TI Programmable 58/59

Applied Statistics

Using the power of your *Solid State Software*™ module

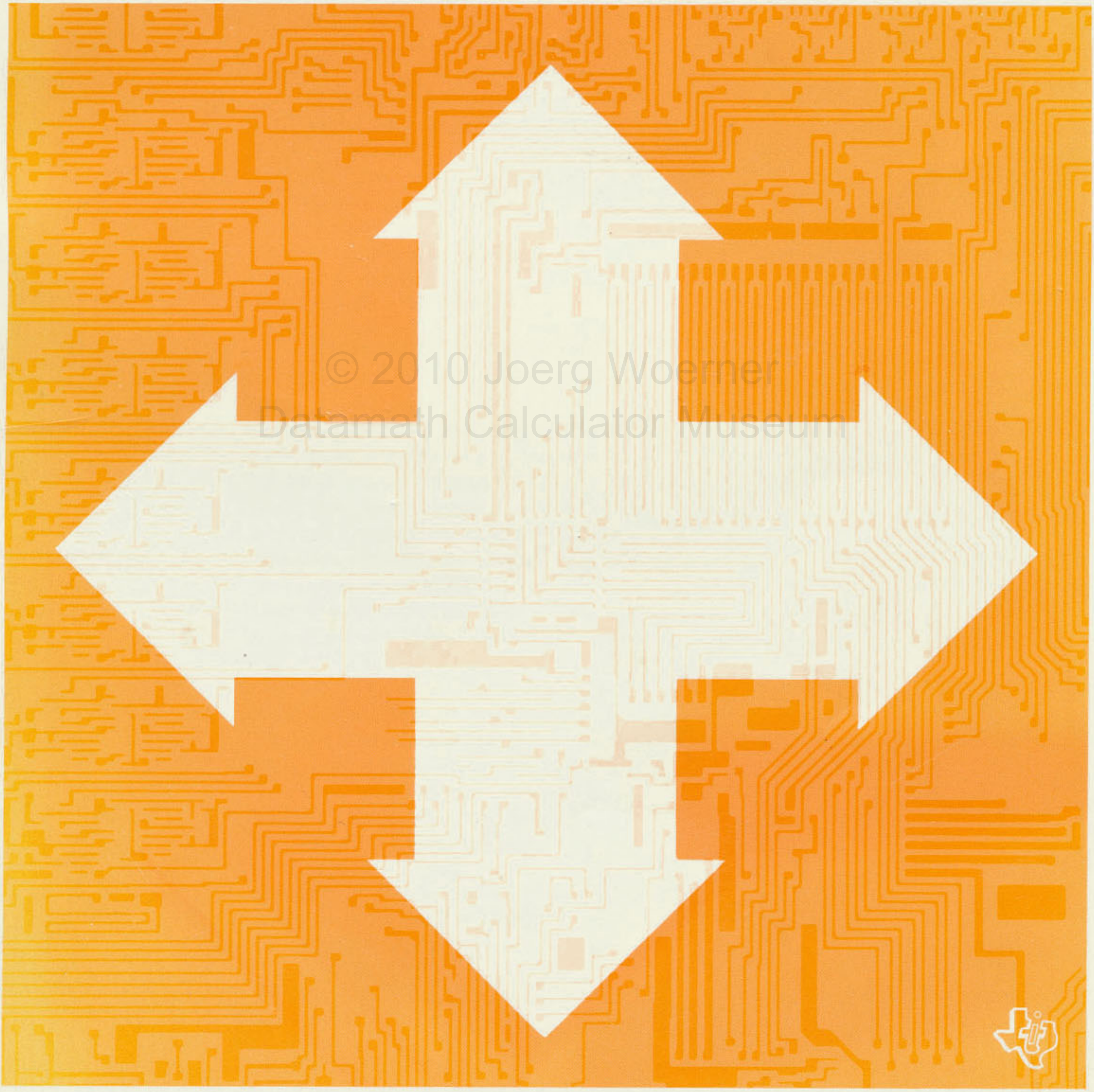


TABLE OF CONTENTS

I. CALCULATOR STATISTICS

Professionals from many fields could make better decisions using statistics — but they have needed a way to make it easy. In the past, the use of statistics required tedious manual calculations or access to expensive computers. But those days are gone forever. Because now, our programmable calculators with their *Solid State Software** libraries place these powerful tools literally into the palm of your hand!

Students, you too can benefit from this library by using it as a valuable learning aid. Don't expect your calculator to replace classroom instruction and self study. But you can use it to gain considerable experience in the practical use of statistics without becoming bogged down in complicated calculations.

No programming skill is needed. However, even though special care has been taken to make this library easy to read and understand, a basic knowledge of statistics is expected. Those of you needing a brief review may find TI's publication, *Calculating Better Decisions,* to be helpful. This text may also help you to extend the capabilities of this library by designing your own programs if you wish.

USING THIS LIBRARY

Your calculator contains a removable *Solid State Software* module which places a large library with a variety of programs at your fingertips the instant you turn the calculator on. Each *Solid State Software* module contains up to 5000 program steps. Within seconds, you can replace the Master Library Module with an optional module, ranging from Applied Statistics to Aviation, to tailor your calculator to solve a series of professional problems with minimal effort. Your *Solid State Software* library does not take up valuable memory space needed for your own programs. In fact, you can call a library program as a subroutine from a program of your own without interruption.

The library programs are discussed in the order that you will need them as each section in this manual is designed to correspond to one of several steps leading to efficient problem solving.

Program documentation includes a complete description of each program and a brief explanation of how the programmed techniques may be applied. The primary reference point for each program is the User Instructions. These instructions are also available for you in the handy pocket guide furnished with your library. When you first run a program, study the program description and complete the sample problems to help you understand the full capabilities and limitations of the program.

*Trademark of Texas Instruments

II. SAMPLE DESIGN

Section II helps you to design and collect your sample. A program used to generate random numbers is of special interest here.

III. DATA ENTRY

Seven programs are described in this section. They are used to enter and organize your data for use by other programs and any other purpose you can think of.

IV. DATA EVALUATION

Now that you have gathered your data you are ready to put it to work. This section describes various programs used for making statistical tests.

V. MODEL FITTING

You may often find the need to fit your data to a model. This section includes a program that compares a theoretical histogram to your data to help you determine the shape of your sample distribution. Various curve fitting procedures are also described.

VI. THEORETICAL DISTRIBUTIONS

The programs described in this section are provided to help you evaluate your test statistic and determine whether to accept or reject your hypothesis.

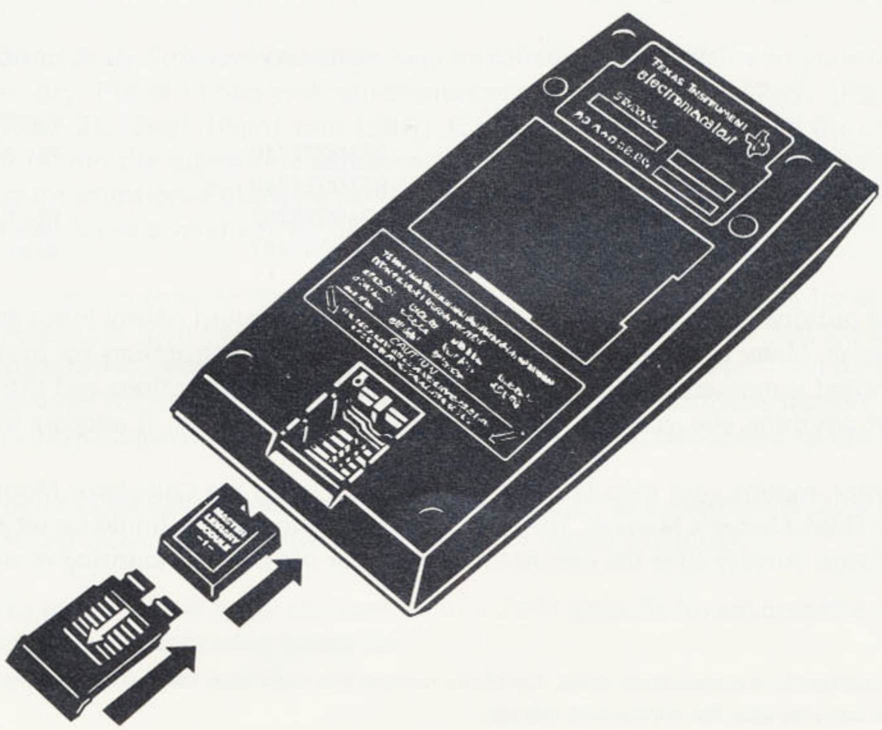
REMOVING AND INSTALLING MODULES

The Master Library module is installed in the calculator at the factory, but can easily be removed or replaced with another. It is a good idea to leave the module in place in the calculator except when replacing it with another module. Be sure to follow these instructions when you need to remove or replace a module.

CAUTION

Be sure to touch some metal object before handling a module to prevent possible damage by static electricity.

1. **Turn the calculator OFF.** Loading or unloading the module with the calculator ON may cause the keyboard or display to lock out. Also, shorting the contacts can damage the module or calculator.
2. Slide out the small panel covering the module compartment at the bottom of the back of the calculator. (See Diagram below.)
3. Remove the module. You may turn the calculator over and let the module fall out into your hand.
4. Insert the module, notched end first with the labeled side up into the compartment. The module should slip into place effortlessly.
5. Replace the cover panel, securing the module against the contacts.



Don't touch the contacts inside the module compartment as damage can result.

RUNNING *SOLID STATE SOFTWARE* PROGRAMS

The Statistics Library contains a variety of useful programs. To help you get started in using the *Solid State Software* programs, install your Statistics Library Module and follow us through a couple of brief examples.

First of all, to eliminate any possibility of having any pending operations or previous results interferring with your current program, turn your calculator off for a couple of seconds, and back on again. This off/on sequence is the assumed starting point for each example problem in this manual. Now press the key sequence [2nd] [Pgm] [0] [1] [SBR] [=] to call and run the *"diagnostic"* program. Notice the display goes blank except for a faint "C " at the far left which indicates that calculations are taking place. After about 15 seconds, "2." will appear in the display. This displayed number indicates that the Statistics Library Module is installed in the calculator and that the calculator and module are operating properly. If the display is flashing after the diagnostic, refer to "In Case of Difficulty" in the SERVICE INFORMATION Appendix of the Owner's Manual.

The diagnostic program is a highly specialized one that works internally to check the operation of your software library. Once you're sure things are working, you can continue with another program in the library. A complete description of this program is found at the end of this section.

Suppose that you need to know the area under the standard normal curve between z = −1.96 and z = 1.96. Look through the nonmagnetic black and gold label cards* and find card ST-19 titled NORMAL DISTRIBUTION. Insert this card in the window above the top row of keys on your calculator. You can now see that the area under the curve from negative infinity to z, P(z), may be found by pressing the [B] key. Now to solve the problem:

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 19		Call Program 19
1.96	[B]	.9750021748	P(1.96)
	[−]	.9750021748	
1.96	[+/−] [B]	.0249978252	P(−1.96)
	[=]	.9500043497	Area

If you have the optional PC-100A printer** you may obtain a record of any input and output data that you wish. Many of the programs in this library include instructions for printing data. Data that is printed is marked by a dagger "†" in both the User Instructions and the examples. You may print anything else you wish by pressing [2nd] [Prt] when it appears in the display.

To use the printer, mount your calculator on the PC-100A using the Calculator Mounting procedure in the PC-100A Owner's Manual. The switch called out in Step 2 should be set to "OTHER" for your calculator. Always turn the calculator and printer off before mounting or unmounting the calculator.

*The cards are supplied in a prepunched sheet. Carefully remove the individual cards from the sheet and insert them in the card carrying case for convenient storage.

**Note: The TI Programmable 58 and TI Programmable 59 will not operate on the PC-100 print cradle.

Before you begin using the Statistics Library programs on your own, here are a few things to keep clearly in mind until you become familiar with your calculator.

1. Press [CLR] before running a program if you are not sure of the status of the calculator. (To be completely sure of calculator status, turn it off and on again – but remember that this clears the program memory.)
2. There is no visual indication of which *Solid State Software* program has been called. If you have any doubts, the safest method is to call the desired program with [2nd] [Pgm] mm, where mm is the two-digit program number. The calculator remains at this program number until another program is called, [RST] is pressed or the calculator is turned off.
3. A flashing display normally indicates an improper key sequence or that a numerical limit has been exceeded. When this occurs, always repeat the program sequence and check that each step is performed as directed by the User Instructions. Any unusual limits of a program are given in the User Instructions or related notes. The In Case of Difficulty portion of Appendix A in the Owner's Manual may be helpful in isolating a problem.
4. Some of the *Solid State Software* programs may run for several minutes depending on input data. If you desire to halt a running program, press the [RST] key. This is considered as an emergency halt operation which returns control to the main memory. A program must be recalled to be run again.

USING *SOLID STATE SOFTWARE* PROGRAMS AS SUBROUTINES

Any of the *Solid State Software* programs may be called as a subroutine to your own program in the main memory. Either of two program sequences may be used: 1) [2nd] [Pgm] mm (User Defined Key) or 2) [2nd] [Pgm] mm [SBR] (Common Label). Both send the program control to program mm, run the subroutine sequence, and then automatically return to the main program without interruption. Following [2nd] [Pgm] mm with anything other than [SBR] or a user-defined key is not a valid key sequence and can cause unwanted results.

It is very important to consider the Program Reference Data in Appendix A for any program called as a subroutine. You must plan and write your own program such that the data registers, flags, subroutine levels, parentheses levels, T-register, angular mode, etc., used by the called subroutine are allowed for in your program. In addition, a Register Contents section of each program description provides a guide to determine where data is or must be located to run the program.

A sample program that calls a *Solid State Software* program as a subroutine is provided in the *PROGRAMMING CONSIDERATIONS* section of the Owner's Manual.

If you need to examine and study the content of a *Solid State Software* program, you can download as described in the following paragraph.

DOWNLOADING *SOLID STATE SOFTWARE* PROGRAMS

If you need to examine a *Solid State Software* program, it can be downloaded into the main program memory.* This allows you to single step through a program in or out of the learn mode. It also allows using the program list or trace features of the optional printer. The only requirement for downloading a *Solid State Software* program is that the memory partition be set so there is sufficient space in the main program memory to receive the downloaded program. The key sequence to download a program is [2nd] [Pgm] mm [2nd] [Op] 09, where mm is the program number to be downloaded. This procedure places the requested program into program memory beginning at program location 000. The downloaded program writes over any instructions previously stored in that part of program memory. Remember to press [RST] before running or tracing the downloaded program.

If you own a TI Programmable 59 the partitioning established when you turn your calculator on is sufficient to download any program in this library. If your calculator is a TI Programmable 58, see Appendix A to determine if you need to repartition your calculator after powering up to download the program you want to examine. Programs which are no longer than 240 steps may be downloaded with the calculator set at the initial partitioning. For longer programs use the key sequence shown on the right to repartition your calculator according to the length of the program you want to download.

Steps	Key Sequence (Needed for 58 Only)
241 - 320	2 [2nd] [Op] 17
321 - 400	1 [2nd] [Op] 17
401 - 480	0 [2nd] [Op] 17

ADDITIONAL NOTES

While every effort has been made to ensure the accuracy of these programs, in the final analysis, *you* must assess program results in light of all available information. Program output that clearly doesn't square with other data should be treated with caution. Your programmable calculator, as a powerful computational instrument, can relieve you of much drudgery. But it can never relieve you of the obligation to exercise your own judgment.

NOTATION

The following list is provided to acquaint you with the notation used in this manual.

- μ - mean of total population or theoretical distribution
- $\overline{x}$ - sample mean
- σ - standard deviation of total population or theoretical distribution
- s - sample standard deviation
- σ^2 - variance of total population or theoretical distribution

*Unless the library is a protected, special-purpose library.

s^2 - sample variance
$\hat{s}$ - standard error of estimate based on sample
ν - degrees of freedom
r - correlation coefficient
r^2 - coefficient of determination
$F(x)$ - continuous distribution function
$F(k)$ - discrete distribution function
$f(x)$ - continuous density function
$f(k)$ - discrete density function
Pr - probability statement
n - sample size
N - population size

REFERENCES FOR FURTHER STUDY

Data evaluation and mathematical statistics:

Brunk, H.D., *An Introduction to Mathematical Statistics,* Blaisdell Publishing Company, Waltham, Mass., 1965

Chou, Ya-lun, *Statistical Analysis,* Holt, Rinehart, and Winston, New York, 1975

DeGroot, Moris H., *Probability and Statistics,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1975

Hoel, Paul G., *Introduction to Mathematical Statistics,* John Wiley and Sons, Inc., New York, 1962

Hoel, Paul G., Port, Sidney C., and Stone, Charles J., *Introduction to Probability Theory* and *Introduction to Statistical Theory,* Houghton Mifflin Company, Boston, 1971

Sample sizing, random sampling, stratified sampling:

Cochran, William G., *Sampling Techniques,* John Wiley and Sons, Inc., New York, 1963

Synthetic sampling and Monte Carlo simulation:

Naylor, Thomas J., et. al., *Computer Simulation Techniques,* John Wiley and Sons, Inc., New York, 1966

STATISTICS LIBRARY DIAGNOSTIC PROGRAM

This program performs the following functions separately.

1. Diagnostic/Library Module Check
2. Linear Regression Initialization

DIAGNOSTIC/LIBRARY MODULE CHECK

This routine checks the operation of your calculator and most of its functions, including conversion and statistics functions that are preprogrammed in the calculator, trigonometric functions, data register operations, program transfers, and comparisons. It also uses other Statistics Library programs to verify that the module is connected and operating correctly. If this diagnostic routine runs successfully, in approximately 15 seconds the numeral 2. is displayed. If the calculator is attached to a PC-100A print cradle, the following is printed:

```
STATISTICS
     2.
```


If there is a malfunction in the calculator or the *Solid State Software* module, a flashing number is displayed. Refer to Appendix A of the Owner's Manual for an explanation of the various procedures to be followed when you have difficulties.

When you simply want to know which of your *Solid State Software* modules is in the calculator without physically looking at it, you can call the Library Module check portion of the routine directly. If the Statistics Library Module is in the calculator, the number 2. is displayed. This number is unique to the Statistics Library (other optional libraries use other identifying digits).

LINEAR REGRESSION INITIALIZATION

This routine initializes the calculator for linear regression by clearing registers R_{01} through R_{06} and the T-register. It should be used whenever linear regression or other built-in statistics functions are to be started. You can also use the routine at any time to clear these registers selectively without disturbing any other registers.

Solid State Software	TI ©1977
STATISTICS LIBRARY DIAGNOSTIC	ST-01
DIAGNOSTIC: [SBR] [=]	
LINEAR REGRESSION INITIALIZATION: [SBR] [CLR]	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	Diagnostic/Module Check			
1a	Select Program		[2nd] [Pgm] 01	
1b	Run Diagnostic		[SBR] [=]	2.[1]
	or			
1c	Library Module Check		[SBR] [2nd] [R/S]	2.[2]
	Initialize Linear Regression			
2a	Select Program		[2nd] [Pgm] 01	
2b	Initialize Linear Regression		[SBR] [CLR]	0.

NOTES:
1. This output is obtained if the calculator is operating properly
2. The number 2 indicates the Statistics Library.
3. The Statistics Library programs are numbered 1 through 22. Program number 0 is the calculator's program memory.

Register Contents

R_{00}	R_{05} LR Init	R_{10}	R_{15}	R_{20}	R_{25}
R_{01} LR Init	R_{06} LR Init	R_{11}	R_{16}	R_{21}	R_{26}
R_{02} LR Init	R_{07}	R_{12}	R_{17}	R_{22}	R_{27}
R_{03} LR Init	R_{08}	R_{13}	R_{18}	R_{23}	R_{28}
R_{04} LR Init	R_{09} Used	R_{14}	R_{19}	R_{24}	R_{29}

Example 1:

Diagnostic

PRESS	DISPLAY	OPTIONAL PRINTOUT
[2nd] [Pgm] 01 [SBR] [=]	2.	STATISTICS 2.

Example 2:

Library Module Check

PRESS	DISPLAY	OPTIONAL PRINTOUT
[2nd] [Pgm] 01 [SBR] [2nd] [R/S]	2.	STATISTICS 2.

Example 3:

Initialize Linear Regression

PRESS	DISPLAY
[2nd] [Pgm] 01 [SBR] [CLR]	0.

II. SAMPLE DESIGN

For reasons of economy most statistical projects begin with the design of a sample from which the characteristics of a population are to be estimated. Selecting a good sample is probably the most important step in the designing of your experiment. Some constraints, such as cost and time limitations, may be beyond your control. However, factors you can control such as how many members of the population should be sampled and which members you should include in your sample, must be carefully determined. Your calculator may be used as a valuable aid in making these decisions.

SAMPLE SIZING

If a sample is random, the precision of population parameter estimates are related to sample size by the equations shown in Table 2.1. You may use these equations to determine a sample size that should give you the precision you need in your estimation.

When using any of the methods outlined in Table 2.1 you must first make an initial estimate of the population parameter indicated in the center column. This estimate may be based on theory, judgment, prior data, or pilot samples. The next step is to use your estimate of this parameter in the formula on the right to find the standard error of the estimate, $\hat{s}$. To determine how large your sample needs to be, simply evaluate $\hat{s}$ for increasing sample sizes until this value is reduced to a satisfactory level. Then once you have found an appropriate sample size, you may take a sample and use the formula on the left to estimate the indicated parameter. (Note that n represents the sample size while N is the population size.)

Table 2.1

Purpose of Sample	Method of Estimating Sample Size	Formula for Estimating Sample Size
Estimate's a population's mean as $\hat{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.	Estimate the population's standard deviation by any method you desire. Then use this estimate and the equation on the right to find a sufficient sample size as described above.	$\hat{s} = \sqrt{\frac{\hat{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}$
Estimate a population's total value Y as $\hat{Y} = N\bar{y}$.	Same as the above.	$\hat{s} = \sqrt{\frac{N^2\hat{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}$
Estimate the ratio of two population values X and Y as $\hat{R} = \bar{x}/\bar{y}$.	Take a small pilot sample and make an initial estimate of R as $\hat{R} = \bar{x}/\bar{y}$. Then use this value and the formula on the right to determine an appropriate sample size as described above.	$\hat{s} = \sqrt{\left(\frac{1 - n/N}{n\,\bar{x}^2}\right)\left(\frac{\sum_{i=1}^{n} (y_i - \hat{R}x_i)^2}{n-1}\right)}$

(continued on next page)

Purpose of Sample	Method of Estimating Sample Size	Formula for Estimating Sample Size
Estimate the proportion of a population possessing an attribute A as $\hat{P} = \overline{y}$.	Assuming $y_i = 1$ if it possesses an attribute A and $y_i = 0$ if it does not, take a small pilot sample and let p equal $\overline{y}$. Then use p in the equation on the right to determine how large of a sample is needed as described above.	$\hat{s} = \sqrt{\left(\frac{N-n}{(n-1)N}\right) p(1-p)}$
Estimate μ_y indirectly through linear regression $\hat{\mu}_{y\,(lr)} = \overline{y} + b(\mu_x - \overline{x}) + \overline{e}$ Where μ_x is known and b is the estimate of the change in y when x is increased by 1. (A population's total value Y may be estimated as $\hat{Y} = N\hat{\mu}_{y\,(lr)}$).	Take a small pilot sample and estimate σ_e as $\hat{\sigma}_e = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[(y_i-\overline{y})-b(x_i-\overline{x})]^2}$. Then use this result and the equation on the right to find a sufficient sample size as described above.	If the x's are normally distributed: $\hat{s} = \sqrt{\frac{\hat{\sigma}_e^{\,2}}{n}\left(1 + \frac{1}{n-3}\right)}$ Otherwise: $\hat{s} = \sqrt{\frac{\hat{\sigma}_e^{\,2}}{n}\left(1 + \frac{1}{n} + \frac{3+Sk}{n^2}\right)}$ where Sk is the measure of the relative skewness of the distribution of the x's.

Example:

As an educational administrator, you want to know how many students would attend a new school supporting a district of 500 households. You would like to survey a minimum number of homes to determine the average number of students eligible per home and multiply by 500 to obtain your estimate. Your estimate should be accurate within 30% at a 95% confidence level (within about 2 standard deviations). That is, the standard error of your estimate should be less than $.15 \times 2 \times 500 = 150$.

