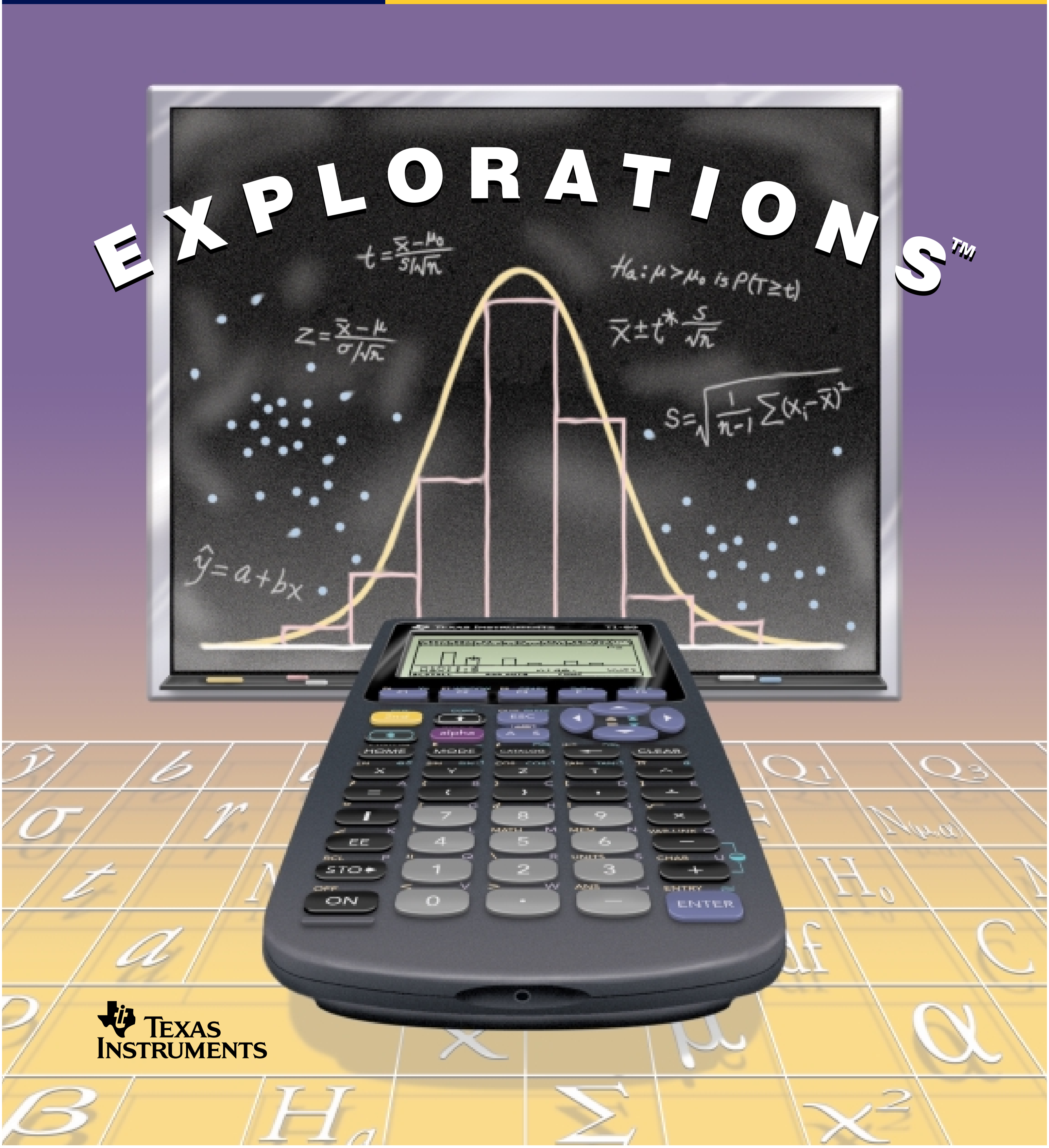
TI-89

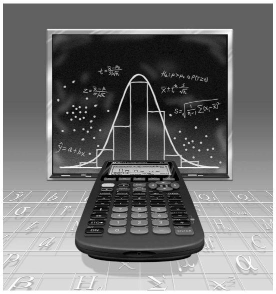
Statistics with List Editor Application Software

Advanced Placement Statistics with the TI-89

Larry Morgan
Roseanne Hofmann
Charles Hofmann



TEXAS INSTRUMENTS



Extensions for Advanced Placement Statistics with the TI-89

Larry Morgan
Montgomery County Community College
Blue Bell, PA

Roseanne Hofmann
Montgomery County Community College
Blue Bell, PA

Charles Hofmann
La Salle University
Philadelphia, PA

Texas Instruments Incorporated
7800 Banner Drive, M/S 3918
Dallas, TX 75251

Attention: Manager, Business Services

Printed in the United States of America.

ISBN: 1-886309-37-X
89/ESB/1L1/2A

We invite your comments and suggestions about this book. Call us at **1-800-TI-CARES** or send e-mail to **ti-cares@ti.com**. Also, you can call or send e-mail to request information about other current and future publications from Texas Instruments.

Visit the TI World Wide Web home page. The web address is: **http://education.ti.com**

Contents

Preface

Advanced Placement Statistics with the TI-89 book is intended to facilitate the use of the TI-89 graphing calculator with most introductory statistics texts. It includes the activities for the topics and four themes of the *Advanced Placement Program Course Description for Statistics*. A fifth theme, *Extensions*, gives additional topics that could be used in a first or second course in statistics. These extension topics include Analysis of Variance, Multiple Regression, Forecasting, and Nonparametric methods.

This book consists of 50 topics. Students should be familiar with Topic 1 (TI-89 Prerequisites), and Topics 2, 3, and 4 (graphical displays) before beginning the remaining topics. The order in which you cover the other topics will depend on your course and textbook.

The TI-89's powerful statistical features allow you to concentrate on concepts rather than on mechanical computations. This book demonstrates, with examples, how to use the features of the TI-89 to solve problems or to clarify important ideas of statistics. For example, Topic 22 shows how to calculate a confidence interval with the TI-89 and uses a simulation to explain the meaning of the term *confidence interval*.

Sometimes there are alternative approaches for getting a result on the TI-89. We do not try to show all alternatives, but we select a procedure that is most useful for the particular activity.

The table on the back cover provides quick reference to the capabilities of the TI-89 and the topics within this book.

Finally, a word of thanks to our students and colleagues who have been supportive of using calculators in the classroom. Thanks to the reviewers of this book for their helpful comments and suggestions. Thanks to our colleagues at Texas Instruments who helped make this book possible.

— *Larry Morgan*

— *Roseanne Hofmann*

— *Charles Hofmann*

The Operating System and Application Versions used for this book

- TI-89 Hardware Version 2.00
- Advanced Mathematics Software Operating System; v2.05; July 5, 2000
- Statistics with List Editor application; v1.01; June 16, 2000

About the Authors

LARRY MORGAN began his professional career as an engineer collecting and analyzing data at Arnold Air Force Station's test facilities in Tennessee. Working with the Peace Corps led to a teaching career that included four years in Ghana, Africa and two years in St. Louis, Missouri. He has been teaching at Montgomery County Community College in Blue Bell, Pennsylvania since 1970.

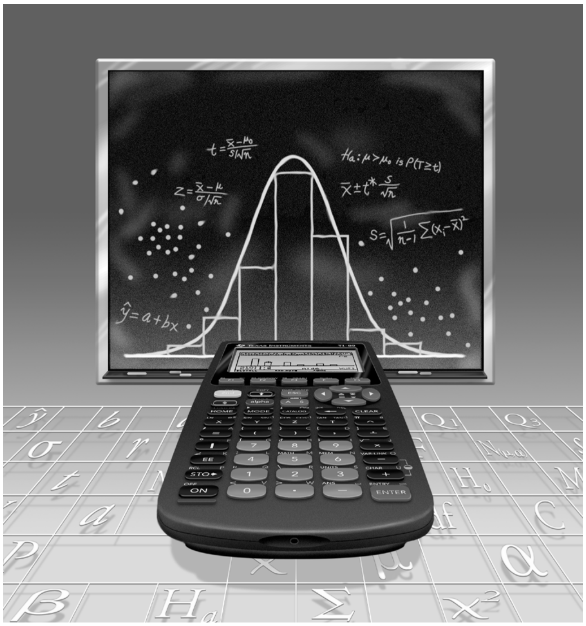
Working with computer statistical packages and his own programs led to "the solution" when the TI-80 series calculators became his ideal computer for his students. He wrote a package of programs that covered the procedures required for a two-semester introductory statistics course and shared them at conferences and workshops. He has published TI-82 supplements to three popular statistics texts, written articles for *Teaching Statistics* and *The Mathematics Teacher*, and authored the *Statistics Handbook for the TI-83* (which is also appropriate for the TI-83 Plus).

ROSEANNE HOFMANN is a professor of mathematics at Montgomery County Community College. She has devoted teaching and professional activities to developing the use of technology in the mathematics classroom to provide visualization, reduce computational burdens, and increase time for problem solving. Professional activities have focused on efforts to teach mathematics teachers how to integrate technology for maximum benefit in the classroom. She has developed workshops and software, and has co-authored two textbooks and two manuals. Dr. Hofmann was awarded several grants, including three NSF grants. She was the recipient of five faculty recognition awards and the Teaching Excellence Award 2000 from Pennsylvania State Mathematics Association of Two Year Colleges (PSMATYC). This semester, Dr. Hofmann is offering the first online statistics course at Montgomery County Community College.

CHARLES E. HOFMANN's teaching responsibilities at La Salle University have included undergraduate and graduate computer science, mathematics, and mathematics education courses. In addition to over 35 years of classroom teaching, Dr. Hofmann has also served as department chair and director of the graduate program in computer information science. Recently, his responsibilities as a consultant have taken him to the larger urban school districts of the United States, where he assists in the design of teacher training programs that implement systemic change in mathematics and science. His mathematics expertise includes the use of calculator and computer technology to enhance the teaching/learning paradigm in the mathematics classroom. To this end, he has given over 200 lectures and workshops in the last several years to in-service and pre-service teachers. He co-authored *Time*, *Value*, *Money: Applications on the TI-83* with Roseanne S. Hofmann.

EXPLORATIONS

Part 5: Extensions



EXPLORATIONS

Chapter 14

More Probability Distributions

You will discover how to do probability calculations for the Poisson distribution with the built-in functions Poisson Pdf and Poisson Cdf. While the hypergeometric and exponential distributions do not have their own built-in functions, these are easily created with other features of the TI-89. Examples for these distributions will also be given.

Topic 34—Hypergeometric Distribution

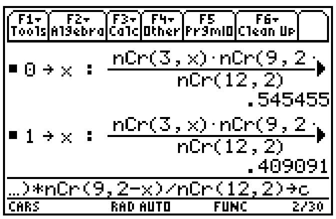
Example: Use folder **CARS** for this topic. In a bag of twelve apples, three have worms. If you randomly selected two apples (without replacement):

1. What is the probability that neither apple will have worms?

   a. Take the number of ways you can pick none of the wormy apples and two good apples, **nCr(3,0)** ∗ **nCr(9,2)** and divide by the number of ways you can pick any two apples, **nCr(12,2)** for **p(0) = .545455**.

   b. From the Home screen, type and paste: **0→x: nCr(3,x)** ∗ **nCr(9,2-x)/nCr(12,2)→a**. (**nCr** is available from [CATALOG].)

   c. Press [ENTER]. If the output is 6/11, press [◆] [ENTER] to convert to decimal form (first computation in screen 1).

(1) 

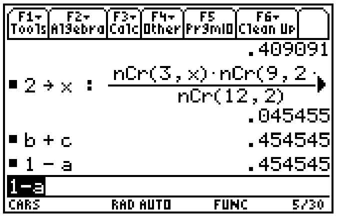
***Note:** From the Home screen with* nCr *from* [CATALOG].

2. What is the probability that at least one apple has a worm in it?

   a. Repeat steps 1b and c on the previous page, except change the zero to 1 and store the result in **b** so **p(0) = .545455** and **p(1) = .409091** (top of screen 1).

   b. Repeat steps 1b and c again, and change the **1** to **2** and store the result, **p(2) = .045455** in **c** (top of screen 2).

   c. Observe that **p(1) + p(2) = 0.454545** by entering **b + c** on the status line. Also observe the complement **1 - p(0) = 0.454545** by entering **1 - a**.

(2)


Simulation

From the Home screen:

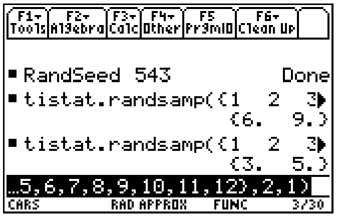
1. Set **RandSeed 543** and then press [ENTER] (top of screen 3).

2. Press [CATALOG], [F3] **Flash Apps** for **randSamp**, and then complete the input line with **tistat.randsamp({1,2,3,4,5,6,7,8,9,10,11,12},2,1)**, indicating a random sample of integers from 1 to 12. The **2** indicates that two values are to be selected, and the final **1** indicates sampling without replacement.

3. Press [ENTER] for the first two apples **{6, 9}**, which are worm free (middle of screen 3). Assume that apples **{1, 2, 3}** have worms, while apples **4** through **12** do not.

4. Press [ENTER] for another selection of 2 from the 12 for **{3, 5}** or the first wormy apple, the second not (bottom of screen 3).

5. Repeat this sequence of steps 18 more times for the results in the table. These results are similar to what you would expect in the long run with many more tries — not with just 20.

(3)


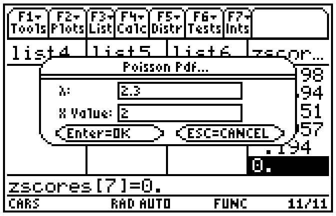
Simulation	Theory
p(0) ≈ 11/20 = 0.55	0.545
p(1) ≈ 8/20 = 0.40	0.409
p(2) ≈ 1/20 = 0.05	0.045

Topic 35—Poisson Distribution

Example: Assume that the number of accidents per month at a given intersection has a Poisson distribution and the mean number of accidents is 2.3 per month.

1. What is the probability of having exactly two accidents for a given month?

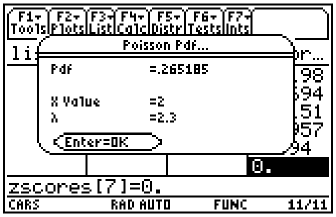
   a. In the Stats/List Editor, press [F5] **Distr**, **D:Poisson Pdf**, with λ: **2.3** and X Value: **2** (screen 4).

(4) 

   b. Press [ENTER] to display screen 5 with Pdf **= .265185**. The Poisson probability formula is:

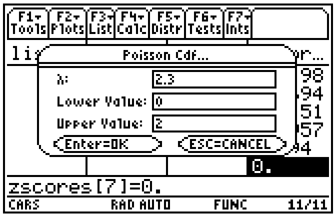
$$p(x) = \frac{\lambda^x * e^{-\lambda}}{x!}$$

   Substituting the given values λ = **2.3** and X Value = **2**, you get $p(2) = (2.3)^2 * e^{-2.3}/2!) = .265185$.

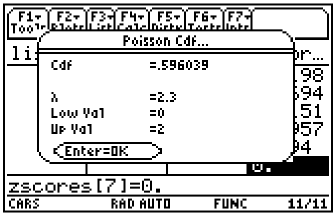
(5) 

2. What is the probability of having at most two accidents for a given month, or **p(0) + p(1) + p(2)**?

   a. Press [F5] **Distr**, **E:Poisson Cdf**, with λ: **2.3**, Lower Value: **0**, and Upper Value: **2** (screen 6).

(6) 

   b. Press [ENTER] to display screen 7 with Cdf **= .596039** (p(0) + p(1) + p(2) = .596039).

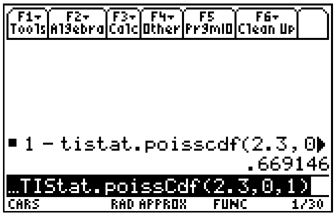
(7) 

3. What is the probability of having at least two accidents for a given month, or
   p(2) + p(3) + p(4) + … = 1- (p(0) + p(1))?

   a. You could calculate **p(0) + p(1)** as done previously and subtract this from 1.

   b. From the Home screen, type and paste
      1 - tistat.poissCdf (2.3,0,1).

   c. Press [ENTER] to display screen 8 with the results **.669146**, or (p(2) + p(3) + p(4) + . . .).

(8) 

Note: **tistat.poissCdf** *(*λ, Lower Value, Upper Value*).*

Topic 36—Exponential Distribution

Example: Assume that the shelf life of cake is exponentially distributed with a mean of μ = 4 days. What proportion of the cakes put on the shelf today would you expect to still be fresh five days from now?

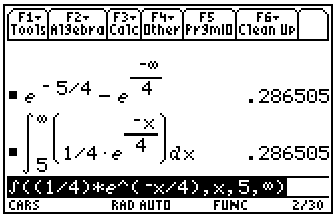
Exponential Pdf: f(x) = (1/μ) ∗ e^(-x/μ)

$p(a \le x \le b) = e$^(-a/μ) - e^(-b/μ) or

$$\int_a^b \left(\frac{1}{\mu}\right) e^{\frac{-x}{\mu}}\, dx$$

1. From the Home screen, calculate the probability using a = 5 days, b = ∞, and μ = 4 days.
   Type **e^(-5/4) - e^(-∞/4)**. The result is .286505 (screen 9).

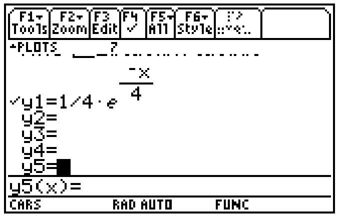
2. Another method of calculating this probability is by using the integral formula given above.
   Type [2nd] [∫] **1/4 ∗ e^(-x/4),x,5,∞)** (screen 9).

(9) 

Plotting and Shading

Enter the **pdf** in the y= editor.

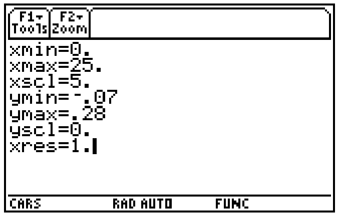
1. Press [◆] [Y=].
2. Highlight **y1**, and press [CLEAR] to clear the previous data if necessary.
3. Enter **1/4** [◆] *e* ˣ **(-x/4)** and press [ENTER] to display screen 10.



(10)

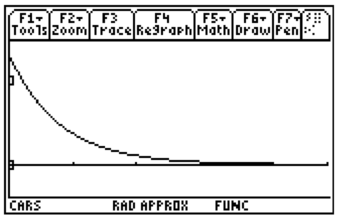
4. Set up the window using [◆] [WINDOW] with the following entries:
   - **xmin = 0**
   - **xmax = 25**
   - **xscl = 5**
   - **ymin = -.07**
   - **ymax = .28**
   - **yscl = 0**
   - **xres = 1**

   (See screen 11.)

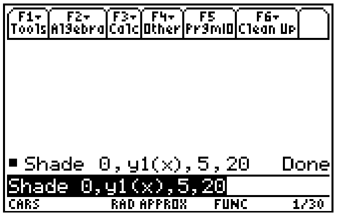


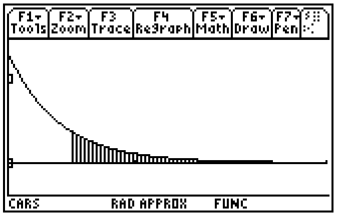
(11)

5. Press [◆] [GRAPH] to display screen 12 without the shading.

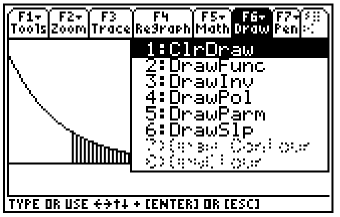

(12)

6. From the Home screen, press [CATALOG], type **Shade 0,y1(x),5,20**, and press [ENTER] (screens 13 and 14).

(13) 

(14) 

7. Press [2nd] [F6] **Draw**, **1:ClrDraw** to remove the shading (screen 15).

(15) 

EXPLORATIONS

Chapter 15

Inference for the Standard Deviation and Variance of Normal Populations for One and Two Samples (The F Distribution)

The first part of Topic 37 shows how to calculate a confidence interval for the standard deviation (σ) and variance (σ^2) of a normal population given a sample standard deviation (s). The remainder of Topic 37 shows how to test a hypothesis with such a sample. The Inverse Chi-square function is used to calculate critical values, while Chi-square Cdf is used to calculate *p*-values.

Topic 38 uses the built-in 2-SampFTest to test the null hypothesis of equal variance of two normal populations.

