

Minitab@Oneonta.Manual:

Selected Introductory Statistical and Data Manipulation Procedures

Gordon & Johnson
2002 Minitab version 13.0

Minitab@Oneonta.Manual: Selected Introductory Statistical and Data Manipulation Procedures

This manual was donated by the authors to support the James A. Fraley Scholarship Fund.

Authors:

Lisa A. Gordon, Adjunct Lecturer: Statistics, Department of Mathematical Sciences, SUNY Oneonta

Steven D. Johnson, Senior Staff Associate, Institutional Research & Adjunct Lecturer: Statistics, Department of Mathematical Sciences, SUNY Oneonta

Comments & Suggestions:

Comments and suggestions may be forwarded to: GORDONLA@ONEONTA.EDU or JOHNSOSD@ONEONTA.EDU.

James A. Fraley Scholarship:

This manual was donated to support the James A. Fraley Scholarship at the SUNY College at Oneonta. All proceeds aid in the growth of the scholarship's endowment. The Fraley scholarship assists students majoring in the field of statistics at SUNY Oneonta. For information about the Fraley Scholarship contact the Department of Mathematical Sciences, Fizzelle Hall, SUNY, Oneonta NY 13820 (607-436-3708).

Printing:

First Printing: August 2000. Revised/reprinted July 2001, August 2002

Disclaimer:

This document is intended for the use by students in STAT 101: Introductory Statistics and STAT 141: Statistical Software: Minitab, both courses offered at SUNY Oneonta. It is not intended to be a comprehensive documentation of all that Minitab can provide with regard to a particular statistic or test. The content of the document may change as needed to meet its function as a reference manual for STAT 101/141. No living trees were harmed in the printing of this manual.

Copyright:

Copyright © 2000, 2001, 2002 by the authors

No rights reserved. Any part of this publication may be reproduced for educational purposes provided that appropriate credit is given to its authors and SPSS, Inc. Use on the SUNY Oneonta campus must be coordinated through the authors and the Department of Mathematical Sciences.

MINITAB™ is a trademark of Minitab, Inc. and is used herein with the owner's permission. The inclusion of Minitab Statistical Software input and output contained in this manual is done with permission of Minitab, Inc. Screen shots were taken from Minitab version 13.0. For additional documentation, FAQ's, etc. go to Minitab's web site at <http://www.minitab.com>.

Tempus Fugit. Carpe Diem. Some assembly required.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
INTRODUCTION	3

THE MINITAB ENVIRONMENT:

GETTING STARTED WITH MINITAB.....	6
PROJECTS AND WORKSHEETS	9

MINITAB STATISTICAL PROCEDURES:

BAR CHART	11
BASIC STATISTICS (Descriptive Statistics).....	12
BINOMIAL RANDOM VARIABLES (Calculating Probabilities).....	14
BINOMIAL RANDOM VARIABLES (Obtaining a Probability Distribution).....	16
BOXPLOT (BOX-AND-WHISKER PLOT)	17
CHI-SQUARE (χ^2): Test of Independence.....	19
COEFFICIENT OF DETERMINATION (r^2) and the SUMS OF SQUARES IN LINEAR REGRESSION	20
CONFIDENCE INTERVAL FOR ONE POPULATION PROPORTION	22
CONFIDENCE INTERVAL FOR μ (σ UNKNOWN): ONE-SAMPLE t	23
CONFIDENCE INTERVAL FOR μ (σ KNOWN): ONE-SAMPLE Z.....	24
CROSS TABULATION TABLE (Contingency Table).....	25
DISCRETE RANDOM VARIABLES (Obtaining a Probability Histogram)	27
DISCRETE RANDOM VARIABLES (Simulation).....	29
DOTPLOT	30
FITTED LINE PLOT(Linear Regression).....	31
FIVE-NUMBER SUMMARY.....	33
FREQUENCY POLYGON/AREA GRAPH.....	34
HISTOGRAM.....	35
HYPOTHESIS TEST FOR ONE POPULATION PROPORTION	36
HYPOTHESIS TEST FOR μ (σ UNKNOWN): ONE-SAMPLE t.....	37
HYPOTHESIS TEST OF MEAN (σ KNOWN): ONE-SAMPLE Z	38
LINE GRAPH.....	39