From Table 2.1 we see that

$$\hat{s} = \sqrt{(N^2\hat{\sigma}^2/n)(1 - n/N)}$$

where:

$\hat{s}$ = the standard error of the estimate ($\hat{s} < 150$ is desired),

N = the population size (500 households),

n = the sample size (to be determined),

σ = the population standard deviation (estimated below).

From prior experience you know that more than 6 students in a family is very rare and that most families have only 2. Since population extremes are usually no more than 2.5 standard deviations from the mean you may estimate the population standard deviation as

$$\hat{\sigma} = (x_{max} - \overline{x})/2.5 = (6 - 2)/2.5 = 1.6.$$

Now program your calculator with the first equation and determine $\hat{s}$ for various sample sizes. You may use the sequence shown below.

	[2nd] [CP]		[1]	(020)	[RCL]
	[LRN]	(010)	[·]		[1]
(000)	[2nd] [Lbl]		[6]		[÷]
	[A]		[x^2]		[5]
	[STO]		[÷]		[0]
	[0] [1]		[RCL]	(025)	[0]
	[5]	(015)	[1]		[ )]
(005)	[0]		[×]		[=]
	[0]		[( ]		[$\sqrt{x}$]
	[x^2]		[1]		[R/S]
	[×]		[−]	(030)	[LRN]

ENTER	PRESS	DISPLAY
15	[A]	203.437132
20	[A]	175.2712184
25	[A]	155.9487095
30	[A]	141.6097925

A sample size of 25 yields an adequate value of $\hat{s}$.

SAMPLE SELECTION

You now need to determine which members of the population should be included in your sample. Your sample should be randomly selected to avoid accidental bias. The *Random Number Generator Program* at the end of this section is designed to help you here. Simply generate 500 random numbers between 0 and 1 using this program and assign one number to each household in the district. Now, since you want to sample 5% of the population (25 households out of 500), include only those households to which you assigned a number that is less than or equal to 0.05.

SAMPLE EVALUATION

Suppose that the sample selected above yielded the following sample of students per family:

3, 4, 1, 1, 0, 1, 1, 1, 2, 2, 1, 0, 5, 2, 1, 3, 2, 2, 1, 1, 2, 1, 1, 0, 0.

Based upon this data you may now reestimate the population mean and standard deviation as 1.52 for the mean and about 1.23 for the standard deviation. You may use either the statistical functions built into your calculator or the *Means and Moments Program* in Section IV, to perform these calculations. (The answers given above were found using the built-in features [2nd] [$\bar{x}$] and [INV] [2nd] [$\bar{x}$].)

As the resulting estimate of the standard deviation is within the limits we established earlier you may now estimate the number of students in the district as 1.52 × 500 = 760.

STRATIFIED SAMPLING *(Improving Sample Sensitivity)*

Recall from the sample sizing formula that if a population has a large intrinsic variance, it takes many sample points to estimate its parameters. In many cases, you may be able to separate a population into subgroups which have differing variability. You should then collect more data from the groups with high variability where you need it the most. This technique is known as stratified sampling. You may use it to obtain a more exact estimate of the population mean than you can find by using a simple random sample. When using this technique, you should make the number of data points collected in each subgroup or strata proportional to the population of the subgroup and the size of its variance.

Example:

Suppose that in the previous example you decided that you didn't have time to survey 25 households. However, using a simple random sample with fewer points wouldn't give you the needed accuracy. But, based upon the wealth and birthrate patterns in the school district you were able to break the population into two strata as shown in Table 2.2.

Table 2.2

	Population (N)	Std Dev (σ)	$N \times \sigma$	Per Cent of Data Points Allocated
Low Income Families	250	1.5	375	75%
High Income Families	250	0.5	125	25%
Total Population	500	1.0	500	100%

The next step is to use the following equation (see Table 2.1) to determine the size of your sample. Remember, you want the standard error of your estimate to be less than 150.

$$\hat{s}^2 = (N^2 \sigma^2 / n)(1 - n/N).$$

Table 2.3

Sample	Sample Sizes	Low Income Variance $\hat{s}_L^2$	High Income Variance $\hat{s}_H^2$	District Variance $\hat{s}^2$	Precision $\hat{s}$
I	19 + 6 = 25	6,839	2,542	9,381	97
II	12 + 4 = 16	11,156	3,844	15,000	122
III	9 + 3 = 12	15,063	5,146	20,209	142
IV	7 + 2 = 9	19,527	7,750	27,277	165

Sample III presents the best combination of precision and economy of the experiment. Note that since the strata are independent the district variance is simply the sum of the strata variances.

You may now use the *Random Number Generator Program* to choose the households for your survey and estimate the number of students in the district as in the last example except that only 9 of the 250 low income and 3 of the 250 high income families need be included.

MONTE CARLO SIMULATION *(Synthetic Sampling)*

There are experiments in which you cannot obtain a particular set of data by directly sampling the population. However, if partial data is available you may be able to predict or synthesize sample points.

Example:

The school board is considering going to a new policy of charging for textbooks on a per capita cost recovery basis. However, there is some concern that the cost may be too high for large families. Due to this concern, you have been asked to determine what proportion of families with children in school will have to spend over $250 for books.

This task may appear difficult at first because the students have no idea of what their books will cost. As a result, you cannot obtain data for your estimate by simply sampling the population. But what do you know about the situation? Based on your experience as a school administrator you know how much you've had to spend on books for your students in the past. Using this knowledge you may estimate the distribution of your population. Let's assume that you estimated costs per student to be normally distributed with a mean of $100 and a standard deviation of $35.

Now, with this information at hand, you may use Monte Carlo simulation to predict your data points. Assign each student in your original sample (See *Sample Evaluation.*) a charge by using the *Random Number Generator Program* and the parameters estimated above to generate normally distributed charges. Suppose that the following data resulted from this process.

Table 2.4

	$	Individual Payment per Student					Family Payment
		1	2	3	4	5	
	1	62	104	59	-	-	225
	2	108	132	129	28	-	397
	3	94	-	-	-	-	94
	4	62	-	-	-	-	62
	5	112	-	-	-	-	112
	6	63	-	-	-	-	63
	7	76	-	-	-	-	76
	8	134	110	-	-	-	244
	9	65	180	-	-	-	245
	10	85	-	-	-	-	85
Family	11	49	90	128	127	53	447
	12	79	102	-	-	-	181
	13	58	-	-	-	-	58
	14	98	117	63	-	-	278
	15	97	78	-	-	-	175
	16	74	89	-	-	-	163
	17	79	-	-	-	-	79
	18	69	-	-	-	-	69
	19	69	108	-	-	-	177
	20	80	-	-	-	-	80
	21	138	-	-	-	-	138

From Table 2.4 you can see that 14% (3 of 21) of the families will have to pay more than $250 for books. Note that families without children are not considered in this example.

RANDOM NUMBER GENERATOR PROGRAM

This program generates sequences of uniformly or normally distributed random numbers. To use this program, simply enter a seed number (0 to 199017), select the distribution, and enter its parameters.

For uniformly distributed random numbers, enter the upper and lower limits of the range you desire. Uniformly distributed random numbers are then generated using the following formula.

$$x_{i+1} = [(24298x_i + 99991) \bmod 199017]\ (x_{max} - x_{min})/199017 + x_{min}$$

where x_0 is your seed.

If you choose to generate normally distributed numbers, enter the mean and standard deviation that you want the generated numbers to have. Each random number is then generated by first generating a pair of uniform random numbers (u_1, u_2). The normally distributed random number is then found by

$$x = \sqrt{-2 \ln u_1}\ \cos(2\pi u_2)\sigma + \mu.$$

As an added feature, this program uses the [2nd] [Σ+] instruction to compile statistical data and allow you to compute the actual mean and standard deviation of the generated numbers. Also, subroutine [D.MS] may be used to generate uniformly distributed random numbers on the interval (0, 1) at any time. However, the statistical data feature does not apply to this routine.

Solid State Software			TI ©1977	
RANDOM NUMBER GENERATOR				ST-02
				Initialize
x_{min} / μ	x_{max} / σ	→ Uniform	→ Normal	Seed

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 02	No Change
2	Initialize		[2nd] [E']	0.
3	Enter random number seed ($0 \leqslant$ Seed $\leqslant 199017$)	Seed†	[E]	Seed
	For Uniform Distribution			
4	Enter lower limit	x_{min}†	[A]	x_{min}
5	Enter upper limit	x_{max}†	[B]	x_{max}
6	Generate random number[1] (Repeat Step 6 as needed)		[C]	Random Number†
	For Normal Distribution			
7	Enter desired mean	μ†	[A]	μ
8	Enter desired standard deviation	σ†	[B]	σ
9	Generate random number[1] (Repeat Step 9 as needed)		[D]	Random Number†
	For Either Distribution			
10	Compute actual mean of generated numbers		[2nd] [$\bar{x}$]	$\bar{x}$
11	Compute actual standard deviation of generated numbers		[INV] [2nd] [$\bar{x}$]	s
12	Display number of random numbers generated		[RCL] 03	n
	For Range of (0, 1)			
13	Generate random number[1] (Repeat Step 13 as needed)		[SBR] [2nd] [D.MS]	Random Number

NOTES: 1. Only the first five digits may be considered random.
† Printed when PC-100A is used.

Register Contents

R_{00}	R_{05} Σx^2	R_{10} Used	R_{15}	R_{20}	R_{25}
R_{01} Σy	R_{06} Σxy	R_{11} Used	R_{16}	R_{21}	R_{26}
R_{02} Σy^2	R_{07}	R_{12} x_{min}, μ	R_{17}	R_{22}	R_{27}
R_{03} n	R_{08}	R_{13} x_{max}, σ	R_{18}	R_{23}	R_{28}
R_{04} Σx	R_{09} Seed	R_{14}	R_{19}	R_{24}	R_{29}

SAMPLE DESIGN

Example:

Compute five uniformly distributed random numbers on the interval (1, 10). Use .32 as your seed.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 02		Select Random Number Program
	[2nd] [E′]	0.	Initialize
.32†	[E]	0.32	Seed
1.†	[A]	1.	x_{min}
10.†	[B]	10.	x_{max}
	[C]	5.87341†	Random No.
	[C]	7.34635†	Random No.
	[C]	3.5911†	Random No.
	[C]	1.63531†	Random No.
	[C]	9.05329†	Random No.

Example:

Compute five normally distributed random numbers with desired mean 5.84 and standard deviation 2.12. Also find the actual mean and standard deviation of the generated numbers. Use 1 as your seed.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 02		Select Random Number Program
	[2nd] [E′]	0.	Initialize
1†	[E]	1.	Seed
5.84†	[A]	5.84	μ
2.12†	[B]	2.12	σ
	[D]	7.8171433†	Random No.
	[D]	7.290557451†	Random No.
	[D]	3.075542923†	Random No.
	[D]	5.109539381†	Random No.
	[D]	3.323206704†	Random No.
	[2nd] [$\overline{x}$]	5.323197952	$\overline{x}$
	[INV] [2nd] [$\overline{x}$]	2.190196047	s
	[RCL] 03	5.	n

† Printed when PC-100A is connected.

III. DATA ENTRY

The programs described in this section are used as general input and storage routines to allow various data evaluation, model fitting, and testing procedures to be applied without the labor of data reentry. Besides collecting the data in an organized fashion, these routines generate commonly used intermediate results. Separate routines are used to assimilate data for various data storage formats as illustrated in Table 3.1. See Table 3.2 for the limits of the raw data base.

Table 3.1a — Intermediate Data

Register	Pgm 03 Univariate Data		Pgm 04 Bivariate Data	Pgm 05 Trivariate Data	Pgm 06		Pgm 07 Histogram Data
	Ungrouped	Grouped			One-Way AOV Data	Two-Way AOV Data	
00	Used	Used	Used	Used	Used	Used	Cell No.
01			Σy	Σy	$\Sigma\Sigma x$	Rows	x_{min}
02			Σy^2	Σy^2	$\Sigma\Sigma x^2$	Columns	Width
03	n	Σf	n_y	n	n	n	n
04	Σx	Σfx	Σx	Σx	Σx	Σx	Σx
05	Σx^2	Σfx^2	Σx^2	Σx^2	Σx^2	Σx^2	Σx^2
06	Used	Used	Σxy	Σxy	Mean	Mean	Cell 1 Cnt
07	Σx^3	Σfx^3	Σx^3	Σz	Variance	Variance	Cell 2 Cnt
08	Σx^4	Σfx^4	Σx^4	Last z	Used	Used	.
09	Last x	Last x	Last x	Last x	Σn		.
10	1	Last f	Last y	Last y	$\Sigma(\Sigma x)^2/n$		.
11	πx	πx^f	$\Sigma x^2 y$	Σxz	Used		.
12	Low x	Low x	$\Sigma(x - y)$	Σyz		Σx_{i1}	.
13	High x	High x	$\Sigma(x - y)^2$	Σz^2		Σx_{i2}	.
14	Mid x	Mid x	$\Sigma(x - y)^2/y$	Used		.	.
15	$\Sigma x_i x_{i+1}$	$\Sigma x_i x_{i+1}$	n_x			.	.
16	$\Sigma(x_{i+1} - x_i)$	$\Sigma(x_{i+1} - x_i)$	Σy^3			(See Note)	.
17	$\Sigma 1/x$	$\Sigma f/x$	Σy^4			.	Cell 12 Cnt
18	Used	Used				.	Last x
19						Σx_{iC}	Cells
20						Σx_{1j}	
21						Σx_{2j}	
22						.	
23						.	
24						.	
25						.	
26	Used	Used	Used	Used	Used	.	Used
27				z Count		Σx_{Rj}	
28	Used	Used	y Count	y Count	i Count	i Count	Used
29	x Count	x Count	x Count	x Count	j Count	j Count	x Count

NOTE: For *Two-Way AOV Data* the actual sums stored in registers R_{12}–R_{27} will vary depending upon the number of rows and columns that your data requires. If R + C is less than 16 then registers $R_{(12+R+C)}$–R_{27} are not used.

Table 3.1b – Raw Data

Register	Pgm 03 Univariate Data		Pgm 04 Bivariate Data	Pgm 05 Trivariate Data	Pgm 06		Pgm 07 Histogram Data
	Ungrouped	Grouped			One-Way AOV Data	Two-Way AOV Data	
30	Pointer				Pointer		Pointer
31	x_1	Pointer	Pointer		x_1		x_1
32	x_2	x_1	x_1	Pointer	x_2		x_2
33	x_3	f_1	y_1	x_1	x_3		x_3
34	x_4	x_2	x_2	y_1	x_4		x_4
35	x_5	f_2	y_2	z_1	x_5		x_5
36	x_6	x_3	x_3	x_2	x_6		x_6
37	x_7	f_3	y_3	y_2	x_7		x_7
.	.	.	.	.	.		.
.	.	.	.	.	.		.
.	.	.	.	.	.		.

As you can see, the data base is divided into two sections. Registers R_{00} – R_{29} make up the intermediate data base. This is where information for use in other programs is stored. The real advantage of this system is the fact that you may also write your own programs to use this intermediate data in any way you wish.

The raw data base is not used by any other program in this library except the *Rank-Sum Tests Program.* However, you may use it to great advantage. One feature of the data entry programs is that there are two ways to accumulate or compile intermediate results in the intermediate data base.

- As you enter raw data from the keyboard it is stored in the raw data base and compiled in the intermediate data base at the same time.
- If your raw data has already been stored you can compile the intermediate data base using a single user-defined key. The advantages of this feature are described below.

Naturally, the first time you enter a collection of data you must use the keyboard method. But what if you would like to use this same data again later? With the TI Programmable 59 you can record your raw data on magnetic cards. Then, when you come back to your calculator, all you have to do is read the data cards and compile the intermediate data base using the second method. Now, suppose that you make a wrong data entry. How do you correct it? The answer is simple if you follow these easy steps. (Detailed procedures are found in the user instructions.)

1. Delete the bad data from the raw data base.
2. Reinitialize the intermediate data base.
3. Compile a new intermediate data base from the raw data base.
4. Enter the correct data and continue entering new data.

Each program uses an initialization routine to set a raw data pointer and prepare the data base for input. These initialization routines also partition the calculators storage area to:

- 480 program locations and 60 data registers in the TI Programmable 59.
- Zero program locations and 60 data registers in the TI Programmable 58.

This provides up to 29 registers for storing raw data. With the TI Programmable 59 you may repartition the storage area to allow as many as 69 registers for raw data storage. Naturally, if you need more program memory space, you may repartition the calculator to allow as few as 40 data registers for the entire data base. To repartition your calculator simply press n [2nd] [Op] **17**. This sequence gives you n × 10 data registers. See your owner's manual for a complete explanation of partitioning.

If you fill the raw data base in either calculator the display flashes when you attempt to enter additional data. If you wish to record this data on magnetic cards you may do so at this time. Then, to enter additional data, reposition the raw data pointer and write over the raw data base with new data. The user instructions detail how this is done. Note, however; that the old raw data is lost unless it is first recorded on magnetic cards.

The following list illustrates the data entry routines required by the programs in this library. Programs that are not listed have their own data entry routines or call one of the data entry routines themselves.

Program	Data Entry Routine
Means and Moments (ST-08)	Univariate Data (ST-03)
Histogram Contruction (ST-09)	Histogram Data (ST-07)
Theoretical Histogram (ST-10)	Histogram Data (ST-07)
t-Statistic Evaluation (ST-13)	Bivariate Data (ST-04)
One-Way AOV (ST-15)	AOV Data (ST-06)
Two-Way AOV (ST-16)	AOV Data (ST-06)
Rank Sum Tests (ST-17)	Bivariate Data (ST-04)
Multiple Linear Regression (ST-18)	Trivariate Data (ST-05)

See the discussions of the programs on the left for data entry examples.

DATA ENTRY NOTES

The following notes apply to all *Data Entry* programs.

1. Initialization affects only the intermediate data base and the raw data pointer. The raw data base is not disturbed. However, you may want to clear these registers using the [CMs] key *before initialization* (check partitioning). Initialization also provides 60 data registers as described earlier.

2. The calculator ignores data entered after the raw data base is filled. This condition is indicated by a flashing display. You may determine how many pieces of data you can store in the raw data base using the following table. This table gives the upper limit of complete sets of data that may be stored for the indicated partitioning.

Table 3.2.

Upper Limit of Data	Partitioning (n) 4	5	6	7	8	9	10
Univariate Data (Ungrouped), Analysis of Variance Data, and Histogram Data	x_9	x_{19}	x_{29}	x_{39}	x_{49}	x_{59}	x_{69}
Univariate Data (Grouped)	x_4 f_4	x_9 f_9	x_{14} f_{14}	x_{19} f_{19}	x_{24} f_{24}	x_{29} f_{29}	x_{34} f_{34}
Bivariate Data	x_4 y_4	x_9 y_9	x_{14} y_{14}	x_{19} y_{19}	x_{24} y_{24}	x_{29} y_{29}	x_{34} y_{34}
Trivariate Date	x_2 y_2 z_2	x_5 y_5 z_5	x_9 y_9 z_9	x_{12} y_{12} z_{12}	x_{15} y_{15} z_{15}	x_{19} y_{19} z_{19}	x_{22} y_{22} z_{22}

3. Follow these steps to record the data base on magnetic cards.
    1. Place the bank number of the registers you wish to record in the display.
    2. Press [2nd] [Write].
    3. Insert magnetic card in card slot.

   The bank number of the intermediate data base ($R_{00} - R_{29}$) is 4. The bank numbers of the raw data base are given below.

Registers	Bank Number
$R_{30} - R_{59}$	3
$R_{60} - R_{89}$	2
$R_{90} - R_{99}$	1

   Note that bank 1 includes program memory.

4. Resetting the raw data pointer to the beginning of the raw data base allows you to continue entering new raw data after filling the raw data base by writing over previously entered data. Although the intermediate data base is not affected, overwritten raw data is lost unless first stored on a magnetic card. Note that you may obtain a hardcopy printer listing of the data registers by entering the number of the first register you want listed and pressing [INV] [2nd] [List]. Then press [R/S] when you want to stop.

5. If you have already compiled your intermediate data base and recorded it on a magnetic card, simply read that card and go on to the data evaluation programs. You don't even have to call the Data Entry program to do this.

6. The length of execution time increases with the number of data points when the intermediate data base is compiled directly from the raw data base.

7. Data must be deleted in the same form it is entered in (e.g., pairs, triplets, etc.). Data that has been overwritten may not be deleted. If the calculator cannot find the data you have asked it to delete, nines are flashed in the display. This process may take several seconds to complete.

UNIVARIATE DATA (UNGROUPED)

Solid State Software			TI ©1977	
UNIVARIATE DATA				ST-03
Delete x	Delete f	Pointer (G)	Compile G	Init G
x	f	Pointer (UG)	Compile UG	Init UG

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 03	No Change
2	Initialize Data Base[1]		[E]	1.
3	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
4	Enter data (repeat for each x_i)[2]	x_i†	[A]	i
	If Raw Data Base is filled:			
5a	Record raw data on magnetic card(s) if desired[3]			
5b	Reset Raw Data Pointer[4] and go to Step 4 to enter additional data		[C]	31.
6	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
7	Read raw data card(s)		[CLR]	0.
		Card		Bank No.
8	Reset Raw Data Pointer[4]		[C]	31.
9	Compile Intermediate Data Base (raw data is printed)[6]		[D]	Last i
10	For additional data cards – go to Step 7			
	To delete data[7]**:**			
11	Enter unwanted data	x_i	[2nd] [A′]	x_i
12	Initialize Data Base[1]		[E]	1.
13	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
14	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[D]	Last i
15	Reenter raw data that has been overwritten using either Steps 4-6 or 7-10			
16	Continue entering new data			

NOTES: See Data Entry Notes.

*For TI-58, repartition by pressing 6 [2nd] [Op] 17.

UNIVARIATE DATA (GROUPED)

Solid State Software				TI ©1977
UNIVARIATE DATA				ST-03
Delete x	Delete f	Pointer (G)	Compile G	Init G
x	f	Pointer (UG)	Compile UG	Init UG

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 03	No Change
2	Initialize Data Base[1]		[2nd] [E']	1.
3	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
4a	Enter frequency[8]	f_i †	[B]	f_i
4b	Enter data (repeat Step 4 for each x_i)[2]	x_i †	[A]	i
	If Raw Data Base if filled:			
5a	Record raw data on magnetic card(s) if desired[3]			
5b	Reset Raw Data Pointer[4] and go to Step 4 to enter additional data		[2nd] [C']	32.
6	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
7	Read raw data card(s)	Card	[CLR]	0 Bank No.
8	Reset Raw Data Pointer		[2nd] [C']	32.
9	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D']	Last i
10	For additional data cards – go to Step 7			
	To delete data[7]:			
11a	Enter frequency	f_i	[2nd] [B']	f_i
11b	Enter unwanted x_i	x_i	[2nd] [A']	x_i
12	Initialize Data Base[1]		[2nd] [E']	1.
13	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
14	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D']	Last i
15	Reenter raw data that has been overwritten using either Steps 4-6 or 7-10			
16	Continue entering new data			

NOTES: See Data Entry Notes for 1-7.

8. The frequency should be a positive integer. The display flashes for negative entries and zero; but no test is made for noninteger entries.

† Printed when PC-100A is used.

*For TI-58, repartition by pressing 6 [2nd] [Op] 17.

BIVARIATE DATA

Solid State Software				TI ©1977
BIVARIATE DATA				ST-04
Delete x	Delete y		Compile	Initialize
x	y		Pointer	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 04	No Change
2	Initialize Data Base[1]		[2nd] [E′]	0.
3	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
4a	Enter x_i	x_i†	[A]	i
4b	Enter y_i	y_i†	[B]	i
	(Repeat Step 4 for each data pair)[2]			
	If Raw Data Base is filled:			
5a	Record raw data on magnetic card(s) if desired[3]			
5b	Reset Raw Data Pointer[4] and go to Step 4 to enter additional data		[D]	32.
6	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
7	Read raw data card(s)	Card	[CLR]	0 Bank No.
8	Reset Raw Data Pointer		[D]	32.
9	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D′]	Last i
10	For additional data cards – go to Step 7			
	To delete data[7]:			
11a	Enter unwanted x_i	x_i	[2nd] [A′]	x_i
11b	Enter unwanted y_i	y_i	[2nd] [B′]	y_i
12	Initialize Data Base[1]		[2nd] [E′]	0.
13	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
14	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D′]	Last i
15	Reenter raw data that has been overwritten using either Steps 4-6 or 7-10			
16	Continue entering new data			

NOTES: See Data Entry Notes.

† Printed when PC-100A is used.

*For TI-58, repartition by pressing 6 **[2nd] [Op] 17**.