Topic 37—Inference for σ and σ^2

Estimating the Population Standard Deviation σ of a Normal Random Variable

Example: A random sample of size 4 from a population that is normally distributed is as follows:

{54.54, 43.44, 54.11, 46.88} → list1

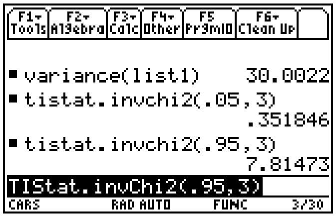
Find the 90% confidence interval for the population standard deviation.

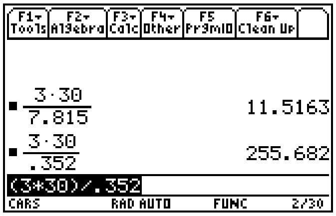
The confidence interval is based on the sample variance (s^2) and the χ^2 distribution, since the sampling distribution of **$(n - 1)\, s^2/n^2$** is Chi-square distributed with n - 1 degrees of freedom for samples from normal populations.

From the Home screen:

1. Enter the list of data from the previous page.
2. Press [CATALOG] for **variance(list1) = 30.0022** (screen 1).

   You need a left and right critical value with .05 in each tail with n - 1 = 4 - 1 = 3 degrees of freedom.
3. Press [CATALOG], [F3] **Flash Apps**, **invChi2(0.05,3)** to get the χ_T^2 value of **0.351846**, and **invChi2(0.95,3)** to get the χ_R^2 value of **7.81473** (screen 1).
4. Confidence intervals for variance = 11.52 < σ^2 < 255.68, or with Lower Value: $(n-1)S_X^2 \div x_R^2 = 3 * 30 \div 7.815 = 11.52$ and Upper Value: $(n-1)S_X^2 \div x_L^2 = 3 * 30 \div 0.352 = 255.68$ (screen 2).
5. Confidence intervals for standard deviation: Since $\sigma = \sqrt{\text{variance}}$, the lower value for $\sigma = \sqrt{11.52} = 3.39$. The upper value for $\sigma = \sqrt{255.68} = 15.99$. So, $3.39 < \sigma < 15.99$.

(1)


(2)


Testing Hypothesis of a Normal Population Standard Deviation, σ

Example: A battery charger gives a charge with a mean of 50 hours and a standard deviation of 10 hours. It is thought that an adjustment to the charger reduced the variability of the length of the charge. To test the standard deviation of the length of the charge in the battery, the following random sample **{54.54, 43.44, 54.11, 46.88}** was taken and stored in **list1**.

__Note:__ If s^2 *is close to* σ^2, $(n - 1)s^2/\sigma^2$ *will be close in value to the degrees of freedom* n - 1 (4 – 1 = 3 *in this case).*

This sample was generated using **randNorm** with **σ = 10** and **σ^2 = 100**.

These values (σ = 10 and σ^2 = 100) do lie in the intervals calculated. The intervals are quite wide, but they are based on a very small sample.

Test: H_0: $\sigma = 10$

H_a: $\sigma < 10$

with significance level $\alpha = 0.05$.

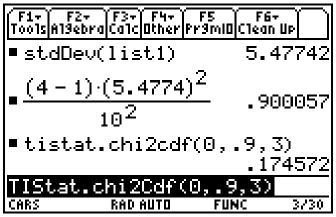
The length of the charge is normally distributed, so the sampling distribution of $(n - 1)s^2/\sigma^2$ is Chi-square distributed.

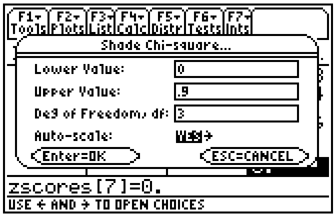
From the Home screen:

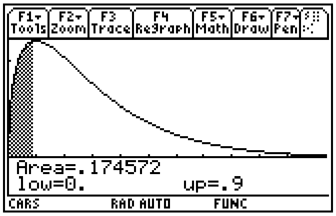
1. Press CATALOG to calculate the standard deviation s_x with **stdDev(list1) = 5.4774** (screen 3).
2. Calculate the Chi-square statistic $(n - 1)s^2/\sigma_0^2$ = **(4 – 1) 5.47742/10^2** = 0.90 (screen 3).
3. Find the p-value, or area, in the left tail of the Chi-square curve with Deg of Freedom, df: **3** by typing or pasting **tistat.Chi2Cdf(0,0.90,3)** = 0.174572 (screen 3).

   (In the Stats/List Editor, the Chi-square Cdf could have been used with F5 **Distr**, **1:Shade**, **3:Shade Chi-square**, with Lower Value: **0**, Upper Value: **0.90**, Deg of Freedom, df: **3**, and Auto-scale: **YES** (screens 4 and 5)).

   Since 0.174572 > .05, you do not reject the null hypothesis. There is insufficient evidence from your sample of size 4 to conclude that the variability in the length of the charge has decreased. The above sample was, in fact, randomly generated from a normal distribution with mean 50 and standard deviation 10.

(3)


(4)


(5)


Note: *From screen 1, the critical value = 0.35. Since the test statistic* 0.90 > .35*, you are not in the critical region on the left so you fail to reject the null hypothesis.*

Topic 38—Testing Hypothesis of Standard Deviations of Two Normal Populations (2-Sample F Test)

Example: To test if there is any difference between the variability of the diastolic blood pressure in a population of men and in a population of women, a random sample from each population was taken and recorded below. Assuming that the populations of diastolic blood pressures are normally distributed, test:

H_0: $\sigma_1 = \sigma_2$ or $(\sigma_1^2 \div \sigma_2^2 = 1)$

H_a: $\sigma_1 \neq \sigma_2$ or $(\sigma_1^2 \div \sigma_2^2 \neq 1)$

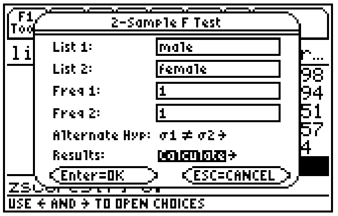
1. From the Home screen, enter the following statements:

   {76, 74, 70, 80, 68, 90, 70, 72, 76, 68, 96} → Male

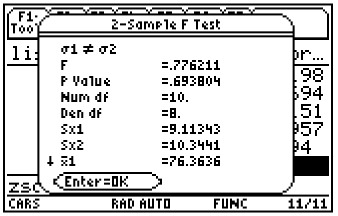
   {70, 82, 90, 68, 60, 62, 80, 74, 62} → Female

	Mean	s_x	n
Male	76.36	9.11	11
Female	72	10.34	9

2. From the Stats/List Editor, press [2nd] [F6] **Tests**, **9:2-SampFTest**, and use Data Input Method: **Data**, with List 1: **male**, List 2: **female**, Freq 1: **1**, Freq 2: **1**, Alternate Hyp: $\sigma_1 \neq \sigma_2$, and Results: **Calculate** (screen 6).

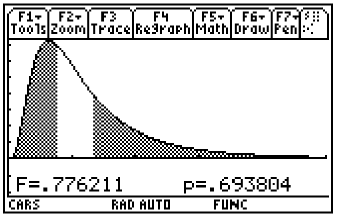
(6)


3. Press [ENTER] for the results shown in screen 7.

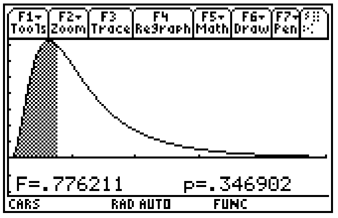
(7)


4. Repeat step 2, except select **Draw** instead of **Calculate** (screen 8).

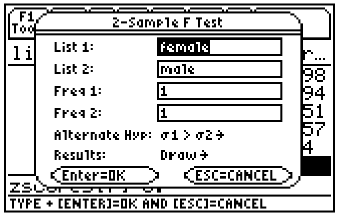
   With a p-value of **0.693804** > .05, you cannot reject the null hypothesis. The ratio of variances is not significantly different from 1, so there is no significant difference in the variability of male and female diastolic blood pressure.

(8)


If there were a one-tailed test ($\sigma_1 < \sigma_2$) with $F = 9.11^2 \div 10.34^2 = 0.776211$, you would see that the p-value is half the value shown in screen 8, or a p-value **= 0.346902**, as shown in screen 9. To see this result, repeat step 2 with Alternate Hyp: $\sigma_1 < \sigma_2$.

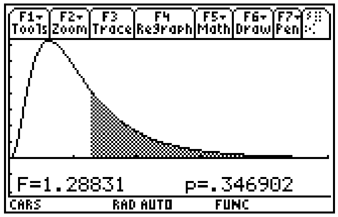
(9)


Before the ready availability of computers or the TI-89, textbooks suggested that the largest sample variance be put in the numerator to force everyone to the right tail so it would be easier to use tables in the back of these textbooks. Notice that in screen 10 (List1: **female** with the largest standard deviation entered first), and screen 11 the sample *p*-value (**0.346902**) is the same as in screen 9. It is no longer necessary to force the larger value into the numerator. If the larger value is in the numerator, the *F* distribution ratio will be larger than 1; if it is not, the ratio will be less than 1.



(10)

Notice that **1/1.28831** from screen 11 equals **0.776211** from screen 9.



(11)

EXPLORATIONS

Chapter 16

Analysis of Variance (ANOVA), One- and Two-Way

Topic 39 uses [2nd] [F6] **Tests**, **C:ANOVA** with lists of data gathered from a completely randomized design. How to use summary statistics for input is also shown. Topics 40 and 41 use [2nd] [F6] **Tests**, **D:ANOVA 2-Way** to analyze a randomized block design and a two-factor factorial experiment.

The idea for the examples used in these topics is from McClave/Benson/Sincich, *Statistics For Business and Economics*, 7th edition, Prentice Hall, 1998. Reprinted with permission of the publisher.

Topic 39—Completely Randomized Designs

Example: To compare the distance traveled by three different brands of golf balls when struck by a driver, you use a completely randomized design. A robotic golfer, using a driver, is set up to hit a random sample of 23 balls (eight each for Brand A and B, but only seven for Brand C) in a random sequence. The distance is recorded for each hit, and the results are shown in the table, organized by brand. Since a robotic golfer is being used and is hitting a random sample of 23 balls, you can assume normality of each sample.

(List) Brand	(branda) A	(brandb) B	(brandc) C
Distance	264.3	262.9	241.9
	258.6	259.9	238.6
	266.4	264.7	244.9
	256.5	254	236.2
	182.7	191.2	167.3
	181	189	165.9
	177.6	185.5	162.4
	187.3	192.1	
Mean	221.8	224.9	208.2
stdDev	42.58	38.08	40.31
n	8	8	7

Note: *The means and standard deviations can all be obtained at one time with* 1Var Stats *from the Home screen, as shown in Topic 9, screens 6 and 7.*

The purpose of an ANOVA test is to determine the existence (or nonexistence) of a statistically significant difference among the three group means, in this case, the mean distance traveled by the three brands of balls. The two assumptions that you must make are (1) each sample comes from a normal distribution, and (2) the standard deviation of each sample is close (42.58, 38.08, and 40.31).

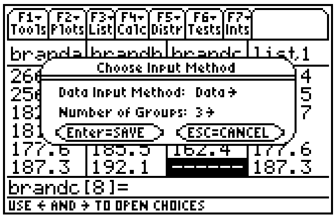
Test the hypothesis H_0: $\mu_A = \mu_B = \mu_C$ with significance level $\alpha = 0.05$ against the alternative that at least one mean differs from the others.

Create lists **branda**, **brandb**, and **brandc** using the data in the table.

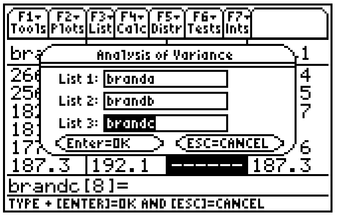
From the Stats/List Editor:

1. Press [2nd] [F6] **Tests**, **C:ANOVA** and use Data Input Method: **Data** and Number of Groups: **3** (screen 1).

(1)



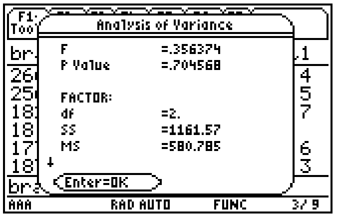
2. Press [ENTER] to display screen 2, with List 1: **branda**, List 2: **brandb**, and List 3: **brandc**, typed or pasted from [2nd] [VAR-LINK].

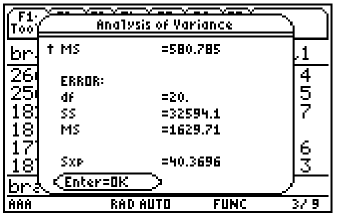
(2) 

Tip: *After typing or pasting* **List 1 branda**, *press* [ENTER] *so* **branda** *is highlighted. Press* [◆] [COPY], *then* [◆] [PASTE] *for* **List 2** *and* **List 3**, *with the* **a** *easily changed to* **b** *or* **c**.

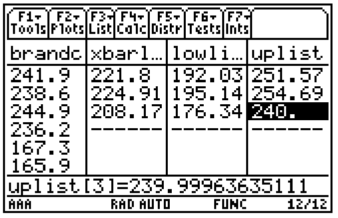
3. Press [ENTER] to display screen 3 with the ANOVA table output. Press ⊝ several times to view the remaining output (screen 4)

$S_{xp} = \sqrt{MSE} = \sqrt{1629.71} = 40.3696.$

(3) 

(4) 

4. Press [ENTER] again to observe the means (xbarlist) and the lower and upper values of 95% confidence intervals for these means (screen 5).

(5) 

Here is the computation for the first 95% confidence interval with $\bar{x}$ = **221.8**, t value = **2.08596**, s_{xp} = **40.3696**, and n = **8**.

$$221.8 \pm 2.08596 * \frac{40.3696}{\sqrt{8}} = 221.8 \pm 29.7724 \rightarrow 192.03 \text{ to}$$

251.57. 2.08596 can be calculated by typing **tistat.inv_t(1 - .05/2, 20)** with df error = 20.

All three of the intervals (192.03, 251.57), (195.14, 254.69), and (176.34, 240) overlap. For example 200 is in each interval, so this supports the *p*-value = **.704568** > .05 conclusion that the data show no significant difference between the mean distance traveled by the three brands of balls. You do not reject the null hypothesis.

Multiple Comparison Procedure

Since you do not reject the null hypothesis, there is no significant difference between any of the means. A multiple comparison procedure is not needed or appropriate. Topic 40 gives an example of the Bonferroni comparison procedure and relates it back to the above example. The confidence intervals above are also a multiple comparison procedure. If they did not all overlap that would suggest how the means may differ.

Data Input Method: Stats

You can repeat the previous hypothesis test comparing the three brands of golf balls by using the Stats option for data input. This can be done only if you already know the sample size, mean, and standard deviation of all samples.

From the Stats/List Editor:

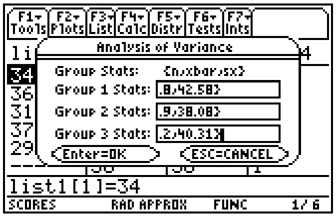
1. Press [2nd] [F6] **Tests**, **C:ANOVA**, and use Data Input Method: **Stats** and Number of Groups: **3**.
2. Enter the data from the last three rows of the given table, using the format displayed in screen 6.

   Group Stats: **{*n*, xbar, s_x}**

   Group 1 Stats: **{8, 221.8, 42.58}**

   Group 2 Stats: **{8, 224.9, 38.08}**

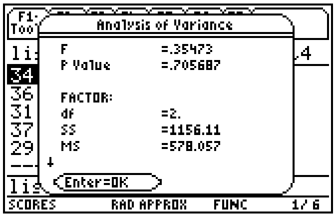
   Group 3 Stats: **{7, 208.2, 40.31}**



(6)

Note: *All the data cannot be seen because of scrolling.*

The output in screen 7 is similar to the output in screen 3 except for rounding differences of the input.



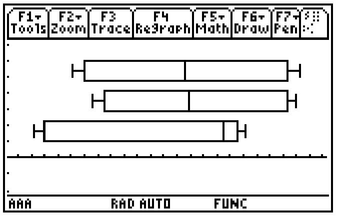
(7)

Assumptions

The assumptions of sampling from normal populations with equal variance are not so restrictive as long as there are no long tails or outliers for the first requirement, and if the sample sizes are the same for the second requirement of equal variance.

The second requirement is easily met since the sample standard deviations are very close (42.58, 38.08, and 40.31) and the sample sizes are nearly the same.

Boxplots of the three lists in screen 8 show the first requirement seems to be met for Brand A and Brand B, but Brand C at the bottom does not look as symmetric and could be skewed to the left. You will discuss this more in the last section of Topic 41. The boxplots do confirm that there are no significant differences in the means with their overlap. Also, note the short tails.

(8)


Topic 40—Randomized Block Design

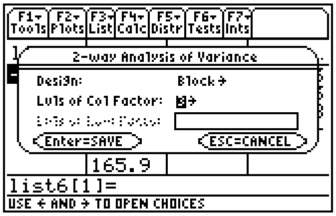
Example: Suppose eight golfers are randomly selected and each golfer hits three balls, one of each brand, in a random sequence. The distance is measured and recorded as shown in the table. Each row is the data for one of the golfers. Test the null hypothesis H_0: $\mu_A = \mu_B = \mu_C$ against the alternate hypothesis that at least one mean differs from the others.

Note: *The data is the same as in Topic 39, but an eighth value of* 172.5 *was added to Brand C. Each list must be of equal length or a "dimension mismatch" error will result. The assumptions are the same as in Topic 39 and will be discussed at the end of Topic 41.*

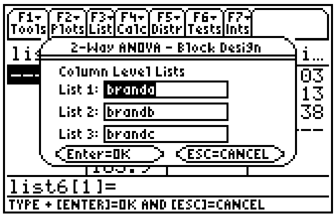
Golfer (Block)	(Column Factor)				
(List) Brand	(branda) A	(brandb) B	(brandc) C	Mean	stdDev
Distance	264.3	262.9	241.9	256.4	12.5
	258.6	259.9	238.6	252.4	11.9
	266.4	264.7	244.9	258.7	12.0
	256.5	254	236.2	248.9	11.1
	182.7	191.2	167.3	180.4	12.1
	181	189	165.9	178.6	11.7
	177.6	185.5	162.4	175.2	11.7
	187.3	192.1	*172.5	184.0	10.2
Mean	221.8	224.9	203.7		
stdDev	42.58	38.08	39.4		
n	8	8	8		

* denotes new value

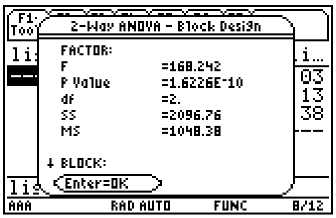
1. Press [2nd] [F6] **Tests**, **D:ANOVA 2-Way**, with Design: **Block**, and Lvls of Col Factor: **3** for the three brands (screen 9).

(9)


2. Press [ENTER] after entering the three lists, **branda**, **brandb**, and **brandc** (screen 10).

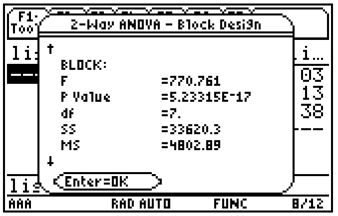
(10)


3. Press [ENTER] to display screen 11.

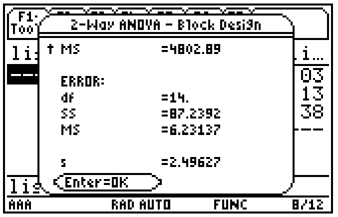
(11)


The factor *p*-value = **1.6E-10 = .00000000016** < .05 leads you to reject the null hypothesis and conclude that the mean distances are not all the same for the three brands of balls.

4. Scroll down using ⊙ to see the additional two screens of output (screens 12 and 13).



(12)



(13)

Bonferroni Multiple Comparison Procedure

You will use the Bonferroni Procedure to see which means differ.

Note: *There are many multiple comparison procedures. Only one procedure is discussed here.*

1. The mean driving distance for each brand is given below:

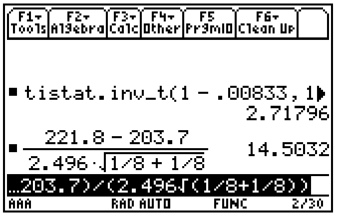
Brand	**A**	**B**	**C**
Mean	221.8	224.9	203.7

2. There are **nCr(3,2) = 3** ways of picking pairs. These are **CA**, **CB**, and **AB**.

3. To do multiple t tests such as:

   H_0: $\mu_C = \mu_A$, H_0: $\mu_C = \mu_B$, and H_0: $\mu_A = \mu_B$, and hold the overall experimental significance level to α = 0.05, you will need to use a significance level = 0.05/3 for each pairwise comparison. Because you are doing a two-tailed test, you must divide by 2 for 0.05/6 = 0.00833 in each tail.