LINEAR CORRELATION.....	40
LINEAR REGRESSION	41
MEASURES OF CENTER	43
MEASURES OF VARIATION.....	44
NORMAL DISTRIBUTION (Probabilities for a Normally Distributed Variable).....	45
NORMAL PROBABILITY PLOT.....	46
NORMALLY DISTRIBUTED VARIABLE (Determining an Observation Corresponding to a Specified Percentile)	48
NORMALLY DISTRIBUTED VARIABLE (Simulating a Normal Distribution).....	49
PARETO CHART	50
PIE CHART	51
SAMPLING DISTRIBUTION OF THE MEAN (Obtaining Means for Generated Samples)	53
SAMPLING DISTRIBUTION OF THE MEAN (Simulation)	54
SCATTERPLOT.....	55
SIMPLE RANDOM SAMPLE (SRS).....	56
STEM-AND-LEAF PLOT.....	58
TALLY	60

MINITAB DATA MANIPULATION:

CALCULATE.....	62
CODE.....	65
DATA ENTRY	67
DISPLAY DATA.....	72
EDITING A WORKSHEET.....	73
EDITING CHARTS AND GRAPHS: TOOL AND ATTRIBUTE PALETTES.....	76
EXPORTING TABLES AND GRAPHS	78
PATTERNED DATA	79
PROJECT MANAGER	82
REPORTPAD	84
SPLIT WORKSHEET	85
APPENDIX A: Data File Descriptions.....	86
APPENDIX B: CONDUCTING A HYPOTHESIS TEST.....	87
INDEX	88

INTRODUCTION

This manual is intended as a reference guide to the Minitab statistical software package used by students in the introductory statistics classes at SUNY Oneonta. The manual provides a "how to" approach to each topic presented. It does not discuss every option available for a given topic, but rather provides sufficient instruction to obtain the basic output one might need at an introductory level. At the same time, this is not a statistics text. Therefore, the user will need to refer to a statistics text for definitions and further clarification of the procedures presented herein.

DOCUMENT OVERVIEW

Most topics are presented on no more than one or two pages and include a brief overview, a **PATH** that lists the steps to follow, and an **EXAMPLE** of how to use the statistical topic. Within each procedure's Example, Minitab screen shots show how to access and complete the desired Minitab command. Minitab output is presented so that the user may see the results of the command selected. In many cases the output is "enhanced" through superimposed notes, which point out various components of the output.

CONVENTIONS USED IN THE MANUAL

There are a number of conventions or means of presenting information that are constant throughout the manual. As one initiates a Minitab command, a **Dialog Box** will open. It is within these Dialog Boxes that one will enter all the information needed to produce the desired output.

PATH: All bold print, the PATH indicates all steps needed to successfully complete the minimum requirements for the Minitab command to provide output.

OPTIONS: With the selected Minitab command, using the Options noted below the PATH will provide additional output or enhance the output.

ADDITIONAL REQUIREMENTS: For some Minitab features it is necessary to go to a second Dialog Box and enter additional information BEFORE selecting the Ok button.

RELATED GRAPHS: For some statistical procedures a reference may be made to a related graph.

RELATED STATISTICAL PROCEDURE: For some Tables and Graphs a related statistical procedure may be noted.

PROCESS: A PROCESS notation includes steps that do not directly produce output, but rather serve as an intermediate step leading to output.

'**VARIABLES**' used in the examples appear in bold, all capital letters within single quotation marks.

"**Cells**" into which one would enter information are presented in bold lettering within double quotation marks.

Words in bold italics are cross-referenced through the document index.

Underlined words or numbers are meant as values to be entered into an identified cell of a dialog box.

ORDERING OF TOPICS

The contents of this manual have been broken into three sections.

Section 1: The Minitab Environment - provides information on how to get started using Minitab.

Section 2: Statistical Procedures - provides an overview of many statistics and graphs encountered in an introductory statistics course.