TRIVARIATE DATA

Solid State Software			TI ©1977	
TRIVARIATE DATA				ST-05
Delete x	Delete y	Delete z	Compile	Initialize
x	y	z	Pointer	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 05	No Change
2	Initialize Data Base[1]		[2nd] [E′]	0.
3	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
4a	Enter x_i	x_i†	[A]	i
4b	Enter y_i	y_i†	[B]	i
4c	Enter z_i	z_i†	[C]	i
	(Repeat Step 4 for each data triplet)[2]			
	If Raw Data Base is filled:			
5a	Record raw data on magnetic card(s) if desired[3]			
5b	Reset Raw Data Pointer[4] and go to Step 4 to enter additional data		[D]	33.
6	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
7	Read raw data card(s)	Card	[CLR]	0 Bank No.
8	Reset Raw Data Pointer		[D]	33.
9	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D′]	Last i
10	For additional data cards — go to Step 7			
	To delete data[7]:			
11a	Enter unwanted x_i	x_i	[2nd] [A′]	x_i
11b	Enter unwanted y_i	y_i	[2nd] [B′]	y_i
11c	Enter unwanted z_i	z_i	[2nd] [C′]	z_i
12	Initialize Data Base[1]		[2nd] [E′]	0.
13	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
14	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D′]	Last i
15	Reenter raw data that has been overwritten using either Steps 4-6 or 7-10			
16	Continue entering new data			

NOTES: See Data Entry Notes.

† Printed when PC-100A is used.

*For TI-58, repartition by pressing 6 **[2nd] [Op] 17.**

ONE-WAY ANALYSIS OF VARIANCE DATA

Solid State Software		TI ©1977		
ANALYSIS OF VARIANCE DATA				ST-06
Delete x	$\bar{x}$	s^2	Compile	Init 1-Way
x	R	C	Pointer	Init 2-Way

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 06	No Change
2	Initialize Data Base[1]		[2nd] [E′]	0.
3	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
4	Enter data for Treatment Group i (repeat for each j)[2]	x_{ij}†	[A]	j
	If Raw Data Base is filled:			
5a	Record raw data on magnetic card(s) if desired[3]			
5b	Reset Raw Data Pointer[4] and go to Step 4 to complete entry of data for Current Treatment Group		[D]	31.
6	Calculate $\bar{x}$ for Current Treatment Group		[2nd] [B′]	$\bar{x}$†
7	Display s^2 for Current Treatment Group		[2nd] [C′]	s^2†
8	Go to Step 4 for Next Treatment Group[8]			
9	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
10	Read raw data card(s) for Treatment Group i	Card	[CLR]	0. Bank No.
11	Reset Raw Data Pointer[4]		[D]	31.
12	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D′]	Last j
13	To enter additional data cards for Current Treatment Group — go to Step 10			
14	Calculate $\bar{x}$ for Current Treatment Group		[2nd] [B′]	$\bar{x}$†
15	Display s^2 for Current Treatment Group		[2nd] [C′]	s^2†
16	Go to Step 10 for Next Treatment Group			

*For TI-58, repartition by pressing 6 **[2nd] [Op] 17**.

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	To delete data[7]:			
17	Enter unwanted data	x_{ij}	[2nd] [A']	x_{ij}
18	Initialize Data Base[1]		[2nd] [E']	0.
19	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
20	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D']	Last j
21	Continue entering data for Current Treatment Group			
22	Reenter data for Current Treatment Group that has been overwritten			
23	Calculate $\overline{x}$ for Current Treatment Group		[2nd] [B']	$\overline{x}$ †
24	Display s^2 for Current Treatment Group		[2nd] [C']	s^2 †
25	Reenter raw data for previous Treatment Groups using either Steps 4-9 or 10-16			
26	Enter data for New Treatment Groups			

NOTES: See Data Entry Notes for 1-7.

8. If you are recording your raw data on magnetic cards, each Treatment Group should be recorded on separate sets of cards. To do this, simply reset the Raw Data Pointer here. Also, data deletion procedures are invalidated unless this pointer is reset.

† Printed when PC-100A is used.

TWO-WAY ANALYSIS OF VARIANCE DATA

Solid State Software			TI ©1977	
ANALYSIS OF VARIANCE DATA			ST-06	
Delete x	$\bar{x}$	s^2	Compile	Init 1-Way
x	R	C	Pointer	Init 2-Way

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 06	No Change
2	Initialize Data Base[1]		[E]	0.
3	Enter number of rows[8]	R†	[B]	R
4	Enter number of columns[8]	C†	[C]	C
5	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
6	Enter data for row i (repeat for each j)[2]	x_{ij}†	[A]	j
	If Raw Data Base is filled:			
7a	Record raw data on magnetic card(s) if desired[3]			
7b	Reset Raw Data Pointer[4] and go to Step 6 to enter additional data		[D]	31.
	(Repeat 6-7 for each row)[9]			
8	Calculate $\bar{x}$[10]		[2nd] [B′]	$\bar{x}$†
9	Display s^2		[2nd] [C′]	s^2†
10	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
11	Read raw data card(s)		[CLR]	0
		Card		Bank No.
12	Reset Raw Data Pointer[4]		[D]	31.
13	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D′]	Last j
14	For additional data cards — go to Step 11			
15	Calculate $\bar{x}$[10]		[2nd] [B′]	$\bar{x}$†
16	Display s^2		[2nd] [C′]	s^2†

*For TI-58, repartition by pressing 6 [2nd] [Op] 17.

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	To delete data[7]:			
17	Enter unwanted data[9]	x_{ij}	[2nd] [A′]	x_{ij}
18	Initialize Data Base[1]		[E]	0.
19	Enter number of rows[8]	R†	[B]	R
20	Enter number of columns[8]	C†	[C]	C
21	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
22	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D′]	Last j
23	Reenter raw data that has been overwritten using either Steps 6-9 or 11-16			
24	Continue entering new data			

NOTES: See Data Entry Notes for 1-7.

8. R + C may not exceed 15.
9. Data deletion procedures may be invalidated unless the Raw Data Pointer is reset between rows.
10. This step may be performed only after all raw data is entered.

† Printed when PC-100A is used.

HISTOGRAM DATA

Solid State Software			TI ©1977	
HISTOGRAM DATA				ST-07
Delete x	x_{min}		Compile	Initialize
x	Cells	Width	Pointer	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program*		[2nd] [Pgm] 07	No Change
2	Initialize Data Base[1]		[2nd] [E']	0.
3	Enter number of cells[8]	Cells†	[B]	Cells
4	Enter lower limit	x_{min}†	[2nd] [B']	x_{min}
5	Enter cell width	Width†	[C]	Width
6	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
	Enter data using either I or II			
I	**KEYBOARD ENTRY**			
7	Enter data (repeat for each x_i)[2]	x_i†	[A]	i
	If Raw Data Base is filled:			
8a	Record raw data on magnetic card(s) if desired[3]			
8b	Reset Raw Data Pointer[4] and go to Step 7 to enter additional data			
9	Record intermediate data on magnetic card if desired[3]			
II	**MAGNETIC CARD ENTRY**[5]			
10	Read raw data card(s)	 Card	[CLR]	0. Bank No.
11	Reset Raw Data Pointer[4]		[D]	31.
12	Compile Intermediate Data Base (raw data is printed)[6]		[2nd] [D']	Last i
13	For additional data card(s) – go to Step 10			
	To delete data[7]**:**			
14	Enter unwanted data	x_i	[2nd] [A']	x_i
15	Initialize Data Base[1]		[2nd] [E']	0.
16	Enter number of cells	Cells†	[B]	Cells
17	Enter lower limit	x_{min}†	[2nd] [B']	x_{min}
18	Enter cell width	Width†	[C]	Width
19	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
20	Recompile raw data currently stored in Raw Data Base (raw data is printed)[6]		[2nd] [D']	Last i
21	Reenter raw data that has been overwritten using either Steps 7-9 or 10-13			
22	Continue entering new data			

NOTES: See Data Entry Notes for 1-7.

8. The number of cells may not exceed 12.

† Printed when PC-100A is used.

* For TI-58, repartition by pressing 6 [2nd] [Op] 17.

DATA TRANSFORM PROGRAMS

You may often find the need to replace your data with computed quantities before you can go on to perform your evaluations and tests. Two programs included in this library are specifically designed for this purpose. These programs are actually data entry routines that first transform your data to the form that you desire. They then assimilate a data base by calling an appropriate data entry program to enter the transformed data.

UNIVARIATE DATA TRANSFORMS

This program may be used wherever the *Univariate Data (Ungrouped) Program* is called for. Two prewritten transform routines are included.

- An exponential transform converts any x you enter to exp(x) before compiling the data in the intermediate data base.
- A logarithmic transform converts any x you enter to ln x before compiling the data in the intermediate data base.

A third routine is provided to transform your data into any form that you wish. All you have to do is store your transform in program memory under label [2nd] [A']. Then select the user-defined transform as explained in the user instructions and enter your data. The only restrictions are that you may not use [=], [CLR], or [RST] in your transform routine. Note that when your subroutine is called x is in the display register. When your subroutine ends the display register should contain f(x). Remember to end your routine with [INV] [SBR].

BIVARIATE DATA TRANSFORMS

This program includes three prewritten routines for transforming data pairs before assimilating the data base.

- (x, y) becomes (x, ln y).
- (x, y) becomes (ln x, ln y).
- (x, y) becomes (ln x, y).

A fourth routine is provided to transform your data into any form that you wish. Simply store your transform for x in program memory under label [2nd] [A']. Then store your transform for y in program memory under label [2nd] [B']. Now select the user-defined transform as explained in the user instructions and enter your data. Again, you may not use [=], [CLR], or [RST] in your transform routines. Also, you must end your routines with [INV] [SBR].

This program may be used wherever the *Bivariate Data Program* is called for. See Section V for further applications of this program.

Solid State Software				TI ©1977
UNIVARIATE DATA TRANSFORMS				ST-11
Exp	Ln	User		Initialize
x				

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	For Preprogrammed Transform			
1	Select Program*		[2nd] [Pgm] 11	No Change
2	Initialize[1]		[2nd] [E′]	1.
3	Repartition if desired	n	[2nd] [Op] 17	Steps. Regs
4	Choose Transform: Exponential, Logarithmic		 [2nd] [A′] [2nd] [B′]	 No Change No Change
5	Enter data (repeat for each x_i)[2]	x_i	[A]	i
	For User-Defined Transform			
6	Enter Transform into program memory (do not use [=], [CLR], or [RST])	f(x)	[2nd] [CP] [LRN] [2nd] [Lbl] [2nd] [A′] [INV] [SBR] [LRN]	
7	Select Program		[2nd] [Pgm] 11	No Change
8	Initialize[1]		[2nd] [E′]	1.
9	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
10	Select User-Defined Transform mode		[2nd] [C′]	No Change
11	Enter data (repeat for each x_i)[2]	x_i	[A]	i

NOTES:

1. Initialization uses routine [E] of the Univariate Data (Ungrouped) program.
2. Once the data is transformed, it is entered using routine [A] of the Univarite Data (Ungrouped) program. See the User Instructions of that program for data deletion procedures and limitations of the Raw Data Base. f(x) is printed when the PC-100A is used.
3. This program uses the same data registers as ST-03.

*For TI-58, repartition by pressing 6 [2nd] [Op] 17.

Solid State Software			TI ©1977	
BIVARIATE DATA TRANSFORMS				ST-12
Exp (x, ln y)	Pwr (ln x, ln y)	Ln (ln x, y)	User	Initialize
x	y	→ b; m (Pgm)	x → y′ (Pgm)	y → x′ (Pgm)

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	For Preprogrammed Transforms			
1	Select Program*		[2nd] [Pgm] 12	No Change
2	Initialize[1]		[2nd] [E′]	0.
3	Repartition if desired	n	[2nd] [Op] 17	Steps. Regs
4	Choose Transform: (x, ln y), (ln x, ln y), (ln x, y)		 [2nd] [A′] [2nd] [B′] [2nd] [C′]	 No Change No Change No Change
5a	Enter x_i[2]	x_i	[A]	i
5b	Enter y_i[2]	y_i	[B]	i
	(Repeat Step 5 for each data pair)			
	For User-Defined Transforms			
6	Enter Transforms into program memory (do not use [=], [CLR], or [RST])	 f(x) g(y)	[2nd] [CP] [LRN] [2nd] [Lbl] [2nd] [A′] [INV] [SBR] [2nd] [Lbl] [2nd] [B′] [INV] [SBR] [LRN]	
7	Select Program		[2nd] [Pgm] 12	No Change
8	Initialize[1]		[2nd] [E′]	0.
9	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
10	Choose User-Defined Transform		[2nd] [D′]	No Change
11a	Enter x_i[2]	x_i	[A]	i
11b	Enter y_i[2]	y_i	[B]	i
	(Repeat Step 11 for each data pair)			

NOTES:

1. Initialization uses routine [2nd] [E′] of the Bivariate Data program.
2. Once the data is transformed, f(x) is entered using routine [A] of the Bivariate Data program and g(y) is entered using routine [B]. Data must be entered in pairs. See the Bivariate Data User Instructions for data deletion procedures and limitations of the Raw Data Base. f(x) and g(y) are printed when the PC-100A is used.
3. This program uses the same data registers as ST-04.

*For TI-58, repartition by pressing 6 [2nd] [Op] 17.

IV. DATA EVALUATION

The first two programs in this section are designed to help you interpret your data. For example, the *Means and Moments Program* can be used to determine the shape of your sample distribution. You may then use this information in evaluating your data.

Statistical data is evaluated in order to make a decision. This decision usually involves choosing between two or more alternatives called hypotheses. For example, what if you want to determine whether or not a coin is balanced? The hypotheses that you wish to test are

H_0: the coin in balanced ($p = 0.5$)

against

H_1: the coin is unbalanced ($p \neq 0.5$).

H_0 is the null hypothesis. If you reject H_0 you would then accept the alternative hypothesis H_1. In the above, p is the probability of heads on any toss of the coin.

To determine whether you should accept or reject H_0 your first step is to obtain a sample. You should then use your sample data to derive a test statistic. Since the binomial distribution (see Section VI) is to be used as your probability model, your test statistic for this experiment is k where k is the number of heads occurring in n tosses of the coin.

Confidence limits for a test statistic are the usual criteria established for accepting or rejecting a hypothesis. Let's suppose that if you accept H_0 you want to be 95% certain that you are right. To achieve this degree of confidence for the type of test described above you would normally construct an acceptance region for your test statistic (k_1, k_2). These values are determined such that $F(k_2) - F(k_1) = 0.95$ where the midpoint of the interval is the expected value of k given n and p. Then, if your test statistic k is such that $k_1 \leqslant k \leqslant k_2$, you would accept H_0.

This may seem like a complicated process and it often is. However, the *Theoretical Distributions* programs (Section VI) make it easy to determine whether to accept or reject your hypothesis.

Most of the programs in this section follow the same pattern of development discussed above. First, a test is described. If this test meets your needs you may use the program to compute a statistic. You can then plug this data into an appropriate *Theoretical Distributions* program (Section VI) to determine whether to accept or reject your hypothesis.

You should realize that the test described above is for a special case. Here, we tested

H_0: $p = 0.5$ against H_1: $p \neq 0.5$.

If we wanted to test hypotheses such as

H_0: $p \leqslant 0.5$ against H_1: $p > 0.5$,

a different test with another type of acceptance region would be required.

The examples in this section offer the acceptance regions of the tests without proof. These examples should give you further insight into confidence levels and acceptance regions. However, a complete discussion is left until Section VI.

MEANS AND MOMENTS PROGRAM

The means and moments of a sample can tell us a lot about the shape of its distribution. For a given set of input data $\{x_1, x_2, \ldots, x_n\}$ with associated frequencies $\{f_1, f_2, \ldots, f_n\}$, you may use this program to calculate the following means, moments, skewness and kurtosis of a sample distribution. To use this program you must enter your data using one of the *Univariate Data* programs found in Section III. If $f_i = 1$ for all i, the calculations are for ungrouped data, otherwise they are for grouped data.

In the following discussion

$$N = \sum_{i=1}^{n} f_i.$$

The arithmetic mean,

$$\overline{x} = 1/N \sum_{i=1}^{n} f_i x_i,$$

is simply the average value of the sample data. For ungrouped data this value is known as the simple arithmetic mean. For grouped data it is called the weighted arithmetic mean.

The geometric mean,

$$g = \prod_{i=1}^{n} (x_i^{f_i})^{1/N},$$

is another measure of central tendency. It is especially useful in averaging ratios, percentages, and rates of change.

The harmonic mean,

$$h = N \div \sum_{i=1}^{n} f_i/x_i ,$$

is primarily used when dealing with ratio data having physical dimensions such as miles per hour.

The second moment,

$$m_2 = 1/N \sum_{i=1}^{n} f_i(x_i - \overline{x})^2$$

($\overline{x}$ is the first moment), is more often called variance. It is the mean of squared deviations of the sample data from $\overline{x}$. This moment is used to measure the variability or dispersion of a population.

The third moment,

$$m_3 = 1/N \sum_{i=1}^{n} f_i(x_i - \overline{x})^3 ,$$

is used to determine whether a distribution is symmetric or skewed about $\overline{x}$. Negative and positive deviations cancel each other out since all deviations in the equation for m_3 are cubed. Therefore, m_3 is equal to zero when the distribution is symmetric about its mean. A distribution is said to be right or positively skewed when m_3 is positive and left or negatively skewed when m_3 is negative. You can also use this program to calculate a relative measure of skewness eliminating any influence by the units your variables are measured in.

$$\text{Skewness} = m_3/(m_2)^{3/2}$$

You may consider your distribution to be symmetric when $-0.5 < \text{Skewness} < 0.5$. The distribution is highly skewed when this value exceeds ± 1.

Figure 4.1

The fourth moment,

$$m_4 = 1/N \sum_{i=1}^{n} f_i(x_i - \overline{x})^4,$$

is used to interpret the flatness or peakedness of a distribution curve. You may use another relative measure known as the kurtosis of distribution to obtain this information.

$$\text{Kurtosis} = m_4/(m_2)^2$$

The kurtosis of a normal distribution is around three (see Section VI). For values less than three the curve flattens out and for values greater than three it becomes more peaked.

Solid State Software			TI ©1977	
MEANS AND MOMENTS				ST-08
→ m_2; m_3; m_4				
→ $\bar{x}$	→ g	→ h	→ Kurtosis	→ Skewness

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 03	No Change
2	Enter Univariate Data according to User Instructions found in Section III.			
3	Select Program		[2nd] [Pgm] 08	No Change
4	Calculate arithmetic mean		[A]	$\bar{x}$†
5	Calculate geometric mean[1]		[B]	g†
6	Calculate harmonic mean		[C]	h†
7a	Calculate second moment		[2nd] [A′]	m_2†
7b	Calculate third moment		[R/S]	m_3†
7c	Calculate fourth moment		[R/S]	m_4†
8	Calculate Kurtosis[2]		[D]	Kurtosis†
9	Calculate Skewness[2]		[E]	Skewness†

NOTES:
1. The geometric mean is not valid for negative values of x.
2. Step 7 must be performed before calculating Kurtosis or Skewness.

† Printed when PC-100A is used.

Register Contents

R_{00}	R_{05}	R_{10}	R_{15}	R_{20} m_3	R_{25}
R_{01}	R_{06}	R_{11}	R_{16}	R_{21} m_2	R_{26}
R_{02}	R_{07}	R_{12}	R_{17}	R_{22}	R_{27}
R_{03}	R_{08}	R_{13}	R_{18}	R_{23}	R_{28}
R_{04}	R_{09}	R_{14}	R_{19} m_4	R_{24}	R_{29}

Example:

The following lists show the respective heights in inches of a random sample of 10 men over age 45 and their adult sons.

Fathers: 67.2, 65.0, 68.3, 69.9, 66.3, 69.7, 69.5, 72.9, 70.2, 74.1.

Sons: 68.4, 65.3, 66.5, 69.0, 73.6, 75.9, 69.7, 69.8, 71.0, 70.8, 67.7, 74.4, 69.9, 71.5, 71.1.

Compare the distribution of the heights of the fathers against that of the sons.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 03		Select Univariate Data Program
	[E]	1.	Initialize Ungrouped Data Entry
67.2†	[A]	1.	x_1
65.0†	[A]	2.	x_2
68.3†	[A]	3.	x_3
69.9†	[A]	4.	x_4
66.3†	[A]	5.	x_5 Heights
69.7†	[A]	6.	x_6 of
69.5†	[A]	7.	x_7 Fathers
72.9†	[A]	8.	x_8
70.2†	[A]	9.	x_9
74.1†	[A]	10.	x_{10}
	[2nd] [Pgm] 08	10.	Select Means and Moments Program
	[A]	69.31†	$\overline{x}$
	[B]	69.25951514†	g
	[C]	69.20924469†	h
	[2nd] [A']	7.0269†	m_2
	[R/S]	3.938892†	m_3
	[R/S]	114.13275†	m_4
	[D]	2.31144059†	Kurtosis
	[E]	.2114600903†	Skewness
	[2nd] [Pgm] 03	.2114600903	Select Univariate Data Program
	[E]	1.	Initialize Ungrouped Data Entry
68.4†	[A]	1.	x_1
65.3†	[A]	2.	x_2
66.5†	[A]	3.	x_3
69.0†	[A]	4.	x_4
73.6†	[A]	5.	x_5
75.9†	[A]	6.	x_6
69.7†	[A]	7.	x_7 Heights
69.8†	[A]	8.	x_8 of
71.0†	[A]	9.	x_9 Sons
70.8†	[A]	10.	x_{10}
67.7†	[A]	11.	x_{11}
74.4†	[A]	12.	x_{12}
69.9†	[A]	13.	x_{13}
71.5†	[A]	14.	x_{14}
71.1†	[A]	15.	x_{15}

† Printed when PC-100A is used.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 08	15.	Select Means and Moments Program
	[A]	70.30666667†	$\overline{x}$
	[B]	70.25279208†	g
	[C]	70.19915809†	h
	[2nd] [A']	7.609955561†	m_2
	[R/S]	4.932666†	m_3
	[R/S]	152.04783†	m_4
	[D]	2.625523954†	Kurtosis
	[E]	.2349678963†	Skewness

† Printed when PC-100A is used.

Summary:

As you can see, these samples come from distributions which are both slightly skewed to about the same degree. The peakedness of the distributions are also nearly equal. Based on this information it is safe to assume that these samples come from distributions having the same shape. The only difference is that the average height of the sons is about 1 inch more than that of the fathers. That is, the distribution of the heights of the sons may be shifted to the right when compared with that of the fathers. This hypothesis is tested in the *Rank-Sum Test Program.*

Example:

When large amounts of data are involved you may find it easier to group your data before entering it. The sample heights of the sons data from the last example is grouped in Table 4.1.

Table 4.1

Range	$64 \leqslant x < 66$	$66 \leqslant x < 68$	$68 \leqslant x < 70$	$70 \leqslant x < 72$	$72 \leqslant x < 74$	$74 \leqslant x < 76$
Frequency	1	2	5	4	1	2
Assigned Value	65	67	69	71	73	75

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 03		Select Univariate Data Program
	[2nd] [E']	1.	Initialize Grouped Data Entry
1†	[B]	1.	f_1
65†	[A]	1.	x_1
2†	[B]	2.	f_2
67†	[A]	2.	x_2
5†	[B]	5.	f_3
69†	[A]	3.	x_3
4†	[B]	4.	f_4
71†	[A]	4.	x_4
1†	[B]	1.	f_5
73†	[A]	5.	x_5
2†	[B]	2.	f_6
75†	[A]	6.	x_6
	[2nd] [Pgm] 08	6.	Select Means and Moments Program
	[A]	70.06666667†	$\overline{x}$
	[B]	70.01414879†	g
	[C]	69.96188555†	h
	[2nd] [A']	7.395555561†	m_2
	[R/S]	4.987258†	m_3
	[R/S]	140.27288†	m_4
	[D]	2.564673624†	Kurtosis
	[E]	.2479737041†	Skewness

† Printed when PC-100A is used.

Summary:

Comparing these outputs to those of the last example it is easy to see that little loss of accuracy occurs when data is grouped; but considerable data entry time is saved. The next program illustrates how you may use your calculator to group your data instead of doing it by hand.

HISTOGRAM CONSTRUCTION PROGRAM

A histogram is constructed to help interpret a set of data points. Each data point is assigned to a class interval or cell of the histogram depending upon its magnitude.