4. The critical t value can be found on the Home screen by typing **tistat.inv_t(1 -.00833, 14) = 2.71796 ≈ 2.72** (screen 14). The **14** is the degrees of freedom for error from screen 13.



(14)

5. From the Home screen, calculate for each comparison.

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with s = 2.49627 in screen 13.

For **AC** (as in screen 14):

$$t = (221.8 - 203.7) \div \left(2.49627 * \sqrt{(1/8 + 1/8)}\right) = 14.503$$

This value is greater than the critical t value of 2.72, so you must reject H_0 and conclude that $\mu_A \neq \mu_C$.

For **BC** (as in screen 15):

$$t = (224.9 - 203.7) \div \left(2.49627 * \sqrt{(1/8 + 1/8)}\right) = 16.987$$

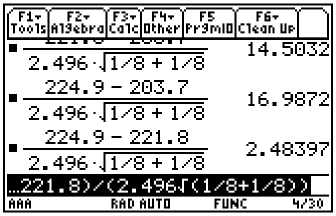
This value is also greater than the critical t value of 2.72, so you must reject H_0 and conclude that $\mu_B \neq \mu_C$.

For **AB** (as in screen 15):

$$t = (224.9 - 221.8) \div \left(2.49627 * \sqrt{(1/8 + 1/8)}\right) = 2.484$$

This value is less than the critical t value of 2.72, so you fail to reject H_0 and conclude that $\mu_A = \mu_B$.

(15)



Notice that Brand C has a significantly smaller mean distance than either Brand A or Brand B, but Brands A and B do not have significantly different means. You show this with a line under A and B, but not C: $C\ \underline{AB}$.

Bonferroni for Completely Randomized Design

In Topic 39 you did not do the multiple comparisons procedure with the example because there was no significant difference between any of the means. If you could have rejected the null hypothesis, then you would have done the Bonferroni Procedure as it was done previously, but with $s = s_{xp} = 40.3696$ as in Topic 39, screen 4 with error df = 20 (which leads to a critical t value of 2.61277). The t statistic between the largest difference between Brand B and Brand C is

$$t = (224.9 - 208.2) \div \left(40.3696 * \sqrt{(1/8 + 1/7)}\right) = .799 < 2.6.$$

There is, therefore, no significant difference.

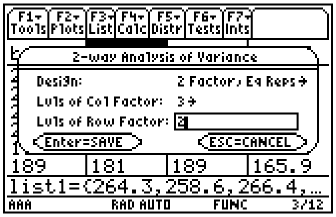
Although the data is the same, there is a large difference in the Mean Squared Error in Topic 39 and Topic 40 (1629.71 compared to 6.23). Because of the different power with which different golfers could hit the ball, you were able to block out much of the variability in Topic 40. All golfers were not able to hit Brand C as far as the other two brands. There was no such blocking possible in Topic 39, where the variability was due to other unknown causes.

Topic 41—Two Factor Designs with Equal Numbers of Replicates

Example: Test three brands of golf balls and two different clubs (drivers and five-iron) in a randomized design. Each of the six ball-club combinations is randomly and independently assigned to four experimental units, each of which consists of a specific position in the sequence of hits by a golf robot. The distance response is recorded for each of the 24 hits, and the results are shown in the table. Test to see if there is any difference in the brands or in the clubs.

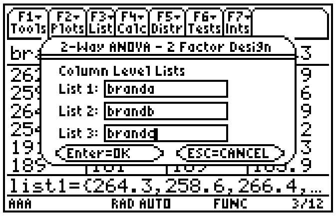
(Row Factor)	(Column Factor)		
Club	**Brand A**	**Brand B**	**Brand C**
Driver	264.3	262.9	241.9
	258.6	259.9	238.6
	266.4	264.7	244.9
	256.5	254	236.2
Five Iron	182.7	191.2	167.3
	181	189	165.9
	177.6	185.5	162.4
	187.3	192.1	172.5

1. Press [2nd] [F6] **Tests**, **D:ANOVA 2-Way**, with Design: **2 Factor**, **Eq Reps**, Lvls of Col Factor (or brands): **3**, and Lvls of Row Factor (or clubs): **2** (screen 16).



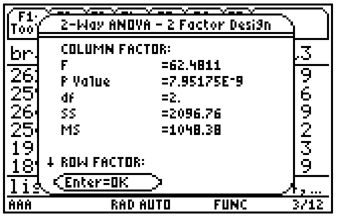
(16)

2. Enter List 1: **branda**, List 2: **brandb**, and List 3: **brandc**. Press [ENTER] to display screen 17.

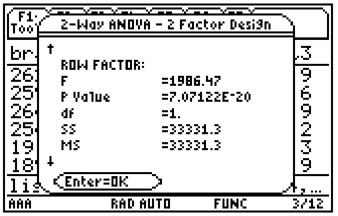


(17)

3. Press [ENTER] to display screen 18.

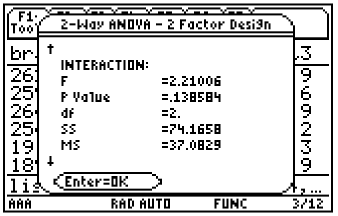
(18) 

4. Scroll down to view the remaining output (screens 19 and 20).

(19) 

Interaction

You see from the results that there is *no* significant *interaction* in screen 20, with F = **2.21006** and *p*-value = **0.138584** > .05. This is also clear from the nearly parallel x,y line plots (screen 22) of the means for each of the six ball-club combinations recorded in the table below.

(20) 

Column Factor			
Row	**Brand A**	**Brand B**	**Brand C**
Driver	261.45	260.38	240.4
Five-Iron	182.15	189.45	167.03

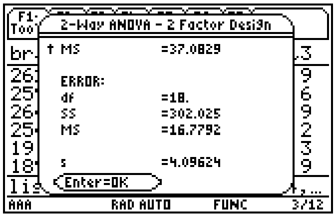
The top plot is for the driver (squares) with the bottom for the five-iron (boxes).

The plots are obtained as follows:

1. Store **1**, **2**, **3** in **list1**, the three driver means in **list2**, and the three five-iron means in **list3**.

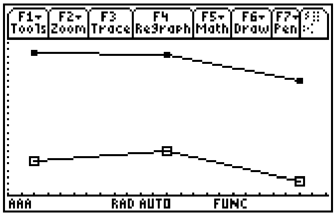
2. Set up and define **Plot 1** as Plot Type: **XYLine**, Mark: **Square**, x: **list1**, y: **list2**, and **Plot 2** as Plot Type: **XYLine**, Mark: **Box**, x: **list1**, and y: **list3**.

(21)



3. From the Plot Setup screen, press F5 **ZoomData** (screen 22).

(22)



It is clear from the nearly parallel plots that there is no *interaction* of the means for each of the six ball-club combinations.

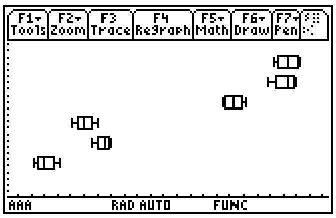
Club Differences (Row Factor)

With p-value = 7.07 E-20 < .05 in screen 19, it is clear that the driver drives the balls farther on average than the five-iron, just as you would expect. Screen 22 makes it clear that this is true for all three brands of balls. The dark squares represent the driver-hit balls, and lower boxes represent the five-iron ones.

Brand Differences (Column Factor)

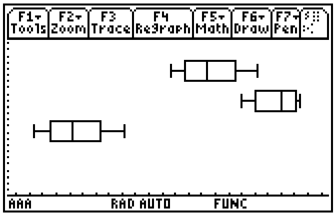
With p-value = 7.95 E-9 = 0.000000008 < .05 in screen 18, there are significant brand differences. Screen 23 shows modified boxplots of the driver data on the upper right, with Brand A at the top followed by Brands B and C. There seems to be no difference between Brand A and Brand B when hit with the driver. Brand C has significantly shorter distances. Five-iron data on the lower left again show Brands A and B on top and Brand C with the lowest values to the extreme left. It appears as if Brand B might be better than Brand A when hit with a five-iron. A re-plot of the five-iron data is given in screen 24 and discussed next.

(23)



Bonferroni Multiple Comparison Procedure

In screen 24, there is some overlap in the tails of Brand A at the top and Brand B in the middle (with Brand C having significantly lower values for distance hit than either Brand A or B). You will now do a Bonferroni comparison to see if the mean of Brand A and Brand B are significantly different with the five-iron, using the means (182.15 and 189.45) given in the table with screen 20.

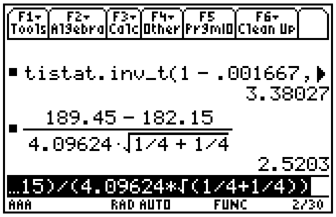
(24)


- There are nCr(6,2) = 6 ∗ 5/2 = 15 possible pairs when comparing the six means in the table.
- The significance level for these pairwise comparisons is (α = 0.05 overall) = 0.05/(15 ∗ 2) = 0.001667 in each tail.
- With ERROR: df = **18** (from screen 21), the critical value is **tistat.inv_t (1-.001667, 18)** = 3.38027 (screen 25).
- With s = **4.09624** (the last line of screen 21).

  As a reminder, the formula is: $t = \dfrac{\bar{x}_2 - \bar{x}_1}{s * \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

  $t = (189.45 - 182.15) \div \left(4.09624 * \sqrt{(1/4 + 1/4)}\right) = 2.5203$

  as calculated in screen 25.

(25)


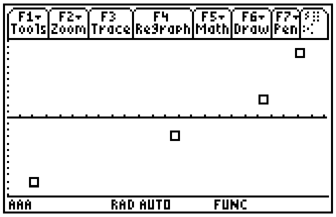
Since t = 2.5203 < 3.38027, you fail to reject the hypothesis that the mean distances for the two brands are different when hit with the five-iron. This confirms your interaction results that brands behaved the same way for the drivers and five-irons.

Note: *The p-value for the above t test is* **tCdf(2.5203, ∞ ,18) ∗ 2 = 0.021385 > 0.001667** *but still small. If the Brand B ball costs less than Brand A, or even if they cost the same, you would buy Brand B.*

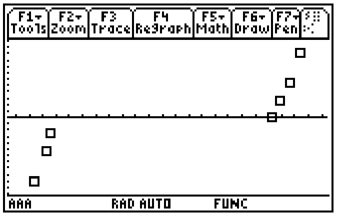
Assumptions

The assumption of equal variance across groups is not so restrictive if the sample sizes are the same (which they are). Notice also that the boxplots of screens 23 and 24 all have about the same spread. The assumption of samples from a normal population is met if there are not long tails, outliers, or obvious skewness. All the boxplots look fairly symmetrical with no outliers, except perhaps for the middle plot in screen 24 (second from the bottom in screen 23).

Screen 26 shows the normal probability plot for this data. Since the plot is based on just four points, it looks satisfactory.

(26)


In Topic 39, screen 8, the normal probability plot for the lower boxplot of Brand C data displays screen 27.

(27)


These are clearly not seven values from a normal distribution (more like three and four values from two different normal distributions). A normal probability plot of Brand A and Brand B review the same patterns. Something caused a downward shift in the robot's power after hitting the first four balls for each brand. It was later discovered that the club changed from a driver to a five-iron, so the Topic 41 analysis was the correct one.

In Topic 40, the shift in distance can be explained by the power of the different golfers, with the first four being more powerful than the last four.

EXPLORATIONS

Chapter 17

Multiple Linear Regression

This topic will show how to use [2nd] [F6] **Tests**, **B:MultRegTests** to build a model and test its validity for making estimations using [2nd] [F7] **Ints**, **8:MultRegInt**. Two examples will be used to show some possibilities.

Topic 42—Multiple Linear Regression with Inference and Residual Diagnostics

In Topics 12 and 33, you used folder **CARS** for information on 13 automatic transmission cars, including their city gas mileage in miles per gallon (mpg) and their weights (wt) in pounds. Topic 33 did inference with a simple linear regression least-squares fit line. The output there could be compared here by entering just one independent variable, **wt**. Topic 12 also fitted a quadratic to the data using [F4] **Calc**, **3:Regressions**, **4:QuadReg**, but no inference was available. You will correct this in Example 1 below.

Example 1: Using the data discussed above and given below:

1. Fit a quadratic function to the data using MultRegTests with y = **mpg** and x_1 = **wt** and x_2 = **wt ∗ wt = x_1^2**.

   ($y = c + bx_1 + ax_1^2 = B_0 + B_1x_1 + B_2x_2$ with $x_2 = x_1^2$)

2. Use the result to predict the gas mileage for a car weighing 3200 pounds. (95% intervals.)

3. Discuss the residuals and diagnostics.

Change folders to **CARS**, and create a new list named **wtwt** by using the command **wt*wt → wtwt**.

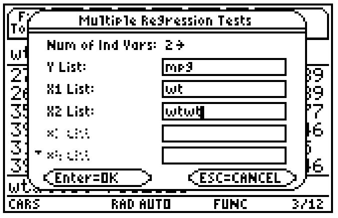
mpg (y)	wt (x1)	wt*wt (x2)
23	2795	7812025
23	2600	6760000
19	3515	12355225
17	3930	15444900
20	3115	9703225
17	3995	15960025
22	3115	9703225
17	4020	16160400
19	3175	10080625
19	3225	10400625
17	3985	15880225
29	2500	6250000
28	2290	5244100

Testing H_0: $\beta_1 = \beta_2 = 0$

From the Stats/List Editor:

1. Press [2nd] [F6] **Tests**, **B:MultRegTests**, with Num of Ind Vars: **2**, Y List: **mpg**, X1 List: **wt**, and X2 List: **wtwt** (screen 1).

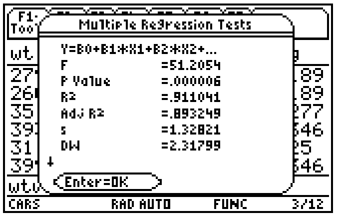
(1)



2. Press [ENTER] to display screen 2 with r^2 = **.911041** (similar to Topic 12, screen 25).

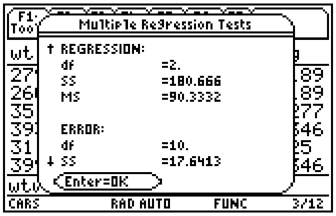
   The F = **51.2054** with *p*-value = **0.000006** < .05 is for testing the hypothesis H_0: $\beta_1 = \beta_2 = 0$. You reject H_0 and conclude at least one or both β_1 and/or β_2 are not zero.

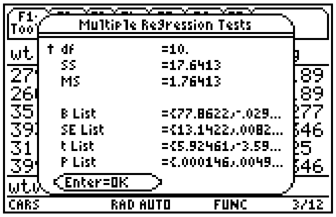
(2)



3. Scroll down to see the ANOVA table output (screen 3).

   After the ANOVA table is displayed, four lists are given. These lists are: **blist**, **selist**, **tlist**, and **plist** (screen 4).

(3)


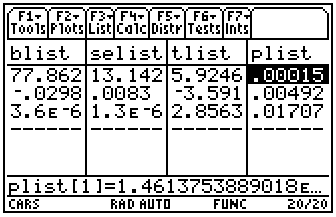
(4)


4. Press ENTER to return to the Stats/List Editor with the completed **blist**, **selist**, **tlist**, and **plist** now accessible in the editor (screen 5).

   The **blist** provides the coefficients b_0, b_1, b_2 for the model $\hat{y} = B_0 + B_1x_1 + B_2x_2$. So the **blist** in screen 5 gives mpg = 77.862 - 0.0298wt + 0.00000365wt^2 (also represented in Topic 12, screen 25, with variables *a*, *b*, and *c*, rather than β_0, β_1, and β_2).

   To test H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$, calculate

   $t = \dfrac{B_1 - 0}{SE} = \dfrac{-0.0298 - 0}{.0083} = -3.59$, using **SE** from the **selist**.

(5)


Note: *This* ***t*** **value** *is the second entry of the* **tlist***.*

5. To get the *p*-value in the **plist**, execute **tistat.tcdf(-∞, -3.59, 10) ∗ 2 = 0.0049 < .05**, where **10** is the degrees of freedom error from screen 3.

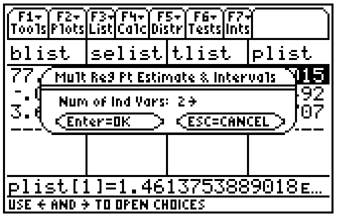
   With a *p*-value so small, you reject the H_0 and conclude that $\beta_1 \neq 0$. To test for H_0: $\beta_2 = 0$ with a *p*-value = **0.01707** (the last value in the **plist**) < .05, you would reject that $\beta_2 \neq 0$. With both rejections and the high r^2 value, it makes sense to use the multiple regression model to estimate gas mileage for a given car's weight.

Note: *There is more output from the Multiple Regression Test. These include the residuals and residuals diagnostics that you will discuss after the next section on interval estimates.*

Finding the 95% Confidence and Prediction Intervals

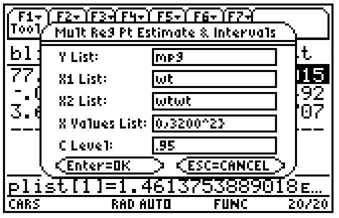
To find the 95% confidence interval for the mean mpg for cars that weigh 3200 lbs, and the 95% prediction interval for the mpg of a particular car of that weight:

1. Press [2nd] [F7] **Ints**, **8:MultRegInt**, with Num of Ind Vars: **2** (screen 6).



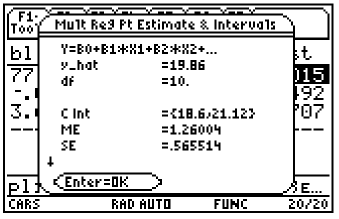
(6)

2. Press [ENTER] with Y List: **mpg**, X1 List: **wt**, X2 List: **wtwt**, X Values List: **{3200, 3200^2}**, and C Level: **0.95** (screen 7).



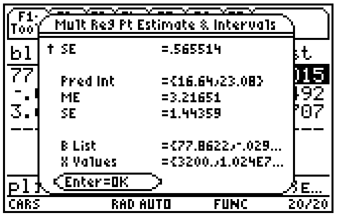
(7)

3. Press [ENTER] to display screen 8. Observe that y_hat = 19.86 and the 95% confidence interval for μ_{mpg}: y_hat ± ME = 19.86 ± 1.26, or 18.6 < μ_{mpg} < 21.12.



(8)

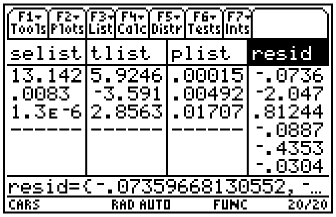
4. Scroll down to see the remaining output values including the 95% prediction interval y_hat ± ME = 19.86 ± 3.22, or 16.6 < *mpg* < 23.1 (screen 9).



(9)

Note: *Two cars in your sample weigh close to 3200 lbs. (3175 and 3225), and both get 19 mpg, which is contained in both intervals.*

Residual Analysis Including Outliers and Influential Points



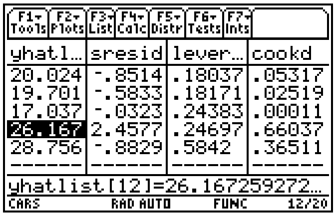
(10)

1. After finishing with screen 9, press [ENTER] to return to a screen like screen 5, but with the **list statvars\resid** in the last column as in screen 10. (Check the three values in the **selist**, **tlist**, and **plist**.)
2. If these values are not correct, as in screen 5, then steps 1 through 5 from the previous section, *Testing* H_0: $\beta_1 = \beta_2 = 0$, should be repeated.

Residual Diagnostics

1. From screen 10, press [2nd] ◄ to move to the left in the Stats/List Editor.
2. Continue with the cursor control or arrow key to highlight the value **26.167** in the **yhatlist** (screen 11).

   The **yhatlist[12]** of 26.167 is for the Mazda Protégé, which could be verified by [2nd] ◄ that shows mpg = 29 and wt = 2500, and by the table in Topic 11.



(11)

3. Press [♦] ► to get a resid = **2.833** calculated as observed mpg - yhat = 29 - 26.167 = 2.833.

   This value is still an *outlier*, with the standard residual sresid[12] = 2.4577 > 2 with *leverage*[12] = .24697 and the greatest *Cook's Distance* = cookd[12] = 0.66037, all shown in screen 11. A large Cook's Distance is an indication of a possible influential point (one rule of thumb says if *Cook's Distance* is ≥ 1).

