Y_i - \frac{1}{n}(\Sigma Y_i)^2}{\Sigma(Y_i^2) - \frac{1}{n}(\Sigma Y_i)^2}$$

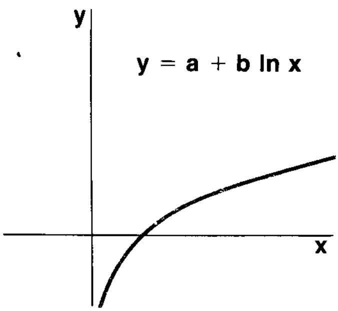
Linear Regression



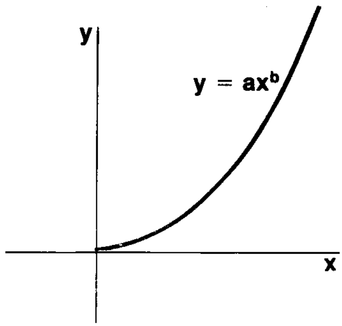
Exponential Curve Fit



Logarithmic Curve Fit



Power Curve Fit



Remarks:

- The program applies the least square method, either to the original equations (straight line and logarithmic curve) or to the transformed equations (exponential curve and power curve).
- Negative and zero values of x_i will cause a machine error for logarithmic curve fits. Negative and zero values of y_i will cause a machine error for exponential curve fits. For power curve fits, both x_i and y_i must be positive, non-zero values.
- As the differences between x and/or y values become small, the accuracy of the regression coefficients will decrease.

SIZE: 016

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
1.	Initialize the program. • for STRAIGHT LINE→ or • for EXPONENTIAL CURVE→ or • for LOGARITHMIC CURVE→ or • for POWER CURVE→		 XEQ ΣLIN XEQ ΣEXP XEQ ΣLOG XEQ ΣPOW	 ΣLIN ΣEXP ΣLOG ΣPOW
2.	Repeat step 2 – 3 for I = 1, 2 ..., n. Input x_i y_i	 x_i y_i	 ENTER↑ A	 (i)
3.	If you made a mistake in inputting x_k and y_k, then correct by	 x_k y_k	 ENTER↑ C	 (k – 1)
4.	Calculate R^2 and regression coefficients a and b.		 E R/S R/S	 R2 = (R^2) a = (a) b = (b)
5.	Calculate estimated y from regression. Input x.	x	R/S	Y. = ($\hat{y}$)
6.	Repeat step 5 for different x's.			
7.	Repeat step 4 if you want the results again.			
8.	To use the same program for another set of data, initialize the program by→ then go to step 2.		▪ A	ΣLIN or ΣEXP or ΣLOG or ΣPOW
9.	To use another program, go to step 1.			

Example 1:

Fit the following set of data into a straight line.

x_i	40.5	38.6	37.9	36.2	35.1	34.6
y_i	104.5	102	100	97.5	95.5	94

Solution:

a = 33.53, b = 1.76

$R^2 = 0.99$

i.e., y = 33.53 + 1.76 x

For x = 37, $\hat{y}$ = 98.65

For x = 35, $\hat{y}$ = 95.13

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 016	
[XEQ] [ALPHA] ΣLIN [ALPHA]	*ΣLIN*
40.5 [ENTER↑] 104.5 [A]	
38.6 [ENTER↑] 102 [A]	
37.9 [ENTER↑] 100 [A]	
36.2 [ENTER↑] 97.5 [A]	
35.2 [ENTER↑] 95.5 [A]	
35.2 [ENTER↑] 95.5 [C]	
35.1 [ENTER↑] 95.5 [A]	
34.6 [ENTER↑] 94 [A]	*6.00*
[E]	*R2=0.99*
[R/S]	*a=33.53*
[R/S]	*b=1.76*
37 [R/S]	*Y.=98.65*
35 [R/S]	*Y.=95.13*

Example

Fit the following set of data into an exponential curve.

x_i	.72	1.31	1.95	2.58	3.14
y_i	2.16	1.61	1.16	.85	0.5

Solution:

a = 3.45, b = −0.58

$y = 3.45\ e^{-0.58x}$

$R^2 = 0.98$

For x = 1.5, $\hat{y}$ = 1.44

For x = 2, $\hat{y}$ = 1.08

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 016	
[XEQ] [ALPHA] ΣEXP [ALPHA]	***ΣEXP***
.72 [ENTER↑] 2.16 [A]	
1.31 [ENTER↑] 1.61 [A]	
1.95 [ENTER↑] 1.16 [A]	
2.58 [ENTER↑] .85 [A]	
3.15 [ENTER↑] .05 [A]	
3.15 [ENTER↑] .05 [C]	
3.14 [ENTER↑] 0.5 [A]	***5.00***
[E]	***R2=0.98***
[R/S]	***a=3.45***
[R/S]	***b=-0.58***
1.5 [R/S]	***Y.=1.44***
2.0 [R/S]	***Y.=1.08***

Example 3:

Fit the following set of data into a logarithmic curve.

x_i	3	4	6	10	12
y_i	1.5	9.3	23.4	45.8	60.1

Solution:

a = −47.02, b = 41.39

$y = -47.02 + 41.39 \ln x$

$R^2 = 0.98$

For x = 8, $\hat{y} = 39.06$

For x = 14.5, $\hat{y} = 63.67$

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 016	
[XEQ] [ALPHA] ΣLOG [ALPHA]	***ΣLOG***
3 [ENTER↑] 1.5 [A]	
4 [ENTER↑] 9.3 [A]	
6 [ENTER↑] 23.4 [A]	
10 [ENTER↑] 45.8 [A]	
12 [ENTER↑] 6.01 [A]	
12 [ENTER↑] 6.01 [C]	
12 [ENTER↑] 60.1 [A]	***5.00***

[E]	*R2=0.98*
[R/S]	*a=-47.02*
[R/S]	*b=41.39*
8 [R/S]	*Y.=39.06*
14.5 [R/S]	*Y.=63.67*

Example 4:

Fit the following set of data into a power curve.

x_i	10	12	15	17	20	22	25	27	30	32	35
y_i	0.95	1.05	1.25	1.41	1.73	2.00	2.53	2.98	3.85	4.59	6.02