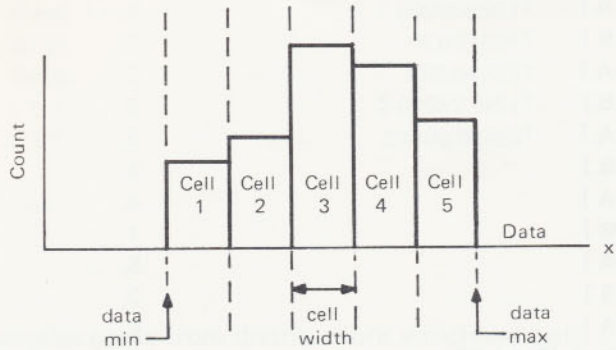


Figure 4.2

To construct a histogram with up to 12 cells, use the *Histogram Data* program found in Section III to enter and assimilate your data points. Remember to specify the number of cells you want and the width of each cell. You should also enter a lower limit for your data if it is different from zero. The program calculates the upper limit of the histogram and discards any data falling outside of the calculated range. This situation is indicated by flashing nines in the display.

Once you have assimilated the histogram data you may use this program to calculate the mean and standard deviation of your data points. You can also determine the count, or number of data points in each cell, and the upper limit of the data placed in a given cell. Note that if a piece of data falls on the upper limit of a cell it is assigned to the next cell. A piece data falling on the point designated data max in Figure 4.2 is discarded.

Solid State Software				TI ©1977
HISTOGRAM CONSTRUCTION				ST-09
→ s				Initialize
→ $\bar{x}$	→ Count	→ x_{max}		

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 07	No Change
2	Enter Histogram Data according to User Instructions found in Section III.			
3	Select Program		[2nd] [Pgm] 09	No Change
4	Initialize		[2nd] [E']	0.
5	Calculate sample mean		[A]	$\bar{x}$†
6	Calculate sample standard deviation[1]		[2nd] [A']	s†
7a	Display count of current cell[2]		[B]	Count†
7b	Calculate upper limit of current cell[3]		[C]	x_{max}†
8	Display accumulation of cell counts		[RCL] 21	Σ Count

NOTES:

1. The n-1 method is used here. You may calculate s^2 using the n method by pressing [2nd] [Op] 11 [x ≷ t].
2. The cell number is incremented by 1 each time [B] is pressed. Divide the count by n to determine the frequency.
3. 7b must be performed immediately following 7a for the cell in question.

† Printed when PC-100A is used.

Register Contents

R_{00}	Cell No.	R_{05}	Σx^2	R_{10}	Cell 5 Count	R_{15}	Cell 10 Count	R_{20}	Used	R_{25}	
R_{01}	x_{min}	R_{06}	Cell 1 Count	R_{11}	Cell 6 Count	R_{16}	Cell 11 Count	R_{21}	Σ Count	R_{26}	
R_{02}	Width	R_{07}	Cell 2 Count	R_{12}	Cell 7 Count	R_{17}	Cell 12 Count	R_{22}		R_{27}	
R_{03}	n	R_{08}	Cell 3 Count	R_{13}	Cell 8 Count	R_{18}		R_{23}		R_{28}	
R_{04}	Σx	R_{09}	Cell 4 Count	R_{14}	Cell 9 Count	R_{19}	Cells	R_{24}		R_{29}	

Example:

Construct a histogram from the second set of data used in the last example. Use 6 cells with a width of 2 and a minimum value of 64. The data is repeated below.

Heights of Sons: 68.4, 65.3, 66.5, 69.0, 73.6, 75.9, 69.7, 69.8, 71.0, 70.8, 67.7, 74.4, 69.9, 71.5, 71.1.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 07		Select Histogram Data Program
	[2nd] [E']	0.	Initialize
6†	[B]	6.	Cells
64†	[2nd] [B']	64.	x_{min}
2†	[C]	2.	Width
68.4†	[A]	1.	x_1
65.3†	[A]	2.	x_2
66.5†	[A]	3.	x_3
69.0†	[A]	4.	x_4
73.6†	[A]	5.	x_5
75.9†	[A]	6.	x_6
69.7†	[A]	7.	x_7 Heights
69.8†	[A]	8.	x_8 of
71.0†	[A]	9.	x_9 Sons
70.8†	[A]	10.	x_{10}
67.7†	[A]	11.	x_{11}
74.4†	[A]	12.	x_{12}
69.9†	[A]	13.	x_{13}
71.5†	[A]	14.	x_{14}
71.1†	[A]	15.	x_{15}
	[2nd] [Pgm] 09	15.	Select Histogram Construction Program
	[2nd] [E']	0.	Initialize
	[A]	70.30666667†	$\overline{x}$
	[2nd] [A']	2.855437586†	s
	[B]	1.†	Cell 1 Count
	[C]	66.†	Cell 1 Max
	[B]	2.†	Cell 2 Count
	[C]	68.†	Cell 2 Max
	[B]	5.†	Cell 3 Count
	[C]	70.†	Cell 3 Max
	[B]	4.†	Cell 4 Count
	[C]	72.†	Cell 4 Max
	[B]	1.†	Cell 5 Count
	[C]	74.†	Cell 5 Max
	[B]	2.†	Cell 6 Count
	[C]	76.†	Cell 6 Max
	[RCL] 21	15.	Total Count

† Printed when PC-100 is used.

Summary:

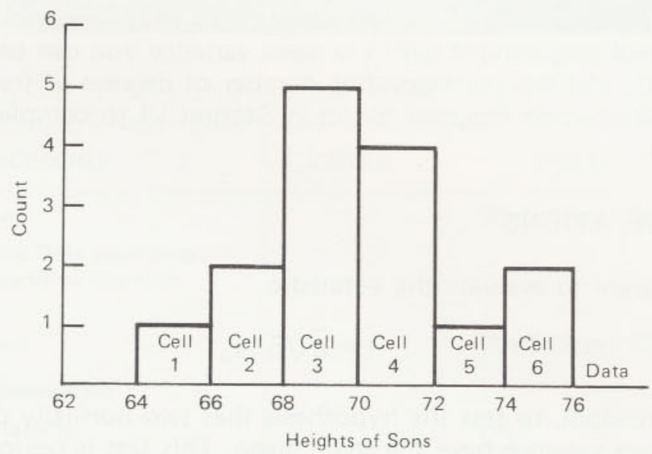


Figure 4.3

FREQUENCY PLOTTING

The probability of a data point falling into a given cell is known as the frequency of that cell. You may calculate the frequency of any cell by dividing the count of that cell by the total number of data points placed in the histogram. However, if you have a PC-100A, you may use the following instruction sequence to plot the frequency function.

Location and Key Code	Key Sequence	Location and Key Code	Key Sequence
000 76	[2nd] [Lbl]	016 55	[÷]
001 11	[A]	017 43	[RCL]
002 05	[5]	018 21	[2] [1]
003 42	[STO]	019 65	[×]
004 26	[2] [6]	020 01	[1]
005 43	[RCL]	021 09	[9]
006 19	[1] [9]	022 95	[=]
007 42	[STO]	023 69	[2nd] [Op]
008 00	[0]	024 07	[0] [7]
009 76	[2nd] [Lbl]	025 97	[2nd] [Dsz]
010 12	[B]	026 00	[0]
011 01	[1]	027 12	[B]
012 44	[SUM]	028 91	[R/S]
013 26	[2] [6]		
014 73	[RCL] [2nd] [Ind]		
015 26	[2] [6]		

To use this routine press [2nd] [CP] [LRN] followed by the keystrokes listed in the key sequence column above. Then press [LRN] again and try it out by running the example on the last page and then pressing [RST] [A]. This routine plots the frequency function lengthwise on the tape with the left position equal to zero and the right equal to one.

t-STATISTIC EVALUATION PROGRAM
(Comparison of Population Means)

The distribution of the t-statistic depends only on the population mean, not the variance. As a result, this statistic is often used to compare the means of different populations. If your samples are from normal populations with the same variance you can use this program to determine the t-statistic and the corresponding number of degrees of freedom. You may then use the *Student's t Distribution Program* found in Section VI to complete your comparison of the population means.

FOR PAIRED OBSERVATIONS

You can use this program to evaluate the t-statistic

$$t = \bar{\Delta}\sqrt{n}\,/s_\Delta$$

with n-1 degrees of freedom, to test the hypothesis that two normally distributed populations with the same unknown variance have the same mean. This test is performed using n paired observations from the two samples. In the above equation:

$\bar{\Delta}$ = the mean of the differences between the paired values.

n = the sample size.

s_Δ = the standard deviation of the differences between the paired values using the n-1 method.

Enter your data using the *Bivariate Data Program* found in Section III.

TWO SAMPLE TEST

You may also use this program to evaluate the t-statistic

$$t = \frac{\bar{x} - \bar{y} - \Delta}{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)^{1/2} \left(\frac{\Sigma x_i^2 - n_x\bar{x}^2 + \Sigma y_j^2 - n_y\bar{y}^2}{n_x + n_y - 2}\right)^{1/2}}$$

with $n_x + n_y - 2$ degrees of freedom, to test the hypothesis that the difference between the means of two normally distributed populations having the same unknown variance is Δ. Again, you must enter your data using the *Bivariate Data Program.* Since the sample sizes, don't have to be the same, you don't have to enter the data in pairs. (See example.) However, data deletion procedures are invalidated when the data isn't entered in pairs.

Solid State Software				TI ©1977
t-STATISTIC EVALUATION				ST-13
$\Delta \rightarrow t$				
$\rightarrow t$	$\rightarrow \nu$	$\rightarrow \bar{\Delta}$	$\rightarrow s_\Delta$	

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 04	No Change
2	Enter Bivariate Data according to User Instructions found in Section III.			
3	Select Program		[2nd] [Pgm] 13	No Change
	For Paired Observation			
4	Compute t-Statistic		[A]	t†
5	Display degrees of freedom		[B]	ν†
6	Display mean of difference between observations		[C]	$\bar{\Delta}$†
7	Display standard deviation of difference between observations		[D]	s_Δ†
	For Two Sample Test			
8	Enter hypothesized difference and compute t-Statistic	Δ	[2nd] [A′]	t†
9	Display degrees of freedom		[B]	ν†

NOTE: † Printed when PC-100A is used.

Register Contents

R_{00}	R_{05} Σx^2	R_{10}	R_{15} n_x	R_{20}	R_{25} ν
R_{01} Σy	R_{06}	R_{11}	R_{16}	R_{21}	R_{26} $\bar{\Delta}$
R_{02} Σy^2	R_{07}	R_{12} $\Sigma(x-y)$	R_{17}	R_{22}	R_{27} s_Δ
R_{03} n_y	R_{08}	R_{13} $\Sigma(x-y)^2$	R_{18}	R_{23} Used	R_{28}
R_{04} Σx	R_{09}	R_{14}	R_{19}	R_{24} Δ	R_{29}

Example:

In an experiment to compare two different diets for pigs, a farmer randomly selects a pair of pigs from each of ten litters. He then chooses one pig from each pair and places them on diet A for a fixed period of time. The remaining pigs are placed on diet B during the same period. At the end of the experiment, the farmer weighs each pig to see how much it had gained. These results are tabulated below.

Table 4.2

	Litter									
	1	2	3	4	5	6	7	8	9	10
Diet A	21.5	18.0	14.7	19.3	21.7	22.9	22.3	19.1	13.3	19.8
Diet B	14.7	16.1	15.2	14.6	17.5	15.6	20.8	20.3	12.0	20.9

DATA EVALUATION

Evaluate the t-statistic for paired observations to test the hypotheses

H_0: the diets cause the same average weight gain ($\mu_A = \mu_B$)

against

H_1: the diets cause different average weight gains ($\mu_A \neq \mu_B$)

at the 95% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 04		Select Bivariate Data Program
	[2nd] [E']	0.	Initalize
21.5†	[A]	1.	x_1
14.7†	[B]	1.	y_1
18.0†	[A]	2.	x_2
16.1†	[B]	2.	y_2
14.7†	[A]	3.	x_3
15.2†	[B]	3.	y_3
19.3†	[A]	4.	x_4
14.6†	[B]	4.	y_4
21.7†	[A]	5.	x_5
17.5†	[B]	5.	y_5
22.9†	[A]	6.	x_6
15.6†	[B]	6.	y_6
22.3†	[A]	7.	x_7
20.8†	[B]	7.	y_7
19.1†	[A]	8.	x_8
20.3†	[B]	8.	y_8
13.3†	[A]	9.	x_9
12.0†	[B]	9.	y_9
19.8†	[A]	10.	x_{10}
20.9†	[B]	10.	y_{10}
	[2nd] [Pgm] 13	10.	Select t-Statistic Evaluation Program
	[A]	2.522310351†	t
	[B]	9.†	ν
	[C]	2.49†	$\overline{\Delta}$
	[D]	3.121769441†	s_Δ

Summary:

When using the t distribution to test hypotheses such as these you should accept H_0 whenever your test statistic falls within a confidence interval about the mean of the distribution (t = 0). The 95% confidence interval or acceptance region for the t-statistic with 9 degrees of freedom is (−2.262, 2.262)*. That is, the probability of a deviation of up to 2.262 is 0.95. Using this confidence interval, you will reject H_0 when it is true in only 5% of your experiments. This is also known as testing at the 5% significance level, the hypotheses that the means are equal against the alternative that they are not. Since the t-statistic calculated in this exercise falls outside of the acceptance region you should reject H_0 in favor of accepting H_1.

†Printed when PC-100A is used.

*You may verify this using the Student's t Distribution Program.

In the above $\overline{\Delta}$ was found to be 2.49. So just for practice, use the two-sample test to test the hypotheses

$$H_0: \ \mu_A - \mu_B = 2.5$$

against

$$H_1: \ \mu_A - \mu_B \neq 2.5.$$

The 95% confidence interval for the t-statistic with 18 degrees of freedom is (–2.101, 2.101). There is no need to reenter your data if you haven't disturbed your calculator since running the last example.

Example:

Two groups of patients at a major hospital are selected for an experiment to compare two drugs used for the relief of pain. One group is given drug x and the other drug y. The resulting number of hours of relief for each patient is given below.

Drug x: 2, 6, 4, 13, 5, 8, 4, 6.
Drug y: 6, 4, 4, 1, 8, 2, 12, 1, 5, 2.

Evaluate the two-sample statistic to test the hypotheses

$$H_0: \ \mu_x - \mu_y = 2 \text{ hours}$$

against

$$H_1: \ \mu_x - \mu_y \neq 2 \text{ hours}$$

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 04		Select Bivariate Data Program
	[2nd] [E']	0.	Initialize
2†	[A]	1.	x_1
6†	[A]	2.	x_2
4†	[A]	3.	x_3
13†	[A]	4.	x_4
5†	[A]	5.	x_5
8†	[A]	6.	x_6
4†	[A]	7.	x_7
6†	[A]	8.	x_8
6†	[B]	1.	y_1
4†	[B]	2.	y_2
4†	[B]	3.	y_3
1†	[B]	4.	y_4
8†	[B]	5.	y_5
2†	[B]	6.	y_6
12†	[B]	7.	y_7
1†	[B]	8.	y_8
5†	[B]	9.	y_9
2†	[B]	10.	y_{10}
	[2nd] [Pgm] 13	10.	Select t-Statistic Evaluation Program
2	[2nd] [A']	–.3087445631†	$\Delta \rightarrow t$
	[B]	16.†	ν

† Printed when PC-100A is used.

Summary:

The 90% confidence interval for the t-statistic with 16 degrees of freedom is (–1.75, 1.75). You may accept the null hypotheses since the value computed for t falls in this range.

Note that in this experiment our null hypothesis is that $\mu_x - \mu_y = 2$. That is, we are assuming that μ_x is larger than μ_y. But what if μ_y is the larger value. In this case you would simply test the hypothesis that $\mu_x - \mu_y = -2$. Try this yourself by entering the sample for drug y using the [A] key and drug x using [B]. You should get the same results when testing for $\Delta = -2$.

CONTINGENCY TABLE ANALYSIS PROGRAM

(Two-Way Classification)

A contingency table is a table in which each observation is classified in two or more ways. The data is Table 4.3 represents number of drivers in various age groups who were involved in zero, one, two, and more than two automobile accidents over a period of 3 years.

Table 4.3

		Age of Driver					
		21 – 30	31 – 40	41 – 50	51 – 60	61 – 70	$N_i.$
Number of Accidents	0	748	821	786	720	672	3747
	1	74	60	51	66	55	306
	2	31	25	22	16	15	109
	>2	9	10	6	5	7	37
	$N._j$	862	916	865	807	749	**4199**

A table such as this, where only two classifications are considered, is known as a two-way contingency table. In analyzing such a table you are often interested in testing the hypothesis that the two classifications are independent of one another. In this example, that means you would like to know if the age of the driver and the number of accidents he is involved in are related. (The hypothesis is that they are not related.)

You may use this program to apply the χ^2 test of independence to the row and column classifications of a contingency table with R rows and C columns where RC ⩽ 25. Calculations are based upon the χ^2 statistic

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} (N_{ij} - E_{ij})^2 / E_{ij}$$

with $(r-1)(c-1)$ degrees of freedom. In the above, N_{ij} is the count, or number observations occurring in cell ij (row i, column j) of the table. E_{ij} is the maximum likelihood estimator of the number of observations that should occur in cell ij when the classifications are independent (i.e., the row and column factors are not related). That is, the probability of an event falling in both row i and column j is the probability of the event falling in row i multiplied by the probability that it falls in column j. This may be expressed as

$$E_{ij} = (N_i.\ N._j)/n$$

where: $N_i.$ = the total number of counts in the i^{th} row.

$N._j$ = the total number of counts in the j^{th} column.

n = the total number of cells, R × C.

The outputs of this program are the χ^2-statistic with $(R-1)(C-1)$ degrees of freedom and the cumulative distribution function of the χ^2-statistic, $P(\chi^2)$. You should accept your null hypothesis whenever $P(\chi^2)$ is less than or equal to the confidence level you are testing at.

Solid State Software			TI ©1977	
CONTINGENCY TABLE ANALYSIS				ST-14
	→ $P(\chi^2)$			Initialize
x	→ χ^2	→ ν	R	C

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 14	No Change
2	Establish correct partitioning	6	[2nd] [Op] 17	Steps. 59
3	Enter number of rows[1]	R†	[D]	R
4	Enter number of columns[1]	C†	[E]	C
5	Initialize data entry routine[2]		[2nd] [E′]	1.
6	Enter data by rows (i.e., $x_{11}, x_{12}, \ldots x_{1C}, x_{21}, \ldots x_{RC}$)[3]	x_{ij}†	[A]	Next j
7	Calculate χ^2-statistic[5]		[B]	χ^2
8	Calculate degrees of freedom if desired		[C]	ν
9	Calculate cumulative distribution function		[2nd] [B′]	$P(\chi^2)$

NOTES:
1. R × C can be no greater than 25.
2. This program uses its own data entry routine.
3. Do not enter negative values. If an error is made, begin again.
4. Perform Steps 1-7 first.
5. Execution time increases with ν.

† Printed when PC-100A is used.

Register Contents

R_{01}+	$N_{.1}, \ldots, N_{.C}, N_{1.}, \ldots, N_{R.}$
R_{26}	χ^2
R_{27-29}	Pointers
R_{30}+	$N_{11}, \ldots, N_{ij}$
R_{57}	n
R_{58}	R
R_{59}	C

Example:

200 voters in a local bond election are randomly selected and asked their opinion on the issue. The voters are then classified according to their answers to this question and whether or not they are property owners as illustrated below.

Table 4.4

	For	Against	Undecided	$N_{i\cdot}$
Property Owner	45	39	21	105
Non-Property Owner	47	26	22	95
$N_{\cdot j}$	92	65	43	**200**

Test the hypothesis that a voters opinion of the bond issue is independent of whether or not he is a property owner at the 90% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 14		Select Contingency Table Program
6	**[2nd]]Op] 17**	Steps . 59	Repartition
2†	[D]	2.	Rows
3†	[E]	3.	Columns
	[2nd] [E′]	1.	Initialize Data Entry
45†	[A]	2.	x_{11}
39†	[A]	3.	x_{12}
21†	[A]	1.	x_{13}
47†	[A]	2.	x_{21}
26†	[A]	3.	x_{22}
22†	[A]	1.	x_{23}
	[C]	2.	ν
	[B]	2.172164486	χ^2
	[2nd] [B′]	.6624637081	$P(\chi^2)$

† Printed when PC-100A is used.

Summary:

The range of the χ^2-statistic is from zero to infinity. In order to evaluate this statistic you would normally have to determine an acceptance region for the computed statistic $(0, \chi_0{}^2)$. $\chi_0{}^2$ depends upon the number of degrees of freedom of your test statistic and the confidence level at which you are testing. At the 90% confidence level, this interval would be (0, 4.61)* for a χ^2-statistic with 2 degrees of freedom. Since the value calculated above for χ^2 falls within this interval, you should accept the original hypothesis. That is, there is no evidence that a voter's opinion on the issue is influenced by whether or not he is a property owner.

Note, however; that this program also computes $P(\chi^2)$. This value ranges over the interval (0, 1). When testing at the 90% confidence level $P(\chi_0{}^2) = 0.90$. This indicates that you should accept H_0 whenever $P(\chi^2) \leqslant 0.90$. Since $P(\chi^2)$ for this example meets this requirement, accept H_0.

Perform this test using the data found in Table 4.3.

*You may verify this using the Chi-Square Distribution Program.

ANALYSIS OF VARIANCE PROGRAMS

When trying to compare the means of several distributions the Student's t-test is no longer applicable. Analysis of Variance is a statistical technique used to test the hypothesis that a number of populations all have the same mean. The test is made by using the sample means to estimate the variance of the population. This estimate is then compared to an estimate of the population variance made from differences between individual elements of the samples. The F distribution is used to perform the actual test.

Two assumptions made when using this technique are:

- The populations are normally distributed.
- The variances of the populations are approximately equal.

ONE-WAY AOV *(Many Distribution Comparisons)*

Sample populations used in one-way analysis of variance are often called treatment groups. Each treatment group i ($i = 1, 2, \ldots, K$) consists of n_i observations x_{ij} ($j = 1, 2, \ldots, n_i$). The different groups need not have the same number of observations.

In the following, let

$$N = \sum_{i=1}^{K} n_i.$$

This program uses the sum of squares among groups to estimate the variance from the sample means. The actual equation is

$$\sigma^2 \doteq \sum_{i=1}^{K} (\overline{x}_i - \overline{x})^2/(K-1) = TSS/(K-1).$$

TSS is known as the treatment sum of squares and for this equation,

$$\overline{x} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}/N.$$

The estimate of the variance made from the individual elements of the samples is accomplished using the sum of squares within groups.

$$\sigma^2 \doteq \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2/(N-K) = ESS/(N-K).$$

ESS is known as the error sum of squares.

The F-statistic

$$F = \frac{TSS/(K-1)}{ESS/(N-K)}$$

with K-1 degrees of freedom in the numerator and N-K degrees of freedom in the denominator is calculated by this program. You can use this data and the F distribution program found in Section VI to test the hypothesis that the means of your populations are equal. Additional outputs are the treatment sum of squares (TSS) and the total sum of squares (SS). Note that though TSS is calculated directly, ESS is found by determining SS and then evaluating the expression ESS = SS − TSS.

Remember to enter your data using the *One-Way AOV Data Program* found in Section III.

Solid State Software			TI ©1977	
1-WAY ANALYSIS OF VARIANCE				ST-15
	→ ν_2			
→ F	→ ν_1	→ ESS	→ TSS	→ SS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 06	No Change
2	Enter One-Way AOV data according to User Instructions found in Section III			
3	Select Program		[2nd] [Pgm] 15	No Change
4	Calculate F-statistic[1]		[A]	F†
5	Display degrees of freedom in numerator		[B]	ν_1†
6	Display degrees of freedom in denominator		[2nd] [B′]	ν_2†
7	Display error sum of squares		[C]	ESS†
8	Display treatment sum of squares		[D]	TSS†
9	Display total sum of squares		[E]	SS†

NOTES: 1. Step 4 must be performed before Steps 5-9.
† Printed when PC-100A is used.

Register Contents

R_{00}	R_{05}	R_{10} $\Sigma(\Sigma x)^2/n$	R_{15} TSS/ν_1	R_{20}	R_{25}
R_{01} $\Sigma\Sigma x$	R_{06}	R_{11} Used	R_{16} ν_2	R_{21}	R_{26}
R_{02} $\Sigma\Sigma x^2$	R_{07}	R_{12} SS	R_{17} ESS/ν_2	R_{22}	R_{27}
R_{03}	R_{08} TSS	R_{13} ESS	R_{18}	R_{23}	R_{28} i Count
R_{04}	R_{09} Σn	R_{14} ν_1	R_{19}	R_{24}	R_{29}

Example:

Let's extend the second t-statistic example to compare the effects of three drugs used for the relief of pain. Assume that three groups of patients are chosen for the experiment and given drugs x, y, and z respectively. The resulting number of hours of relief for each patient is given below.