   The last value of list **leverage[13]** = 0.5842 is the largest and greater than $2(k + 1)/n = 2(2 + 1) \div 13 = 0.4615$, where k is the number of independent variables. It has high leverage, but since its standard residual has a small absolute value of .8829 = | sresid[13] | its Cook's Distance, which is a function of both leverage and sresid, is small at .36511 = cookd[13].

Residual Plot

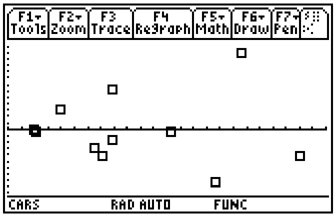
In a full analysis one would look at many plots, but you will plot only **yhat** vs. **sresid**.

1. Set up and define **Plot 1** as Plot Type: **Scatter**, Mark: **Box**, x: **statvars\yhatlist**, and y: **statvars\sresid**.

2. Plot with F5 **ZoomData** (screen 12).

   If the outlier (in the upper right corner) were lower (not an outlier) this plot would be fine.

(12)



Note: **yhatlist** *not* **y-hat**. *Instead of scanning the list in the Stats/List Editor for large Cook's Distances, these could be plotted so they stand out like the outlier in screen 12.*

Durbin-Watson Statistic

To test for autocorrelation of the residuals, the Durbin-Watson Statistic is displayed in screen 2 (DW = **2.31799**).

Example 2: The original data set is extended to include cars with manual transmissions and more variables. The variable **mpg** will now be **cmpg** for city fuel consumption in miles per gallon, and **wt** will now be **wts** to indicate more cars are involved.

New variables listed are:

Variable	Description
hmpg	Highway fuel consumption in miles per gallon
dplace	Engine displacement in liters
trans	0 = automatic transmission 1 = manual
ghg	Greenhouse gases emitted (in tons/yr)
wtswts	wts^2 or wts ∗ wts

Use this data to find a good multiple regression model to predict city fuel consumption.

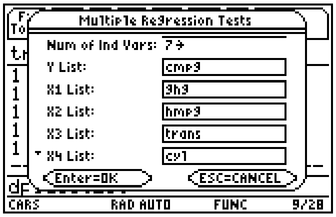
name	cmpg	ghg	hmpg	trans	cyl	dplace	wts	wtswts
Chevrolet (Cavalier)	23	10	31	0	4	2.2	2795	2795^2
Dodge (Neon)	23	10	32	0	4	2.0	2600	2600^2
Ford (Taurus)	19	12	27	0	6	3.0	3515	3515^2
Lincoln (Centurion)	17	14	24	0	8	4.6	3930	3930^2
Mercury (Mystique)	20	12	29	0	6	2.5	3115	3115^2
Olds (Aurora)	17	13	26	0	8	4.0	3995	3995^2
Pontiac (Grand Am)	22	11	30	0	4	2.4	3115	3115^2
Cadillac (Deville)	17	13	26	0	8	4.6	4020	4020^2
Chrysler (Sebring)	19	12	27	0	6	2.5	3175	3175^2
BMW 3-Series (bmw3s)	19	12	27	0	6	2.8	3225	3225^2
Ford Crown (Victoria)	17	14	24	0	8	4.6	3985	3985^2
Mazda (Protégé)	29	9	34	0	4	1.6	2500	2500^2
Hyundai (Accent)	28	9	37	0	4	1.5	2290	2290^2
Chevrolet (Camaro)	19	12	30	1	6	3.8	3545	3545^2
Mitsubishi (Eclipse)	22	10	33	1	4	2.0	3235	3235^2
Toyota (Camry)	23	10	32	1	4	2.2	3240	3240^2
Chevrolet (Corvette)	18	12	28	1	8	5.7	3220	3220^2
Ford (Mustang)	20	12	29	1	6	3.8	3450	3450^2
Honda (Civic)	32	8	37	1	4	1.6	2440	2440^2

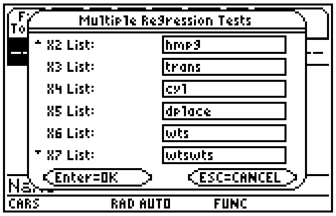
(Source: M. Triola, *Elementary Statistics,* 8th edition (page 803), © 2001 Addison Wesley Longman Inc. Reprinted by permission of Addison Wesley Longman.)

Backward Elimination

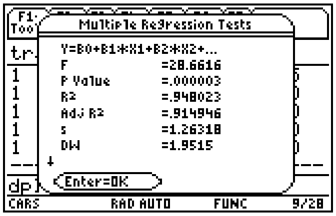
There are a number of ways of selecting variables to include in a model to make it as simple as possible but with minimum SSE or maximum r^2. (You will use Adjusted r^2 to take degrees of freedom into account.)

1. Press [2nd] [F6] **Tests**, **B:MultRegTests**, with Num of Ind Vars: **7**, Y List: **cmpg**, X1 List: **ghg**, X2 List: **hmpg**, X3 List: **trans**, X4 List: **cyl**, X5 List: **dplace**, X6 List: **wts**, and X7 List: **wtswts** (screens 13 and 14).

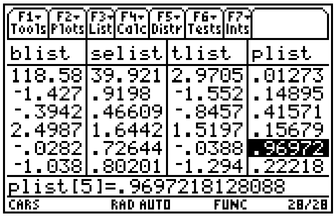
(13) 

(14) 

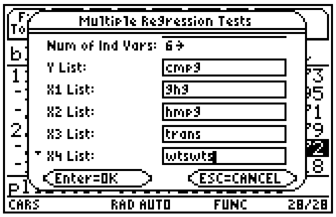
2. Press [ENTER] to display Adj R^2 = **0.914946** (screen 15).

(15) 

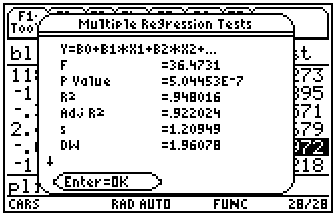
3. Press [ENTER] to display screen 16. The largest p-value is **plist[5] = 0.96972** (**plist[7]** = **0.02496** and **plist[8]** = **0.02487** are not shown). The **blist** starts with B_0, B_1, B_2, B_3, B_4, so you fail to reject the H_0: $\beta_4 = 0$ and will drop the fourth variable, **cyl**, from the model.

(16) 

4. Press [2nd] [F6] **Tests**, **B:MultRegTests**, with Num of Ind Vars: **6** and X4 List: **wtswts** (replacing **cyl** with **wtswts** that was in the last slot in screen 14). (See screen 17.)

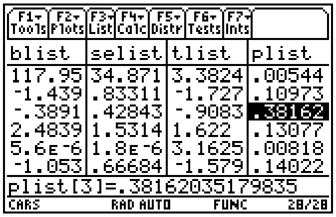
(17) 

5. Press [ENTER] to display Adj R^2 = **0.922024** (screen 18).



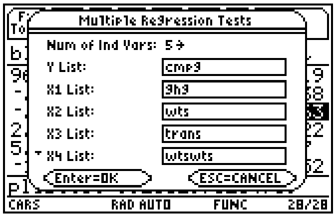
(18)

6. Press [ENTER] to display screen 19. The largest p-value is **plist[3] = .38162** (**plist[7]** = **0.0086** is not shown). You fail to reject H_0: $\beta_2 = 0$ and will drop the second variable, **hmpg**, from the model.



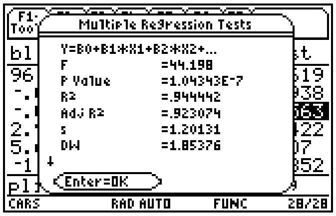
(19)

7. Press [2nd] [F6] **Tests**, **B:MultRegTests**, with Num of Ind Vars: **5** and X2 List: **wts** (replacing **hmpg** with **wts** that was in the last slot). See screen 20 (the fifth variable, **dplace**, is not displayed).



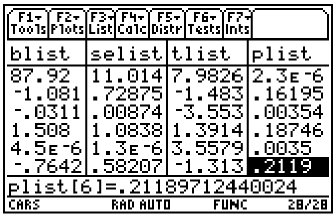
(20)

8. Press [ENTER] to display Adj R^2 = **0.923074** (screen 21).



(21)

9. Press [ENTER] to display screen 22. The largest p-value is **plist[6] = .2119**, which is associated with the fifth variable, **dplace**.



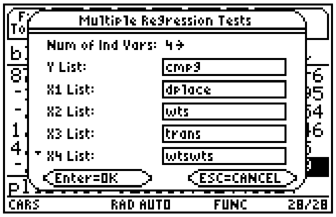
(22)

If you continue until all p-values (other than the first, for B_0) are less than 0.05, three variables remain, **ghg**, **wts**, and **wtswts** with Adj R^2 = **0.922247**, which is not significantly different from the value in screen 21.

Variation on Backward Elimination

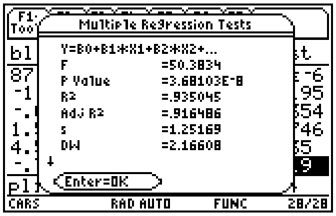
The three p-values in screen 22 that are greater than 0.05 are **0.16195**, **0.18746**, and **0.2119**. Since these are not all that much different, you could choose to replace the first independent variable **ghg** with the fifth variable, **dplace**, since **ghg** is the only variable that is not a physical characteristic, or measurement of the car.

1. Press [2nd] [F6] **Tests**, **B:MultRegTests**, with Num of Ind Vars: **4** and X1 List: **dplace** (replacing **ghg** with **dplace** that was last). Refer to screen 23.



(23)

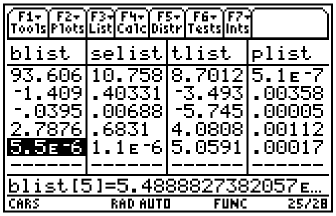
2. Press [ENTER] to display Adj R^2 = **0.916486** (screen 24).



(24)

3. Press [ENTER] to display all the p-values < 0.05 (but you are only interested in the last four), so you reject H_0: $\beta_1 = 0$, H_0: $\beta_2 = 0$, H_0: $\beta_3 = 0$, and H_0: $\beta_4 = 0$ and the regression equation becomes (from the **blist** in screen 25 and the variables in screen 23):

   cmpg = 93.606 - 1.409 ∗ dplace - 0.03954 ∗ wts + 2.7876 ∗ trans + 0.000005489 ∗ wtswts.

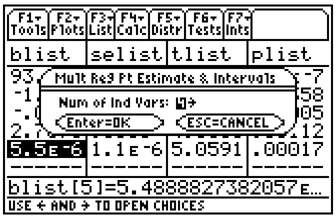


(25)

Confidence and Prediction Intervals

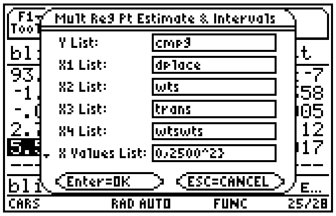
Use the above model to calculate 95% intervals of city fuel consumption for a car like the Mazda Protégé with dplace = **1.6**, wts = **2500**, trans = **0**, and wtswts = **2500^2**.

1. Press [2nd] [F7] **Ints**, **8:MultRegInt**, with Num of Ind Vars: **4** (screen 26).



(26)

2. Press [ENTER] with Y List: **cmpg**, X1 List: **dplace**, X2 List: **wts**, X3 List: **trans**, X4 List: **wtswts**, X Values List: **{1.6, 2500, 0, 2500^2}**, and C Level: **0.95** (not all displayed in screen 27).



(27)

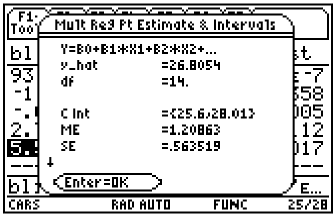
3. Press [ENTER] to display screens 28 and 29 with:

   95% Confidence Interval for μ_{cmpg}:

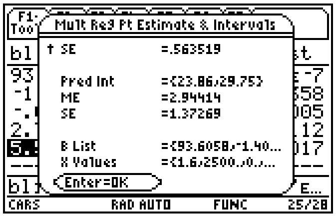
   y_hat ± ME = 26.8 ± 1.2, or $25.6 < \mu_{cmpg} < 28.0$.

   95% Predicted Interval for *cmpg*:

   y_hat ± ME = 26.8 ± 2.9, or 23.86 < *cmpg* < 29.75.



(28)

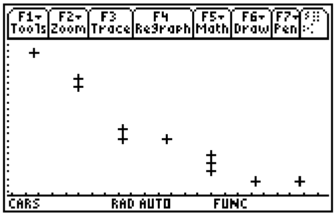


(29)

The data for the Mazda Protégé has been causing a problem in Topics 11, 12, and 33, and in the earlier example. It gets 29 mpg in the city, which is in the above prediction interval. The Mazda Protégé with this model (sresid = **1.9636**, lever = **.20268**, and cookd = **.19603**), is right on the border of being considered an outlier but not an influential point.

Another Model

Screen 30 displays the plot of X List: **ghg**, Y List: **cmpg**, which clearly shows curvature, so a quadratic model makes sense in this example.



(30)

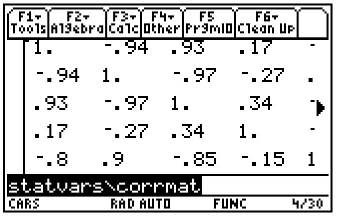
The least efficient cars (lowest mpg) emit the most greenhouse gases.

The quadratic model:
cmpg = **97.622 - 11.46 ∗ ghg + 0.4076 ∗ (ghg)^2** has Adj R^2 = **.956587** with a 95% prediction interval for ghg = **9** (as in the Mazda Protégé) of 27.5 ± 2.07, or 25.43 < *cmpg* < 29.58 with a sresid = **1.8211**.

This is an interesting and simple model that does very well.

Correlation Matrix

Screen 31 displays the correlation matrix for the values in the table at the beginning of this example.

(31) 

To generate the correlation matrix:

1. Set **Display Digits** in the MODE menu to **Float2**.
2. In the Stats/List Editor, press [F4]**Calc, 5:CorrMat**, with input data lists: **cmpg**, **ghg**, **hmpg**, **trans**, **cyl**, **dplace**, **wts**, **wtswts**, and the results to be stored in **cmat**.
3. Press [ENTER] to display **Done.**
4. On the Home screen, type **cmat** and then press [ENTER] (screen 31).

The complete top two rows are displayed below.

	cmpg	ghg	hmpg	trans	cyl	dplace	wts	wtswts
cmpg	1	-.94	.93	.17	-.80	-.79	-.88	-.85
ghg	-.94	1	-.97	-.27	.90	.83	.91	.90

You see that **ghg** is highly correlated with **cmpg** (*r* = -.94), followed closely by **hmpg** (*r* = 0.93). In the second row you see **ghg** is highly correlated with all the variables but **trans** (*r* = -.27). This is probably why it works so well by itself.

EXPLORATIONS

Chapter 18

Forecasting

There are many tools available to help in making forecasts from data, including multiple regression (Topic 42). In this chapter, you will look at two techniques with the aid of programs available from Texas Instruments (listings are given in Appendix A and on the Internet at **http://education.ti.com**). Exponential smoothing with and without a trend component will use program **exsmooth**. Program **forecast** uses a multiplicative time-series model with trend, seasonal, and irregular components. You must download and install these programs before you begin this chapter.

Topic 43—Exponential Smoothing (program exsmooth)

Example: The data used in both Topics 43 and 44 is given in the table and represents the number of sales for a small business over the past four years. (Store the **number of sales** in the list **sales** on your calculator.)

Year 1	List Sales	Year 2	List Sales	Year 3	List Sales	Year 4	List Sales
Quarter 1	593	Quarter 1	680	Quarter 1	708	Quarter 1	736
Quarter 2	512	Quarter 2	623	Quarter 2	662	Quarter 2	692
Quarter 3	705	Quarter 3	785	Quarter 3	853	Quarter 3	908
Quarter 4	756	Quarter 4	846	Quarter 4	884	Quarter 4	945

Simple Exponential Smoothing (trend constant = 0)

You will use the model $F_{t+1} = aY_t + (1 - a)F_t$. This model uses the data for this period (F_t) to forecast for the next period (F_{t+1}). The parameter a is the smoothing constant and can be a value from 0 to 1. Y_t represents the observed sales at time t, while F_{t+1} represents the sales forecast for the following period of time.

To start:

$F_2 = Y_1$

$F_3 = aY_2 + (1 - a)F_2$

.

.

.

$F_{t+1} = aY_t + (1 - a)F_t$

Measures of Fit

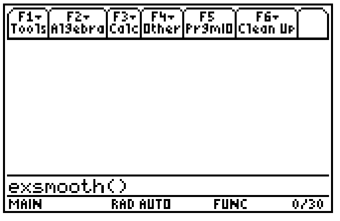
MSE = mean squared error = $\displaystyle\sum_{t=2}^{n} \frac{(Y_t - F_t)^2}{(n-1)}$

MAD = mean absolute difference = $\displaystyle\sum_{t=2}^{n} \frac{abs(Y_t - F_t)}{(n-1)}$

(You drop the first value because it had no real forecast.)

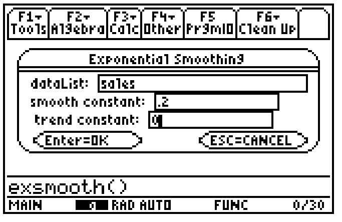
First try: $a = .2$, $b = 0$

1. From the Home screen, type or paste **exsmooth( )** to the input line (screen 1).



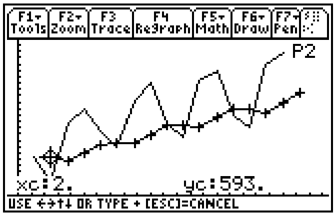
(1)

2. Press ENTER to display screen 2 with dataList: **sales**, smooth constant: **0.2**, and trend constant: **0**.



(2)

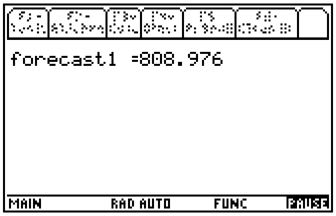
3. Press [ENTER] to display screen 3.



(3)

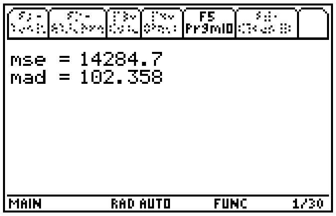
Note: *You are in* **Trace** *mode for the smoothed data. Use ▶ and ◀ to check these values. Use ▼ and ▲ to move to the original data and back.*

4. Press [ENTER] to display screen 4 with the forecast for the first quarter of the fifth year (**808.975 ≈ 809 sales**). You will discuss forecast for other quarters after the next try of $a = .39$, $b = 0$.



(4)

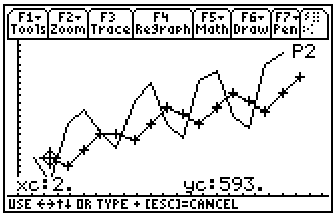
5. Press [ENTER] to display screen 5 with measures of fit MSE = **14284.7** and MAD = **102.358**.



(5)

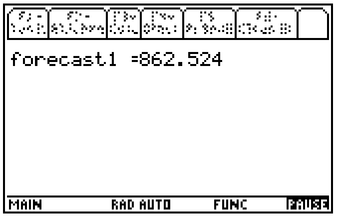
Another try: $a = .39$, $b = 0$

1. From the Home screen, type or paste **exsmooth( )** to the input line.
2. Press [ENTER] to display the screen with dataList: **sales**, smooth constant: **0.39**, and trend constant: **0**.
3. Press [ENTER] to display screen 6.



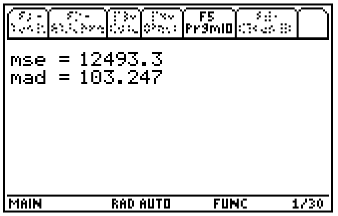
(6)

4. Press [ENTER] to display screen 7 with the forecast for the first quarter of the fifth year (**862.524 ≈ 863 sales**).



(7)

5. Press [ENTER] to display screen 8 with measures of fit MSE = **12493.3** and MAD = **103.247**.



(8)

Try other ***a*** values for practice.

The forecast for the first quarter of the fifth year is also the forecast for all future quarters using this technique.

This is a reasonable procedure and easy to use if there is not too much trend to the data and you want a smoother forecast curve and forecast down the middle of the data.

The following example is fairly smooth but the data have a steep trend, so the above method gives poor results.

Exponential Smoothing with Trend (Holt-Winters Method)

Example: Dow Jones Industrial Average (1980 – 1999).