Section 3: Data Manipulation - provides information on many way to enhance/modify Minitab worksheet data files. These procedures are not necessary to obtain statistical output, but do facilitate data analysis.

DATA

The data used for examples throughout this manual were selected from worksheets that are distributed with Minitab. They may be accessed from within Minitab using the following **PATH: File => Open Worksheet => set the "Look in:" cell to the Data folder => select the desired data file => Open**. Minitab will indicate that it is opening the worksheet in the current project. Select OK and the worksheet will be opened in the project.

The data sets used in this manual are noted below. A key to their content is located in the manual's Appendix A. Descriptions for all Minitab sample worksheets may be found by using the following **PATH: Help => Search Help => Index => enter data file name => select file => Display**.

DATA FILES:

Data: Market.MTW

Data: Pulse.MTW

Data: Schools.MTW

Data: Track15.MTW

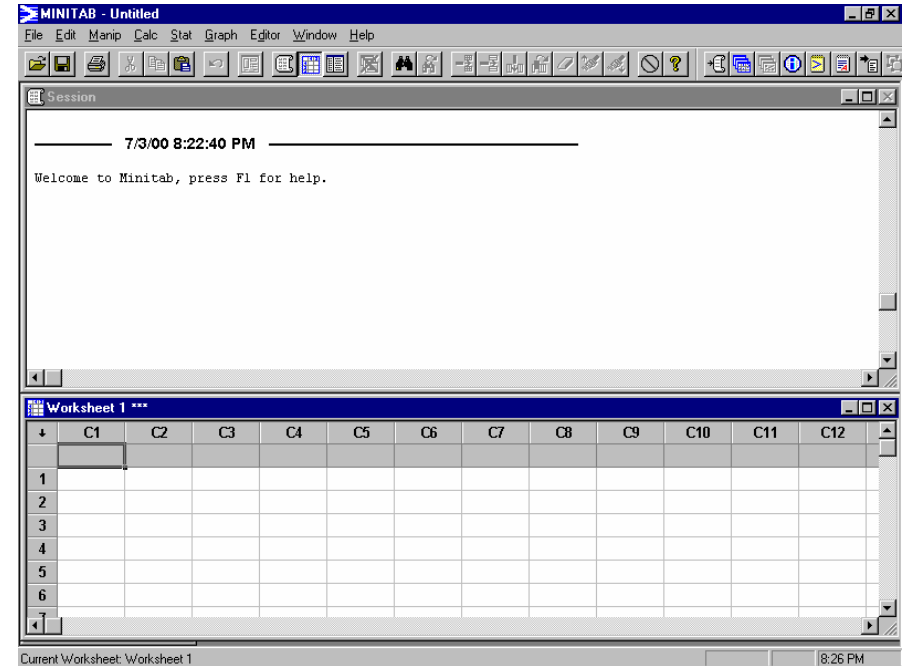
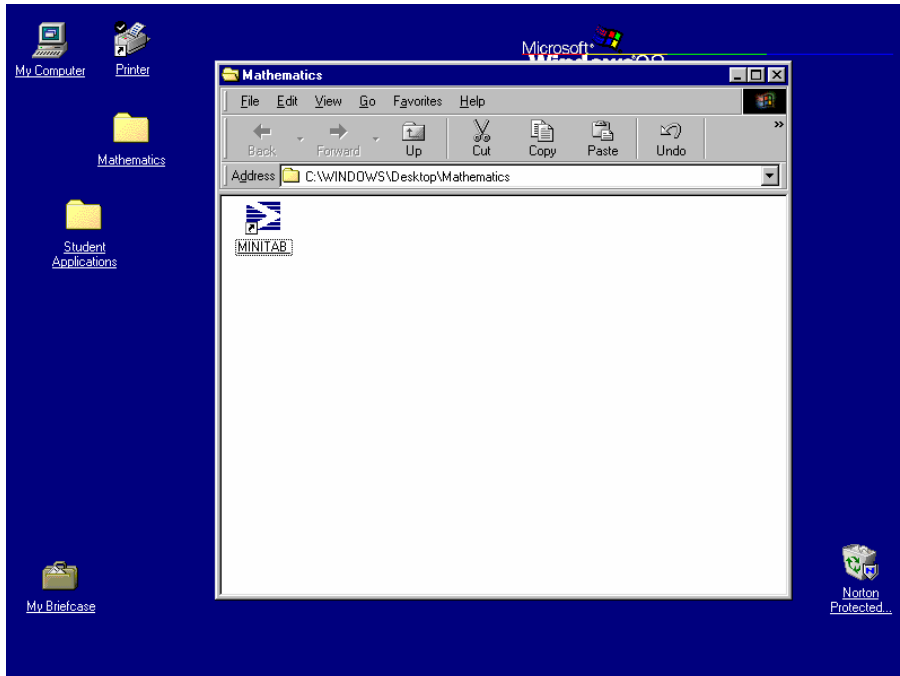
THE MINITAB ENVIRONMENT: HOW TO GET STARTED USING MINITAB