Drug x: 2, 6, 4, 13, 5, 8, 4, 6;
Drug y: 6, 4, 4, 1, 8, 2, 12, 1, 5, 2;
Drug z: 2, 1, 3, 3, 1, 7, 1, 4, 2.

Test the hypotheses

H_0: All of the drugs relieve pain for the same amount of time ($\mu_x = \mu_y = \mu_z$)

against

H_1: Not all of the drugs relieve pain for the same amount of time.

Since the t-test can only be used for comparing two sample populations we'll have to use one-way analysis of variance to perform this test.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 06		Select AOV Data Program
	[2nd] [E′]	0.	Initialize 1-Way AOV
2†	[A]	1.	x_1
6†	[A]	2.	x_2
4†	[A]	3.	x_3
13†	[A]	4.	x_4
5†	[A]	5.	x_5
8†	[A]	6.	x_6
4†	[A]	7.	x_7
6†	[A]	8.	x_8
	[2nd] [B′]	6.†	$\overline{x}$
	[2nd] [C′]	9.75†	s_x^2
6†	[A]	1.	y_1
4†	[A]	2.	y_2
4†	[A]	3.	y_3
1†	[A]	4.	y_4
8†	[A]	5.	y_5
2†	[A]	6.	y_6
12†	[A]	7.	y_7
1†	[A]	8.	y_8
5†	[A]	9.	y_9
2†	[A]	10.	y_{10}

ENTER	PRESS	PRINT	DISPLAY
	[2nd] [B′]	4.5†	$\overline{y}$
	[2nd] [C′]	10.85†	s_y^2
2†	[A]	1.	z_1
1†	[A]	2.	z_2
3†	[A]	3.	z_3
3†	[A]	4.	z_4
1†	[A]	5.	z_5
7†	[A]	6.	z_6
1†	[A]	7.	z_7
4†	[A]	8.	z_8
2†	[A]	9.	z_9
	[2nd] [B′]	2.666666667†	$\overline{z}$
	[2nd] [C′]	3.333333333†	s_z^2
	[2nd] [Pgm] 15	3.333333333	Select 1-Way AOV Program
	[A]	2.632794457†	F
	[B]	2.†	ν_1
	[2nd] [B′]	24.†	ν_2

† Printed when PC-100A is used.

Summary:

As described above, the F-statistic is actually the ratio of two estimates of a population's variance. Since variances are always positive, this ratio can never be less than zero. It should also be evident that F will vary around 1 when H_0 is true. Consequently, you should reject H_0 only if F is significantly greater than 1. The acceptance region for this test is $(0, F_0)$ where the probability that F is less than F_0 is equal to the degree of confidence that you desire. F_0 is also controlled by the degrees of freedom of the estimates. When testing at the 90% confidence level, the acceptance region for the F-statistic with 2 degrees of freedom in the numerator and 24 in the denominator is (0, 2.54)*. Since the F-statistic computed in this exercise exceeds 2.54 you should reject H_0.

TWO-WAY AOV *(Row-Column Effects)*

You can use two-way analysis of variance to evaluate the combined effects of two variables on a third. Table 4.5 expresses the percent of light reflected from five types of plastic surfaces coated with three types of paint.

Table 4.5

		Type of Surface				
		1	2	3	4	5
Type	A	14.5	13.6	16.3	23.2	19.4
of	B	14.6	16.2	14.8	16.8	17.3
Paint	C	16.2	14.0	15.5	18.7	21.0

*You may verify this using the F Distribution Program.

Here, the effects of the paint types on the population mean are known as row effects. Now, to estimate the population variance on the basis of row means, the sum of squares among rows is used.

$$\sigma^2 \doteq C \sum_{i=1}^{R} (\overline{x}_{i.} - \overline{x})^2/(R-1) = RSS/(R-1)$$

In the above equation:

R = the number of rows.

C = the number of columns.

$\overline{x}_{i.}$ = the mean of the sample values in row i.

$\overline{x} = \sum_{i=1}^{R} \sum_{j=1}^{C} x_{ij}/RC.$

RSS = the row sum of squares (corresponds to TSS in one-way AOV).

Similar to the error sum of squares used in one-way AOV, the estimate of the variance made from the individual elements of the sample is found using the residual sum of squares (Res).

$$\sigma^2 \doteq \sum_{i=1}^{R} \sum_{j=1}^{C} (x_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x})^2/(R-1)(C-1) = Res/(R-1)(C-1).$$

The F-statistic

$$F_R = \frac{RSS/(R-1)}{Res/(R-1)(C-1)}$$

with R-1 degrees of freedom in the numerator and $(R-1)(C-1)$ degrees of freedom in the denominator is calculated by this program. And as with one-way AOV, you can use this data and the F-distribution program to test the effects of row variables on the population mean.

Similar to the above, the effects of surface types on the population mean are known as column effects. Here the estimate of the population variance made on the basis of column means is accomplished using the sum of squares among columns.

$$\sigma^2 = R \sum_{j=1}^{C} (\overline{x}_{.j} - \overline{x})^2/(C-1) = CSS/(C-1).$$

CSS is known as the column sum of squares.

The F-statistic used for testing column effects is

$$F_C = \frac{CSS/(C-1)}{Res/(R-1)(C-1)}$$

with C-1 degrees of freedom in the numerator and $(R-1)(C-1)$ degrees of freedom in the denominator.

In addition to the F-statistics and their corresponding degrees of freedom, this program yields the row sum of squares (RSS), the column sum of squares (CSS), and the total sum of squares (SS). RSS and CSS are computed directly. However, Res is evaluated as Res = SS − RSS − CSS.

Enter your data using the *Two-Way AOV Data Program* found in Section III. Remember, R+C cannot exceed 16.

Solid State Software		TI ©1977		
2-WAY ANALYSIS OF VARIANCE				ST-16
	→ RSS	→ F_R		
→ SS	→ CSS	→ F_C	→ $\nu_1;\nu_2$	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 06	No Change
2	Enter Two-Way AOV data according to User Instructions found in Section III[1]			
3	Select Program		[2nd] [Pgm] 16	No Change
4	Calculate total sum of squares		[A]	SS†
5	Calculate column sum of squares		[B]	CSS†
6	Calculate row sum of squares[2]		[2nd] [B′]	RSS†
	For Column Effects			
7	Calculate F-Statistic		[C]	F_C†
8	Calculate degrees of freedom in numerator		[D]	ν_1†
9	Calculate degrees of freedom in denominator		[R/S]	ν_2†
	For Row Effects[3]			
10	Calculate F-Statistic		[2nd] [C′]	F_R†
11	Calculate degrees of freedom in numerator		[D]	ν_1†
12	Calculate degrees of freedom in denominator		[R/S]	ν_2†

NOTES:
1. R + C cannot exceed 16.
2. Perform Step 5 before Step 6.
3. Perform Steps 7-9 first.

† Printed when PC-100A is used.

Register Contents

R_{00}	R_{05}	R_{10} RSS	R_{15} *	R_{20} *	R_{25} *
R_{01} Rows	R_{06} Mean	R_{11} Used	R_{16} *	R_{21} *	R_{26} *
R_{02} Columns	R_{07} Variance	R_{12} *	R_{17} *	R_{22} *	R_{27} *
R_{03} n	R_{08} SS	R_{13} *	R_{18} *	R_{23} *	R_{28}
R_{04}	R_{09} CSS	R_{14} *	R_{19} *	R_{24} *	R_{29} j Count

*See Table 3.1.

Example:

Similar to the first t-statistic example, let's suppose that three pigs are randomly chosen from each of nine litters. Then, one pig from each litter is placed on diet A, a second on diet B, and the third on diet C for a fixed period of time. At the end of the experiment the pigs are weighed to see how much weight they have gained. These results are tabulated below.

Table 4.6

	Litter								
	1	2	3	4	5	6	7	8	9
Diet A	14.7	14.5	19.2	19.2	14.7	19.6	20.0	16.1	16.5
Diet B	18.8	19.2	19.9	19.0	20.3	18.8	20.3	22.0	16.7
Diet C	14.2	13.2	14.2	19.3	18.4	15.7	18.8	17.6	14.0

In the first experiment we were concerned with determining if each diet caused the same average weight gain. We would like to solve the same problem here; but since we are dealing with more than two samples the t-test can no longer be used. To perform this evaluation we again turn to analysis of variance. One-way AOV would tell us whether or not there is any significant difference between the diets. But what if we also want to compare the litters at the same time? For this purpose we may use 2-way AOV. Test the hypotheses

H_0: each diet causes the same average weight gain ($\mu_A = \mu_B = \mu_C$)

against

H_1: not all the diets cause the same weight gain

and

H_0: the litters all had the same average weight gain

against

H_1: not all the litters had the same average weight gain.

Perform these tests at the 95% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 06		Select AOV Data Program
	[E]	0.	Initialize 2-Way AOV
3†	[B]	3.	Rows
9†	[C]	9.	Columns
14.7†	[A]	1.	x_{11}
14.5†	[A]	2.	x_{12}
19.2†	[A]	3.	x_{13}
19.2†	[A]	4.	x_{14}
14.7†	[A]	5.	x_{15}
19.6†	[A]	6.	x_{16}

ENTER	PRESS	DISPLAY	COMMENTS
20.0†	[A]	7.	x_{17}
16.1†	[A]	8.	x_{18}
16.5†	[A]	9.	x_{19}
18.8†	[A]	1.	x_{21}
19.2†	[A]	2.	x_{22}
19.9†	[A]	3.	x_{23}
19.0†	[A]	4.	x_{24}
20.3†	[A]	5.	x_{25}
18.8†	[A]	6.	x_{26}
20.3†	[A]	7.	x_{27}
22.0†	[A]	8.	x_{28}
16.7†	[A]	9.	x_{29}
14.2†	[A]	1.	x_{31}
13.2†	[A]	2.	x_{32}
14.2†	[A]	3.	x_{33}
19.3†	[A]	4.	x_{34}
18.4†	[A]	5.	x_{35}
15.7†	[A]	6.	x_{36}
18.8†	[A]	7.	x_{37}
17.6†	[A]	8.	x_{38}
14.0†	[A]	9.	x_{39}
	[2nd] [B′]	17.58888889†	$\overline{x}$
	[2nd] [C′]	5.786172841†	s^2
	[2nd] [Pgm] 16	5.786172841	Select 2-Way AOV Program
	[A]	156.2266667†	SS
	[B]	54.88666667†	CSS
	[2nd] [B′]	51.08222222†	RSS
	[C]	2.18420587†	F_C } Column Effects
	[D]	8.†	ν_1 } Column Effects
	[R/S]	16.†	ν_2 } Column Effects
	[2nd] [C′]	8.131234517†	F_R } Row Effects
	[D]	2.†	ν_1 } Row Effects
	[R/S]	16.†	ν_2 } Row Effects

† Printed when PC-100A is used.

Summary:

The acceptance region for the F-statistic with 2 degrees of freedom in the denominator and 16 in the numerator at the 95% confidence level is (0, 3.63)*. Since F_R falls outside of this range you should reject the hypotheses that there is no difference between the diets. In fact, since F_R is not even close to the acceptance region, the probability of a significant difference between the diets is extremely high.

At the same confidence level, the acceptance region for the F-statistic with 8 and 16 degrees of freedom is (0, 2.59)*. Since F_C is in this interval there is no indication of a significant difference between the litters.

Just for practice, try the paint example using the data given in Table 4.5. Testing at the 95% confidence level, your acceptance regions should be (0, 4.46) for F_R and (0, 3.84) for F_C.

*You may verify this using the F Distribution Program.

RANK-SUM TESTS PROGRAM

The Wilcoxon-Mann-Whitney rank-sum test is a procedure used to compare the means of two populations having the same distribution. You do not have to know the form of the distribution to use this test. However, assuming the distributions of the cumulative density functions f(x) and g(x) take the same form, you may use this program to test

H_0: the means of the populations are equal, f(x) = g(x)

against

H_1: the means of the populations differ by an unknown constant c, f(x + c) = g(x).

In Figure 4.4, H_1 indicates that g(x) has shifted to the right by an amount c.

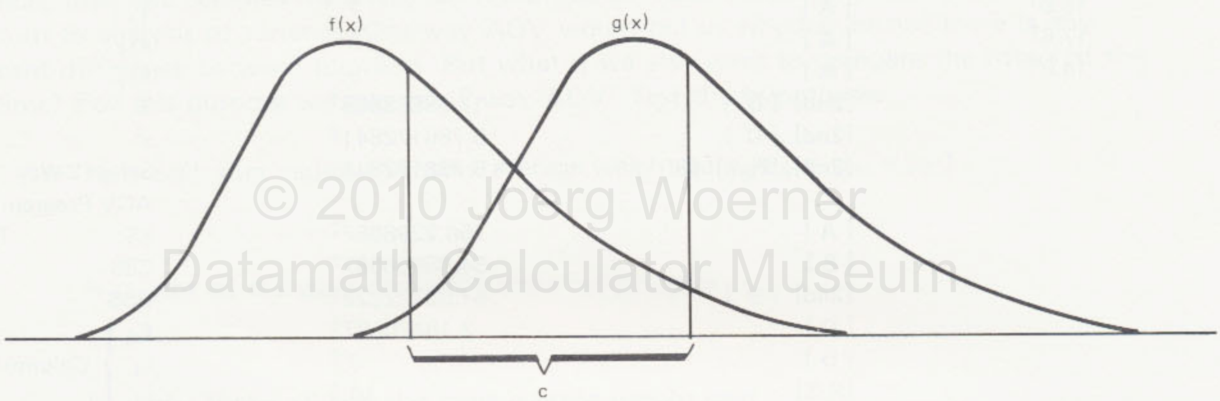


Figure 4.4

Suppose that X is a random sample of size m and Y is a random sample of size n. Assuming that these samples are from populations having the same distribution, you may use this program to compare their means. The first step is to order or rank the values in your two samples from smallest to largest. (The program does this for you.) For example, if your samples were

$$X = \{2, 5, 6\} \text{ and } Y = \{3, 6, 7\},$$

the combined ordered set would be

$$\{x_1, y_1, x_2, x_3, y_2, y_3\}.$$

Note that when ties occur the program ranks the x value first. Naturally, you should use a larger sample than in this demonstration.

When H_1 is true for a positive c, the y values tend to be larger than the x values. A statistic taking advantage of this fact is the sum of the ranks of the y's in the combined ordered set. That is, if $R(y_i)$ is the rank of y_i, then

$$T_y = \sum_{i=1}^{n} R(y_i).$$

In the example given above $T_y = 2 + 5 + 6 = 13$.

Now, if H_0 is true and m, n $\geqslant$ 10, then T_y possesses an approximate normal distribution. Consequently, the Mann-Whitney statistic for y

$$w_y = mn - T_y + n(n+1)/2$$

is also nearly normally distributed. The mean and variance of w_y are given as

$$\overline{w} = mn/2 \quad \text{and} \quad s_w{}^2 = mn(m+n+1)/12.$$

After the program converts this to standard normal form we have

$$z_y = (w_y - \overline{w})/s_w.$$

You may now plug this data into the *Normal Distribution Program* and perform a lower-tailed test to evaluate your hypotheses.

If you suspect that H_1 may be true for a negative c you should use T_x as the basis of your test. Calculations proceed similarly to the above except that an upper-tailed test is used in the final evaluation. Upper and lower-tailed tests are described in Section VI.

To use this program, enter your data using the *Bivariate Data Program* found in Section III. Since the data is not paired enter all the x values first. Then enter all the y values. (You should note, however; that this method of entry invalidates data deletion procedures.) There is no need for the sample sizes to be the same. The only restriction is that registers R_{00} through $R_{(31+m+n)}$ must be left available for program use.

Solid State Software				TI ©1977
RANK SUM TESTS				ST-17
	$\rightarrow T_y$			$\rightarrow s_{\bar{w}}^2$
Rank Data	$\rightarrow T_x$	$\rightarrow w$	$\rightarrow z$	$\rightarrow \bar{w}$

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 04	No Change
2	Enter Bivariate Data according to User Instructions found in Section III[1]			
3	Select Program		[2nd] [Pgm] 17	No Change
4	Rank data[2]		[A]	Ordered data is flashed in display and printed.
5	Calculate rank sum of x[3,5]		[B]	T_x †
6	Calculate Mann-Whitney for x		[C]	w_x †
7	Calculate normal deviate for x		[D]	z_x †
8	Calculate rank sum of y[4]		[2nd] [B′]	T_y †
9	Calculate Mann-Whitney for y		[C]	w_y †
10	Calculate normal deviate for y		[D]	z_y †
11	Display rank mean		[E]	$\bar{w}$ †
12	Display rank variance		[2nd] [E′]	s_w^2 †

NOTES:
1. Enter all the x values first, then enter the y values. This invalidates data deletion procedures.
2. Perform this step before 5 or 8.
3. Perform this step before 6 and 7.
4. Perform this step before 9 and 10.
5. Execution time increases with the number of data points.

† Printed when PC-100A is used.

Register Contents

R_{00}	R_{05}	R_{10}	R_{15} n_x	R_{20}	R_{25} T	R_{30} and above are used in ranking the raw data.
R_{01}	R_{06}	R_{11}	R_{16}	R_{21}	R_{26} Used	
R_{02}	R_{07}	R_{12}	R_{17}	R_{22}	R_{27} Used	
R_{03} n_y	R_{08}	R_{13}	R_{18}	R_{23} $\bar{w}$	R_{28}	
R_{04}	R_{09}	R_{14}	R_{19}	R_{24} s_w^2	R_{29}	

Example:

In the *Means and Moments* example we showed that the distribution of the heights of a group of men had the same shape as the distribution of the heights of their sons. The only difference detected was that the second distribution appeared to be shifted slightly to the right. The data used in that example is repeated below.

Fathers: 67.2, 65.0, 68.3, 69.9, 66.3, 69.7, 69.5, 72.9, 70.2, 74.1.

Sons: 68.4, 65.3, 66.5, 69.0, 73.6, 75.9, 69.7, 69.8, 71.0, 70.8, 67.7, 74.4, 69.9, 71.5, 71.1.

Test the hypotheses

H_0: the means of the distributions are equal, $f(x) = g(x)$

against

H_1: the means of the distributions differ by a positive constant c, $f(x + c) = g(x)$.

Perform this test at the 95% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 04		Select Bivariate Data Program
	[2nd] [E′]	0.	Initialize
67.2†	[A]	1.	x_1
65.0†	[A]	2.	x_2
68.3†	[A]	3.	x_3
69.9†	[A]	4.	x_4
66.3†	[A]	5.	x_5 Heights of Fathers
69.7†	[A]	6.	x_6
69.5†	[A]	7.	x_7
72.9†	[A]	8.	x_8
70.2†	[A]	9.	x_9
74.1†	[A]	10.	x_{10}
68.4†	[B]	1.	x_1
65.3†	[B]	2.	x_2
66.5†	[B]	3.	x_3
69.0†	[B]	4.	x_4
73.6†	[B]	5.	x_5
75.9†	[B]	6.	x_6
69.7†	[B]	7.	x_7
69.8†	[B]	8.	x_8 Heights of Sons
71.0†	[B]	9.	x_9
70.8†	[B]	10.	x_{10}
67.7†	[B]	11.	x_{11}
74.4†	[B]	12.	x_{12}
69.9†	[B]	13.	x_{13}
71.5†	[B]	14.	x_{14}
71.1†	[B]	15.	x_{15}

DATA EVALUATION

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 17		Select Rank-Sum Tests Program
	[A]	65.†	Rank Data – Data for Group x is Ordered First – Ordered Data is Briefly Displayed and Printed.
		66.3†	
		67.2†	
		68.3†	
		69.5†	
		69.7†	
		69.9†	
		70.2†	
		72.9†	
		74.1†	
		65.3†	
		66.5†	
		67.7†	
		68.4†	
		69.†	
		69.7†	
		69.8†	
		69.9†	
		70.8†	
		71.†	
		71.1†	
		71.5†	
		73.6†	
		74.4†	
		75.9†	
		0.	
	[B]*	111.†	T_x
	[C]	94.†	w_x
	[D]	1.053930373†	z_x
	[2nd] [B′]	214.†	T_y
	[C]	56.†	w_y
	[D]	−1.053930373†	z_y
	[E]	75.†	$\overline{w}$
	[2nd] [E′]	325.†	$s_w^{\ 2}$

†Printed when PC-100A is used.

*Requires approximately 25 seconds.

Summary:

Since H_1 assumes that c is positive, z_y is the statistic that we are interested in. At the 95% confidence level the acceptance region for z_y is $(-1.64, \infty)$*. Since z_y is within this range, accept H_0. That is, there is no indication that the sons are taller than their fathers.

*You may verify this using the Normal Distribution Program.

V. MODEL FITTING

THEORETICAL HISTOGRAM PROGRAM

To determine the shape of your sample distribution, use this program to compare a histogram constructed from your sample data points to various histogram models. To begin, simply use your data and the *Histogram Data Program* found in Section III to construct a histogram. Next, choose a theoretical distribution that you think might fit your data. Then write a subroutine that calculates the probability function of the distribution you've selected and store it in program memory under label [2nd] [A']. Remember to end your routine with [INV] [SBR] and don't use [=], [CLR], or [RST]. Registers 21-26 are available for your use. Note that x is in the display register when your subroutine is called. When your subroutine ends, the display register should contain f(x).

Once you have complete the above use this program to calculate the theoretically expected count of each cell. As you do this the program compiles a χ^2-statistic for a goodness of fit test where

$$\chi^2 = \sum_{i=1}^{n} \frac{(\text{expected count} - \text{observed count})^2}{\text{expected count}}$$

with $N-1$ degrees of freedom. (N is the number of cells in your histogram.) To complete the goodness of fit test simply calculate $Q(\chi^2)$ using this program.

$Q(\chi^2)$ is the upper tail area of the chi-square curve. If you wish to test the hypothesis that the distribution of your sample is the theoretical distribution you have chosen at say the 90% confidence level (also known as the 10% significance level), then you may accept this hypothesis whenever $0.10 \leqslant Q(\chi^2) \leqslant 1$. In order to ensure the validity of the goodness of fit test you should collect sufficient data for the theoretically expected count of each cell to be no less than 5. Also, your histograms should be made up of at least three cells.

Solid State Software				TI ©1977
THEORETICAL HISTOGRAM				ST-10
→ Cnt (Disc)				Init Disc
→ Cnt (Cont)		→ $Q(\chi^2)$		Init Cont

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 07	No Change
2	Enter Histogram data according to User Instructions found in Section III			
3	Calculate sample mean if desired[1,6]		[2nd] [$\bar{x}$] [x≷t]	$\bar{x}$
4	Calculate sample standard deviation if desired[1,6]		[INV] [2nd] [$\bar{x}$] [x≷t]	Ignore s
	For Continuous Distribution			
5	Enter continuous probability function into program memory (do not use [=], [CLR], or [RST])[2]	f(x)	[2nd] [CP] [LRN] [2nd] [Lbl] [2nd] [A′] [INV] [SBR] [LRN]	
6	Select Program		[2nd] [Pgm] 10	No Change
7	Initialize		[E]	0.
8	Calculate theoretically expected count of cell i[3] (repeat for each cell)		[A]	Count†
9	Calculate chi-square goodness of fit test[4]		[C]	$Q(\chi^2)$†
	For Discrete Distribution[5]			
10	Enter discrete probability function into program memory (do not use [=], [CLR, or [RST])[2]	f(k)	[2nd] [CP] [LRN] [2nd] [Lbl] [2nd] [A′] [INV] [SBR] [LRN]	
11	Select Program		[2nd] [Pgm] 10	No Change
12	Initialize		[2nd] [E′]	0.
13	Calculate theoretically expected count of cell i[3] (repeat for each cell)		[2nd] [A′]	Count†
14	Calculate chi-square goodness of fit test[4]		[C]	$Q(\chi^2)$†

NOTES:

1. Initialization of the Theoretical Histogram program destroys the data needed to compute $\bar{x}$ and s. Note that if you need to know the observed counts of the cells in your Histogram you may perform the Histogram Construction program at this time.
2. Initialization of the Histogram Data program provides 60 data registers. If you own a TI Programmable 58 you will have to repartition your calculator before entering your subroutine. Observe that the prewritten library routines calculating f(x) for the normal and binomial distributions may be called by your subroutine. However, due to conflicting register assignments, the chi-square and student's -t routines may not be used. Initialization may take as long as a minute depending on the length of your subroutine.
3. The cell number is incremented by 1 each time you press [A] or [2nd] [A′]. Calculation of the expected count may take as long as a minute to complete. A count of zero causes a flashing display indicating invalid results. Press [RCL] 20 to display $\Sigma\chi^2$ for current cell.