Year	List djia	Year	List djia
1980	964	1990	2634
1981	875	1991	3169
1982	1047	1992	3301
1983	1259	1993	3754
1984	1212	1994	3834
1985	1547	1995	5117
1986	1896	1996	6448
1987	1939	1997	7908
1988	2169	1998	9181
1989	2753	1999	11497

(Source: Dow Jones Indexes. Reprinted with permission from Dow Jones & Company. © 2000 Dow Jones & Company. All Rights Reserved.)

Model:

Where a is the smooth constant, as before, and b is a trend constant. $0 < a < 1$, and $0 < b < 1$. Y_t represents the observed sales for time period t as noted earlier. S_t represents the level of the smoothed series associated with time period t, and T_t represents the trend component of the forecast.

$S_2 = Y_1$

$T_2 = Y_2 - Y_1$

$F_3 = S_2 + T_2 = Y_2$

$S_3 = aY_3 + (1 - a)(S_2 + T_2)$

$T_3 = b(S_3 - S_2) + (1 - b)T_2$

$F_4 = S_3 + T_3$

.

.

.

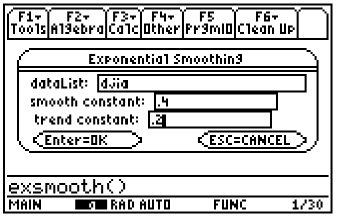
$S_t = aY_t + (1 - a)(S_{t-1} + T_{t-1})$

$T_t = b(S_t - S_{t-1}) + (1 - b)T_{t-1}$

$F_{t+1} = S_t + T_t$

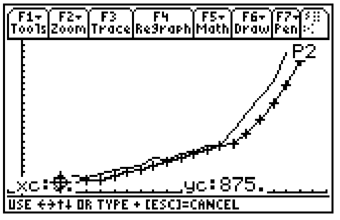
First try: $a = .4$, $b = .2$

1. Run the **exsmooth** program with dataList: **djia**, smooth constant: **0.4**, and trend constant: **0.2** (screen 9).



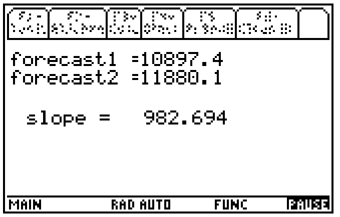
(9)

2. Press [ENTER] to display screen 10 with the smoothing line lagging behind the data.

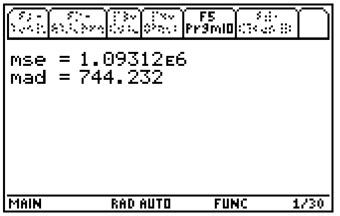


(10)

3. Press [ENTER] to display screen 11 with forecast for 2000 of **10897.4**, which is lagging behind the 1999 value of 11,497. The 2001 forecast, or forecast2, is **11880.1**, or 982.694 (the slope) greater than the 2000 forecast. The 2002 forecast could also be obtained as **11880.1 + 982.7 = 12,862.8**, and so on.

(11) 

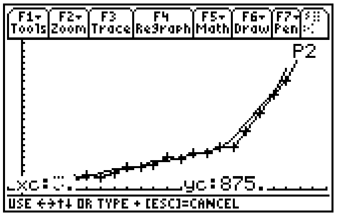
4. Press [ENTER] to display screen 12 with MSE ≈ **1.09312E6** and MAD = **744.232** as a measure of the fit.

(12) 

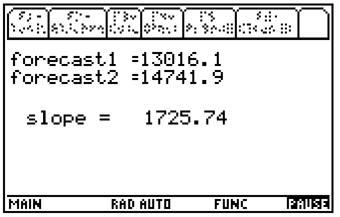
Another try: $a = .8$, $b = .5$

1. Run the **exsmooth** program with dataList: **djia**, smooth constant: **0.8**, and trend constant: **0.5**.

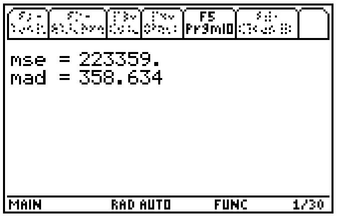
2. Press [ENTER] to display screen 13 with the smoothing line lagging behind the data.

(13) 

3. Press [ENTER] to display screen 14. This is a better fit (MSE = **223,359 < 1,093,120**), with a steeper slope and a higher forecast for 2000 of 13,016. If it is really a better forecast, time will tell.

(14) 

4. Press [ENTER] to display screen 15. This is a better fit (MSE = **223,359 < 1,093,120**), with a steeper slope (1725.74) and a higher forecast for 2000 of 13,016.1.

(15) 

Topic 44—Forecasting with Seasonality (program forecast)

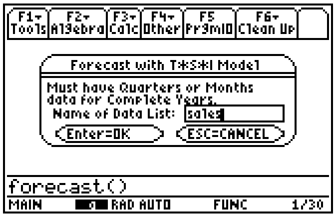
Multiplicative Model (program forecast)

For the multiplicative time-series model, you assume the **Y** data value is made up of three components: **Y = T ∗ S ∗ I**, where **T**, **S**, and **I** represent the *trend*, *seasonal*, and *irregular* components. (There is no attempt to isolate a cyclical component other than being part of the trend.)

Example: The sales data (quarterly data) given as the first example of Topic 43.

1. Run the **forecast** program with Name of Data List: **sales** (screen 16).

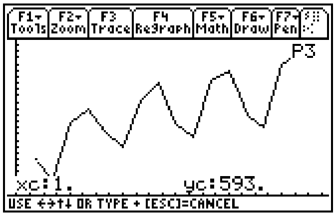
You are reminded that you "must have quarters or months data for complete years." You have quarters for four complete years for 16 data values. The name of the dataList: is **sales**.



(16)

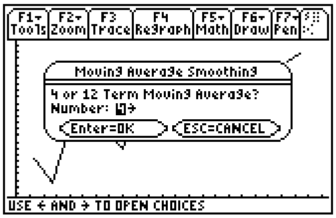
Note: *When the BUSY note is in the lower right corner of your screen, be patient for the results to be calculated.*

2. Press [ENTER] in **Trace** mode so you can use the arrow keys ▶ and ◀ to check the data. There is a definite seasonal component to the data and an upward trend (screen 17).



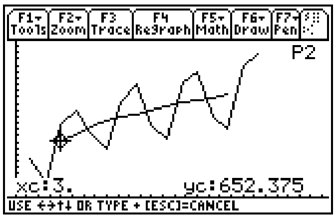
(17)

3. Press [ENTER] to display screen 18 where you select **4 term centered moving average (MA)** since you have quarterly data — four quarters each year.



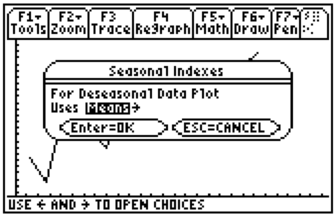
(18)

4. Press [ENTER] to display screen 19 with the smoothed moving average down the middle of the data. This plot represents the upward trend of the data (**MA ≈ T**) and possibly some cycle component.



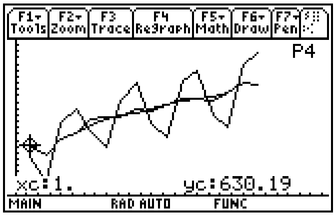
(19)

5. Press [ENTER] to display screen 20 with the option of using the means or medians in calculating the seasonal indices used to deseasonalize the data. (You have picked the mean.)



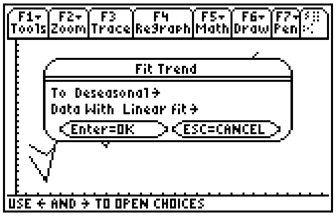
(20)

6. Press [ENTER] to display screen 21 with the deseasonalized data plot added to screen 19.



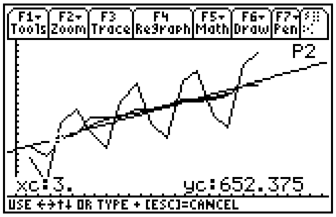
(21)

7. Press [ENTER] to display screen 22 where you have two choices. Select the To Deseasonal option for data to fit a linear regression line.



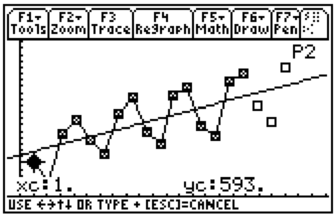
(22)

8. Press [ENTER] to display screen 23. The linear regression trend line is filled to the data on the deseasonlized data of screen 21.



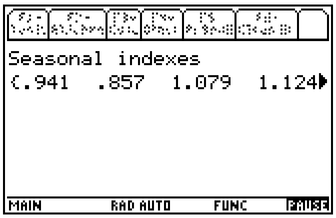
(23)

9. Press [ENTER] to display a plot of the data points, plus the forecast for the next four quarters (without a line joining them). You are in **Trace** mode so you can check this (screen 24).



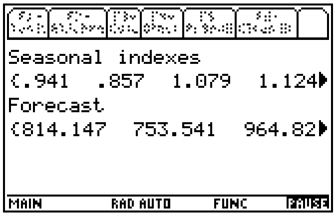
(24)

10. Press [ENTER] to view the seasonal indices, the first two quarters below the trend (value less than 1), and the other two quarters above the trend (screen 25).

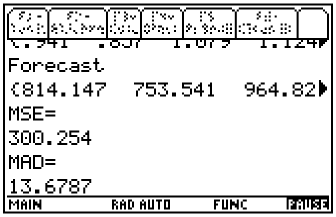


(25)

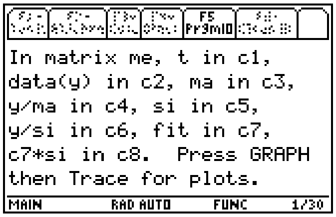
11. Press [ENTER] to display screen 26 with the forecast for the next four quarters.

(26)


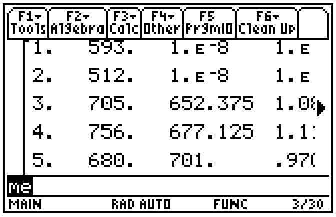
12. Press [ENTER] to display the forecast with MSE = **300.254** and MAD = **13.6787**, compared to simple exponential smoothing values of MSE = **12,493** and MAD = **103** from Topic 43, screen 8 (screen 27).

(27)


13. Press [ENTER] to display matrix **me** with **t** from 1 to 20 (or 16 + 4) in column 1, the sales data in column 2, the centered moving average in column 3, the **S ∗ I** components in column 4 (the **T** and **C** divided out with **MA**), seasonal indices in column 5, the deseasonalized data in column 6, the trend fit in column 7, and the forecast in column 8. You can use this matrix to check the calculations (screen 28).

(28)


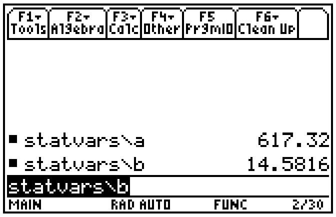
14. From the Home screen, enter **me** (screen 29).

(29)


Note: *Very small values in a cell* (.00000001 = 1.E-8) *indicate that cell is not used for calculations. The third value in column 3 (or* 652.375*) is the centered four term moving average for the third quarter of the first year. Centered moving averages cannot be calculated for the first two quarters of the first year because of the lack of data from the last two quarters of the previous year.*

15. Pasting **a** and **b** from the **statvars** folder indicates that the trend fit is Y_t = **617.32 + 14.58t**, or a slope of 14.58 ≈15 (screen 30).

16. To forecast for the first quarter of the fifth year (t = **17**) get the trend component Y_n = **617.32 + 14.58 ∗ 17 = 865.18**. Multiply this by the seasonal index for the first quarter of **0.941** (screen 25), or **865.18 ∗ 0.941 ≈ 814.1**, as in screen 26.

(30)


EXPLORATIONS

Chapter 19

Some Nonparametric Procedures

There are many nonparametric methods. Many of these use the ranks of the sample data obtained by sorting the data. You will cover five topics to give an idea of what is possible with the TI-89. The Sign Test and the Wilcoxon Signed-Ranks Test are the nonparametric counterparts to the matched pairs *t* test in Topic 32. The Wilcoxon (Mann-Whitney) Test is the nonparametric counterpart of the two-sample *t* test for independent samples in Topic 32. Topic 46 uses a program to introduce a computer intensive method of testing hypothesis. (This example is the computer counterpart of the Wilcoxon (Mann-Whitney) Test of Topic 45.) Topic 47 shows how to construct a confidence interval for the median of a population.

Topic 45—Two Sample (Independent and Matched Pairs) Nonparametric Procedures - the Sign Test and Tests Using Ranks

Sign Test for Matched Pairs

Example: To test the claim that a blood pressure medication reduces the diastolic blood pressure, a random sample of ten people with high blood pressure had their blood pressures recorded. After a few weeks on the medication, their blood pressures were recorded again. See the data in the table on the next page.

Subject	Before (list1)	After (list2)	list3 = list1 - list2
1	94	87	+7
2	87	88	-1
3	105	93	+12
4	92	87	+5
5	102	92	+10
6	85	88	-3
7	110	96	+14
8	95	87	+8
9	92	92	0
10	93	86	+7

Assigning p to represent the proportion in the population whose blood pressure would be reduced by medication (+ in the above samples), you want to test the following hypothesis:

H_0: $p = 0.5$

H_a: $p > 0.5$

If the medication has no effect on the blood pressure reading, you would expect about half of the sample population to do better after taking the medication and half of the sample population to do worse. The zero difference does not add any information; therefore, you see that of the nine subjects, there are two negative signs and seven positive signs. If the null hypothesis is true, you would expect 4.5 of each (the average in the long run). This is a binomial distribution with
***n* = 9**, ***p* = 0.5**, ***n* ∗ *p* = 9 ∗ 0.5 = 4.5 =** μ.

P-Value and Conclusion

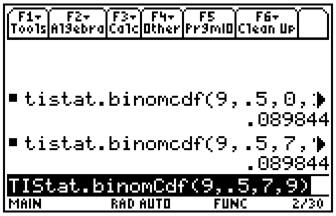
From the Home screen:

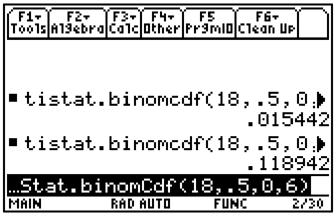
1. Press [CATALOG], [F3] **Flash Apps binomCdf(** and then press **9**, **.5**, **0**, and **2**.

2. Press [ENTER] for *p*-value = **0.089844**, which is the probability of getting two or fewer (0 to 2) negative signs (screen 1). You could also have found the probability of getting seven or more (7 to 9) positive signs, also shown in screen 1.

   You do not have strong evidence to reject the null hypothesis that the medication has no effect. (Not significant at the α = 0.05 level.)

   If the sample size were doubled to 18 and the number of negatives doubled to 4, you would reach the conclusion to reject H_0 with *p*-value = **.015442** < .05 (screen 2).

(1) 

(2) 

Note: *With the larger sample size, however, you could also see a larger proportion of negatives (for example, six or more), and again, you would not be able to reject the null hypothesis (p-value =* **0.118942***, also in screen 2).*

The Sign Test is easy to use and explain. As always, not rejecting the null hypothesis does not mean that the null hypothesis is true; you just have insufficient evidence to say it is false.

The next test uses more of the information available than does the Sign Test, so perhaps it will lead to a different conclusion.

Wilcoxon Signed-Ranks Test for Matched Pairs

Example: Use the data from the *Sign Test for Matched Pairs* section, with the reduction of blood pressure stored in **list3**.

From the Home screen, type

{7, -1, 12, 5, 10, -3, 14, 8, 0, 7} → list3.

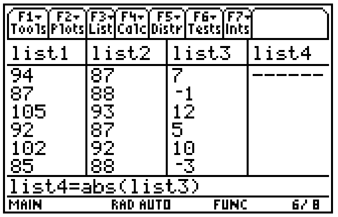
Use the following procedure to test the following hypotheses:

H_0: The population of high blood pressures is unchanged after taking medication.

H_a: The population of high blood pressures is decreased (shifted to the left) after taking medication.

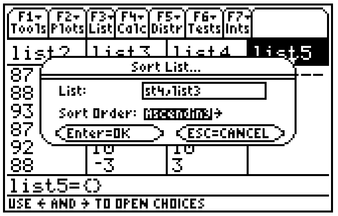
In the Stats/List Editor with the differences in **list3** (screen 3):

1. Highlight **list4** and select **abs**(**list3**) (screen 3).
2. Press [ENTER] and all the values in **list4** are positive.



(3)

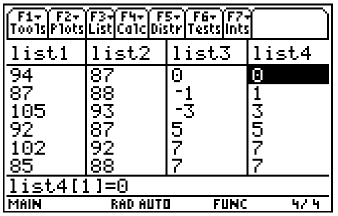
3. Press [F3] **List**, **2:Ops 1:Sort List**, with List: **list4**, **list3** (in that order), and Sort Order: **Ascending**. This sorts **list4** in order and carries along **list3** (screen 4).



(4)

4. Press [ENTER] to display screen 5.

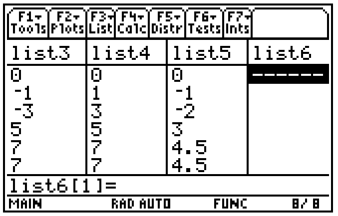
   The values in **list4** are sorted with the smallest value, zero, displayed first. **List3** will take along the corresponding values (0 comes first even though -3 is the smallest number in **list3**).



(5)

5. Put the ranks of the data in **list4** into **list5** as follows:
   a. Rank 0 as 0.
   b. Rank 1 and 3 as -1 and -2 (negative because they are from negative differences in **list3**).
   c. Rank 5 as 3.
   d. Rank the two 7's by averaging the fourth and fifth ranks (or (4 + 5)/2 = 4.5).
   e. Rank the other values, which are all different and positive, as 6, 7, 8, and 9.

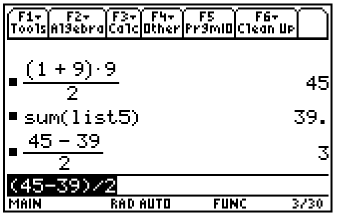
See screen 6 under the **list5** heading.



(6)

6. List the sum of the integers from 1 to $n = (1 + n) * n/2$. Thus, the sum of the integers from 1 to 9 = (1 + 9) ∗ 9/2 = 45 and T = **sum(list5) = 39** is your test statistic (screen 7).

   As a check on your work, note that half the difference between the above two values is the absolute value of the sum of the negative ranks, or **abs(-1 + -2) = 3**.

(7) 

P-Value and Conclusion

What is the probability of getting a sum of 39 by chance if there is no change in blood pressure? All possible ranks from 1 to 9, positive or negative, are possible. This is 2 ^ 9 = 512 possibilities.

The five ways of getting a sum of 39 or more are listed in the table below. Thus, the *p*-value = **5/512 = 0.0097765 ≈ 0.01** < .05. There is strong evidence that the blood pressure has decreased.

Samples with signed ranks that add to 39 or more

Sample	±	±	±	±	±	±	±	±	±	Sum
1	-1	-2	3	4	5	6	7	8	9	39
2	1	2	-3	4	5	6	7	8	9	39
3	1	-2	3	4	5	6	7	8	9	41
4	-1	2	3	4	5	6	7	8	9	43
5	1	2	3	4	5	6	7	8	9	45

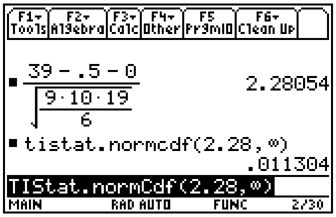
Charts in Text and Normal Approximation

As a sample size increases, it can be tedious to calculate all of the possibilities. Texts will have a chart for small sample sizes. (Some texts base their chart on the smallest sum of the positive or negatives, -3 in your example.) For sizes that the charts cannot handle, the normal approximation below gives good results.

Mean: $\mu_T = 0$

Standard Deviation: $\sigma_T = \sqrt{\left(n(n+1)(2n+1) \div 6\right)}$

If the medication had no effect, you would expect an equal number of positive and negative ranks of their differences. Thus, on average, the sum of the ranks should be zero. $Z = (T \pm 0.5 - \mu_T) \div \sigma_T = (39 - 0.5 - 0) \div \sqrt{(9*10*10 \div 6)} = 2.28$, where **0.5** is a continuity correction. For the right tail, you start at 38.5 to include the class of 39. The area in the right tail of the normal distribution is a *p*-value of **0.0113 ≈ 0.01** as in screen 8, using **tistat.normalCdf(2.28, ∞)**. This is very close to the exact value above.