GETTING STARTED WITH MINITAB

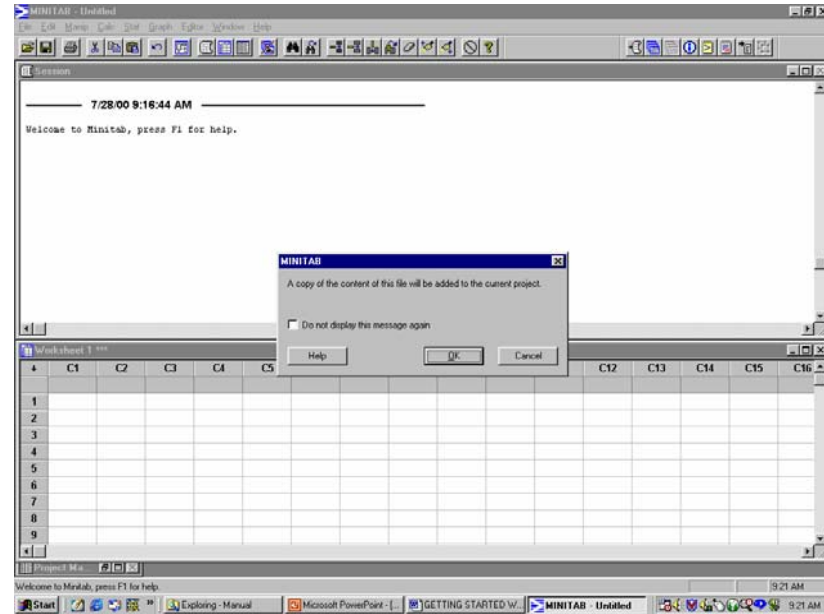
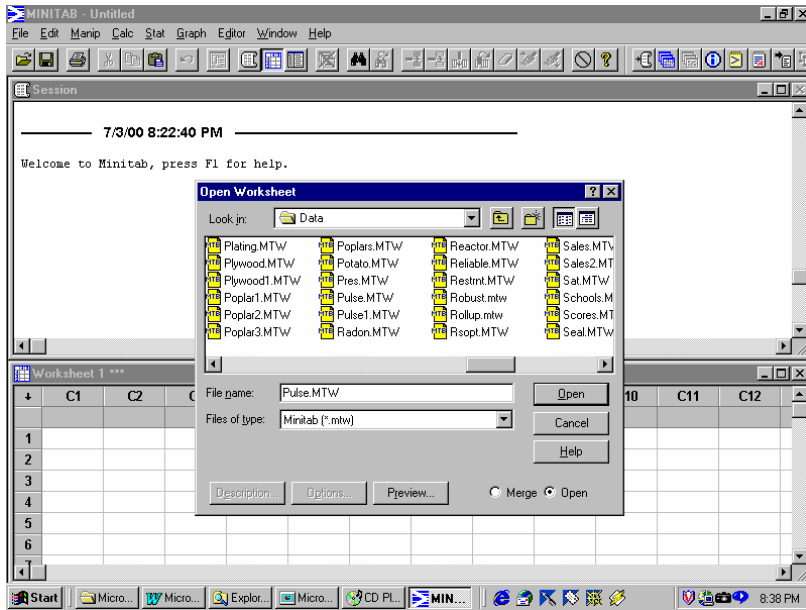
WHERE TO FIND MINITAB: Minitab can be accessed from virtually any PC-based computer lab on campus. For a listing of the on-campus computer labs and hours of availability, go to the following URL: http://www.oneonta.edu/ocs/sti/computer_map.html

HOW TO ACCESS MINITAB: The location of the Minitab icon may vary from one lab to another. In most cases, it will be located on the desktop in a folder named **Mathematics** or **Student Applications**.