NOTES:

4. This step must be performed last and can only be done once. Execution time increases with ν. To run a new problem, recompile your data.
5. For discrete distributions, x_{min} and the cell width must be integers.
6. If the display flashes simply press [CE] and continue. The error condition is cuased when the calculator attempts to calculate the mean and std. deviation of the "x" data normally summed into $R_{04}-R_{05}$. Since the program uses these registers for other purposes bad data may have been stored there.

† Printed when PC-100A is used.

Register Contents

R_{00}	Used	R_{05}	Used	R_{10}	Cell 5 Count	R_{15}	Cell 10 Count	R_{20}	Σx^2	R_{25}	
R_{01}	x_{min}	R_{06}	Cell 1 Count	R_{11}	Cell 6 Count	R_{16}	Cell 11 Count	R_{21}		R_{26}	
R_{02}	Width	R_{07}	Cell 2 Count	R_{12}	Cell 7 Count	R_{17}	Cell 12 Count	R_{22}		R_{27}	Used
R_{03}	n	R_{08}	Cell 3 Count	R_{13}	Cell 8 Count	R_{18}	Used	R_{23}		R_{28}	Upper Limit
R_{04}	Lower Limit	R_{09}	Cell 4 Count	R_{14}	Cell 9 Count	R_{19}	Cells	R_{24}		R_{29}	Used

Example:

In the *Means and Moments Program* example we discovered that the shapes of the sample distributions under consideration there are the same. Then, in the *Rank-Sum Tests Program* example we found that there is no significant difference between the means of these samples. Consequently, we may assume that the two samples come from populations having the same distribution. Combine these samples and build a histogram from the data. Then compare this histogram to a theoretical histogram based on the normal distribution. The data is repeated below

Men's Heights: 67.2, 65.0, 68.3, 69.9, 66.3, 69.7, 69.5, 72.9, 70.2, 74.1, 68.4, 65.3, 66.5, 69.0, 73.6, 75.9, 69.7, 69.8, 71.0, 70.8, 67.7, 74.4, 69.9, 71.5, 71.1.

Note that in order to have at least five data points in each cell our histogram may contain no more than three cells.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 07		Select Histogram Data Program
	[2nd] [E']	0.	Initialize
3†	[B]	3.	Cells
64†	[2nd] [B']	64.	x_{min}
4†	[C]	4.	Width
67.2†	[A]	1.	x_1
65.0†	[A]	2.	x_2
68.3†	[A]	3.	x_3
69.9†	[A]	4.	x_4
66.3†	[A]	5.	x_5
69.7†	[A]	6.	x_6
69.5†	[A]	7.	x_7
72.9†	[A]	8.	x_8
70.2†	[A]	9.	x_9
74.1†	[A]	10.	x_{10}
68.4†	[A]	11.	x_{11}
65.3†	[A]	12.	x_{12}
66.5†	[A]	13.	x_{13}
69.0†	[A]	14.	x_{14}
73.6†	[A]	15.	x_{15}
75.9†	[A]	16.	x_{16}
69.7†	[A]	17.	x_{17}
69.8†	[A]	18.	x_{18}
71.0†	[A]	19.	x_{19}
70.8†	[A]	20.	x_{20}
67.7†	[A]	21.	x_{21}
74.4†	[A]	22.	x_{22}
69.9†	[A]	23.	x_{23}
71.5†	[A]	24.	x_{24}
71.1†	[A]	25.	x_{25}
	[2nd] [$\bar{x}$] [$x \gtrless t$]	69.908	$\bar{x}$
	[INV] [2nd] [$\bar{x}$]	2.58069758*	Ignore
	[CE] [$x \gtrless t$]	2.816457586	s

*See Note 6 of the User Instructions.
†Printed when PC-100A is used.

(Now store your subroutine for computing f(x) in program memory. Note that x must be normalized before calling the library routine described in Normal Distribution Program to calculate the probability density function. For simplicity, just use $\overline{x} = 69.9$ and $s = 2.8$ as estimates of μ and σ.)

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [CP] [LRN]	000 00	(Key Codes)
	[2nd] [Lbl]	001 00	000 76
	[2nd] [A′]	002 00	001 16
	[( ]	003 00	002 53
	[( ]	004 00	003 53
	[CE]	005 00	004 24
	[–]	006 00	005 75
	[6]	007 00	006 06
	[9]	008 00	007 09
	[·]	009 00	008 93
	[9]	010 00	009 09
	[ )]	011 00	010 54
	[÷]	012 00	011 55
	[2]	013 00	012 02
	[·]	014 00	013 93
	[8]	015 00	014 08
	[ )]	016 00	015 54
	[2nd] [Pgm]	017 00	016 36
	[1] [9]	018 00	017 19
	[A]	019 00	018 11
	[INV] [SBR]	020 00	019 92
	[LRN]	2.814296833	

(Now compare your histogram against the theoretical histogram.)

	[2nd] [Pgm] 10	2.814196833	Select Theoretical Histogram Program
	[E]	0.	Initialize Continuous Distribution
	[A]*	6.057795645†	Cell 1
	[A]*	13.62110728†	Cell 2
	[A]*	5.560044745†	Cell 3
	[C]*	.9668120413†	$Q(\chi^2)$

†Printed when PC-100A is used.
*Requires 10 – 30 seconds.

Summary:

The value obtained for $Q(\chi^2)$ indicates a good fit for any reasonable confidence level that you choose to test at. However, these results should be treated with caution since the histogram contains only three cells. Note that the histogram constructed from the sample data actually has 6 data points in cell 1, 14 in cell 2, and 5 in cell 3. You could have used the *Histogram Construction Program* to obtain this information before you compared your histogram to the theoretical model.

LINEAR REGRESSION MODELS

In analyzing multivariate data you are often interested in finding a mathematical relationship between your variables. That is, you would like to know how changes in one variable affect another. A straight-line linear model for bivariate data is expressed as

$$y = b + mx.$$

In the above, y is the mathematically or functionally dependent variable and x is the independent variable. (Don't confuse this with statistical dependence – it isn't the same.) A line fitted to the sample data points (x, y) would have b as its y-intercept and a slope of m.

The linear regression capabilities of your calculator make it easy to determine appropriate values for m and b when you enter a set of sample data points. The best way to illustrate this feature is with an example. Let's say your company has recently started advertising in a new medium (say a series of magazines), on a weekly basis. The marketing manager has a record of the amount spent on advertising each week (*x*) and the corresponding sales volume (*y*) and there seems to be a fairly good relationship. His question of you is: what would the expected sales volume be if $4750 is spent on magazine advertising next week?

Amount Spent on Advertising (*x*)	Weekly Sales Volume (*y*)
$1000	101,000
$1250	116,000
$1500	165,000
$2000	209,000
$2500	264,000
$4750	???

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 01		Select Diagnostic Program
	[SBR] [CLR]	0.	Initialize Linear Regression Data Registers
	[RST]	0.	Return from Program 01
1000	[x ≷ t]	0.	x_1
101000	[2nd] [Σ+]	1.	y_1
1250	[x ≷ t]	1001.	x_2
116000	[2nd] [Σ+]	2.	y_2
1500	[x ≷ t]	1251.	x_3
165000	[2nd] [Σ+]	3.	y_3
2000	[x ≷ t]	1501.	x_4
209000	[2nd] [Σ+]	4.	y_4
2500	[x ≷ t]	2001.	x_5
264000	[2nd] [Σ+]	5.	y_5
4750	[2nd] [Op] 14	514672.4138	$x \rightarrow y'$
	[2nd] [Op] 12	−11922.41379	b
	[x ≷t]	110.862069	m
	[2nd] [Op] 13	.9935283439	r

Summary:

Observe that data for this type of problem is entered by placing the x value in the T-register, the y value in the display, and pressing [2nd] [Σ+]. Then, the answer to your problem is found by entering the amount you want to spend on advertising (x = 4750) and predicting the resulting sales volume (y = 514,672) by pressing [2nd] [Op] **14**. Additional special operating codes are used to compute the slope, intercept, and correlation coefficient. See your Owner's Manual for a complete explanation on all the built-in statistical features.

ABOUT THE CORRELATION COEFFICIENT

Pressing [2nd] [Op] **13** displays the correlation coefficient of the two sets of data, r. A value close to plus 1 indicates a high positive correlation and a value near minus 1 indicates a high negative correlation. A value of about zero indicates that the two sets of data are not related.

For example, suppose your company gives two tests to new employees—*Test A* and *Test B.* If there is a high positive correlation between the two tests, then you can predict that an employee who scores high (or low) on *Test A* will also score high (or low) on *Test B.* On the other hand, if there is a high negative correlation between the two tests, you can predict that an employee who scores high (or low) on *Test A* will score low (or high) on *Test B.* If there is no correlation (correlation coefficient equals 0), then you can say nothing about how an employee's performance on *Test A* relates to his or her performance on *Test B.*

In the above example then, the value we computed for r tells us there is a very high positive correlation between our samples. However, you should note an important point here. Strictly speaking all we've shown in this example is that a definite *relationship* exists between advertising and sales. Be careful about drawing conclusions about *cause and effect.* In this case, you can probably be pretty sure that your advertising is pushing your sales up—but in other cases, the "cause and effect" relation may not be so obvious. Two variables that are related to a *third* can show a relation to each other—without a "cause and effect" relation between them.

In this example, we're predicting the future based on only five data points from the past—and that's not much to go on. In general, the less data you have to go on, the more "chancy" your prediction will be. As it turns out there's a quick way to get a measure of how valid your correlation factor is under different data conditions.

This test may be used to test the hypotheses

$$H_0: \ r = 0 \quad \text{against} \quad H_1: \ r \neq 0.$$

First, let $\nu = n - 2$ where n is the number of sample data points you have. Then calculate a t-statistic using the formula

$$t = \sqrt{\nu r^2/(1-r^2)}.$$

For our example,

$$\nu = 3 \quad \text{and} \quad t \doteq 15.15.$$

The 95% confidence interval for the t-statistic with 3 degrees of freedom is (−3.18, 3.18). Since the t-statistic computed above is not even close to the acceptance region you may consider r to be highly significant.

BIVARIATE CURVE FITTING PROGRAM

It should be clear that a straight-line model is not applicable in all situations. Population growth for example has traditionally followed an exponential curve,

$$y = be^{mx}$$

In order to use the built-in statistical functions to fit an exponential curve to a set of sample data points you would first have to convert the above to the linear equation

$$\ln y = \ln(be^{mx}) = \ln b + mx.$$

This expression indicates that a straight line fitted to the data points (x, ln y) would have a y-intercept of ln b and a slope of m. So all you have to do is convert your data to this form and use the linear regression feature of your calculator to fit a straight line to the transformed points. As mentioned above, this operation would give you the values of ln b and m. To find b, simply press [INV] [ln x] when ln b appears in the display.

You may perform these conversions manually if you wish—but you don't have to. You can let the *Bivariate Data Transforms Program* do it for you. This program was first introduced in Section III.

As you should already know, *Bivariate Data Transforms* can be used to transform your data points (x, y) to any of the following forms:

(x, ln y), (ln x, ln y), (ln x, y).

You have seen that the first set of points can be used for exponential curve fit—but how may the others be used?

The expression for a power curve takes the form

$$y = bx^m.$$

We can easily convert this to

$$\ln y = \ln(bx^m) = \ln b + m \ln x.$$

This means that a straight line fitted to the data points (ln x, ln y) would have ln b as its y-intercept and m as its slope.

A logarithmic curve is expressed as

$$y = b + m \ln x.$$

Since this equation is already in our standard format we can quickly see that a line fitted to the data points (ln x, y) would have a y-intercept of b and a slope of m.

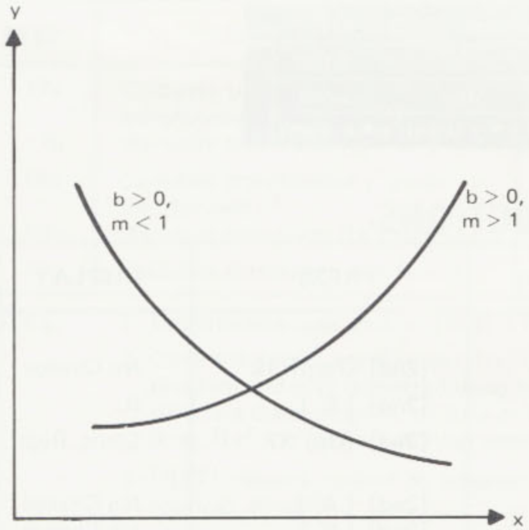


Figure 5.1 — Exponential Curves

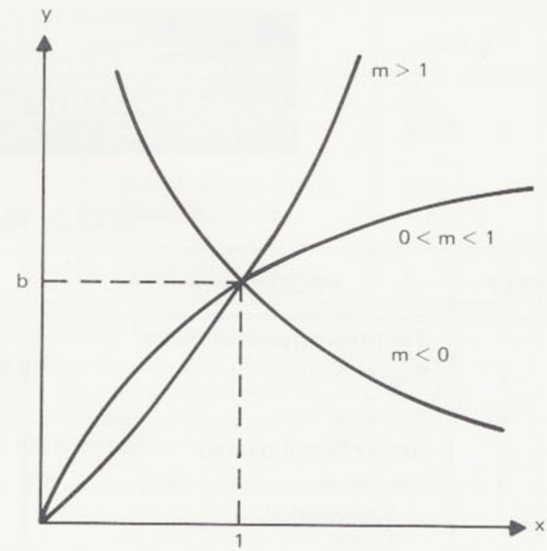


Figure 5.2 — Power Curves

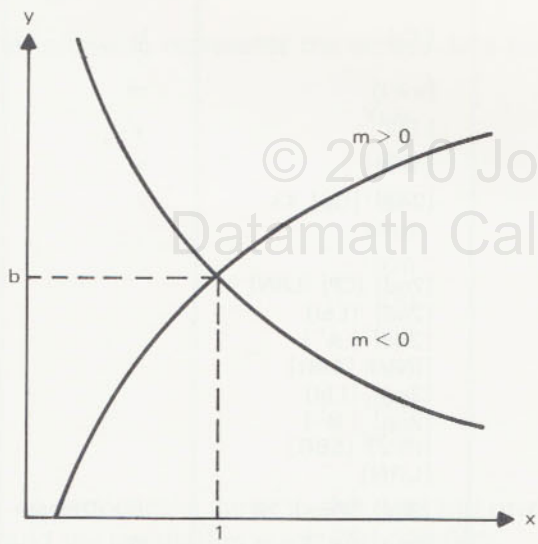


Figure 5.3 — Logarithmic Curves

Note that these graphs represent x versus y, not the transformed data points.

If you are fitting one of these curves to your data, the *Bivariate Data Transforms Program* includes routines for determining the correct values of b and m and predicting x given y or y given x. For example, even though the built-in calculator feature can only find ln b for an exponential curve fit, this program automatically transforms ln b to b before displaying the result.

You also have the option of defining your own transforms for x and y as described in Section III. This allows you to easily fit almost any curve you wish to your sample data. However, this program recognizes user-defined transforms for input data only. To obtain output data, just use the built-in calculator features to compute the transformed results. Then convert this data to the correct form yourself. Remember to enter transformed values of x and y when predicting new points. The user instructions include a complete explanation of this procedure.

Solid State Software			TI ©1977	
BIVARIATE DATA TRANSFORMS				ST-12
Exp (x, ln y)	Pwr (ln x, ln y)	Ln (ln x, y)	User	Initialize
x	y	→ b; m (Pgm)	x → y′ (Pgm)	y → x′ (Pgm)

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
	For Preprogrammed Curve			
1	Select Program		[2nd] [Pgm] 12	No Change
2	Initialize[1]		[2nd] [E′]	0.
3	Repartition if desired	n	[2nd] [Op] 17	Steps. Regs
4	Choose curve: Exponential, Power, Logarithmic		 [2nd] [A′] [2nd] [B′] [2nd] [C′]	 No Change No Change No Change
5a	Enter x_i[2]	x_i	[A]	i
5b	Enter y_i[2]	y_i	[B]	i
	(Repeat Step 5 for each data pair)			
6a	Calculate y-intercept and slope of line fitted to data points		[C]	b
6b	Display slope		[x≷t]	m
7	Calculate y′ given x	x	[D]	y′
8	Calculate x′ given y	y	[E]	x′
9	Calculate correlation coefficient		[2nd] [Op] 13	r
	For User-Defined Curve			
10	Enter transforms into program memory (do not use [=], [CLR], or [RST])	 f(x) g(y)	[2nd] [CP] [LRN] [2nd] [Lbl] [2nd] [A′] [INV] [SBR] [2nd] [Lbl] [2nd] [B′] [INV] [SBR] [LRN]	
11	Select Program		[2nd] [Pgm] 12	No Change
12	Initialize[1]		[2nd] [E′]	0.
13	Repartition if needed	n	[2nd] [Op] 17	Steps. Regs
14	Select User-Defined Curve Mode		[2nd] [D′]	No Change
15a	Enter x_i[2]		[A]	i
15b	Enter y_i[2]		[B]	i
	(Repeat Step 15 for each data pair)			
16a	Calculate slope and intercept of straight line fitted to transformed data		[C]	b
16b	Manually transform b to correct form			
16c	Display transformed m		[x≷t]	m
16d	Manually transform m to correct form			

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
17a	Calculate transformed y′ given transformed x^3	x	[RST] [2nd] [A′] [2nd] [Op] 14	f(x) g(y′)
17b	Manually transform g(y′) to y′			
18a	Calculate transformed y′ given transformed y^3	y	[RST] [2nd] [B′] [2nd] [Op] 15	g(y) f(x′)
18b	Manually transform f(x′) to x′			
19	Calculate correlation coefficient		[2nd] [Op] 13	r

NOTES:

1. Initialization uses routine [2nd] [E′] of the Bivariate Data program.
2. Once the data is transformed, f(x) is entered using routine [A] of the Bivariate Data program and g(y) is entered using routine [B]. Data must be entered in pairs. See the Bivariate Data program for data deletion procedures and limitation of the Raw Data Base. f(x) and g(y) are printed when the PC-100A is used.
3. [RST] returns control to program memory and allows you to use your transform routines directly.
4. See Table 3.1 for register contents.

Example:

The following represents the census data for the United States for the years 1890 - 1970.

Year	Population
1890	62,947,714
1900	75,994,575
1910	91,972,266
1920	105,710,620
1930	122,775,046
1940	131,669,275
1950	150,697,361
1960	179,323,175
1970	203,235,298

Fit an exponential curve to this data and predict the population in the year 2000. In what year should the population reach 300,000,000?

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 12		Select Bivariate Transforms Program
	[2nd] [E′]	0.	Initialize
	[2nd] [A′]	0.	Choose Exponential Curve Fit
1890†	[A]	1.	x_1
62947714†	[B]	1.	y_1
1900†	[A]	2.	x_2
75994575†	[B]	2.	y_2
1910†	[A]	3.	x_3
91972266†	[B]	3.	y_3
1920†	[A]	4.	x_4
105710620†	[B]	4.	y_4
1930†	[A]	5.	x_5
122775046†	[B [	5.	y_5
1940†	[A]	6.	x_6
131669275†	[B]	6.	y_6
1950†	[A]	7.	x_7
150697361†	[B]	7.	y_7
1960†	[A]	8.	x_8
179323175†	[B]	8.	y_8
1970†	[A]	9.	x_9
203235298†	[B]	9.	y_9
	[C]	.0001716447	b
	[x ⪌ t]	.0141183066	m
	[2nd] [Op] 13	.9957238762	r
2000	[D]	314510832.3	$x \rightarrow y'$
300000000	[E]	1996.65427	$y \rightarrow x'$

Summary:

The resulting curve is

$$y = 0.0001716447\, e^{(0.0141183066)x}.$$

You may evaluate the t-statistic for r if you wish. However, due to the high value of r, there is really no need to perform this test. As you can see, the population should exceed 314 million in the year 2000 and reach 300 million in mid-1996.

Example:

In attempting to determine an appropriate price to charge for his product, a manufacturer obtained the following data in units sold is on a monthly basis.

Table 5.1

Price in Dollars	21	22	23	24	25	26
Thousands of Units Sold	20.3	17.8	15.9	14.1	12.4	10.8

Fit a reciprocal curve to this data where

$$y = b + m\ (1/x).$$

† Printed when PC-100A is used. (If data is transformed then the transformed values are printed instead).

Then determine what price the manufacturer should charge if he wants to sell 15,000 units per month. How many units could he sell if he only charged $18?

[2nd] [CP]
[LRN]
[2nd] [Lbl]
[2nd] [A']
[1/x]
[INV] [SBR]
[2nd] [Lbl]
[2nd] [B']
[INV] [SBR]
[LRN]

Subroutine [2nd] [A'] transforms x to 1/x. Subroutine [2nd] [B'], even though it just returns y in its original form, is needed in order to prevent an error condition. Now run this problem.

Note that if you have a TI Programmable 58 you will have to repartition your calculator after initialization.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 12		Select Bivariate Transform Program
	[2nd] [E']	0.	Initialize
5	[2nd] [Op] 17	79.49	Repartition (58 Only)
	[2nd] [D']		Select User-Defined Curve
21†	[A]	1.	x_1
20.3†	[B]	1.	y_1
22†	[A]	2.	x_2
17.8†	[B]	2.	y_2
23†	[A]	3.	x_3
15.9†	[B]	3.	y_3
24†	[A]	4.	x_4
14.1†	[B]	4.	y_4
25†	[A]	5.	x_5
12.4†	[B]	5.	y_5
26†	[A]	6.	x_6
10.8†	[B]	6.	y_6
	[C]	−28.62600532	b
	[x ≷ t]	1024.840156	m
	[2nd] [Op] 13	.9996511874	r
18	[1/x] [2nd] [Op] 14	28.30955893	$x \rightarrow y'$
15	[2nd] [Op] 15 [1/x]	23.49149662	$y \rightarrow x'$

† Printed when PC-100A is used. (If data is transformed – transformed values are printed.)

Summary:

The resulting curve is approximately

$$y = -28.6 + 1024.8/x.$$

If the manufacturer wants to sell 15,000 units he may charge $23.49 for his product. If he only charged $18 he could sell over 28,000 units.

MULTIPLE LINEAR REGRESSION PROGRAM

When working with trivariate data you can use this program to fit an equation of the form

$$z = a_0 + a_1 x + a_2 y$$

to your sample data triplets (x, y, z).

The regression coefficients are calculated using the least-squares method. You can also use this program to find the multiple coefficient of determination and to predict values of z for a given x and y. You must enter your data using the *Trivariate Data Program* found in Section III.

Solid State Software			TI ©1977	
MULTIPLE LINEAR REGRESSION				ST-18
x	$y \rightarrow z'$			
$\rightarrow a_0$	$\rightarrow a_1$	$\rightarrow a_2$	$\rightarrow R^2$	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 05	No Change
2	Enter Trivariate Data according to User Instructions found in Section III			
3	Select Program		[2nd] [Pgm] 18	No Change
4a	Calculate coefficients and display a_0		[A]	a_0 †
4b	Display a_1		[B]	a_1 †
4c	Display a_2		[C]	a_2 †
5	Calculate coefficient of determination		[D]	R^2 †
6	Calculate z′ for a given x and y	x† y†	[2nd] [A′] [2nd] [B′]	x z′†

NOTE: † Printed when PC-100A is used.

Register Contents

R_{00}	R_{05} Σx^2	R_{10}	R_{15} Used	R_{20} a_1	R_{25}
R_{01} Σy	R_{06} Σxy	R_{11} Σxz	R_{16} Used	R_{21} a_2	R_{26}
R_{02} Σy^2	R_{07} Σz	R_{12} Σyz	R_{17} Used	R_{22} Used	R_{27}
R_{03} n	R_{08}	R_{13} Σz^2	R_{18} Used	R_{23} Used	R_{28}
R_{04} Σx	R_{09} x	R_{14} Used	R_{19} a_0	R_{24}	R_{29}

Example:

In an experiment to study the effects of two gasoline additives on the gas mileage of a specific car the following data is obtained.