Solution:

a = .03, b = 1.46

$y = .03x^{1.46}$

$R^2 = 0.94$

For x = 18, $\hat{y}$ = 1.76

For x = 23, $\hat{y}$ = 2.52

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 016	
[XEQ] [ALPHA] ΣPOW [ALPHA]	*ΣPOW*
10 [ENTER↑] 0.95 [A]	
12 [ENTER↑] 1.05 [A]	
15 [ENTER↑] 1.25 [A]	
17 [ENTER↑] 1.41 [A]	
20 [ENTER↑] 1.73 [A]	
22 [ENTER↑] 2.00 [A]	
25 [ENTER↑] 2.53 [A]	
27 [ENTER↑] 2.98 [A]	
30 [ENTER↑] 3.85 [A]	
32 [ENTER↑] 4.59 [A]	
35 [ENTER↑] 60.2 [A]	
35 [ENTER↑] 60.2 [C]	
35 [ENTER↑] 6.02 [A]	*11.00*
[E]	*R2=0.94*
[R/S]	*a=0.03*
[R/S]	*b=1.46*
18 [R/S]	*Y.=1.76*
23 [R/S]	*Y.=2.52*

MULTIPLE LINEAR REGRESSION

Three Independent Variables

For a set of data points $\{(x_i, y_i, z_i, t_i), i = 1, 2, \ldots, n\}$, this program fits a linear equation of the form:

$$t = a + bx + cy + dz$$

by the least squares method.

Regression coefficients a, b, c, and d are calculated by solving the following system of equations. Gauss's elimination method with partial pivoting is used.

$$\begin{bmatrix} n & \Sigma x_i & \Sigma y_i & \Sigma z_i \\ \Sigma x_i & \Sigma(x_i)^2 & \Sigma(x_i y_i) & \Sigma(x_i z_i) \\ \Sigma y_i & \Sigma(y_i x_i) & \Sigma(y_i)^2 & \Sigma(y_i z_i) \\ \Sigma z_i & \Sigma(x_i z_i) & \Sigma(y_i z_i) & \Sigma(z_i)^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \Sigma t_i \\ \Sigma x_i t_i \\ \Sigma y_i t_i \\ \Sigma z_i t_i \end{bmatrix}$$

The coefficient of determination R^2 is defined as:

$$R^2 = \frac{a\Sigma t_i + b\Sigma x_i t_i + c\Sigma y_i t_i + d\Sigma z_i t_i - \frac{1}{n}(\Sigma t_i)^2}{\Sigma(t_i^2) - \frac{1}{n}(\Sigma t_i)^2}$$

Two Independent Variables

For a set of data points $\{(x_i, y_i, t_i), i = 1, 2, \ldots, n\}$, this program fits a linear equation of the form:

$$t = a + bx + cy$$

by the least squares method.

Regression coefficients a, b, and c are calculated by solving the following system of equations. Gauss's elimination method with partial pivoting is used.

$$\begin{bmatrix} n & \Sigma x_i & \Sigma y_i \\ \Sigma x_i & \Sigma(x_i)^2 & \Sigma x_i y_i \\ \Sigma y_i & \Sigma y_i x_i & \Sigma(y_i)^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \Sigma t_i \\ \Sigma x_i t_i \\ \Sigma y_i t_i \end{bmatrix}$$

The coefficient of determination R^2 is defined as:

$$R^2 = \frac{a\Sigma t_i + b\Sigma x_i t_i + c\Sigma y_i t_i - \frac{1}{n}(\Sigma t_i)^2}{\Sigma(t_i^2) - \frac{1}{n}(\Sigma t_i)^2}$$

Remarks:

- If the coefficient matrix has determinant equal to zero, indicating no solution or more than one solution, "DATA ERROR" will be displayed.
- There is no restriction on the maximum number of data points n, but the following minimum condition for n must be satisfied:

  $n \geqslant 3$ for the case of two independent variables
  $n \geqslant 4$ for the case of three independent variables

Refernce:

HP-67/97 Math Pac I, program MA1-07

				SIZE: 045
STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
	Three Independent Variables			
1.	Initialize the program.		[XEQ] ΣMLRXYZ	ΣMLRXYZ
2.	Repeat step 2~3 for i=1,2,...,n.			
	Input: x_i	x_i	[ENTER↑]	
	y_i	y_i	[ENTER↑]	
	z_i	z_i	[ENTER↑]	
	t_i	t_i	[A]	(i)
3.	If you made a mistake in inputting x_k, y_k, z_k, and t_k, then correct by	x_k	[ENTER↑]	
		y_k	[ENTER↑]	
		z_k	[ENTER↑]	
		t_k	[C]	(k−1)
4.	Calculate R^2 and regression coefficients a,b,c, and d.		[E]	R2=(R^2)
			[R/S]	a=(a)
			[R/S]	b=(b)
			[R/S]	c=(c)
			[R/S]	d=(d)
5.	Calculate estimated t from regression. Input: x	x	[ENTER↑]	
	y	y	[ENTER↑]	
	z	z	[R/S]	T.=($\hat{t}$)
6.	Repeat step 5 for different (x,y,z)'s.			
7.	To recall sums used in calculation:			
	Σx_i		[RCL] 32	(Σx_i)
	Σy_i		[RCL] 33	(Σy_i)
	Σz_i		[RCL] 34	(Σz_i)
	Σt_i		[RCL] 41	(Σt_i)
	Σx_i^2		[RCL] 35	(Σx_i^2)
	Σy_i^2		[RCL] 38	(Σy_i^2)
	Σz_i^2		[RCL] 40	(Σz_i^2)
	Σt_i^2		[RCL] 30	(Σt_i^2)
	$\Sigma x_i y_i$		[RCL] 36	($\Sigma x_i y_i$)
	$\Sigma x_i z_i$		[RCL] 37	($\Sigma x_i z_i$)
	$\Sigma x_i t_i$		[RCL] 42	($\Sigma x_i t_i$)
	$\Sigma y_i z_i$		[RCL] 39	($\Sigma y_i z_i$)
	$\Sigma y_i t_i$		[RCL] 43	($\Sigma y_i t_i$)
	$\Sigma z_i t_i$		[RCL] 44	($\Sigma z_i t_i$)
8.	Repeat step 4 if you want the results again.			
9.	To use the program for another set of data, initialize the program by → then go to step 2.		■ [A]	ΣMLRXYZ

Example 1:

For the following set of data, find the regression line with three independent variables. i.e. t = a + bx + cy + dz

i	1	2	3	4	5
x_i	7	1	11	11	7
y_i	25	29	56	31	52
z_i	6	15	8	8	6
t_i	60	52	20	47	33

Solution:

The regression line is described by t = 103.45 − 1.28x − 1.04y − 1.34z.