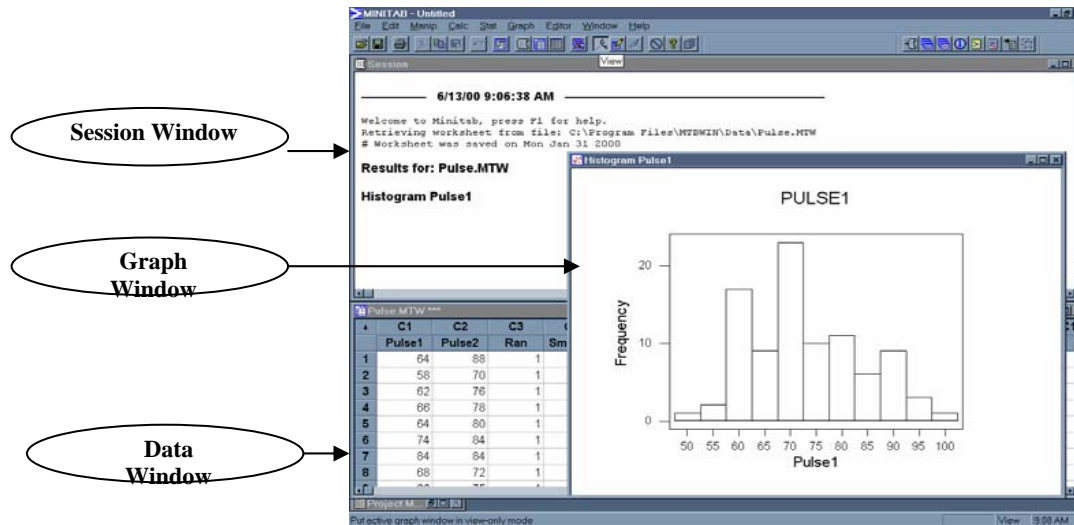
TO START MINITAB: double-click on the Minitab icon. The following will appear:



To open a Worksheet or Project: PATH: File => Open Worksheet or Open Project => select a data file => Open.



Minitab Windows:



The **Session Window** displays text output of the analysis for an active worksheet.

The **Graph Windows** displays all graphs (up to 15) that are created. Each graph will appear in its own window.

Data Window (Worksheet) displays the worksheet(s) being used.

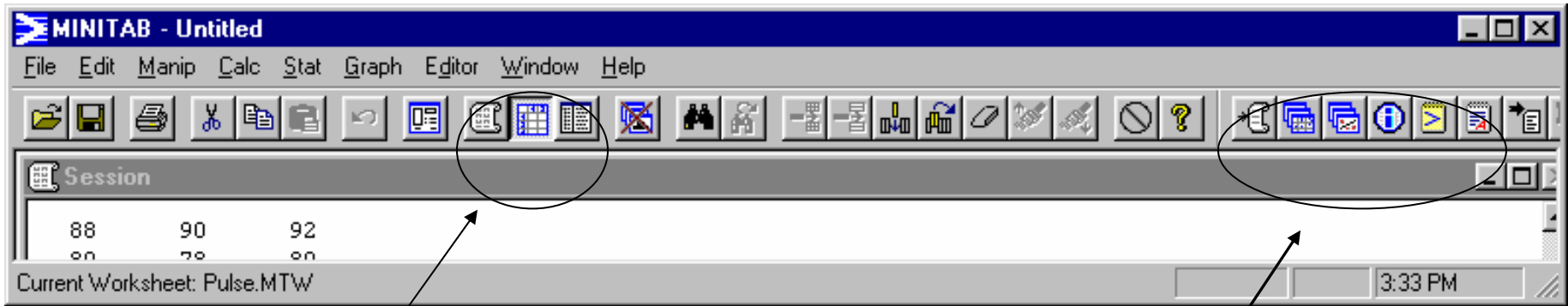
Project Manager: A Project Manager serves as framework within which the previously noted windows and those noted below are organized. Additional windows include:

Info Window which contains a listing of variables and their characteristics.