Table 5.2

Units of Additive x	0	0	0	1	1	1	2	2	2
Units of Additive y	0	1	2	0	1	2	0	1	2
Gas Mileage (z)	17.3	18.1	18.7	18.6	19.1	19.5	19.6	19.9	20.3

Fit an equation of the form

$$z = a_0 + a_1 x + a_2 y$$

to this data and predict the gas mileage for x = 1, y = 0.5.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 05		Select Trivariate Data Program
	[2nd] [E′]	0.	Initialize
0†	[A]	1.	x_1
0†	[B]	1.	y_1
17.3†	[C]	1.	z_1
0†	[A]	2.	x_2
1†	[B]	2.	y_2
18.1†	[C]	2.	z_2
0†	[A]	3.	x_3
2†	[B]	3.	y_3
18.7†	[C]	3.	z_3
1†	[A]	4.	x_4
0†	[B]	4.	y_4
18.6†	[C]	4.	z_4
1†	[A]	5.	x_5
1†	[B]	5.	y_5
19.1†	[C]	5.	z_5
1†	[A]	6.	x_6
2†	[B]	6.	y_6
19.5†	[C]	6.	z_6
2†	[A]	7.	x_7
0†	[B]	7.	y_7
19.6†	[C]	7.	z_7
2†	[A]	8.	x_8
1†	[B]	8.	y_8
19.9†	[C]	8.	z_8
2†	[A]	9.	x_9
2†	[B]	9.	y_9
20.3†	[C]	9.	z_9

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 18		Select Multiple Regression Program
	[A]	17.56111111†	a_0
	[B]	0.95†	a_1
	[C]	0.5†	a_2
	[D]	.9782301163†	R^2
1†	[2nd] [A']	1.	x
.5†	[2nd] [B']	18.76111111†	$y \rightarrow z'$

† Printed when PC-100A is used.

Summary:

The resulting equation is approximately

$$z = 17.56 + 0.95x + 0.5y.$$

Clearly, additive x is twice as effective as y. Note that a_0 corresponds to the gas mileage of the car when no additive is entered.

NONLINEAR REGRESSION

Just as with bivariate models, multivariate regression models may assume a nonlinear form. For example you may want to fit an equation of the form

$$z = a_0 x^{a_1} y^{a_2}$$

to your sample data. To do this simply convert this equation to the standard format given earlier. The new equation is

$$\ln z = \ln a_0 + a_1 \ln x + a_2 \ln y.$$

Now, instead of entering x, y, and z into the program, enter ln x, ln y, and ln z. The resulting regression coefficients are ln a_0, a_1, and a_2. To find a_0 simply press [INV] [ln x] when ln a_0 appears in the display. And if you need to predict new values for z, enter ln x and ln y to obtain ln z. Then convert ln z to z by pressing [INV] [ln x].

This is just one example of fitting a nonlinear equation to sample data. Naturally, there are many equations that you may choose to use. Another example is given below

Example:

A production manager believes that a workman's performance is related to the number of hours that he works in a week according to the quadratic equation

$$y = a_0 + a_1 x + a_2 x^2.$$

The following data is based on a scale devised by the manager.

Table 5.3

Hours Worked (x)	20	30	40	50	60	70
Performance (y)	7.8	10.6	11.7	10.5	7.9	5.1

Fit this equation to the sample data and predict the scaled performance for a 45-hour week.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 05		Select Trivariate Data Program
	[2nd] [E′]	0.	Initialize
20†	[A]	1.	x_1
20†	[x^2] [B]	1.	x_1^2
7.8†	[C]	1.	y_1
30†	[A]	2.	x_2
30†	[x^2] [B]	2.	x_2^2
10.6†	[C]	2.	y_2
40†	[A]	3.	x_3
40†	[x^2] [B]	3.	x_3^2
11.7†	[C]	3.	y_3
50†	[A]	4.	x_4
50†	[x^2] [B]	4.	x_4^2
10.5†	[C]	4.	y_4
60†	[A]	5.	x_5
60†	[x^2] [B]	5.	x_5^2
7.9†	[C]	5.	y_5
70†	[A]	6.	x_6
70†	[x^2] [B]	6.	x_6^2
5.1†	[C]	6.	y_6
	[2nd] [Pgm] 18	6.	Select Multiple Regression Program
	[A]	−1.382857143†	a_0
	[B]	.6227142857†	a_1
	[C]	−.0076428571†	a_2
	[D]	.9766337894†	R^2
45†	[2nd] [A′]	45.	x
45†	[x^2] [2nd] [B′]	11.1625†	$x^2 \rightarrow y$

†Printed when PC-100A is used. (Note that the squared value is printed when the data is squared before entry.)

Summary:

The resulting equation is approximately

$$y = -1.3829 + .6227x - .0076x^2.$$

The coefficient of determination calculated by the program indicates that this is a good fit. The predicted performance for a 45-hour week is 11.16.

VI. THEORETICAL DISTRIBUTIONS

As you have seen in *Data Evaluation,* a statistical experiment leads to the evaluation of a test statistic. A test statistic is nothing more than a random variable with a known frequency of occurrence or distribution. Since you know the distribution this statistic you can determine the range of values that it may take when your null hypothesis is true. This range is called the acceptance region. If your value falls outside of the acceptance region it is in what is known as the critical region. The sizes of these regions depend upon the confidence level at which you wish to make your test.

There is always the possibility of chance variations occuring when you sample a population. Confidence levels are used to take these chance variations into consideration so as to protect against rejecting your null hypothesis (H_0) when it is actually true. For example, if you are testing at the 95% confidence level and the test statistic falls inside the critical region, then you are 95% sure that this result is significant and not simply caused by chance variations in sampling.

Testing at the 95% confidence level is also known as testing at the 5% level of significance. That is, you will reject H_0 when it is true in only 5% of your experiments.

Let's take a look at the upper-tailed test. The hypotheses tested in this case are

$$H_0: \ S \leqslant S_0 \qquad \text{against} \qquad H_1: \ S > S_0 .$$

In the above, S is the test statistic and S_0 is the critical point or dividing line between the acceptance and critical regions. If we were testing at the 5% significance level, our frequency distribution curve might look like Figure 6.1.

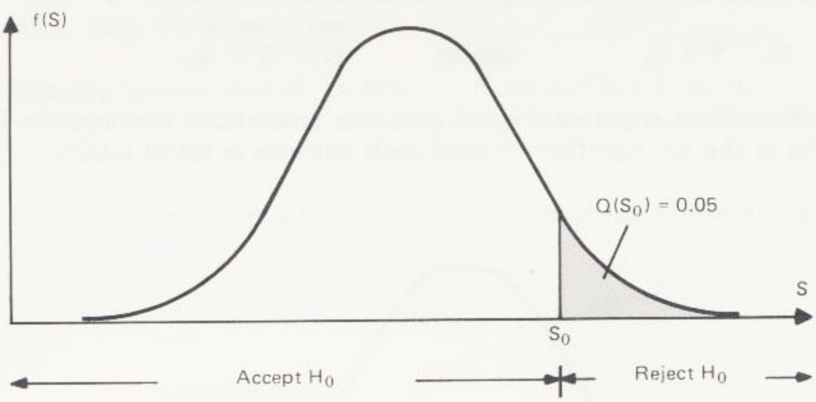


Figure 6.1

The shaded area designated $Q(S_0)$ is the upper-tail area of the distribution curve. Since we are testing at the 5% significance level S_0 has been chosen such that this area equals 0.05. Now, if our test statistic does not exceed S_0 we will accept the null hypothesis. The critical region for this test is (S_0, ∞).

To determine if your statistic falls within the critical region simply calculate Q(S) using an appropriate program from this section. In this case, if Q(S) is no less than 0.05 we would accept H_0 since this would indicate that S does not exceed S_0. This is the same as accepting H_0 when P(S) is no greater than 0.95.

In the case of a lower-tailed test the hypotheses being tested are

$$H_0: \ S \geqslant S_0 \qquad \text{against} \qquad H_1: \ S < S_0.$$

The lower-tail area of a frequency distribution curve is designated by P(S).

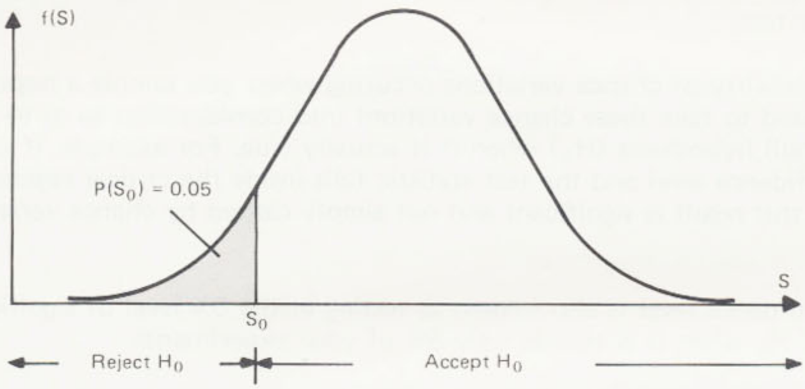


Figure 6.2

The critical region for this test is $(-\infty, S_0)$. To complete this test simply calculate P(S) for your statistic and accept H_0 if this value is 0.05 or greater. Or if you wish, you may determine Q(S) and accept H_0 if this value does not exceed 0.95.

The hypotheses of a two-tailed test are

$$H_0: \ S = S_0 \qquad \text{against} \qquad H_1: \ S \neq S_0.$$

For this type of test the critical region is divided into two equal parts as shown in Figure 6.3. Note that when testing at the 5% significance level each tail area is set at 0.025.

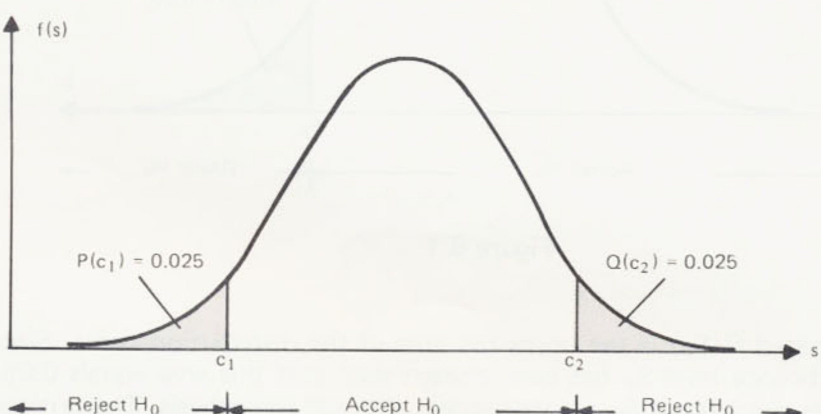


Figure 6.3

In this case we have two critical points c_1 and c_2. These points establish what is known as a confidence interval for S_0. Specifically, (c_1, c_2) is a 95% confidence interval for S_0. If our statistic is such that $c_1 \leqslant S \leqslant c_2$ we would accept the null hypotheses. To determine if S is in this range simply calculate P(S) or Q(S). Then if $0.025 \leqslant P(S) \leqslant 0.975$ or $0.025 \leqslant Q(S) \leqslant 0.975$, accept H_0. (Note that when one of these conditions is met the other is also true.)

SELECTING A CONFIDENCE LEVEL

When you select a confidence level for your experiment it is important that you realize how the statistical process works. Remember, the amount of information you have in your sample does not change. Therefore, if you select a high level of confidence, then what you are confident of is less definite.

Since this last statement may be a little vague, here's an example. Suppose that a mechanic looks at your car and tells you he is pretty sure it will cost between $80 and $100 to fix it. However, if you tell him that he has to be 99.9% sure of his estimate, he will probably estimate a wider range, say $50 to $200. If your experiment requires more confidence over a smaller range you may need to take a larger sample.

As we have pointed out before, testing at the 95% confidence level (5% level of significance) means that the probability of rejecting H_0 when it is true is 0.05. But what about the probability of accepting H_0 when it is false? Calculating this probability is beyond the scope of this material. However, it should be evident that as you decrease the probability of rejecting a true H_0, you increase the probability of accepting a false H_0. The following guidelines should help you to determine what level you should test at.

- If you strongly believe that H_0 is true, or if rejecting a true H_0 would be costly or serious, select of small probability of rejecting a true H_0 by testing at a low level of significance. (Say 1% or even less.)

- If you strongly believe that H_0 is false, or if accepting a false H_0 would be costly or serious, select a low probability of accepting a false H_0 by testing at a high level of significance. (Say 10%, or even as high 25%.)

- If you have no convictions either way, test at a moderate level of significance. (Say 5%.)

NORMAL DISTRIBUTION PROGRAM

The normal distribution is the most often used probability model in modern statistics. As a general rule, you may use this model whenever the distribution of the population you are sampling from takes on one of the shapes illustrated below. However, for the best results make sure that your parent population has at least 100 elements and that your sample size is greater than thirty. The probability density function of a random variable having normal distribution is given by

$$f(x) = (1/\sigma\sqrt{2\pi})\ \exp[-(x - \mu)^2/2\sigma^2].$$

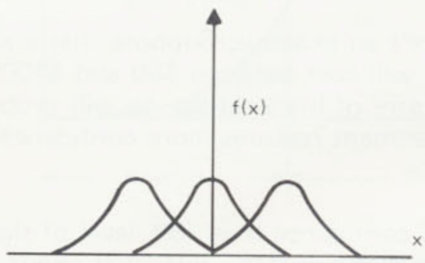


Figure 6.4a

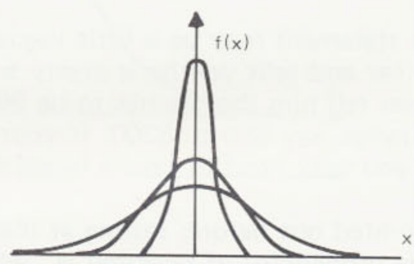


Figure 6.4b

Figure 6.4a illustrates three normal curves with the same standard deviation and different means. Figure 6.4b shows three normal curves with the same mean and different standard deviations. (The normal distribution actually has an infinite range. That is, the density curve never reaches the x-axis even though it may appear to do so in the diagrams.)

For the sake of convenience, a normal curve is said to be in standard form when it has a mean of zero and a standard deviation of one. Now, since any linear transformation of a normal variable yields a new normal variable, any normal variable x may be transformed to standard form or normalized using the equation

$$z = (x - \mu)/\sigma.$$

Here μ is the population mean and σ is its standard deviation. The probability density function of the standard normal variable then becomes

$$\phi(z) = (1/\sqrt{2\pi})\ \exp(-z^2/2).$$

Now consider the case where you would like to use the normal model when your population isn't normally distributed. If you are working with a large sample, and if you know the mean and standard deviation of your parent population, you may use the central limit theorem to make inferences about the mean of your sample, $\overline{x}$. Specifically, if n is your sample size, then

$$f(\overline{x}) = \phi[\sqrt{n}\ (\overline{x} - \mu)/\sigma].$$

Once you have normalized your random variable you may use this program to evaluate the probabilities listed below.

$$\Pr(Z \leqslant z) = P(z).$$
$$\Pr(Z > z) = Q(z).$$
$$\Pr(Z \leqslant |z|) = A(z).$$

Calculations are performed using a series expansion to approximate

$$Q(z) = \int_z^{\infty} \phi(u)du.$$

This approximation is accurate to $\pm 7.5 \times 10^{-8}$ for $z \leqslant 4.7$. The remaining probabilities are calculated from Q(z).

$$P(z) = 1 - Q(z). \qquad A(z) = 1 - 2Q(z). \qquad P(-z) = Q(z).$$

Solid State Software			TI ©1977	
NORMAL DISTRIBUTION				ST-19
$z \rightarrow \phi(z)$	$z \rightarrow P(z)$	$z \rightarrow Q(z)$	$z \rightarrow A(z)$	

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 19	No Change
2	Calculate standard normal density of z	z	[A]	$\phi(z)$
3	Calculate $\Pr(Z \leqslant z)$	z	[B]	P(z)
4	Calculate $\Pr(Z > z)$	z	[C]	Q(z)
5	Calculate $\Pr(Z \leqslant |z|)$	z	[D]	A(z)

Register Contents

R_{00}	R_{05}	R_{10}	R_{15}	R_{20}	R_{25} Used
R_{01}	R_{06}	R_{11}	R_{16}	R_{21}	R_{26} Used
R_{02}	R_{07}	R_{12}	R_{17}	R_{22}	R_{27}
R_{03}	R_{08}	R_{13}	R_{18}	R_{23}	R_{28}
R_{04}	R_{09}	R_{14}	R_{19}	R_{24}	R_{29}

Examples:

1. Find the area under the standard normal curve between z = 0 and z = 1.2.

2. If the life of a flashlight battery is normally distributed with mean μ = 120 hours and standard deviation σ = 10 hours, find the probability that the battery will last for more than 100 hours.

3. Prove that the null hypothesis in the *Rank Sum Tests* example should be accepted ($z_y \doteq -1.05$).

4. A steel company produces steel beams with a mean weight of 1245 pounds and a standard deviation of 10 pounds. Find the probability that a shipment of 20 beams will exceed a 25,000 pound weight limit. ($\overline{x}$ = 25,000/20 = 1250.)

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 19		Select Normal Distribution Program
1.2	[B]	.8849302684	P(z)
	[–]	.8849302684	
0	[B]	.5000000005	P(z)
	[=]	.3849302679	Area
100	[–]	100.	x
120	[=] [÷]	–20.	$x-\mu$
10	[=]	–2.	$(x-\mu)/\sigma = z$
	[C]	0.977249938	Q(z)
1.05	[+/–] [B]	.1468590807	$P(z_y)$
1250	[–]	1250.	$\overline{x}$
1245	[=] [÷] [( ]	5.	$\overline{x}-\mu$
10	[÷]	10.	σ
20	[$\sqrt{x}$] [ )] [=]	2.236067977	$(\overline{x}-\mu/(\sigma/\sqrt{n})$
	[C]	.0126736174	Q(z)

Summary:

1. This area is found by simply performing the following calculations:

$$\Pr(Z \leqslant 1.2) - \Pr(Z \leqslant 0) = P(1.2) - P(0) = \text{Area}.$$

2. In this example x is first converted to a standard normal variable using the expression

$$z = (x-\mu)/\sigma.$$

Once this conversion is made, the probability of the battery lasting more than 100 hours is found as $\Pr(Z > z) = Q(z)$.

3. Since H_1 assumes that the distribution of the heights of the sons has shifted to the right when compared with that of the fathers we may restate our hypotheses as

$$H_0: \mu_x = \mu_y \qquad \text{against} \qquad H_1: \mu_x < \mu_y.$$

   In this case a lower-tailed test is needed to complete the evaluation. That is, we should accept H_0 only if $Q(z_y) \leqslant 0.95$ when testing at the 95% confidence level. Since $Q(-1.05) < 0.95$, accept H_0.

4. The central limit theorem is used to calculate $Pr(\overline{x} > 1250)$. As you can see, this implies that the probability of exceeding the weight limit is extremely low.

BINOMIAL DISTRIBUTION PROGRAM

The binomial distribution may be used when your sample is obtained using either of the following techniques.

- Sampling from an infinite population.
- Sampling from a finite population with replacement.

Bascially, you may be dealing with binomial variables when these conditions are satisfied.

- The experiment consists of a fixed number of statistically independent trials.
- Each trial results in either success or failure.
- Each trial has identical probabilities of success p, and failure 1 − p.

The probability function of a binomial distribution is given by

$$f(k;\, n,\, p) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, 2, \ldots, n; \\ 0, & \text{elsewhere} \end{cases}$$

where n is the number of trials and k is the number of successes in the experiment.

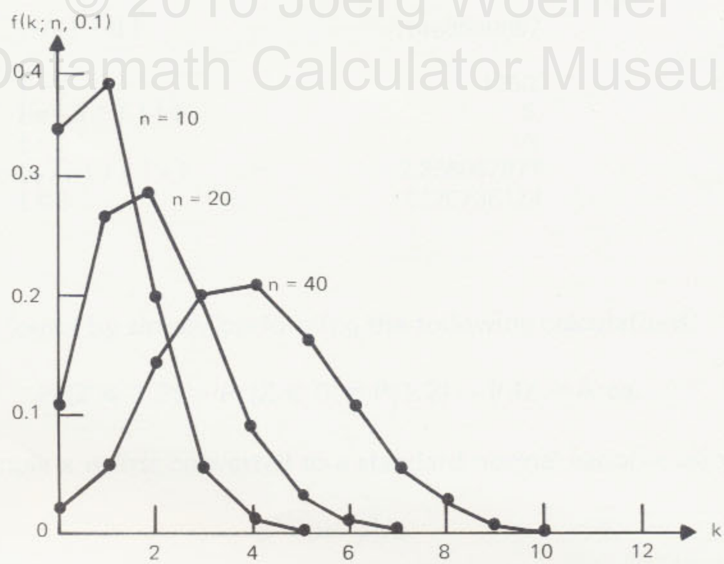


Figure 6.5

As you can see, the normal curve is a good approximation for the binomial distribution when n is large. This relationship becomes more apparent as p approaches 0.5.

The cumulative distribution function of a random variable with binomial distribution is simply

$$F(k;\, n,\, p) = \sum_{j=0}^{k} f(j;\, n,\, p).$$

You can use this program to calculate the following probabilities for n trials of an experiment. Calculations are based on the above summation.

Probability of exactly k successes = $f(k;\, n,\, p) = f(k)$.

Probability of k or fewer successes = $F(k;\, n,\, p) = P(k)$.

Probability of more than k success = $1 - F(k;\, n,\, p) = Q(k)$.

You may also use this program to determine the mean and standard deviation of a binomially distributed population where

$$\mu = np \qquad \text{and} \qquad \sigma = \sqrt{np(1-p)}.$$

Solid State Software				TI ©1977
BINOMIAL DISTRIBUTION				ST-20
$\rightarrow\mu$	$\rightarrow\sigma$			
n	p	k→f(k)	k→P(k)	k→Q(k)

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 20	No Change
2	Enter number of trials	n	[A]	n
3	Enter probability of success on each trial	p	[B]	p
4	Calculate mean		[2nd] [A′]	μ
5	Calculate standard deviation		[2nd] [B′]	σ
6	Calculate probability of k successes	k	[C]	f(k)
7	Calculate probability of k or fewer successes	k	[D]	P(k)
8	Calculate probability of more than k successes	k	[E]	Q(k)

NOTES:
1. Steps 4-8 may be performed at any time and in any order following Steps 1-3.
2. If an output flashes in the display the calculator probably overflowed in calculation. Disregard results. (This only occurs for large n and small k.)

Register Contents

R_{00}	R_{05}	R_{10}	R_{15}	R_{20}	R_{25} Used
R_{01}	R_{06}	R_{11}	R_{16}	R_{21} N	R_{26} f(k)
R_{02}	R_{07}	R_{12}	R_{17}	R_{22} p	R_{27}
R_{03}	R_{08}	R_{13}	R_{18}	R_{23} $1-p$	R_{28}
R_{04}	R_{09}	R_{14}	R_{19}	R_{24} P(k)	R_{29}

Example:

Suppose that you tossed a coin 50 times and obtained 28 heads. Test the hypotheses

H_0: the coin is balanced ($p = 0.5$)

against

H_1: the coin is unbalanced ($p \neq 0.5$)

at the 95% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 20		Select Binomial Distribution Program
50	[A]	50.	n
.5	[B]	0.5	p
	[2nd] [A']	25.	μ
	[2nd] [B']	3.535533906	σ
28	[C]*	.0788256707	f(k)
28	[D]*	.8388818398	P(k)
28	[E]*	.1611181602	Q(k)

*Requires approximately 25 seconds.

Summary:

A two-tailed test is required for this exercise. You should accept H_0 if $0.025 \leqslant P(k) \leqslant 0.975$. Since the value found above falls in this range, you may say that the coin is balanced. The other outputs are informative, but they are not needed in this example. Therefore, you could complete this exercise by simply entering n and p and evaluating P(k).

ANALYZING WITH SMALL SAMPLES

As mentioned earlier, the normal distribution is the most commonly used probability model for large samples. But what if you could only collect a sample of 5 or 10 elements? Or what if you didn't know the mean or standard deviation of your parent population? Well, you would probably use one of the families of distributions discussed on the following pages. Statistics for small sample sizes often assume one of these distributions. Each of these distributions is related to the standard normal probability model. These relationships are based on the number of degrees of freedom of the sample distribution.