$R^2 = 1.00$

For x = 7, y = 25, z = 6, $\hat{t}$ = 60.50

For x = 1, y = 29, z = 15, $\hat{t}$ = 52.00

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 045	
[XEQ] [ALPHA] ΣMLRXYZ [ALPHA]	***ΣMLRXYZ***
7 [ENTER↑] 25 [ENTER↑]	
6 [ENTER↑] 60 [A]	
1 [ENTER↑] 29 [ENTER↑]	
15 [ENTER↑] 52 [A]	
11 [ENTER↑] 56 [ENTER↑]	
8 [ENTER↑] 20 [A]	
11 [ENTER↑] 31 [ENTER↑]	
8 [ENTER↑] 47 [A]	
7 [ENTER↑] 53 [ENTER↑]	
6 [ENTER↑] 33 [A]	
7 [ENTER↑] 53 [ENTER↑]	
6 [ENTER↑] 33 [C]	
7 [ENTER↑] 52 [ENTER↑]	
6 [ENTER↑] 33 [A]	***5.00***
[E]	***R2=1.00***
[R/S]	***a=103.45***
[R/S]	***b=-1.28***
[R/S]	***c=-1.04***
[R/S]	***d=-1.34***
7 [ENTER↑] 25 [ENTER↑] 6 [R/S]	***T.=60.50***
1 [ENTER↑] 29 [ENTER↑] 15 [R/S]	***T.=52.00***

				SIZE: 045
STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
	Two Independent Variables			
1.	Initialize the program.		[XEQ]	ΣMLRXY
2.	Repeat step 2~3 for i=1,2,...,n.			
	Input: x_i	x_i	[ENTER↑]	
	y_i	y_i	[ENTER↑]	
	t_i	t_i	[A]	(i)
3.	If you made a mistake in inputting x_k,y_k, and t_k, then correct by	x_k	[ENTER↑]	
		y_k	[ENTER↑]	
		t_k	[C]	(k−1)
4.	Calculate R^2 and regression coefficients a, b, and c.		[E]	R2=(R^2)
			[R/S]	a=(a)
			[R/S]	b=(b)
			[R/S]	c=(c)
5.	Calculate estimated t from regression. Input: x	x	[ENTER↑]	
	y	y	[R/S]	T.=($\hat{t}$)
6.	Repeat step 5 for different (x,y)'s.			
7.	To recall sums used in calculation:			
	Σx_i		[RCL] 32	(Σx_i)
	Σy_i		[RCL] 33	(Σy_i)
	Σt_i		[RCL] 41	(Σt_i)
	Σx_i^2		[RCL] 35	(Σx_i^2)
	Σy_i^2		[RCL] 38	(Σy_i^2)
	Σt_i^2		[RCL] 30	(Σt_i^2)
	$\Sigma x_i y_i$		[RCL] 36	($\Sigma x_i y_i$)
	$\Sigma x_i t_i$		[RCL] 42	($\Sigma x_i t_i$)
	$\Sigma y_i t_i$		[RCL] 43	($\Sigma y_i t_i$)
8.	Repeat step 4 if you want the results again.			
9.	To use the program for another set of data, initialize the program by → then go to step 2.		[A]	ΣMLRXY

Example 2:

For the following set of data, find the regression line with two independent variables. i.e. t = a + bx + cy

i	1	2	3	4
x_i	1.5	0.45	1.8	2.8
y_i	0.7	2.3	1.6	4.5
t_i	2.1	4.0	4.1	9.4

Solution:

The regression line is $t = -0.10 + 0.79x + 1.63y$

$R^2 = 1.00$

For x = 2, y = 3, $\hat{t} = 6.37$

For x = 1.5, y = 0.7 $\hat{t} = 2.23$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 045	
[XEQ] [ALPHA] ΣMLRXY [ALPHA]	ΣMLRXY
1.5 [ENTER↑] 0.7 [ENTER↑] 2.1 [A]	
0.46 [ENTER↑] 2.3 [ENTER↑] 4.0 [A]	
0.46 [ENTER↑] 2.3 [ENTER↑] 4.0 [C]	
0.45 [ENTER↑] 2.3 [ENTER↑] 4.0 [A]	
1.8 [ENTER↑] 1.6 [ENTER↑] 4.1 [A]	
2.8 [ENTER↑] 4.5 [ENTER↑] 9.4 [A]	4.00
[E]	R2=1.00
[R/S]	a=-0.10
[R/S]	b=0.79
[R/S]	c=1.63
2 [ENTER↑] 3 [R/S]	T.=6.37
1.5 [ENTER↑] 0.7 [R/S]	T.=2.23

POLYNOMIAL REGRESSION

Cubic Regression

For a set of data points (x_i, y_i), $i = 1, 2, \ldots, n$, this program fit a cubic equation of the form:

$$y = a + bx + cx^2 + dx^3$$

by the least squares method.

Regression coefficients a, b, c and d are calculated by solving the following system of equations. Gauss's elimination method with partial pivoting is used.

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 & \Sigma x_i^5 \\ \Sigma x_i^3 & \Sigma x_i^4 & \Sigma x_i^5 & \Sigma x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \\ \Sigma x_i^3 y_i \end{bmatrix}$$

The coefficient of determination is:

$$R^2 = \frac{a\Sigma y_i + b\Sigma x_i y_i + c\Sigma x_i^2 y_i + d\Sigma x_i^3 y_i - \frac{1}{n}(\Sigma y_i)^2}{\Sigma(y_i^2) - \frac{1}{n}(\Sigma y_i)^2}$$

Parabolic Regression

For a set of data points (x_i, y_i), $i = 1, 2, \ldots, n$, this program fits a parabola of the form:

$$y = a + bx + cx^2$$

by the least squares method.

Regression coefficients a, b, and c are calculated by solving the following system of equations. Gauss's elimination method with partial pivoting is used.

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}$$

The coefficient of determination is:

$$R^2 = \frac{a\Sigma y_i + b\Sigma x_i y_i + c\Sigma x_i^2 y_i - \frac{1}{n}(\Sigma y_i)^2}{\Sigma(y_i^2) - \frac{1}{n}(\Sigma y_i)^2}$$

Remarks:

- If the coefficient matrix has determinant equal to zero, indicating no solution or more than one solution, ***DET=0*** will be displayed.
- There is no restriction on the maximum number of data points n , but the following minimum condition for n must be satisfied:

  $n \geqslant 3$ for Parabolic Regression
  $n \geqslant 4$ for Cubic Regression

Reference:

HP-67/97 Math Pac I, program MA1-07

STEP	INSTRUCTIONS	INPUT	FUNCTION	SIZE: 045 DISPLAY
	Cubic Regression			
1.	Initialize the program.		[XEQ] ΣPOLYC	ΣPOLYC
2.	Repeat step 2~3 for i=1,2,...,n. Input: x_i y_i	 x_i y_i	 [ENTER↑] [A]	 (i)
3.	If you made a mistake in inputting x_k and y_k, then correct by	 x_k y_k	 [ENTER↑] [C]	 (k−1)
4.	Calculate R^2 and regression coefficients a,b,c, and d.		[E] [R/S] [R/S] [R/S] [R/S]	R2=(R^2) a=(a) b=(b) c=(c) d=(d)
5.	Calculate estimated y from regression. Input x.	x	[R/S]	Y.=($\hat{y}$)
6.	Repeat step 5 for different x's.			
7.	To recall sums in calculation: Σx_i Σx_i^2 Σx_i^3 Σx_i^4 Σx_i^5 Σx_i^6 Σy_i $\Sigma x_i y_i$ $\Sigma x_i^2 y_i$ $\Sigma x_i^3 y_i$		 [RCL] 32 [RCL] 33 [RCL] 34 [RCL] 37 [RCL] 39 [RCL] 40 [RCL] 41 [RCL] 42 [RCL] 43 [RCL] 44	 (Σx_i) (Σx_i^2) (Σx_i^3) (Σx_i^4) (Σx_i^5) (Σx_i^6) (Σy_i) ($\Sigma x_i y_i$) ($\Sigma x_i^2 y_i$) ($\Sigma x_i^3 y_i$)
8.	Repeat step 4 if you want the results again.			
9.	To use the program for another set of data, initialize the program by → then go to step 2.		■ [A]	ΣPOLYC

Example 1:

For the following set of data, perform a cubic regression, i.e., find suitable coefficients for:

$$y = a + bx + cx^2 + dx^3$$

i	1	2	3	4	5
x	.8	1	1.2	1.4	1.6
y	24	20	10	13	12

Solution:

$y = 47.94 - 9.76x - 41.07x^2 + 20.83x^3$

$R^2 = 0.87$

For $x = 1$, $\hat{y} = 17.94$

For $x = 1.4$, $\hat{y} = 10.94$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 045	
[XEQ] [ALPHA] ΣPOLYC [ALPHA]	***ΣPOLYC***
.8 [ENTER↑] 24 [A]	
1 [ENTER↑] 20 [A]	
1.3 [ENTER↑] 10 [A]	
1.3 [ENTER↑] 10 [C]	
1.2 [ENTER↑] 10 [A]	
1.4 [ENTER↑] 13 [A]	
1.6 [ENTER↑] 12 [A]	***5.00***
[E]	***R2=0.87***
[R/S]	***a=47.94***
[R/S]	***b=-9.76***
[R/S]	***c=-41.07***
[R/S]	***d=20.83***
1 [R/S]	***Y.=17.94***
1.4 [R/S]	***Y.=10.94***

				SIZE: 045
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
	Parabolic Regression			
1.	Initialize the program.		[XEQ] ΣPOLYP	ΣPOLYP
2.	Repeat step 2~3 for i=1,2,...,n.			
	Input: x_i	x_i	[ENTER↑]	
	y_i	y_i	[A]	(i)
3.	If you made a mistake in inputting x_k and y_k, then correct by	x_k	[ENTER↑]	
		y_k	[C]	(k−1)
4.	Calculate R^2 and regression coefficients a,b, and c		[E]	R2=(R^2)
			[R/S]	a=(a)
			[R/S]	b=(b)
			[R/S]	c=(c)
5.	Calculate estimated y from regression. Input x.	x	[R/S]	Y.=($\hat{y}$)
6.	Repeat step 5 for different x's.			
7.	To recall sums in calculation:			
	Σx_i		[RCL] 32	(Σx_i)
	Σx_i^2		[RCL] 33	(Σx_i^2)
	Σx_i^3		[RCL] 34	(Σx_i^3)
	Σx_i^4		[RCL] 37	(Σx_i^4)
	Σy_i		[RCL] 41	(Σy_i)
	$\Sigma x_i y_i$		[RCL] 42	($\Sigma x_i y_i$)
	$\Sigma x_i^2 y_i$		[RCL] 43	($\Sigma x_i^2 y_i$)
8.	Repeat step 4 if you want the results again.			
9.	To use the program for another set of data, initialize the program by → then go to step 2.		■ [A]	ΣPOLYP

Example 2:

For the following set of data, perform a parabolic regression, i.e., find suitable coefficients for:

$$y = a + bx + cx^2$$

i	1	2	3	4	5	6	7
x	1	2	3	4	5	6	7
y	5	12	34	50	75	84	128

Solution:

$y = -4.00 + 6.64x + 1.64x^2$

$R^2 = 0.98$

For $x = 2$, $\hat{y} = 15.86$

For $x = 4$, $\hat{y} = 48.86$

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 045	
[XEQ] [ALPHA] ΣPOLYP [ALPHA]	ΣPOLYP
1 [ENTER↑] 5 [A]	
2 [ENTER↑] 12 [A]	
3 [ENTER↑] 34 [A]	
4 [ENTER↑] 50 [A]	
5 [ENTER↑] 75 [A]	
6 [ENTER↑] 84 [A]	
7 [ENTER↑] 128 [A]	***7.00***
[E]	***R2=0.98***
[R/S]	***a=-4.00***
[R/S]	***b=6.64***
[R/S]	***c=1.64***
2 [R/S]	***Y.=15.86***
4 [R/S]	***Y.=48.86***

t STATISTICS

Paired t Statistic

Given a set of paired observations from two normal populations with means μ_1, μ_2 (unknown)

x_i	x_1	x_2	...	x_n
y_i	y_1	y_2	...	y_n

let

$$D_i = x_i - y_i$$

$$\overline{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$$

$$s_D = \sqrt{\frac{\Sigma D_i^2 - \frac{1}{n}(\Sigma D_i)^2}{n - 1}}$$

The test statistic

$$t = \frac{\overline{D}}{s_D} \cdot \sqrt{n}$$

which has n − 1 degrees of freedom (df) can be used to test the null hypothesis

$$H_0: \mu_1 = \mu_2$$

Reference:

Statistics in Research, B. Ostle, Iowa State University Press, 1963.

t Statistic For Two Means

Suppose $\{x_1, x_2, \ldots, x_{n1}\}$ and $\{y_1, y_2, \ldots, y_{n2}\}$ are independent random samples from two normal populations having means μ_1, μ_2 (unknown) and the same unknown variance σ^2.

We want to test the null hypothesis

$$H_0: \mu_1 - \mu_2 = d$$

Define

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$$

$$\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i$$

$$t = \frac{\bar{x} - \bar{y} - d}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{\Sigma x_i^2 - n_1\bar{x}^2 + \Sigma y_i^2 - n_2\bar{y}^2}{n_1 + n_2 - 2}}}$$

We can use this t statistic which has the t distribution with $n_1 + n_2 - 2$ degrees of freedom (df) to test the null hypothesis H_0.