(8) 

Wilcoxon (Mann-Whitney) Test for Two Independent Samples

Example: Test the claim that teaching Method B results in higher test scores than Method A based on the scores in the table from random samples of students taught with the two methods.

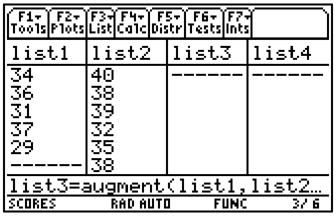
	Method A (list1)	Method B (list2)
	34	40
	36	38
	31	39
	37	32
	29	35
		38
Mean	33.4	37
s_x	3.362	2.966
n	5	6

To determine if the difference between the two sample means (33.4 vs. 37) is significant, you use the one-tailed rank sum test. The null and alternate hypotheses are:

H_0: The two population distributions are identical.

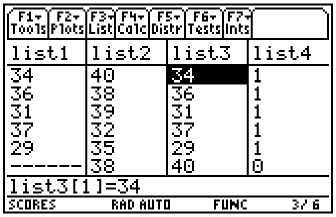
H_a: The center of population B is to the right of the center of population A.

1. Type Method A scores in **list1** and Method B scores in **list2**.
2. Highlight **list3**, press [CATALOG], select **augment(** and then press [ENTER].
3. Type or paste **list1**, **list2** (screen 9).



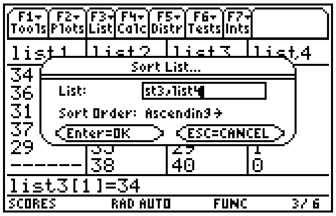
(9)

4. Press [ENTER] with **list1** followed by **list2**, and then **list3**. Fill in **list4** with five 1's, followed by six 0's (screen 10). **List3** consists of the combined scores of both methods, with **list4** indicating which scores are Method A (1's) and which are Method B (0's).



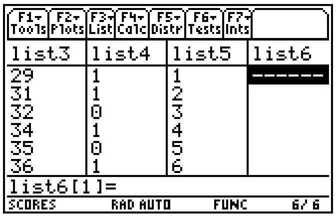
(10)

5. With **list3** or a row in **list3** highlighted, press [F3] **List**, **2:Ops**, **1:Sort List**, with List: **list3**, **list4**, and Sort Order: **Ascending** (screen 11).



(11)

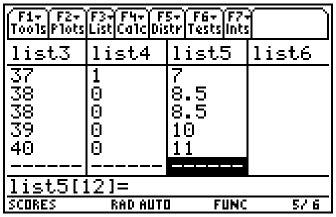
6. Press [ENTER] to display screen 12.



(12)

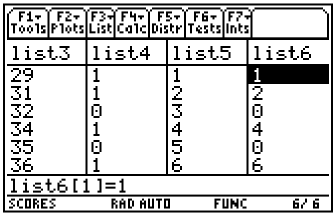
7. Enter **list5** with values: **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8.5**, **8.5**, **10**, and **11** to show the rank of the 11 combined scores (screen 13).

8. Highlight **list6** and then let list6 = **list4** ∗ **list5**.

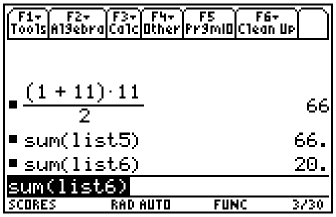
(13) 

Note: **list3***, which is now in order, has tied eighth and ninth entries of* **38***. The rank of each* **38** *is given a value of* **8.5** *in* **list5** *(the mean of* 8 *and* 9*). This is so that one* **38** *is not favored over the other, and yet the sum of ranks is meaningful* (8 + 9 = 8.5 + 8.5 = 17)*.*

9. Press ENTER to display screen 14. This screen shows ranks for Method A (those with 1's next to them in **list4**) in **list6**, along with 0's corresponding to Method B scores.

(14) 

10. To check your work, the sum of n ranks is given by the formulas **(1 +** ***n*****)** ∗ ***n*****/2 = (1 + 11)** ∗ **11/2 = 66**. **Sum(list5) = 66** also verifies the sum of the total ranked scores (screen 15).

(15) 

11. The command **sum(list6) = 20** is in the last line of screen 15. This is the sum of the ranks in Method A. Therefore, the sum for Method B is **66 − 20 = 46**.

To test H_0, you compare the ranked sums for each method by finding the appropriate p-value.

Note: *The average rank value is* **66/11 = 6***, so if there is no difference in methods, Method A should have about* 5 ∗ 6 = 30 > 20 *and Method B should have about* 6 ∗ 6 = 36 < 46*. Is this difference significant?*

P-Value and Conclusion

There are 462 (nCr(11,5) = 462) ways of picking five ranks for Method A and six ranks for Method B. There are 19 possibilities where the sum of the five ranks is 20 or less (see the table on the next page). This p-value = **19/462 = 0.0411** < .05. You reject the null hypothesis and conclude that Method B does significantly better.

Sample A	Rank											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Sum**
1	1	2	3	4	5							15
2	1	2	3	4		6						16
3	1	2	3	4			7					17
4	1	2	3	4				8				18
5	1	2	3	4					9			19
6	1	2	3	4						10		20
7	1	2	3		5	6						17
8	1	2	3		5		7					18
9	1	2	3		5			8				19
10	1	2	3		5				9			20
11	1	2	3			6	7					19
12	1	2	3			6		8				20
13	1	2		4	5	6						18
14	1	2		4	5		7					19
15	1	2		4	5			8				20
16	1	2		4		6	7					20
17	1		3	4	5	6						19
18	1		3	4	5		7					20
19		2	3	4	5	6						20

Charts in Text and Normal Approximation

It is tedious to calculate all the values in the previous table, and it is not necessary because texts covering this topic have charts for your use. If your sample sizes are too large for the table, you can use the following normal approximation.

- Let n be the smaller sample size and m the larger sample size ($n = 5$, $m = 6$).
- Calculate the expected mean sum of ranks for the smaller sample: $\mu_T = n(n + m + 1)/2 = 5(5 + 6 + 1)/2 = 30$.

- Calculate the Standard Deviation:

$$\sigma_T = \sqrt{\left(n*m*(n+m+1)\div 12\right)} = \sqrt{(5*6*12)\div 12} = 5.477$$

- $Z = (T \pm 0.5 - \mu_T) \div \sigma_T = (20 + 0.5 - 30) \div 5.477 = -1.734$

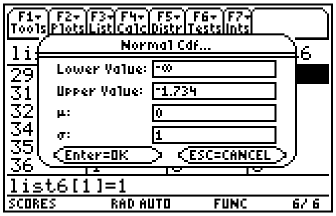
Note: The denominator is always 12.

The 0.5 is the continuity correction, since you want the area to the left, including the class with 20 that ends at 20.5.

The ± sign indicates whether the class of 20 is included or not included. In this example, "+" is used because the class of 20 is included.

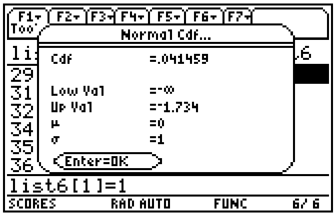
Find the area in the left of the normal distribution, or an approximate p-value = **0.0415** (very close to the exact value of 0.0411 in the *Wilcoxon (Mann-Whitney) Test for Two Independent Samples* section).

1. In the Stats/List Editor, press [F5] **Distr**, **4:normalCdf**, with Lower Value: **-∞** and Upper Value: **-1.734**. Values μ: **0** and σ: **1** could be left blank (screen 16).



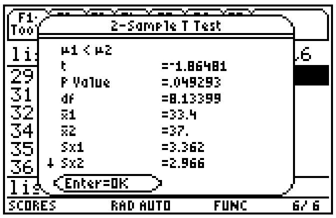
(16)

2. Press [ENTER] to display Cdf = P Value = **0.041459** < .05 (screen 17).



(17)

***Note:** If you could meet the requirement to use* [2nd] [F6] **Tests**, **4:2-SampTTest** *(not pooled) you obtain a p-value =* **0.0493** *(screen 18).*



(18)

Topic 46—Randomization Test (Resampling) for Two Independent Samples

Example: Test the claim that teaching Method B results in higher test scores than Method A, based on the following scores from random samples of students taught with the two methods.

Method A (list1)	Method B (list2)
34	40
36	38
31	39
37	32
29	35
	38

Use the following procedure to test the following hypothesis:

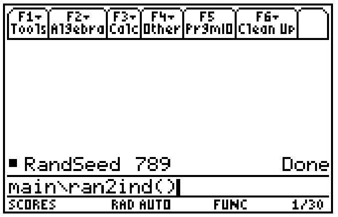
H_0: $\mu_B - \mu_A = 0$ from identical populations, (or $\mu_A = \mu_B$)

H_a: $\mu_B - \mu_A > 0$

1. Enter Method A scores in **list1** and Method B scores in **list2**. You will need program **ran2ind**, which is listed at the end of this topic.

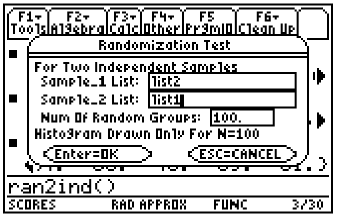
2. For this example, set **RandSeed 789** to repeat the results given below (screen 19).

3. Paste or type program name **ran2ind()** in the input line of screen 19.



(19)

4. Press ENTER, with Sample_1 List: **list2** and Sample_2 List: **list1** (screen 20).

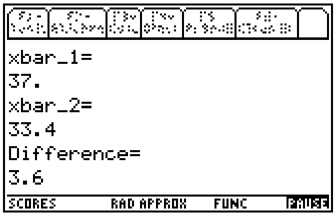


(20)

You do this so that the mean differences will be positive, making the following explanation easier. You are going to sample five values from the 11 given in both samples, and put these values in one sample with the remaining 6 in another sample. You then calculate the differences of their means and store the data in **list lx6**. This will be done 100 times, so you will get an automatic histogram of the differences in means when you finish.

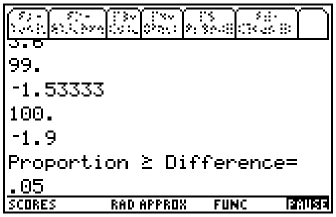
***Note:** If the number of Random Groups or resampling is different from* 100*, you can still plot your own histogram of the results in* **list lx6** *if required.*

5. Press [ENTER] to show the means of each sample and their differences. The program is set up to see how rare it is to get the difference 3.6 or greater (screen 21).

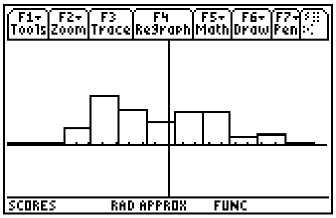
(21) 

6. Press [ENTER] and sampling takes place with the first resampling differences of -2.2667, and continuous printing of the results to the 100th difference of -.1.9, with 0.05 or five of the 100 differences were 3.6 or greater (screen 22).

   Therefore, the *p*-value is approximate **.05**. (Recall 0.04 in the *Wilcoxon (Mann-Whitney) Test for Two Independent Samples* section.) Of course, different seeds and different sample sizes will give different results. These are approximate.

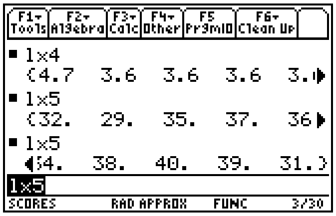
(22) 

7. Press [ENTER] to display the mound shape of the differences (screen 23).

(23) 

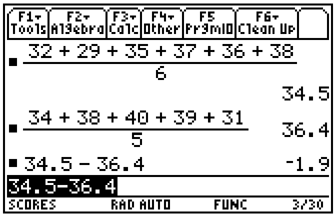
This test is very much like the previous section, *Wilcoxon (Mann-Whitney) Test for Two Independent Samples*, but this test was concerned with all possible totals (or equivalently, means) of size five and six. Here you do the same thing, but rather than theoretically figure out the number of possibilities, you just randomly generate them with the actual sample data rather than their ranks. No need to worry about ties.

If you look at **list lx4**, the differences are sorted in descending order on the Home screen. As in the top lines of screen 24, you see the five greatest values are **4.7**, **3.6**, **3.6**, **3.6**, and **3.6** (all greater or equal to 3.6), with the next lower value of **3.2333** (not shown).



(24)

The bottom of screen 24 shows the last resampling (in **list lx5**) has the first six values (Method B) of **32**, **29**, **35**, **37**, **36**, and **38** (not shown) for a mean **B** or xbar_1 = **34.5**, while the last five values (Method A) are **34**, **38**, **40**, **39**, and **31** for a mean **A** or xbar_2 = **36.4**. The last differences is **xbar_2 - xbar_1 = -1.9**, as in screens 22 and 25.



(7)

The program can be rerun with different values of ***n***, but a histogram will *not* automatically be plotted when $n \neq 100$.

The **ran2ind** program listing follows:

```
()
Prgm
FnOff  :PlotsOff  :ClrIO
setMode("Exact/Approx","APPROXIMATE")
{}→lx5:{}→lx6:Ø→p
⁻25→ymin:5Ø→ymax:Ø→yscl
Dialog
Title "Randomization Test":Text "For Two
Independent Samples"
Request "Sample_1 List",xx1:Request "Sample_2
List",xx2
Request "Num of Random Groups",kk
Text "Histogram Drawn Only For N=1ØØ"
EndDlog
expr(xx1)→lx1:expr(xx2)→lx2:expr(kk)→k
mean(lx1)→a:dim(lx1)→m
a*m→e
mean(lx2)→b:dim(lx2)→n
b*n→f
a-b→d
e+f→t:m+n→l
Disp "xbar_1=",a
Disp "xbar_2=",b
Disp "Difference=",d
Pause :ClrIO
lx1→lx5:m+1→j
For i,1,n
lx2[i]→lx5[j]
1+j→j
EndFor
```

```
For i,1,k
l→c:1→j;Ø→s
For j,1,n
tistat,randint(1,c)→v
s+lx5[v]→slx5[c]→h
lx5[v]→lx5[c]:h→lx5[v]
c-1→c
EndFor
(t-s)/m-s/n→u
u→lx6[i]
If u≥d:1+p→p
Disp i,u
EndFor
Disp "Proportion ≥ Difference="
Disp p/k:Pause
If k≠1ØØ:Stop
lx6→lx4:SortD lx4
NewPlot 3,4,lx6
ZoomData
EndPrgm
```

Topic 47—Estimation (Confidence Intervals) for Medians

Example: A classroom activity where 10 students are picked at random. There can be repeats with the same student selected more than once. Each student turns in a secret ballot stating the amount of money they currently have with them. The following results were obtained (after the data were put in order).

$0.62, 0.62,) 2.75, 6.80, 9.72, 16.15, 19.10, 23.50, (34.50, 56.10

Note: *The parentheses are used to trim off the end values.*

It is stated that you are about 90% confident that the median amount of money in the possession of all the students on this particular day is between $2.75 to $23.50. This is with the two lowest and two highest values trimmed off the sample results.

If you are willing to settle for about 66% confidence, trim one more value from each end for an estimate of from $6.80 to $19.10.

Theory

If **B** stands for a value Below the true median of the distribution of money from all students and **A** for a value Above the true median, the following samples will *not* contain the median of the population after two values are trimmed from each end. There must be at least one B and one A in what remains for it to contain the median:

BB)BBBBBB(BB

BB)BBBBBB(BA

BB)BBBBBB(AA

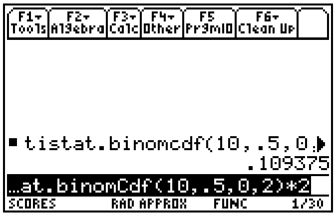
BB)AAAAAA(AA

BA)AAAAAA(AA

AA)AAAAAA(AA

If there is an even number of students in the class (if not, the teacher joins in as part of the group) then half the values will be equal to or below the median, and half the values will be equal to or above the median, so getting an **A** or **B** is like getting a head or tail.

What is the probability of getting **0**, **1**, **2**, **8**, **9**, and **10** heads with a binomial with n = **10**, p = **.5**? The answer is by symmetry
(p(0) + p(1) + p(2)) ∗ 2 = **tistat.binomCdf(10,.5,0,2)** ∗ **2** = .109 (screen 26). So there is 1 - .109 = .891 (or approximately 90%) chance of containing the median in the trimmed interval.



(26)

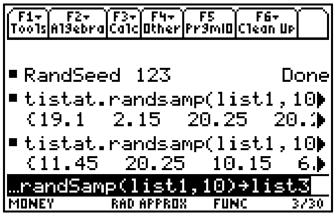
If you trim three, (p(0) + p(1) + p(2) + p(3)) ∗ 2 = 0.34375, so 1 - 0.34 = .66 is the chance of containing the median.

If the class in your example had 24 members with the following money, listed from smallest to largest:

0.05	**0.10**	**0.50**	**0.62**	**1.13**	**1.55**	**2.15**	**2.75**	**5.50**	**6.80**	**9.72**	**10.15**
11.45	**15.10**	**16.10**	**16.15**	**19.10**	**20.25**	**23.25**	**23.50**	**26.10**	**29.04**	**34.50**	**56.10**

1. Store the above 24 values in **list1**. The median of the data is ($10.15 + 11.45)/2 = $10.80.

2. Set **RandSeed 123** to repeat the results displayed below (screen 27).



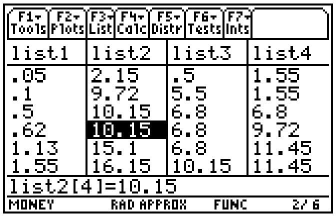
(27)

3. Pick random samples of size 10 from **list1**, with repeats possible, and store in **list2** with **tistat.randsamp(list1,10)→list2**.

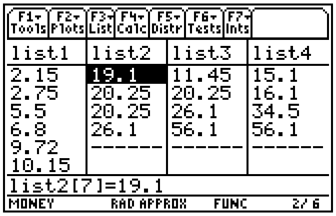
4. Repeat two more times, but store to **list3** and **list4**, with parts of **list2** and **list3** as shown in screen 27.

5. Sort **list2**, **list3**, and **list4** in ascending order with the following results (screens 28 and 29).

(28)



(29)



Trimming three values from each end, you get:

List: (**lower**, **upper**)

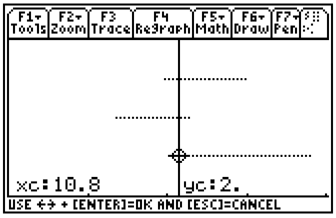
List4: (**9.72**, **15.1**)

List3: (**6.8**, **11.45**)

List2: (**10.15**, **19.1**)

Since these are 66% confidence intervals, you would expect about two of three intervals calculated this way to contain the true median in the long run. All three of these intervals contain the population median of 10.80. See screen 30 with the vertical line at the median (xc: **10.8**) and the lowest line of interval from **list2** (yc: **2**).

(30)



Source: "Estimation with Classroom Data," *Teaching Statistics*, Volume 7, Number 2, May 1985. Used with written permission of the Teaching Statistics Trust.

EXPLORATIONS

Chapter 20

More Fits to Bivariate Data

This book has covered most of the regressions listed in the Stats/List Editor under [F4] **Calc**, **3:Regressions**. (See screens 1 and 2 below.) Although only the QuadReg under polynomial regression (Topics 12 and 42) was covered, CubicReg and QuartReg are handled in a similar manner. For completeness here, Logist83 and Logistic (Topic 48), SinReg (Topic 49), and MedMed (Topic 50) will be covered.

Topic 48—The Logistic Curve

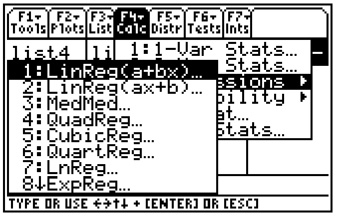
Logist83 Fit

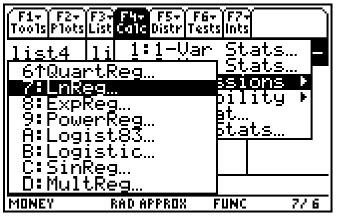
(This is the same fit as the TI-83 logistic fit.)

Example: U.S. Census data (population in millions).

cyear	census	cyear	census
1800	5.31	1900	75.99
1810	7.24	1910	91.97
1820	9.64	1920	105.71
1830	12.87	1930	122.78
1840	17.07	1940	131.67
1850	23.19	1950	151.33
1860	31.44	1960	179.32
1870	39.82	1970	203.21
1880	50.16	1980	226.50
1890	62.95	1990	249.63

Population data from the U. S. Census Bureau.
(U. S. Census Bureau, Statistical Abstract of the United States: 2000; http://www.census.gov/prod/2001pubs/statab/sec01.pdf). Source for U. S. Census data is Nagel and Saff, *Fundamentals of Differential Equations*, 3rd edition, Addison Wesley Longman, 1993. Reprinted by permission of Addison Wesley Longman.