History Window in which an ongoing log of commands, etc. is maintained.

ReportPad Window which serves as a mini word processor for the compilation of related tables, graphs, etc.

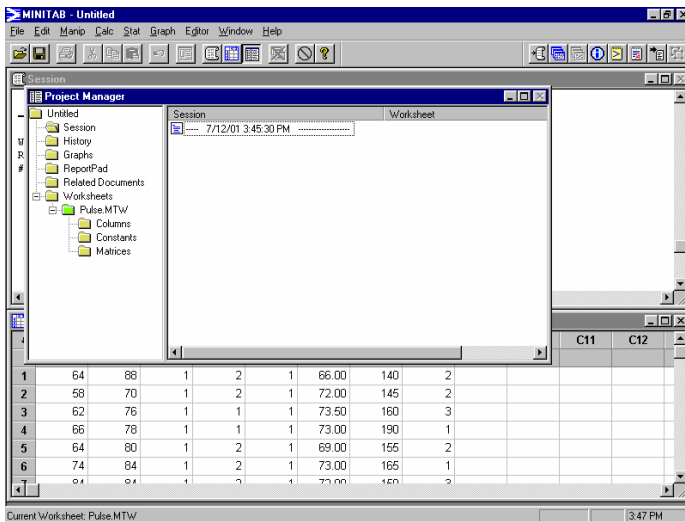
Among the icons listed just below the main menu are a number that represent direct links to these six windows.



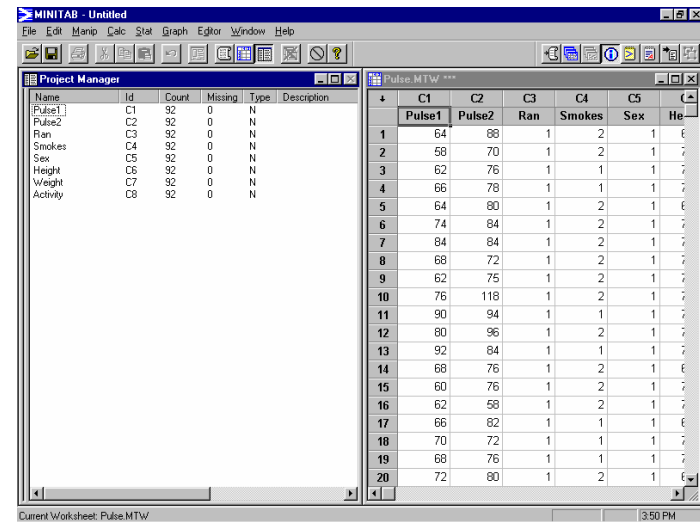
Session & Worksheet toggle icons and Project Manager Icon

Session, Worksheet, Graph, Info, History and ReportPad icons that open within a Project Manager format

Components within the Project manager:



Example showing Info Window and its corresponding Worksheet within the Project Manager framework.



PROJECTS AND WORKSHEETS

Each time Minitab is opened without specifying a file, a new (empty) **Project** will appear. A **Project** contains a Worksheet or Worksheets, Session Window output, and any associated graphs. Saving a **Project** saves all of these components. Upon reopening the project, all windows will reappear as they were when saved. A **Worksheet** contains all the data from a data set. It represents one of several components of a Minitab **Project**. One, or more **Worksheets** may be contained within a Project. **Worksheets** may be saved as individual files.