Reference:

Statistical Theory and Methodology in Science and Engineering, K.A. Brownlee, John Wiley & Sons, 1965.

				SIZE: 015
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
	Paired t Statistic			
1.	Initialize the program.		[XEQ] ΣPTST	ΣPTST
2.	Repeat step 2~3 for i=1,2,...,n. Input: x_i y_i	x_i y_i	[ENTER↑] [A]	 (i)
3.	If you made a mistake in inputting x_k and y_k, then correct by	x_k y_k	[ENTER↑] [C]	 (k−1)
4.	To calculate the test statistic: $\bar{D}$ s_D t df		 [E] [R/S] [R/S] [R/S]	 DBAR=($\bar{D}$) SD=(s_D) T=(t) DF=(df)
5.	Repeat step 4 if you want the results again.			
6.	To use the same program for another set of data, initialize the program by → then go to step 2.		■ [A]	ΣPTST
	t Statistic for Two Means			
7.	Initialize the program.		[XEQ] ΣTSTAT	ΣTSTAT
8.	Repeat step 8~9 for i=1,2,...,n_1 Input x_i .	x_i	[A]	(i)
9.	If you made a mistake in inputting x_k, then correct by	x_k	[C]	(k−1)
10.	Initialize for the 2nd array of data		[R/S]	0.00
11.	Repeat step 11 ~ 12 for j=1, 2, ..., n_2. Input y_j.	y_j	[A]	(j)
12.	If you made a mistake in inputting y_h, then correct by	y_h	[C]	(h−1)
13.	Input d to calculate test statistic: t df	d	 [E] [R/S]	 T=(t) DF=(df)
14.	Repeat step 13 if you want to calculate the test statistic for a different value of d.			
15.	To use the same program for another set of data, initialize the program by → then go to step 8.		■ [A]	ΣTSTAT

Example 1:

x_i	14	17.5	17	17.5	15.4
y_i	17	20.7	21.6	20.9	17.2

$\overline{D} = -3.20$

$s_D = 1.00$

$t = -7.16$

$df = 4.00$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 015	
[XEQ] [ALPHA] ΣPTST [ALPHA]	***ΣPTST***
14 [ENTER↑] 17 [A]	
17.5 [ENTER↑] 20.7 [A]	
17 [ENTER↑] 21.6 [A]	
17 [ENTER↑] 15 [A]	
17 [ENTER↑] 15 [C]	
17.5 [ENTER↑] 20.9 [A]	
15.4 [ENTER↑] 17.2 [A]	***5.00***
[E]	***DBAR=-3.20***
[R/S]	***SD=1.00***
[R/S]	***T=-7.16***
[R/S]	***DF=4.00***

Example 2:

x	79	84	108	114	120	103	122	120		
y	91	103	90	113	108	87	100	80	99	54

$n_1 = 8$

$n_2 = 10$

If $d = 0$ (i.e., H_0: $\mu_1 = \mu_2$)

then $t = 1.73$, $df = 16.00$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 015	
[XEQ] [ALPHA] ΣTSTAT [ALPHA]	***ΣTSTAT***
79 [A] 84 [A]	
99 [A] 99 [C] 108 [A]	
114 [A] 120 [A]	
103 [A] 122 [A] 120 [A]	***8.00***
[R/S]	***0.00***
91 [A] 103 [A] 90 [A] 113 [A]	
108 [A] 87 [A] 100 [A] 80 [A]	
99 [A] 54 [A]	***10.00***
0 [E]	***T=1.73***
[R/S]	***DF=16.00***

CHI-SQUARE EVALUATION

This program calculates the value of the χ^2 statistic for the goodness of fit test by the equation

$$\chi_1^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with df} = n - 1$$

where O_i = observed frequency

E_i = expected frequency

n = number of classes

If the expected values are equal

$$\left(E = E_i = \frac{\Sigma O_i}{n} \text{ for all } i\right)$$

then

$$\chi_2^2 = \frac{n\Sigma O_i^2}{\Sigma O_i} - \Sigma O_i$$

Remarks:

- In order to apply the goodness of fit test to a set of given data, combining some classes may be necessary to make sure that each expected frequency is not too small (say, not less than 5).

Reference:

Mathematical Statistics, J.E. Freund, Prentice Hall, 1962.

SIZE: 008

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
	Unequal Expected Frequency			
1.	Initialize the program.		[XEQ] ΣXSQEV	ΣXSQEV
2.	Repeat step 2~3 for i=1,2,...,n. Input: O_i E_i	O_i E_i	[ENTER↑] [A]	 (i)
3.	If you made a mistake in inputting O_k and E_k, then correct by	O_k E_k	[ENTER↑] [C]	 (k−1)
4.	Calculate χ_1^2.		[E]	XSQ=(χ_1^2)
5.	To use the same program for another set of data, initialize the program by → then go to step 2.		■ [A]	ΣXSQEV
	Equal Expected Frequency			
6.	Initialize the program.		[XEQ] ΣEEFXSQ	ΣEEFXSQ
7.	Repeat step 7~8 for i=1,2,...,n. Input: O_i	O_i	[A]	(i)
8.	If you made a mistake in inputting O_h, then correct by	O_h	[C]	(h−1)
9.	Calculate: χ_2^2 E		[E] [R/S]	XSQ=(χ_2^2) E=(E)
10.	Repeat step 9 if you want the results again.			
11.	To use the same program for another set of data, initialize the program by → then go to step 7.		■ [A]	ΣEEFXSQ

Examples 1:

Find the value of χ^2 statistic for the goodness of fit for the following data set:

O_i	8	50	47	56	5	14
E_i	9.6	46.75	51.85	54.4	8.25	9.15

$$\chi_1^2 = 4.84$$

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 008	
[XEQ] [ALPHA] ΣXSQEV [ALPHA]	***ΣXSQEV***
8 [ENTER↑] 9.6 [A]	
50 [ENTER↑] 46.75 [A]	
47 [ENTER↑] 51.85 [A]	
56 [ENTER↑] 54.4 [A]	
5 [ENTER↑] 8.25 [A]	
88 [ENTER↑] 88 [A]	
88 [ENTER↑] 88 [C]	
14 [ENTER↑] 9.15 [A]	***6.00***
[E]	***XSQ=4.84***

Example 2:

The following table shows the observed frequencies in tossing a die 120 times. χ^2 can be used to test if the die is fair.