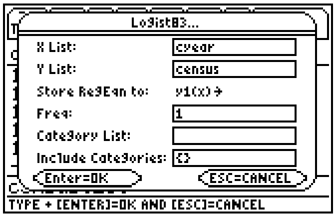
(1)


(2)


Create a new folder named **CENSUS**. Enter the years in a new list named **cyear** and the census data in a new list named **census**.

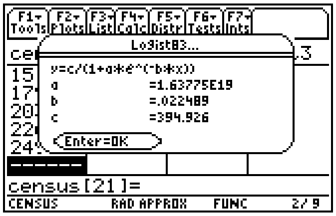
1. From the Stats/List Editor, turn off all functions and plots with [F2] **Plots**, **4:FnOff** and [F2] **Plots**, **3:PlotsOff**.

2. Press [F4] **Calc**, **3:Regressions**, **A:Logist83**, with X List: **cyear**, Y List: **census**, and Store RegEqn to: **y1(x)** (screen 3).



(3)

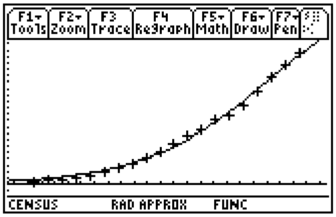
3. Press [ENTER] for $y = \dfrac{394.926}{1+1.63775E19 * e^{-.022489x}}$. This is the logistic regression model fitted to the data in X List and Y List using an iterative least-squares fit (screen 4).



(4)

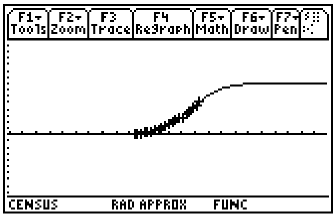
4. From the Stats/List Editor, press [F2] **Plots**, and set up and define **Plot 1** as Plot Type: **Scatter**, Mark: **Plus**, X List: **cyear**, and Y List: **census**.

5. Press [F5] **ZoomData** to display screen 5. Observe how well the logistic curve from step 3 fits the census data.



(5)

6. Press [F2] **Zoom** and **3:ZoomOut**. Respond to *New Center?* by pressing [ENTER]. This zoom command regraphs the curve with an expanded window (screen 6). Observe the S-shaped nature of the curve and how it levels out in the long run.

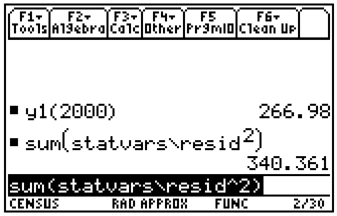


(6)

7. From the Home screen:

   a. Predict the population for the 2000 census by typing **y1(2000)**. Substituting year 2000 into your regression equation stored in **y1** gives 266.98 million.

b. Type or paste **sum(statvars\resid^2)**. The result of 340.361 represents the sum of squares of the error (SSE). You will compare these results with results obtained in step 4 of the next section (screen 7).

(7) 

Theory

The previous model results from assuming three different rates of change occurring simultaneously: (1) the rate at which the population is *increasing* is proportional to the current population size: $\frac{dy}{dt} = k_1 y$,
(2) because of natural causes, the rate at which the population is *decreasing* is also proportional to the current population size: $\frac{dy}{dt} = k_2 y$, and
(3) because the two above rates of change interact with competition, the rate at which the population is *decreasing* is proportional to $p(p - 1)/2$ interactions with competition: $\frac{dy}{dt} = -k_3 \frac{y(y-1)}{2}$.

Since simultaneous variations are additive, you can combine the three rates of change to form the differential equation $\frac{dy}{dt} = k_1 y - k_2 y - k_3 \frac{y(y-1)}{2}$, which can be written in the form $by - \frac{b}{c} y^2$. This equation has a solution:

$y(t) = \frac{c}{1+ae^{-bt}} \approx \frac{394.926}{1+1.63775\text{E}19 * e^{-0.022489t}}$ as the logistic growth curve.

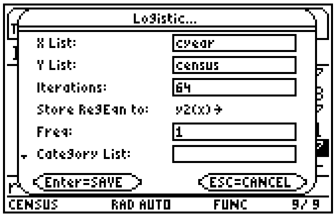
As t gets very large, $y(t)$ approaches c = 394.926 million. Not just populations follow this general S-shape. Number of sales, profits, learning, or a number of variables start increasing at a slow rate, increase more rapidly, then level off for the same S-shape.

Logistic Regression

This regression has the same general shape as Logist83 as explained at the end of this section.

Example: Use the U. S. Census data from the previous section.

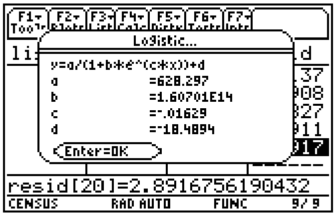
1. From the Stats/List Editor, press [F4] **Calc**, **3:Regressions**, **B:Logistic** with X List: **cyear**, Y List: **census**, Iterations: **64**, and Store RegEqn to: **y2(x)** (screen 8). Typically, large iteration values result in better accuracy but longer execution time, and vice versa.)



(8)

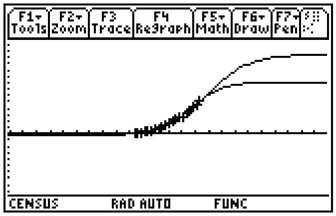
2. Press [ENTER] to display screen 9. Be patient with the BUSY signal in the lower right corner. The desired model is

$$y(t) = \frac{628.297}{1 + 1.60701\text{E}14 * e\hat{}(-0.01629 * t)} - 18.4894.$$



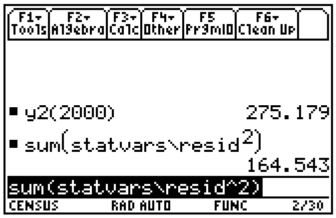
(9)

3. Building on the results of screen 6, press [◆] [GRAPH] to display screen 10, which fits the data similar to screen 6, but levels off at a higher level.



(10)

4. From the Home screen, (screen 11), you see that this model predicts a 2000 population of 275.179 million (higher than the 266.98 of screen 7), and the SSE = **164.543** < 340.361 of screen 7, indicating this model fits the data better than the model used in the previous section since the sum of squares of error is smaller.



(11)

Perhaps you will know the results of the 2000 census as you read this, and know which model gives the best prediction.

In fitting the given data better, the limiting value is higher than the 394.926 million of the previous section, but the curve has the same overall S-shape.

Note: *As t gets very large, p(t) approaches* **a + d = 628.297 - 18.4894 = 609.808** *millions.*

Theory

$y(t) = \dfrac{C}{1 + Ae^{-Bt}}$ is equivalent to $y(t) = \dfrac{a}{1 + be^{cx}}$ if you replace C with a, A with b, and $-B$ with c.

Therefore, the logistic function has the same form as Logist83 and the same S-shape. The added d in $\dfrac{a}{1 + be^{cx}} + d$ translates the curve up or down so this function can have a zero value.

Topic 49—SinReg – Trigonometric Sine Fit

Example: X represents the day of the year (equal intervals of every thirtieth day), and Y the number of daylight hours in Alaska (from the TI-83 Guidebook).

X (day) list1	1	31	61	91	121	151	181	211	241	271	301	331	361
Y (hours) list2	5.5	8	11	13.5	16.5	19	19.5	17	14.5	12.5	8.5	6.5	5.5

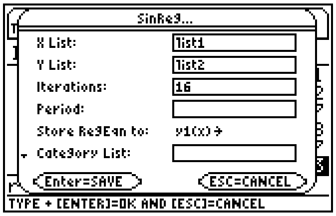
The sine regression fit is particularly accurate for data that is cyclical, such as the seasonal number of daylight hours in Alaska.

Create a new folder named **sinregr**.

With X values stored in **list1** and Y values stored in **list2**:

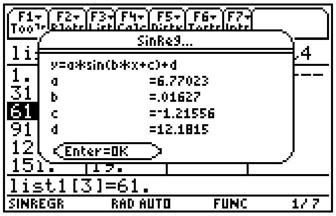
1. From the Stats/List Editor, turn off all functions and plots with [F2] **Plots**, **4:FnOff** and [F2] **Plots**, **3:PlotsOff**.

2. Press [F4] **Calc**, **3:Regressions**, **C:SinReg**, with X List: **list1**, Y List: **list2**, and Iterations: **16** (the maximum number of iterations allowed). (Typically, larger iteration values result in better accuracy but longer execution times, and vice versa.) Set Store RegEqn to: **y1(x)**, and turn other functions off. Leave the Period: field blank (screen 12).



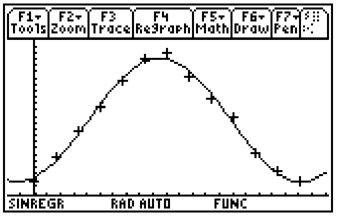
(12)

3. Press [ENTER] to display screen 13. Observe the sine regression for the above data is y = **6.77023 ∗ sin(.01627x – 1.21556) + 12.1815**. It is assumed that **x** is an angle measured in Radians, regardless of the angle mode setting.

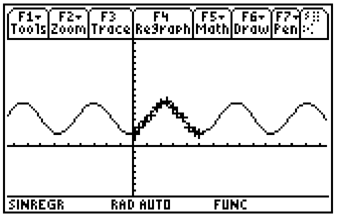


(13)

4. Press [F2] **Plots** to set up and define **Plot 1** as Plot Type: **Scatter**, Mark: **Plus**, x: **list1**, y: **list2**, and then press [F5] **ZoomData** (screen 14).

(14)


5. Press [F2] **Zoom**, **3:ZoomOut**, press [ENTER] and then press [CLEAR]. This expanded view is a better display of the periodic nature of the fit.

(15)


Note: *The* Period: *field in screen 12 specifies an estimated period. If omitted, the difference between values in* **list1** *should be equal and in sequential order. If you specify a period, the differences between the* x *values can be unequal.*

Topic 50—MedMed Fit

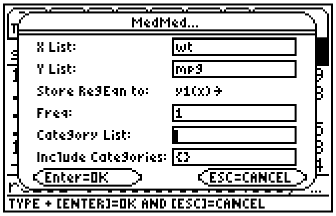
Example: Use the car data from Topic 11 with weight (lbs) and fuel consumption (mpg) for this example.

X Wt (lbs)	2795	2600	3515	3930	3115	3995	3115	4020	3175	3225	3985	2500	2290
Y mpg	23	23	19	17	20	17	22	17	19	19	17	29	28

From the Mode screen, change the current folder to **CARS**. Check to see if you still have the **wt** and **mpg** lists. If not, enter the data from the above table.

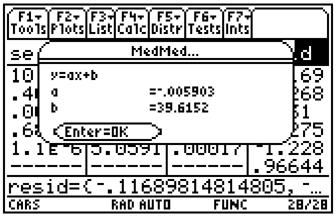
1. From the Stats/List Editor, turn off all functions and plots with [F2] **Plots**, **4:FnOff** and [F2] **Plots**, **3:PlotsOff**.

2. Press [F4] **Calc**, **3:Regressions**, **3:MedMed**, with X List: **wt**, Y List: **mpg**, and Store RegEqn to: **y1(x)** (screen 16).



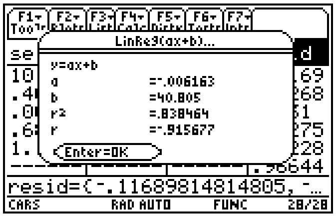
(16)

3. Press [ENTER] to display screen 17. The regression equation for the MedMed line is y = **-.005903x + 39.6152**. Notice that the slope is **-.005903**.



(17)

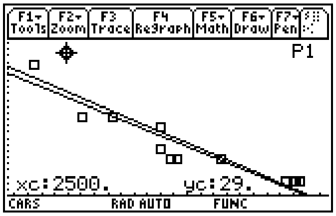
4. Press [F4] **Calc**, **3:Regressions**, **2:LinReg (ax+b)**, on X List: **wt**, Y List: **mpg**, Store RegEqn to: **y2(x)**, and then press [ENTER] (screen 18). The equation of the linear regression is y = **-.006163x + 40.805**. Notice the slope is -.006163, which is more negative and a steeper slope than the -.005903 slope of the MedMed line.



(18)

5. Press [F2] **Plots**, and set up and define **Plot 1** as Plot Type: **Scatter**, Mark: **Box**, X: **wt**, and Y: **mpg**. Then from the Plot Setup screen, press [F5] **ZoomData** (screen 19).

6. Press [F3] **Trace** and ▶ 11 times to highlight the data point corresponding to the Mazda Protégé (screen 19).



(19)

7. Although these lines are not far apart, the MedMed line (the lower line) is not as influenced by the highlighted point at the top of the screen, as discussed in Topic 11, screen 15, with (xc: **2500**, yc: **29**). This is because the MedMed fit uses the medians of batches of data and is resistant to unusual data points.

EXPLORATIONS

Appendix A

Installing the Statistics with List Editor Application

Texas Instruments provides the TI-89 Statistics with List Editor application free of charge. It can be installed on your calculator by:

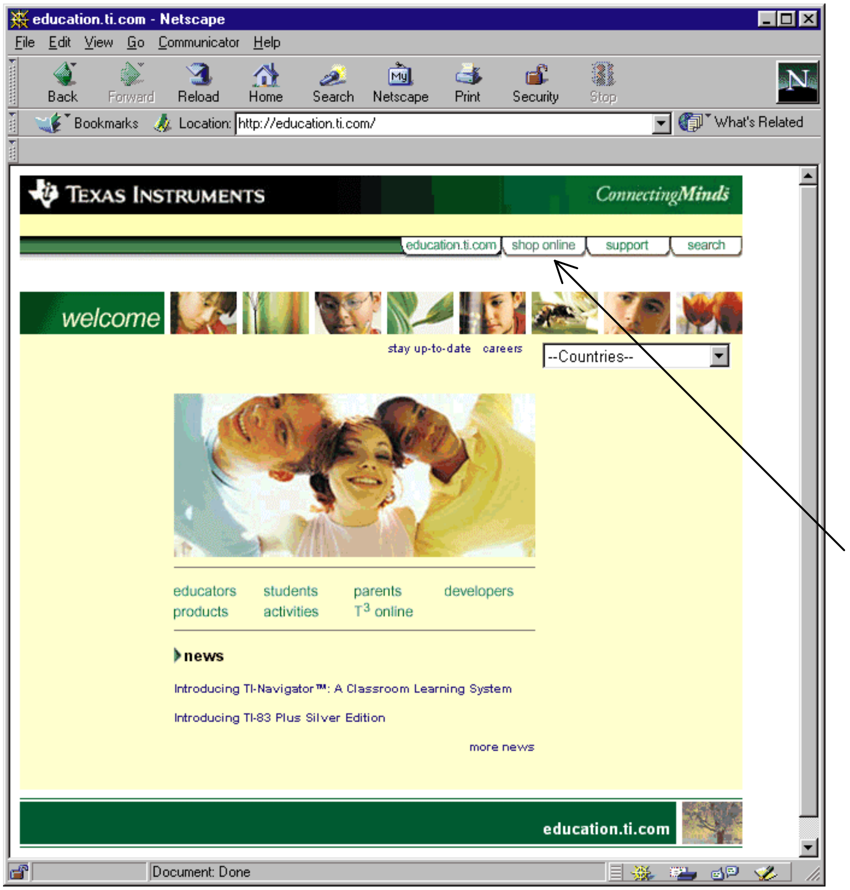
- Downloading from the TI web site.
- Copying from another calculator that has the application.

When you obtain the application from the TI web site, you can download it to a computer and then transfer it to the calculator using a TI-GRAPH LINK ™ cable and TI-GRAPH LINK software. TI-GRAPH LINK software can also be downloaded from the TI web site. The TI-GRAPH LINK cable must be purchased separately. Calculator-to-calculator transfer requires the link cable provided with the calculator.

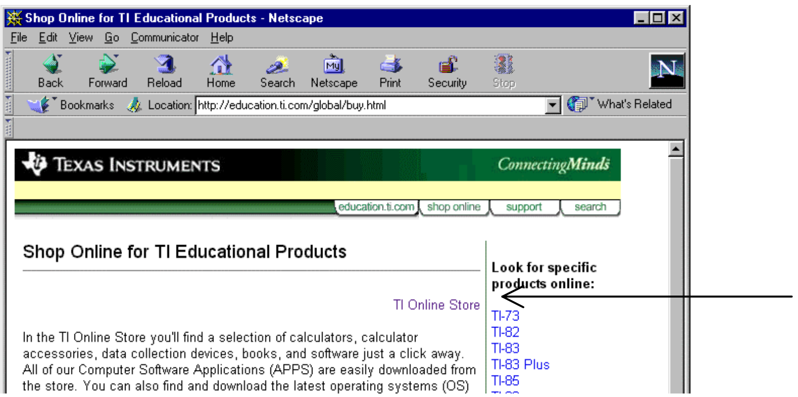
Downloading from the Internet

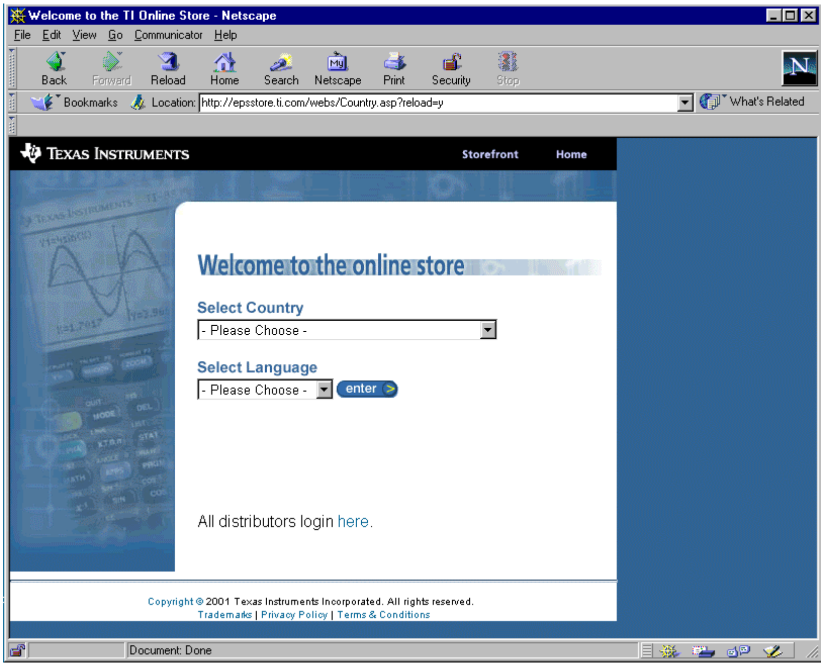
The screens in this section were current at the time this document was created. Since Web sites are modified from time to time, the screens you encounter may differ slightly. Keep in mind that you are looking for applications for the TI-89 calculator with FLASH memory.

Using the browser of your choice, go to **http://education.ti.com**

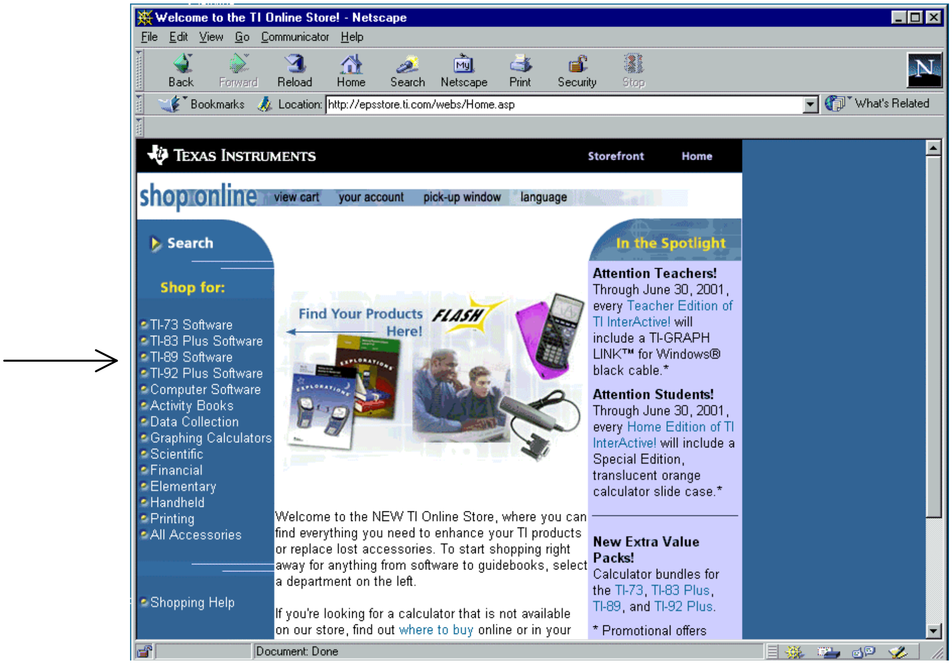


1. On the **education.ti.com** home page, click on **Shop Online** (upper right corner of the page).
2. On the **Shop Online** page, click **TI Online Store**.