The term degrees of freedom is first introduced in Seciton IV where applications of these special distributions are discussed. But no explanation is given there. Basically, you may think of the number of degrees of freedom as the number of free or functionally independent variables in a sample. To clarify this, consider the case of a sample of size one. What if you wanted to estimate the population variance from your sample? It can't be done! You need at least two variables to estimate the variance of a population. Even then the variance can be based on only one item. Thus, one of the variables must be considered as functionally dependent on the other, and we say that the sample distribution has $2 - 1 = 1$ degree of freedom. In general, the number of degrees of freedom is $\nu = n - k$ where k is the number of constraints present in the calculation of a given parameter.

Just how the number of degrees of freedom affects the shape of a distribution curve is illustrated on the following pages.

CHI-SQUARE DISTRIBUTION PROGRAM

Essentially, the chi-square distribution is the distribution of the variance of a normally distributed random variable. That is, if $X_1, X_2, \ldots, X_n$ are independent standard normal random variables, then $X_1^2 + X_2^2 + \ldots + X_n^2 = \chi^2$ has a chi-square distributition with n degrees of freedom. This statistic is most often used to establish confidence intervals for the standard deviation of a population since its distribution depends only on σ.

The shape of a chi-square curve is controlled by its number of degrees of freedom, ν. As you can see in Figure 6.6 the mean of the chi-square distribution is ν. Also, the variance of each curve is equal to 2ν. Clearly, the range of a chi-square variable is from zero to infinity as its distribution is defined as the sum of squared values.

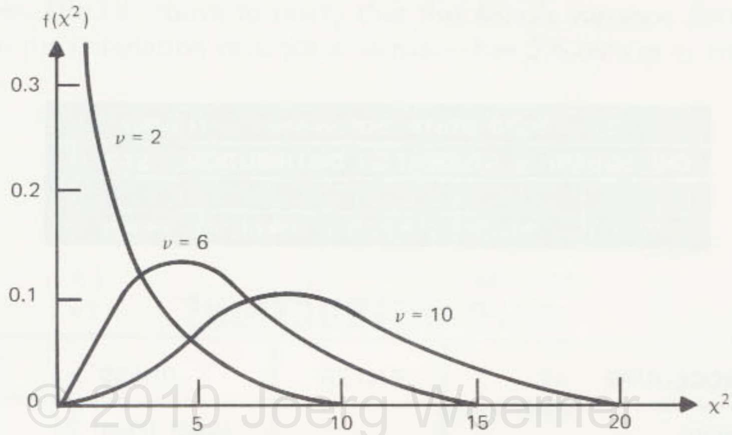


Figure 6.6

For values of ν greater than 30, use is recommended of the fact that, for large ν, the sampling distribution of

$$z = \sqrt{2\chi^2} - \sqrt{2\nu}$$

is approximately the standard unit normal distribution. If this normal approximation is not used for $\nu > 30$, the inequality

$$e^{\chi^2} \text{ and } \left(\frac{\chi^2}{2}\right)^{\nu} \leqslant 9.9999999 \times 10^{99}$$

must be satisfied or erroneous results will be obtained with this program.

The probability density function of a chi-square variable computed directly by this program, is evaluated as

$$f(\chi^2) = [(\chi^2)^{(\nu-2)/2} \exp(-\chi^2/2)] / 2^{\nu/2}\,\Gamma(\nu/2).$$

THEORETICAL DISTRIBUTIONS

A series expansion is used to approximate the cumulative distribution function

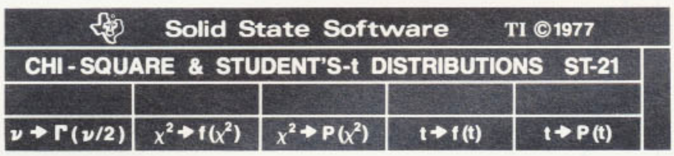
$$P(\chi^2) = \int_0^{\chi^2} f(u)du = Pr(X \leqslant \chi^2).$$

You may calculate the following probability manually.

$$Q(\chi^2) = 1 - P(\chi^2) = Pr(X > \chi^2).$$

An additional output of this program computed when the number of degrees of freedom is entered is the gamma function of $\nu/2$ where

$$\Gamma(\nu/2) = (\nu/2 - 1)!$$



USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 21	No Change
2	Enter degrees of freedom[1]	ν	[A]	$\Gamma(\nu/2)$
3	Enter χ^2-Statistic and calculate density function[2]	χ^2	[B]	$f(\chi^2)$
4	Enter χ^2-Statistic and calculate cumulative distribution function[2]	χ^2	[C]	$P(\chi^2)$

NOTES:
1. Execution time increases with ν.
2. Perform Step 2 first.

Register Contents

R_{00}	R_{05}	R_{10}	R_{15} ν	R_{20} Used	R_{25}
R_{01}	R_{06}	R_{11}	R_{16} χ^2	R_{21} Used	R_{26}
R_{02}	R_{07}	R_{12}	R_{17} $\Gamma[(\nu+2)/2]$	R_{22} Used	R_{27}
R_{03}	R_{08}	R_{13}	R_{18} $\Gamma[(\nu+1)/2]$	R_{23} Used	R_{28}
R_{04}	R_{09}	R_{14}	R_{19} $\Gamma(\nu/2)$	R_{24}	R_{29}

Example:

If s^2 is the sample variance of a sample of size n selected from a normal population, then the distribution of s^2 can be obtained from

$$s^2 = \chi^2 (\sigma^2/n).$$

where χ^2 is the χ^2-statistic with n-1 degrees of freedom and σ^2 is the population variance. This is equivalent to

$$\chi^2 = (n\, s^2/\sigma^2).$$

Now suppose that 20 scales of a certain make are tested for accuracy and yield a sample variance of 3 ounces. Use the above to verify that this sample variance does not contradict the assumption that the population variance is no more than 2.5 ounces at the 95% confidence level.

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 21		Select Chi-Square Distribution Program
19	[A]	119292.462	$\nu \to \Gamma(\nu/2)$
20	[x]	20.	n
3	[÷]	60.	ns^2
2.5	[=]	24.	$ns^2/\sigma^2 = \chi^2$
	[C]	.8038476428	$P(\chi^2)$

Summary:

The actual hypotheses tested in this example are

$$H_0: \ \sigma^2 \leqslant 2.5 \qquad \text{against} \qquad H_1: \ \sigma^2 > 2.5.$$

However, to make this evaluation you must test

$$H_0': \ \chi^2 \leqslant ns^2/\sigma^2 \qquad \text{against} \qquad H_1': \ \chi^2 > ns^2/\sigma^2$$

where χ^2 is the χ^2-statistic with n-1 degrees of freedom.

Due to the nature of the hypotheses, an upper-tailed test is required to complete this evaluation. That is, you may accept H_0 only if $P(\chi^2) \leqslant 0.95$ when testing at the 95% confidence level. Since P(24) with 19 degrees of freedom is within this range, accept H_0.

STUDENT'S t DISTRIBUTION PROGRAM

When small samples are involved, t curves are often used as probability models when it can be assumed that the parent population is approximately normally distributed. The t-statistic with n degrees of freedom is defined as

$$t = X_0/(\sqrt{X_1^2 + X_2^2 + \ldots + X_n^2})/n = X_0/(\sqrt{\chi^2}/n)$$

where $X_0, X_1, \ldots, X_n$ are n+1 independent standard normal variables.

A t curve with $\nu > 1$ degrees of freedom is similar to a standard normal curve in that its mean is always zero*. However, its standard deviation is given as $\sigma = \sqrt{\nu/(\nu-2)}$.* It should be clear that σ converges to one for large values of ν. Therefore, for ν greater than 30, the t curve becomes the standard normal curve.

Figure 6.7

The probability density function of the t distribution is evaluated as

$$f(t) = \frac{\Gamma[(\nu + 1)/2]\,(1 + t^2/\nu)^{-(\nu + 1)/2}}{\sqrt{\pi\nu}\;\Gamma(\nu/2)}$$

The program then uses a series expansion to approximate the cumulative distribution function

$$P(t) = \int_{-\infty}^{t} f(u)du = Pr(T \leqslant t).$$

Using this result, you may calculate the following probabilities manually.

$$Q(t) = 1 - P(t) = Pr(T > t).$$

$$A(t) = 2P(t) - 1 = Pr(T \leqslant |t|).$$

* A random variable with t distribution has no mean for $\nu = 1$ and no standard deviation for $\nu \leqslant 2$.

This program also computes the gamma function of $\nu/2$ when the number of degrees of freedom is entered as in the last program.

Solid State Software TI ©1977				
CHI-SQUARE & STUDENT'S-t DISTRIBUTIONS ST-21				
$\nu \rightarrow \Gamma(\nu/2)$	$x^2 \rightarrow f(x^2)$	$x^2 \rightarrow P(x^2)$	$t \rightarrow f(t)$	$t \rightarrow P(t)$

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 21	No Change
2	Enter degrees of freedom[1]	ν	[A]	$\Gamma(\nu/2)$
3	Enter t-Statistic and calculate density function[2]	t	[D]	f(t)
4	Enter t-Statistic and calculate cumulative distribution function[1,2]	t	[E]	P(t)

NOTES:
1. Execution time increases with ν.
2. Perform Step 2 first.
3. Program operation leaves the calculator in radian mode.

Register Contents

R_{00}	R_{05}	R_{10}	R_{15} ν	R_{20} Used	R_{25}
R_{01}	R_{06}	R_{11}	R_{16} t	R_{21} Used	R_{26}
R_{02}	R_{07}	R_{12}	R_{17} $\Gamma[(\nu + 2)/2]$	R_{22} Used	R_{27}
R_{03}	R_{08}	R_{13}	R_{18} $\Gamma[(\nu + 1)/2]$	R_{23} Used	R_{28}
R_{04}	R_{09}	R_{14}	R_{19} $\Gamma(\nu/2)$	R_{24}	R_{29}

Examples:

1. Find the area under the t curve with 30 degrees of freedom between t = 0 and t = 1.2. Compare this area with the corresponding area for the standard normal curve.

2. Show that the null hypothesis in the first *t-Statistic Evaluation* example should be rejected ($t \doteq 2.5$, $\nu = 9$).

3. Verify that the null hypothesis in the second *t-Statistic Evaluation* example should be accepted ($t \doteq -0.31$, $\nu = 16$).

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 21		Select Student's t Distribution Program
30	[A]	8.7178291 10	$\nu \rightarrow \Gamma(\nu/2)$
1.2	[E]	.8802348246	P(t)
	[–]	.8802348246	
0	[E]	0.5	P(t)
	[=]	.3802348246	Area
9	[A]	11.6317284	$\nu \rightarrow \Gamma(\nu/2)$
2.5	[E]	.9830690862	P(t)
16	[A]	5040.	$\nu \rightarrow \Gamma(\nu/2)$
.31	[+/–] [E]	.3802810836	P(t)

Summary:

1. This area is found using the same procedure described in the Normal Distribution example. Note that the sizes of these areas are the same, demonstrating the fact that the t curve becomes the standard normal curve for $\nu \geqslant 30$.

2. The hypotheses of this example require a two-tailed test. We should therefore accept H_0 if $0.025 \leqslant P(t) \leqslant 0.975$ when testing at the 95% confidence level. Since P(2.5) for 9 degrees of freedom is not in this range, reject H_0.

3. Again, a two-tailed test is required to complete this example, However, since $P(-0.31)$ for 16 degrees of freedom is such that $0.05 \leqslant P(t) \leqslant 0.95$, you should accept H_0 at the 90% confidence level.

F DISTRIBUTION PROGRAM

The F-Statistic is actually the ratio of two variances. Accordingly, you may use this ratio to compare the variances of normal populations. It may also be used to make inferences about the effect of one random variable on another by studying the population variance as is done in *Analysis of Variance* in Section IV.

By definition, if X and Y are independent random variables having chi-square distribution with ν_1 and ν_2 degrees of freedom respectively, then the F-statistic is

$$F_{\nu_1,\,\nu_2} = \frac{X/\nu_1}{Y/\nu_2}$$

with ν_1 degrees of freedom in the numerator and ν_2 degrees of freedom in the denominator. Again, the shape of an F curve is defined by its number of degrees of freedom. The mean and variance are given by

$$\mu = \nu_2/(\nu_2 - 2) \qquad \text{and} \qquad \sigma^2 = \frac{2\,\nu_2^{\,2}\,(\nu_1 + \nu_2 - 2)}{\nu_1\,(\nu_2 - 2)(\nu_2 - 4)}\,.$$

A random variable with F distribution has no mean for $\nu_2 \leqslant 2$ and no variance for $\nu_2 \leqslant 4$.

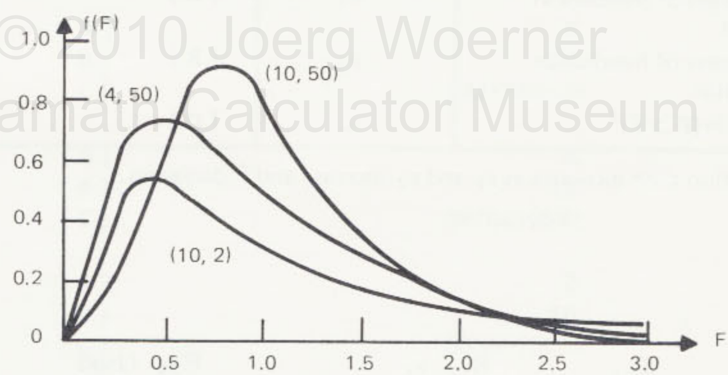


Figure 6.8

The probability density function of a F variable is given by

$$f(F) = \left(\frac{\Gamma[(\nu_1 + \nu_2)/2]\;\nu_1^{(\nu_1/2)}\,\nu_2^{(\nu_2/2)}}{\Gamma(\nu_1/2)\;\Gamma(\nu_2/2)} \right) \left(\frac{F^{(\nu_1/2)-1}}{(\nu_1\,F + \nu_2)^{(\nu_1+\nu_2)/2}} \right).$$

THEORETICAL DISTRIBUTIONS

Once you have entered the degrees of freedom for your variable into the program, the tail area of the appropriate F curve is evaluated as

$$Q(F) = \int_F^{\infty} f(u)du = \Pr(F > F).$$

A series expansion is used to approximate this integral. You may use this information to calculate the following probability.

$$P(F) = 1 - Q(F) = \Pr(F \leqslant F).$$

Note that a reasonable limit for ν_1 and ν_2 is 120.

Solid State Software		TI ©1977		
F DISTRIBUTION				ST-22
ν_1	ν_2	F→Q(F)		

USER INSTRUCTIONS

STEP	PROCEDURE	ENTER	PRESS	DISPLAY
1	Select Program		[2nd] [Pgm] 22	No Change
2	Enter degrees of freedom in numerator	ν_1	[A]	ν_1
3	Enter degrees of freedom in denominator	ν_2	[B]	ν_2
4	Calculate Pr(F > F)[1]	x	[C]	Q(F)

NOTE: 1. Execution time increases as ν_1 and ν_2 increase and F decreases.

Register Contents

R_{00}	R_{05}	R_{10}	R_{15} ν_1	R_{20} Used	R_{25} Used
R_{01}	R_{06}	R_{11}	R_{16} ν_2	R_{21} Used	R_{26} Used
R_{02}	R_{07}	R_{12}	R_{17} Used	R_{22} Used	R_{27} Used
R_{03}	R_{08}	R_{13}	R_{18} Used	R_{23} Used	R_{28}
R_{04}	R_{09}	R_{14}	R_{19} Used	R_{24} Used	R_{29}

Examples:

1. Find Q(F) for $F_{7,20} = 2.5$. Then find P(F) for $F_{20,7} = 1/2.5 = 0.4$. What do you notice about these results?

2. Prove that the null hypothesis in the *One-Way AOV* example should be rejected ($F_{2,24} \doteq 2.63$).

3. Verify the results of the *Two-Way AOV* example ($F_{C(8,16)} \doteq 2.18$, $F_{R(2,16)} \doteq 8.13$).

ENTER	PRESS	DISPLAY	COMMENTS
	[2nd] [Pgm] 22		Select F Distribution Program
7	[A]	7.	ν_1
20	[B]	20.	ν_2
2.5	[C]	.0510167666	Q(F)
20	[A]	20.	ν_1
7	[B]	7.	ν_2
1	[−]		
.4	[C]	.9489832334	Q(F)
	[=]	.0510167666	P(F)
2	[A]	2.	ν_1
24	[B]	24.	ν_2
2.63	[C]	.0927348639	Q(F)
8	[A]	8.	ν_1
16	[B]	16.	ν_2
2.18	[C]	.0878610597	$Q(F_C)$
2	[A]	2.	ν_1
16	[B]	16.	ν_2
8.13	[C]	.0036613813	$Q(F_R)$

Summary:

1. This example illustrates what is known as the reciprocal property of the F distribution. That is,

$$\Pr(F_{\nu_1,\nu_2} \geqslant F) = \Pr(F_{\nu_1,\nu_2} \leqslant 1/F).$$

2. We need to use an upper-tailed test for this example. To test at the 90% confidence level H_0 should be accepted if $P(F) \leqslant 0.90$. This is the same as accepting H_0 when $Q(F) > 0.10$. Since $Q(2.63) < 0.10$ for 2 and 24 degrees of freedom, reject H_0.

3. Since these tests are performed at the 95% confidence level (5% level of significance) you should conclude that there are no row or column effects only when $Q(F) > 0.05$. Since $Q(F_C) > 0.05$ for 8 and 16 degrees of freedom there is no indication of column effects. However, since $Q(F_R) < 0.05$ for 2 and 16 degrees of freedom, there is a strong indication of row effects.

APPENDIX A: PROGRAM REFERENCE DATA

Program Number	Title	No. of Steps	Data Base Used	Data Reg. Used	Flags Used	SBR Levels	Paren. Levels	Calls Pgm.	Special Functions Used	x ≥ t	ABS Address	Angular Model	Print	EE*	Program Number
01	Diagnostic	111		1-6, 9, 10		1	4	2	N/A	X	X	Deg	Output		01
02	Random Number Generator	141		1-6, 9-13		1	3	1	Σ+	CP	X	Rad	In/Out		02
03	Univariate Data	463	Univariate	0, 3-18, 26, 28+	0, 5-7	2	2	4, 7		X	X		Input		03
04	Bivariate Data	282	Bivariate	1-17, 26 28, 29, 31+	7	2	1	3		X	X		Input		04
05	Trivariate Data	276	Trivariate	1-14, 26-29, 32+	7	1	1	4		X	X		Input		05
06	AOV Data	251	AOV	0-30+	0,7	3	1	3, 7	$\overline{X}$ Op 11	X	X		In/Out		06
07	Histogram Data	186	Histogram	0-19 26, 28+	7	1	3	3, 4		X	X		Input		07
08	Means and Moments	154	Univariate	3-5, 7, 8, 11, 17, 19-21		1	2		$\overline{X}$	X			Output		08
09	Histogram Construction	79	Histogram	0-17, 19-21		1	2		$\overline{X}$ INV $\overline{X}$	X	X		Output		09
10	Theoretical Histogram	216	Histogram	1-29		2	3	0, 21		X	X		Output		10
11	Transforms for Univariate Data	49	Univariate	0, 3-18, 26, 28+	0,1, 2,7	3	2	3		X	X		Transformed Input		11
12	Transforms for Bivariate Data	195	Bivariate	1-17, 26, 28, 29, 31+	1,2 3,7	2	3	4		X	X		Transformed Input		12
13	t-Statistic Evaluation	152	Bivariate	1-5, 12, 13 15, 23-27		1	3		$\overline{X}$		X		Output		13
14	Contingency Table	256		1-59		2	2	21		X	X		Input	X	14
15	1-Way AOV	111	AOV	1, 2, 8-17, 28		0	2						Output		15
16	2-Way AOV	178	AOV	1-3, 6-27, 29		1	2			X	X		Output		16
17	Rank-Sum Tests	339	Bivariate	3, 15, 23-27, 30+		1	3			CP	X		Output		17
18	Multiple Linear Regression	268	Trivariate	1-7, 9, 14-23	7	0	3				X		In/Out		18

APPENDIX A: PROGRAM REFERENCE DATA

Program Number	Title	No. of Steps	Data Base Used	Data Reg. Used	Flags Used	SBR Levels	Paren. Levels	Calls Pgm.	Special Functions Used	x ≥ t	ABS Address	Angular Model	Print	EE*	Program Number
19	Normal Distribution	159		25, 26	1	2	2			CP	X				19
20	Binomial Distribution	152		21-26		1	2			CP	X				20
21	Chi-Square and Student's t Dist.	381		15-23		1	2			X	X	Rad			21
22	F Distribution	400		15-27	1,2	1	1			X	X	Rad			22
	Pointers and Counters	201													

*May not be run in engineering mode.

ONE-YEAR LIMITED WARRANTY FOR CALCULATOR AND/OR LIBRARY MODULE

THIS TEXAS INSTRUMENTS ELECTRONIC CALCULATOR WARRANTY EXTENDS TO THE ORIGINAL CONSUMER PURCHASER OF THE CALCULATOR OR MODULE.

WARRANTY DURATION: This calculator and/or module is warranted to the original consumer purchaser for a period of one year from the original purchase date.

WARRANTY COVERAGE: This calculator and/or module is warranted against defective materials or workmanship. **THIS WARRANTY IS VOID IF THE CALCULATOR OR MODULE HAS BEEN DAMAGED BY ACCIDENT, UNREASONABLE USE, NEGLECT, IMPROPER SERVICE OR OTHER CAUSES NOT ARISING OUT OF DEFECTS IN MATERIAL OR WORKMANSHIP.**

WARRANTY DISCLAIMERS: ANY IMPLIED WARRANTIES ARISING OUT OF THIS SALE, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, ARE LIMITED IN DURATION TO THE ABOVE ONE YEAR PERIOD. TEXAS INSTRUMENTS SHALL NOT BE LIABLE FOR LOSS OF USE OF THE CALCULATOR OR OTHER INCIDENTAL OR CONSEQUENTIAL COSTS, EXPENSES, OR DAMAGES INCURRED BY THE CONSUMER OR ANY OTHER USER.

Some states do not allow the exclusion or limitation of implied warranties or consequential damages, so the above limitations or exclusions may not apply to you.

LEGAL REMEDIES: This warranty gives you specific legal rights, and you may also have other rights that vary from state to state.

WARRANTY PERFORMANCE: During the above one year warranty period, your TI calculator or module will either be repaired or replaced with a reconditioned comparable model (at TI's option) when the calculator or module is returned postage prepaid to a Texas Instruments Service Facility listed below.

In the event of replacement with a reconditioned model, the replacement calculator will continue the warranty of the original unit or six months, whichever is longer. Other than the postage requirement, no charge will be made for such repair or replacement of in-warranty calculators or modules unless one of the alternative remedies is chosen.

TI strongly recommends that you insure the product for value, prior to mailing.

TEXAS INSTRUMENTS CONSUMER SERVICE FACILITIES

Texas Instruments Service Facility
P.O. Box 2500
Lubbock, Texas 79408

Texas Instruments Service Facility
41 Shelley Road
Richmond Hill, Ontario, Canada

Consumers in California and Oregon may contact the following Texas Instruments offices for additional assistance or information.

Texas Instruments Consumer Service
831 South Douglas Street
El Segundo, California 90245
(213) 973-1803

Texas Instruments Consumer Service
10700 Southwest Beaverton Highway
Park Plaza West, Suite 565
Beaverton, Oregon 97005 (503) 643-6758

IMPORTANT NOTICE REGARDING PROGRAMS AND BOOK MATERIALS

TEXAS INSTRUMENTS MAKES NO WARRANTY, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR PARTICULAR PURPOSE, REGARDING THESE PROGRAMS OR BOOK MATERIALS OR ANY PROGRAMS DERIVED THEREFROM AND MAKES SUCH MATERIALS AVAILABLE SOLELY ON AN "AS IS" BASIS.

IN NO EVENT SHALL TEXAS INSTRUMENTS BE LIABLE TO ANYONE FOR SPECIAL, COLLATERAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING OUT OF THE PURCHASE OR USE OF THESE BOOK MATERIALS OR PROGRAMS AND THE SOLE AND EXCLUSIVE LIABILITY OF TEXAS INSTRUMENTS, REGARDLESS OF THE FORM OF ACTION, SHALL NOT EXCEED $25.00. MOREOVER, TEXAS INSTRUMENTS SHALL NOT BE LIABLE FOR ANY CLAIM OF ANY KIND WHATSOEVER AGAINST THE USER OF THESE PROGRAMS OR BOOK MATERIALS BY ANY OTHER PARTY.

TEXAS INSTRUMENTS
INCORPORATED
DALLAS, TEXAS

Printed in U.S.A.

1014984-22