Note: Assume that the expected frequencies are equal.

number	1	2	3	4	5	6
frequency O_i	25	17	15	23	24	16

$$\chi_2^2 = 5.00$$

$$E = 20.00$$

Since 5.00 is less than 11.07, the data does not support the statement that the die is "unfair" (5% significance level).

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 008	
[XEQ] [ALPHA] ΣEEFXSQ [ALPHA]	ΣEEFXSQ
25 [A] 17 [A] 15 [A] 22 [A]	
22 [C]	
23 [A] 24 [A] 16 [A]	6.00
[E]	XSQ=5.00
[R/S]	E=20.00

CONTINGENCY TABLE

Contingency tables can be used to test the null hypothesis that two variables are independent.

This program calculates the χ^2 statistic for testing the independence of the two variables. Also Pearson's coefficient of contingency C_c, which measures the degree of association between the two variables, is calculated.

2 x k CONTINGENCY TABLE

i \ j	1	2	...	k	Totals
1	x_{11}	x_{12}	...	x_{1k}	R_1
2	x_{21}	x_{22}	...	x_{2k}	R_2
Totals	C_1	C_2	...	C_k	T

3 x k CONTINGENCY TABLE

i \ j	1	2	...	k	Totals
1	x_{11}	x_{12}	...	x_{1k}	R_1
2	x_{21}	x_{22}	...	x_{2k}	R_2
3	x_{31}	x_{32}	...	x_{3k}	R_3
Totals	C_1	C_2	...	C_k	T

Equations:

$$\text{Row sum } R_i = \sum_{j=1}^{k} x_{ij} \qquad \begin{array}{l} i = 1, 2 \text{ (for } 2 \times k) \\ i = 1, 2, 3 \text{ (for } 3 \times k) \end{array}$$

$$\text{Column sum } C_j = \sum_{i=1}^{n} x_{ij} \qquad \begin{array}{l} j = 1, 2, \ldots, k \\ n = 2 \text{ (for } 2 \times k) \\ n = 3 \text{ (for } 3 \times k) \end{array}$$

$$\text{Total } T = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} \qquad \begin{array}{l} n = 2 \text{ (for } 2 \times k) \\ n = 3 \text{ (for } 3 \times k) \end{array}$$

Chi-square statistic

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{(x_{ij} - E_{ij})^2}{E_{ij}} \text{ with df} = (n - 1)(k - 1)$$

$$= T\left(\sum_{i=1}^{n} \sum_{j=1}^{k} \frac{x_{ij}^2}{R_i C_j}\right) - T \qquad \begin{matrix} n = 2 \text{ (for } 2 \times k) \\ n = 3 \text{ (for } 3 \times k) \end{matrix}$$

Contingency coefficient

$$C_c = \sqrt{\frac{\chi^2}{T + \chi^2}}$$

Reference:

B. Ostle, *Statistics in Research,* Iowa State University Press, 1972.

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
				SIZE: 015
	2×k			
1.	Initialize the program.		[XEQ] ΣCTKK	ΣCTKK
2.	Repeat step 2~5 for j=1,2,...,k. input: x_{1j} x_{2j}	x_{1j} x_{2j}	[ENTER↑] [A]	(j)
3.	(Optional) Calculate column sum C_j.		[R/S]	CS=(C_j)
4.	If you made a mistake in inputting x_{1h} and x_{2h}, then correct by	x_{1h} x_{2h}	[ENTER↑] [C]	(h−1)
5.	(Optional) Calculate column sum C_h (correction).		[R/S]	CS=(−C_h)
6.	Go to step 12 for contingency table calculations.			
	3×k			
7.	Initialize the program.		[XEQ] ΣCTKKK	ΣCTKKK
8.	Repeat step 8~11 for j=1,2,...,k. input: x_{1j} x_{2j} x_{3j}	x_{1j} x_{2j} x_{3j}	[ENTER↑] [ENTER↑] [A]	(j)
9.	(Optional) Calculate column sum C_j.		[R/S]	CS=(C_j)
10.	If you made a mistake in inputting x_{1h}, x_{2h}, and x_{3h}, then correct by	x_{1h} x_{2h} x_{3h}	[ENTER↑] [ENTER↑] [C]	(h−1)

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
11.	(Optional) Calculate column sum C_h (correction).		[R/S]	CS=(−C_h)
12.	Calculate: Test statistic χ^2 Coefficient C_c Row sum 1 R_1 Row sum 2 R_2 Row sum 3 R_3 (3×k only) Total T		 [E] [R/S] [R/S] [R/S] [R/S] [R/S]	 XSQ=(χ^2) CC=(C_c) R1=(R_1) R2=(R_2) R3=(R_3) T=(T)
13.	Repeat step 12 if you want the results again.			
14.	To use the same program for another set of data, initialize by → then go to step 2 or step 8.		■ [A]	ΣCTKK or ΣCTKKK
15.	To use the other program, go to step 1 or step 7.			

Example 1:

Find the test statistic χ^2 and coefficient of contingency C_c for the following set of data.

	1	2	3
A	2	5	4
B	3	8	7

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 015	
[XEQ] [ALPHA] ΣCTKK [ALPHA]	ΣCTKK
2 [ENTER↑] 3 [A]	1.00
[R/S]	CS=5.00
5 [ENTER↑] 8 [A] 4 [ENTER↑] 7 [A]	3.00
[E]	XSQ=0.02
[R/S]	CC=0.03
[R/S]	R1=11.00
[R/S]	R2=18.00
[R/S]	T=29.00

Example 2:

Find test statistic χ^2 and coefficient of contingency C_c for the following set of data.

i \ j	1	2	3	4
1	36	67	49	58
2	31	60	49	54
3	58	87	80	68

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 015	
[XEQ] [ALPHA] ΣCTKKK [ALPHA]	***ΣCTKKK***
36 [ENTER↑] 31 [ENTER↑] 58 [A]	***1.00***
[R/S]	***CS=125.00***
67 [ENTER↑] 60 [ENTER↑] 87 [A]	
4 [ENTER↑] 49 [ENTER↑] 80 [A]	
4 [ENTER↑] 49 [ENTER↑] 80 [C]	
49 [ENTER↑] 49 [ENTER↑] 80 [A]	
58 [ENTER↑] 54 [ENTER↑] 68 [A]	***4.00***
[E]	***XSQ=3.36***
[R/S]	***CC=0.07***
[R/S]	***R1=210.00***
[R/S]	***R2=194.00***
[R/S]	***R3=293.00***
[R/S]	***T=697.00***

SPEARMAN'S RANK CORRELATION COEFFICIENT

Spearman's rank correlation coefficient is a measure of rank correlation under the following circumstance: n individuals are ranked from 1 to n according to some specified characteristic by 2 observers, and we wish to know if the 2 rankings are substantially in agreement with one another.