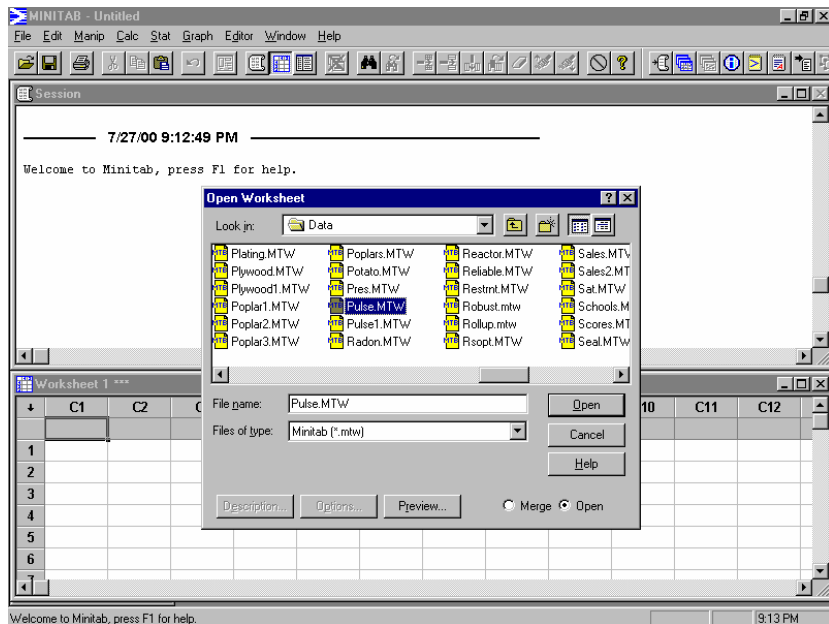
Projects are saved with the extension **.MPJ**. **Worksheets** are saved with the extension **.MTW**.

PATH: File => Open Worksheet => select a worksheet => select “Open.” (Note: A small dialog box will open indicating that a copy of the selected worksheet will be added to the project. Select “OK”.)

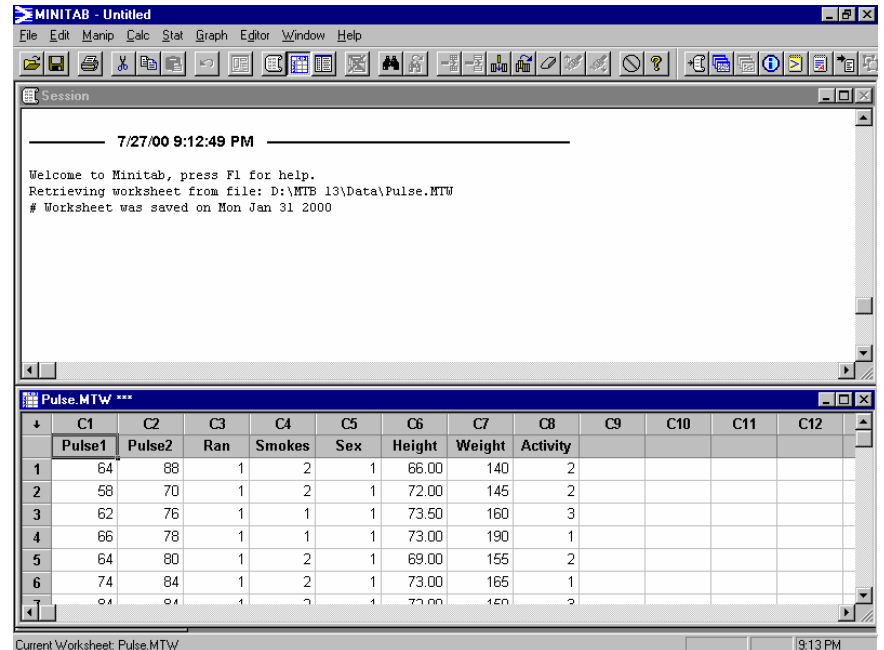
OPTIONS: To open a previously saved Project, follow the path above, substituting “Open Project” in place of “Open Worksheet.” Additionally, a location for where to find the file must be provided by using the “Look in” drop-down box.

EXAMPLE: Open the worksheet Pulse.MTW.

Dialog Box Input: Existing worksheets are displayed in alphabetical order. Select a worksheet.



Minitab Output: The worksheet Pulse.MTW is opened.