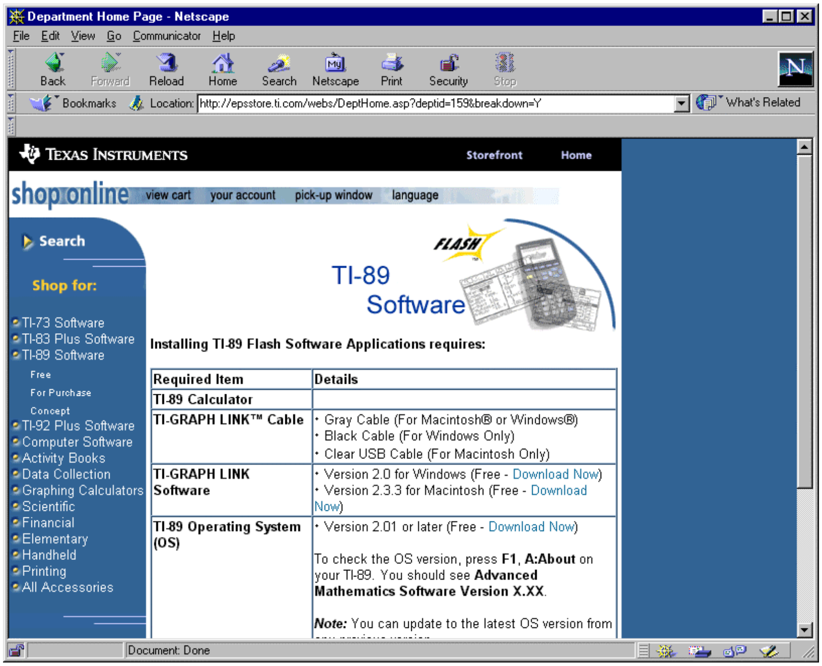




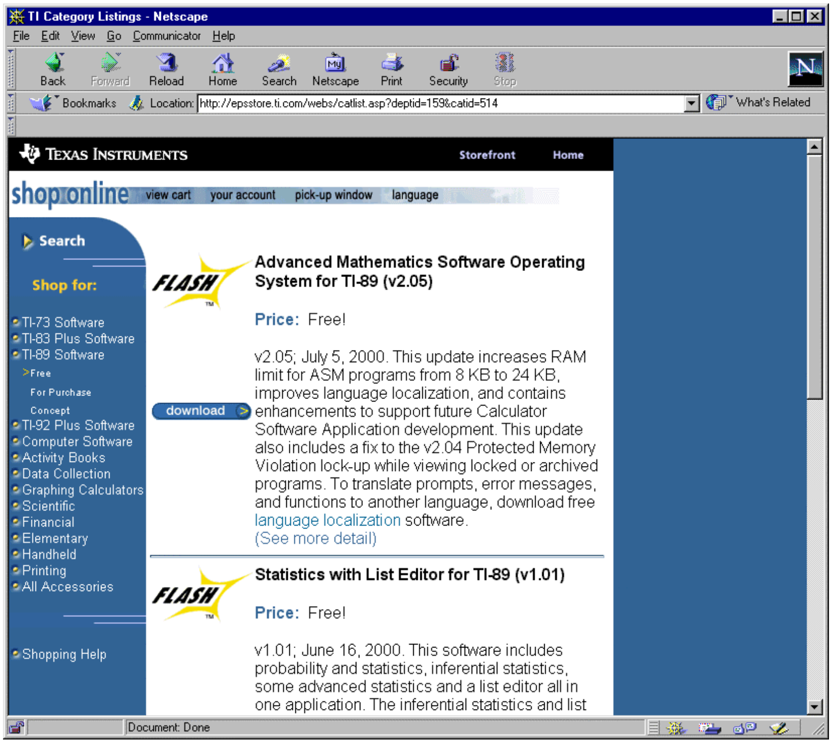
3. Select your country and language.
4. Click **Enter**.



5. In the left **Shop for:** column, click **TI-89 Software**.



6. In the left column, click on **Free**.



7. This screen shows the calculator applications available for the TI-89 calculator with FLASH memory, including an upgrade to the operating system. If you do not have the current

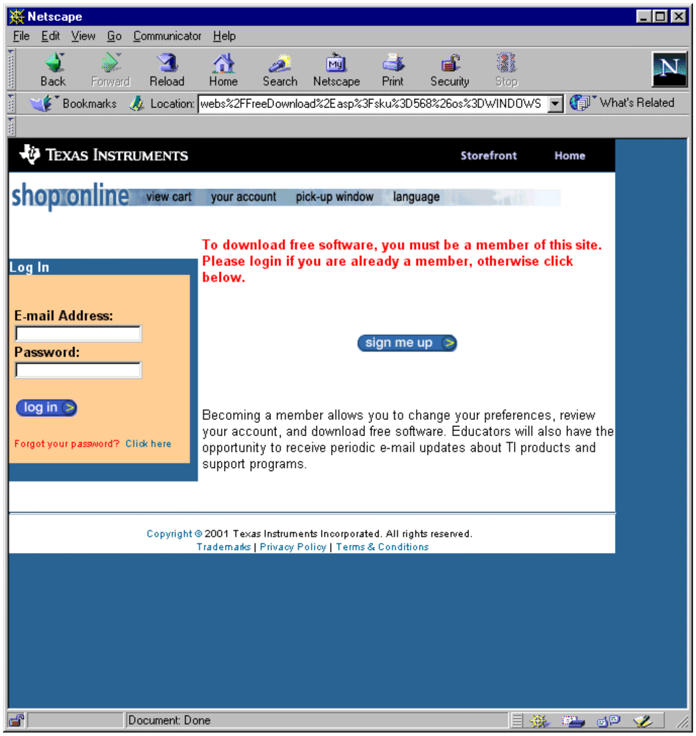
version, we suggest that you upgrade your TI-89 operating system. Check the version on the calculator by using [F1] **About** from the Home screen.

8. To download the Statistics with List Editor app, scroll down the page until you can see the description for **Statistics with List Editor for TI-89**.

9. Click **Download**.

10. Read the license agreement.

11. If you agree, click **Accept**.



12. On the Log In screen:

    a. If you are a member, log in using your e-mail address and password.

    b. If you are not a member, click **sign me up**. You must register here by providing you name, e-mail address, and shipping/billing address.

13. On the **Download/Installation Instructions** page, download the two files. (The first file, **statsle.pdf** is the documentation for the app; the second file, **TIStatle.89k**, is the application software that you will install on your TI-89.)

14. After you download the files, print the **Download/Install Instructions** page. You will need these instructions to install the app on your TI-89.

Copying from Another Calculator that has the Application

Note that these instructions are found on the Internet. You can access them by choosing the link **Another TI-89** following the **Statistics with List Editor** on www.ti.com/calc /flash/89.htm.

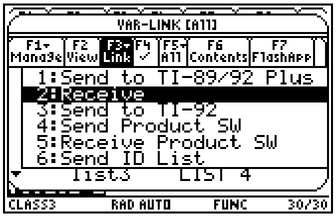
To send applications between two calculators, you must connect the calculators with the unit-to-unit cable. Then set up one calculator to send the applications and the other calculator to receive the applications.

Note: *Do not attempt to transfer an application if a low-battery message appears on either the receiving or sending calculators.*

TI-89 Receiving

To set up the TI-89 calculator to ***receive*** applications:

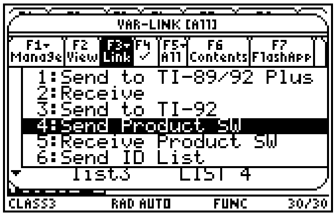
1. Press [2nd] [VAR-LINK] to display the VAR-LINK [All] screen (screen 1).
2. Press [F3] **Link** to display the menu options (screen 1).
3. Select **2:Receive**. The screen displays **VAR-LINK; WAITING TO RECEIVE** and **BUSY** in the status line of the receiving unit.

(1)


TI-89 Sending

To ***send*** the operating system to another calculator:

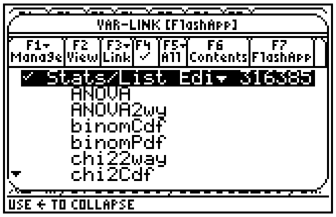
1. Press [2nd] [VAR-LINK] to display the VAR-LINK [All] screen (screen 2).
2. Press [F3] **Link** to display the menu options (screen 2).
3. Select **4:Send Product SW**. A warning message is displayed.
4. Press **ESC** to halt the process, or press [ENTER] to start the transmission.

(2)


To ***select*** applications to send to another calculator:

1. Press [2nd] [VAR-LINK] to display the VAR-LINK [All] screen (screen 3).
2. Press [2nd] [F7] **FlashApp** to display the applications.
3. Press [F4] to select an application. Selected names are marked with a checkmark (screen 3).
4. Press [F3] **Link** to display the menu options.
5. Select **1:Send to TI-92Plus/TI-89** to start the transmission.

(3)



During the transfer, the receiving unit shows how the transfer is progressing. When the transfer is complete, the sending and receiving units return to the VAR-LINK [All] screen.

EXPLORATIONS

Appendix B

Quick Reference for the Statistics with List Editor Application

This section contains a quick reference guide for the Statistics with List Editor application menus.

Using Menus

The table on the next page is a Quick Reference for the Statistics with List Editor menus. Note that when you select menu items ending with an ellipsis (…), a dialog box is displayed. When you select other menu items, the function selected is inserted at the cursor location on the entry line.

Many of the menu items are also available in the **CATALOG** (select [CATALOG] or [CATALOG] followed by **F3:Flash Apps**).

The T## after a menu item indicates that an example of its use is given in the Topic number shown.

Quick Reference

F1: TOOLS
- 3:SetUp Editor… (T01)
- 5:Copy
- 6:Paste
- 7:Clear a-z…
- 8:Clear Editor
- 9:Format… (T01)
- A:About…

F2: PLOTS
- 1:Plot Setup… (T03)
- 2:Normal Prob Plot…(T18)
- 3:PlotsOff (T03)
- 4:FnOff (T03)

F3: LIST
- 1:Names…
- 2:Ops
- 3:Math
- 4:Attach List Formula…
- 5:Delete Item

Ops
- 1:Sort List … (T02)
- 2:Sort List, Adjust All…
- 3:dim(
- 4:Fill…
- 5:seq((T01)
- 6:cumSum((T04)
- 7:ΔList(
- 8:augment((T01)
- 9:left(
- A:mid(
- B:right(

F4: CALCULATIONS
- 1:1-Var Stats… (T05, T09)
- 2:2-Var Stats… (T08)
- 3:Regressions
- 4:Probability
- 5:CorrMat… (T11)
- 6:Show Stats… (T01)

Regressions
- 1:LinReg(a+bx) …
- 2:LinReg(ax+b) …
- 3:MedMed… (T50)
- 4:QuadReg… (T12)
- 5:CubicReg…
- 6:QuartReg…
- 7:LnReg… (T12)
- 8:ExpReg… (T12)
- 9:PowerReg… (T12)
- A:Logist83… (T48)
- B:Logistic… (T48)
- C:SinReg… (T49)
- D:MultReg… (T42)

Math
- 1:min(
- 2:max(
- 3:mean((T01)
- 4:median(
- 5:sum((T11)
- 6:product(
- 7:stdDev((T21)
- 8:variance(
- 9:stdDevPop(
- A:varPop(

F5:DISTRIBUTIONS
- 1:Shade
- 2:Inverse
- 3:Normal Pdf… (T18)
- 4:Normal Cdf… (T18)
- 5:t Pdf… (T31)
- 6:t Cdf… (T31)
- 7:Chi-square Pdf…
- 8:Chi-square Cdf… (T31)
- 9:F Pdf…
- A:F Cdf…
- B:Binomial Pdf… (T16)
- C:Binomial Cdf… (T16)
- D:Poisson Pdf… (T35)
- E:Poisson Cdf… (T35)
- F:Geometric Pdf… (T17)
- G:Geometric Cdf… (T17)

Probability
- 1:rand83((T20)
- 2:nPr(
- 3:nCr((T34)
- 4:!
- 5:randInt((T14)
- 6:.randNorm((T18)
- 7:randBin((T15)
- 8:randSamp((T14)
- 9:rand(
- A:RandSeed… (T14)

F6:TESTS
- 1:Z-Test… (T27)
- 2:T-Test… (T31)
- 3:2-SampZTest… (T29)
- 4:2-SampTTest… (T32)
- 5:1-PropZTest… (T26)
- 6:2-PropZTest… (T28)
- 7:Chi2 GOF… (T30)
- 8:Chi2 2-way… (T30)
- 9:2-SampFTest… (T38)
- A:LinRegTTest… (T33)
- B:MultRegTests… (T42)
- C:ANOVA… (T39)
- D:ANOVA2-Way… (T40)

F7:INTERVALS
- 1:ZInterval… (T23)
- 2:TInterval… (T31)
- 3:2-SampZInt… (T25)
- 4:2-SampTInt… (T32)
- 5:1-PropZInt… (T22)
- 6:2-PropZInt… (T24)
- 7:LinRegTInt… (T33)
- 8:MultRegInt… (T42)

EXPLORATIONS

Appendix C

Program Listings for exsmooth and forecast

You can download the files for these programs from the Internet at **http://education.ti.com.** Search for the program name.

Program exsmooth.89p

```
()
Prgm
FnOff :ClrIO:ClrHome:PlotsOff
PlotsOn 2:PlotsOn 3
setMode("Graph","FUNCTION")
Ø→yscl
Dialog
Title "Exponential Smoothing"
Request "dataList",lyy
Request "smooth constant",aa
Request " trend constant",bb
EndDlog
expr(lyy)→ly:dim(ly)→n
expr(aa)→a:expr(bb)→b
seq(x,x,1,n,1)→lx:lx→lb:{}→lr
{}→lf:{}→lt:ly[1]→lf[1]:2→j:n-1→m
If b=Ø Then:Goto lbl1
Else:ly[2]-ly[1]→lb[1]:ly[1]→lt[1]
    lb[1]+lt[1]→lf[1]:3→j:n-2→m
    Goto lbl2
EndIf
Lbl lbl1
For i,2,n
a*ly[i]+(1-a)*lf[i-1]→lf[i]
EndFor
lf[n]→p
Goto lbl3
Lbl lbl2
For i,2,n-1
a*ly[i+1]+(1-a)*(lt[i-1]+lb[i-1])→lt[i]
b*(lt[i]-lt[i-1])+(1-b)*lb[i-1]→lb[i]
lt[i]+lb[i]→lf[i]
EndFor
lf[n-1]→p
seq(x,x,3,n+1,1)→lxx
Goto lbl4
Lbl lbl3
seq(x,x,2,n+1,1)→lxx
Lbl lbl4
NewPlot 2,2,lxx,lf,,,,3
NewPlot 3,2,lx,ly,,,,5
ZoomData:Trace
Output 1,1,"forecast1 ="
  Output 1,65,p
If b=Ø Then:Goto lbl5
Else
  Output 1Ø,1,"forecast2 =":Output
  1Ø,65,p+lb[n-1]
Output 3Ø,9,"slope ="
  Output 3Ø,67,lb[n-1]
EndIf
Lbl lbl5
Pause :ClrIO
For i,j,n
  ly[i]-lf[i-j+1]→lr[i-j+1]
EndFor
sum(lr^2)/m→d
sum(abs(lr))/m→e
Output 1,1,"mse ="
  Output 1,35,d
Output 1Ø,1,"mad ="
  Output 1Ø,35,e
EndPrgm
```

Program forecast.89p

```
()
Prgm
FnOff :ClrIO:ClrHome:PlotsOff
PlotsOn 3:Ø→yscl
Dialog
Title "Forecast with T*S*I Model"
Text "Must have Quarters or Months"
Text "data for Complete Years."
Request "Name of Data List",xx2
EndDlog
expr(xx2)→x2:dim(x2)→n
seq(x,x,1,n,1)→x1
{"4","12"}→no
NewPlot 3,2,x1,x2,,,,5
ZoomData:Trace
Dialog
Title "Moving Average Smoothing"
Text "4 or 12 Term Moving Average?"
DropDown "Number:",no,aa
EndDlog
If aa=1 Then
4→m
Else
12→m
EndIf
dim(x2)→n
n+m→r
r+Ø.5→xmax
newMat(r,8)→me
Fill 1E-ØØ8,me
For i,1,n
x1[i]→me[i,1]
x2[i]→me[i,2]
EndFor
For i,n+1,r
i→me[i,1]
EndFor
m/2+1→a:n-m→b
n-m/2→c
newList(b)→x3:newList(b)→x4:a→k
For i,1,b
me[k,1]→x3[i]
1+k→k
EndFor
For i,1,b
Ø→t:i→j:j+m-1→e
For d,j,e
t+me[d,2]→t
EndFor
i+1→j:j+m-1→e
For d,j,e
t+me[d,2]→t
EndFor
t/(2*m)→x4[i]
EndFor
PlotsOn 2
NewPlot 2,2,x3,x4,,,,5:Trace
a→k
For i,1,b
x4[i]→me[k,3]
k+1→k
EndFor
a→k
For i,k,c
me[i,2]/(me[i,3])→me[i,4]:EndFor
{"Means","Medians"}→ctr
Dialog
Title "Seasonal Indexes"
Text "For Deseasonal Data Plot"
DropDown "Uses",ctr,ff
EndDlog
a→k:newList(m)→x6:Fill
Ø,x6:newList(m-1)→t6
Fill Ø,t6
For i,1,m
1→k1
For j,k,c,m
me[j,4]→t6[k1]:k1+1→k1
EndFor
If ff=1 Then
mean(t6)→x6[i]
Else
median(t6)→x6[i]
EndIf
k+1→k
EndFor
m/(sum(x6))*x6→x6
For i,1,m/2
For j,1,m-1
x6[j]→h:x6[j+1]→x6[j]:h→x6[j+1]
EndFor
EndFor
1→k
For i,1,r/m
```

```
For  j,1,m
x6[j]→me[k,5]
1+k→k
EndFor
EndFor
For  i,1,n
me[i,2]/(me[i,5])→me[i,6]
EndFor
newList(n)→x5
For  i,1,n
me[i,6]→x5[i]
EndFor
NewPlot  4,2,x1,x5,,,,5
Trace
{"Deseasonal","M.A.Smoothed"}→choices
{"Linear fit","Quadratic fit"}→fits
Dialog
Title  "Fit Trend"
DropDown  "To",choices,bb
DropDown  "Data With",fits,cc
EndDlog
If bb=1 Then
x1→xx1:x5→xx5
Else
x3→xx1:x4→xx5
EndIf
If cc=1 Then
tistat.linregbx(xx1,xx5)
Else
tistat.quadreg(xx1,xx5)
EndIf
statvars\regeqn(x)→y6(x)
Trace
For  x,1,r
y6(x)→me[x,7]:EndFor
For  i,n+1,r
me[i,7]*me[i,5]→me[i,8]
EndFor
seq(a,a,1,r,1)→x3
For  i,1,n
me[i,2]→x4[i]
EndFor
For  i,n+1,r
me[i,8]→x4[i]
EndFor
NewPlot  2,1,x3,x4,,,,1
NewPlot  4,2,x1,x2,,,,2
ZoomData:Trace
Disp  "Seasonal indexes"
Pause  round(x6,3)
newList(Ø)→x6:1→j
For  i,n+1,r
me[i,8]→x6[j]
1+j→j
EndFor
Disp  "Forecast"
Pause  round(x6,5)
newList(n)→x6
For  i,1,n
me[i,5]*me[i,7]→x6[i]
EndFor
x2-x6→x6
sum(x6^2)/n→m
Disp  "MSE=",m
sum(abs(x6))/n→m
Disp  "MAD=",m
Pause
ClrIO
Disp  "In matrix me, t in c1,":
Disp  "data(y) in c2, ma in c3,":
Disp  "y/ma in c4, si in c5,":
Disp  "y/si in c6, fit in c7,":
Disp  "c7*si in c8.  Press GRAPH":
Disp  "then Trace for plots."
EndPrgm
```

Index

L

M

N

O

P

R

S

T

U

V

W

X