Spearman's rank correlation coefficient is defined by

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2 - 1)}$$

where **n = number of paired observations (x_i, y_i)**
D_i = rank (x_i) − rank (y_i) = $R_i - S_i$

If the X and Y random variables from which these n pairs of observations are derived are independent, then r_S has zero mean and a variance equal to

$$\frac{1}{n - 1}$$

A test for the null hypothesis

$$H_o\text{: X, Y are independent}$$

is made using

$$z = r_S\sqrt{n - 1}$$

which is approximately a standardized normal variable (for large n, say $n \geq 10$).

If the null hypothesis of independence is not rejected, we can infer that the population correlation coefficient $\rho(x, y) = 0$, but dependence between the variables does not necessarily imply that $\rho(x, y) \neq 0$.

Note:

$-1 \leq r_S \leq 1$
$r_S = 1$ indicates complete agreement in order of the ranks and $r_S = -1$ indicates complete agreement in the opposite order of the ranks.

Reference:

Nonparametric Statistical Inference, J. D. Gibbons, McGraw Hill, 1971.

STEP	INSTRUCTIONS	INPUT	FUNCTION	DISPLAY
				SIZE: 003
1.	Initialize the program.		[XEQ] ΣSPEAR	ΣSPEAR
2.	Repeat step 2~3 for i=1,2,...,n. Input: R_i S_i	R_i S_i	[ENTER↑] [A]	(i)
3.	If you made a mistake in inputting R_k and S_k, then correct by	R_k S_k	[ENTER↑] [C]	(k−1)
4.	Calculate: r_s z		[E] [R/S]	RS=(r_s) Z=(z)
5.	Repeat step 4 if you want the results again.			
6.	For another set of data, initialize the program by → then go to step 2.		■ [A]	ΣSPEAR

Example:

The following data set is the result of two tests in a class; find r_s and z.

Student	x_i Math Grade	y_i Stat Grade	R_i Rank of x_i	S_i Rank of y_i
1	82	81	6	7
2	67	75	14	11
3	91	85	3	4
4	98	90	1	2
5	74	80	11	8
6	52	60	15	15
7	86	94	4	1
8	95	78	2	9
9	79	83	9	6
10	78	76	10	10
11	84	84	5	5
12	80	69	8	13
13	69	72	13	12
14	81	88	7	3
15	73	61	12	14

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 003	
[XEQ] [ALPHA] ΣSPEAR [ALPHA]	***ΣSPEAR***
6 [ENTER↑] 7 [A]	
14 [ENTER↑] 11 [A]	
3 [ENTER↑] 4 [A]	
1 [ENTER↑] 2 [A]	
11 [ENTER↑] 8 [A]	
5 [ENTER↑] 5 [A]	
5 [ENTER↑] 5 [C]	
15 [ENTER↑] 15 [A]	
4 [ENTER↑] 1 [A]	
2 [ENTER↑] 9 [A]	
9 [ENTER↑] 6 [A]	
10 [ENTER↑] 10 [A]	
5 [ENTER↑] 5 [A]	
8 [ENTER↑] 13 [A]	
13 [ENTER↑] 12 [A]	
7 [ENTER↑] 3 [A]	
12 [ENTER↑] 14 [A]	***15.00***
[E]	***RS=0.76***
[R/S]	***Z=2.85***

Notes

Notes

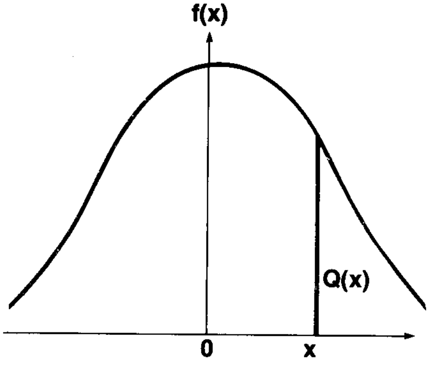
NORMAL AND INVERSE NORMAL DISTRIBUTION

This program evaluates the standard normal density function f(x) and the normal integral Q(x) for given x. If Q is given, x can also be found. The standard normal distribution has mean 0 and standard deviation 1.

Equations:

1. Standard normal density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



2. Normal integral

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt$$

Polynomial approximation is used to calculate Q(x) for given x.

Define $R = f(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \epsilon(x)$
where $|\epsilon(x)| < 7.5 \times 10^{-8}$

$$t = \frac{1}{1 + r|x|}, \qquad r = 0.2316419$$

$b_1 = .319381530, \quad b_2 = -.356563782$

$b_3 = 1.781477937, \quad b_4 = -1.821255978$

$b_5 = 1.330274429$

$$\text{Then } Q(x) = \begin{cases} R & \text{if } x \geq 0 \\ 1 - R & \text{if } x < 0 \end{cases} \quad \text{with error } |\epsilon(x)| < 7.5 \times 10^{-8}$$

3. Inverse normal

For a given $0 < Q < 1$, x can be found such that

$$Q = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$$

The following rational approximation is used:

$$\text{Define } y = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \epsilon(Q)$$

where $|\epsilon(Q)| < 4.5 \times 10^{-4}$

$$t = \begin{cases} \sqrt{\ln \frac{1}{Q^2}} & \text{if } 0 < Q \leq 0.5 \\ \sqrt{\ln \frac{1}{(1-Q)^2}} & \text{if } 0.5 < Q < 1 \end{cases}$$

$c_0 = 2.515517 \quad d_1 = 1.432788$

$c_1 = 0.802853 \quad d_2 = 0.189269$

$c_2 = 0.010328 \quad d_3 = 0.001308$

$$\text{Then } x = \begin{cases} y & \text{if } 0 < Q \leq 0.5 \\ -y & \text{if } 0.5 < Q < 1 \end{cases} \quad \text{with error } |\epsilon(Q)| < 4.5 \times 10^{-4}$$

Reference:

Abramowitz and Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, 1970.

				SIZE: 019
STEP	**INSTRUCTIONS**	**INPUT**	**FUNCTION**	**DISPLAY**
1.	Initialize the program.		[XEQ] ΣNORMD	ΣNORMD
2.	Input x to calculate f(x).	x	[C]	F=(f(x))
3.	Input x to calculate Q(x).	x	[E]	Q=(Q(x))
4.	Input Q(x) to calculate x.	Q(x)	[A]	X=(x)
5.	Repeat any of the above steps if desired.			

Example 1:

Find f(x) and Q(x) for x = 1.18 and x = −2.28.

Keystrokes:	**Display:**
[XEQ] [ALPHA] SIZE [ALPHA] 019	
[XEQ] [ALPHA] ΣNORMD [ALPHA]	***ΣNORMD***
1.18 [C]	***F=0.20***
1.18 [E]	***Q=0.12***
2.28 [CHS] [E]	***Q=0.99***
2.28 [CHS] [C]	***F=0.03***

Example 2:

Given Q = 0.12 and Q = 0.95, find x.

(If you have run through Example 1, then you can proceed; otherwise you have to initialize the program as described in Example 1).

Keystrokes:	**Display:**
0.12 [A]	***X=1.18***
0.95 [A]	***X=-1.65***

Notes

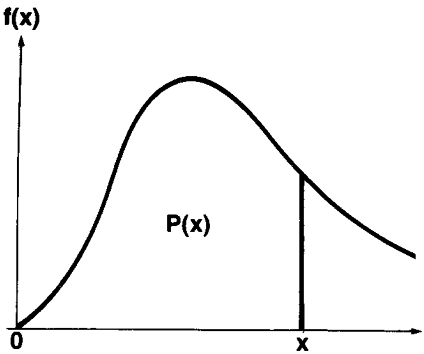
CHI-SQUARE DISTRIBUTION

This program evaluates the chi-square density

$$f(x) = \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2} - 1} e^{-\frac{x}{2}}$$

where $x \geqslant 0$

ν is the degrees of freedom.



Series expansion is used to evaluate the cumulative distribution

$$P(x) = \int_0^x f(t)\, dt$$

$$= \left(\frac{x}{2}\right)^{\frac{\nu}{2}} \frac{e^{-\frac{x}{2}}}{\Gamma\left(\frac{\nu+2}{2}\right)} \left[1 + \sum_{k=1}^{\infty} \frac{x^k}{(\nu+2)(\nu+4)\dots(\nu+2k)}\right]$$

The program calculates successive partial sums of the above series. When two consecutive partial sums are equal, the value is used as the sum of the series.

Remarks:

- Program requires $\nu < 141$. If $\nu > 141$, erroneous overflow will result.
- If both x and ν are large, f(x) may result in an overflow error.
- If ν is even,

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2} - 1\right)!$$

If ν is odd,

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2} - 1\right)\left(\frac{\nu}{2} - 2\right) \cdots \left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Reference:

Abramowitz and Stegun, *Handbook of Mathematical Functions,* National Bureau of Standards, 1970.

STEP	INSTRUCTIONS	INPUT	FUNCTION	SIZE: 007 DISPLAY
1.	Initialize the program.		[XEQ] ΣCHISQD	ΣCHISQD
2.	Input degrees of freedom ν.	ν	[A]	($\Gamma(\nu/2)$)
3.	Input x to calculate f(x).	x	[C]	F=(f(x))
4.	Input x to calculate P(x).	x	[E]	P=(P(x))
5.	Repeat step 3 or step 4 if desired.			
6.	For a different ν, go to step 2.			

Examples:

1. If degrees of freedom $\nu = 20$, find f(x), P(x) for x = 9.6 and x = 15.
2. If $\nu = 3$, find f(x) and P(x) for x = 7.82.

Keystrokes:	Display:
[XEQ] [ALPHA] SIZE [ALPHA] 007	
[XEQ] [ALPHA] ΣCHISQD [ALPHA]	ΣCHISQD
20 [A]	362880.00
9.6 [C]	F=0.02
9.6 [E]	P=0.03
15 [E]	P=0.22
15 [C]	F=0.06
3 [A]	0.89
7.82 [C]	F=0.02
7.82 [E]	P=0.95

APPENDIX A

PROGRAM DATA

	PROGRAM	#REG. TO COPY	DATA REGISTERS	FLAGS	DISPLAY FORMAT
1.	Basic Statistics for Two Variables	50	00 ~ 11	00 ~ 03, 21, 27, 29	FIX 2
2.	Moments, Skewness, and Kurtosis	36	00 ~ 11	00 ~ 03, 21, 27, 29	FIX 2
3.	Analysis of Variance (One Way)	29	00 ~ 19	00 ~ 03, 21, 27, 29	FIX 2
4.	Analysis of Variance (Two Way)	33	00 ~ 17	00 ~ 03, 21, 27, 29	FIX 2
5.	Analysis of Covariance (One Way)	60	00 ~ 25	00 ~ 03, 21, 27, 29	FIX 2
6.	Curve Fitting	34	00 ~ 15	00 ~ 03, 21, 27, 29	FIX 2
7.	Multiple Linear Regression	157	00 ~ 44	00 ~ 03, 21, 27, 29	FIX 2
8.	Polynomial Regression	102	00 ~ 44	00 ~ 03, 21, 27, 29	FIX 2
9.	t Statistics	29	00 ~ 14	00 ~ 03, 21, 27, 29	FIX 2
10.	Chi-Square Evaluation	21	00 ~ 07	00 ~ 03, 21, 27, 29	FIX 2
11.	Contingency Table	33	00 ~ 14	00 ~ 03, 21, 27, 29	FIX 2
12.	Spearman's Rank Correlation Coefficient	13	00 ~ 02	00 ~ 03, 21, 27, 29	FIX 2
13.	Normal and Inverse Normal Distribution	47	00 ~ 18	00 ~ 03, 21, 27, 29	FIX 2
14.	Chi-Square Distribution	21	00 ~ 06	00 ~ 03, 21, 27, 29	FIX 2

APPENDIX B

PROGRAM LABELS

*A
*B
*BE
*C
*MD
*MT
ΣANOCOV
ΣAOVONE
ΣAOVTWO
ΣBSTAT
ΣBSTG
ΣCHISQD
ΣCTKK
ΣCTKKK
ΣEEFXSQ
ΣEXP
ΣLIN
ΣLOG
ΣMLRXY
ΣMLRXYZ
ΣMMTGO
ΣMMTUG
ΣNORMD
ΣPOLYC
ΣPOLYP
ΣPOW
ΣPTST
ΣSPEAR
ΣTSTAT
ΣXSQEV

The labels in this list are not in the same order as they appear in the catalog listing for the module.

[illegible], OR 97330, U.S.A.

European H[illegible]
150, Route du [illegible]
P.O. Box, CH-121[illegible]
Geneva-Sw[illegible]

HP-United Kingdom
(Pinewood)
Gb-Nine Mile Ride, Wokingham
Berkshire RG11 3LL

00041-90030 Rev. E English-IK-IPC

Printed in Singapore 8/84