MINITAB STATISTICAL PROCEDURES: INTRODUCTORY STATISTICS AND GRAPHS

BAR CHART

A *Bar Chart* is a chart used to present a *frequency distribution*. It consists of bars (rectangles) of equal width. The lengths of the bars are proportional to the *frequencies* they represent.

DATA: Pulse.MTW

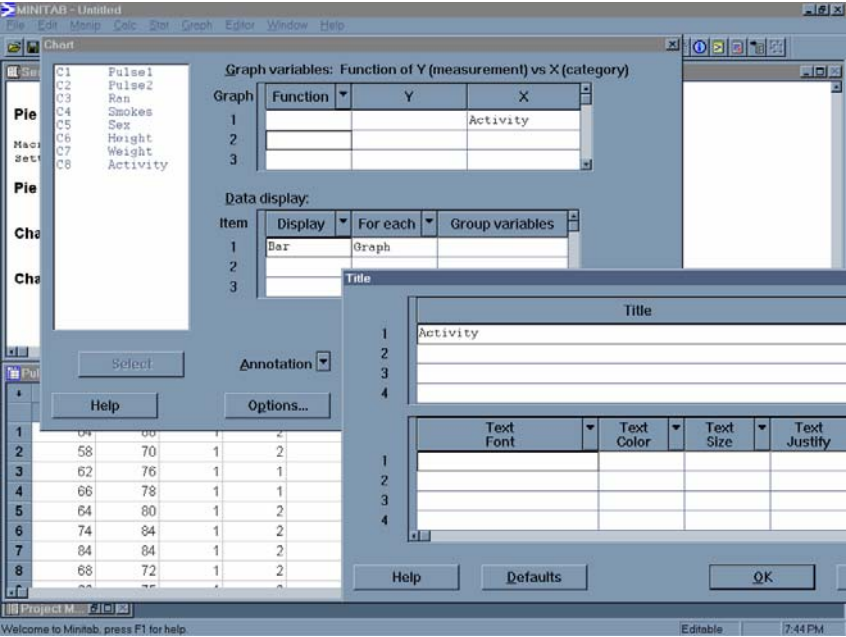
PATH: Graph => Chart => move variable to the “X-variable” cell.

OPTIONS: To add a title, use the “**Annotation**” drop-down box, and select “**Title.**” Type in a title.

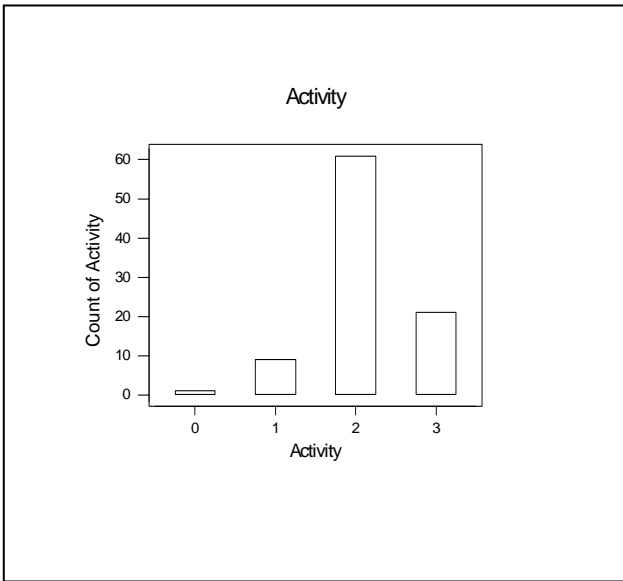
RELATED STATISTICAL PROCEDURE: Tally

EXAMPLE: Obtain a Bar Chart of ‘ACTIVITY’.

Dialog Box: Variable ‘ACTIVITY’ is moved into the “X-Variable cell”
A title has been added.



Minitab Output: Bar Chart of ‘ACTIVITY’



Activity	Count of Activity
0	1
1	10
2	60
3	20

BASIC STATISTICS (Descriptive Statistics)

Descriptive Statistics are used to summarize or describe the important characteristics of a known set of population data. The methods used can be tabular (tables) or graphical (graphs).

DATA: Pulse.

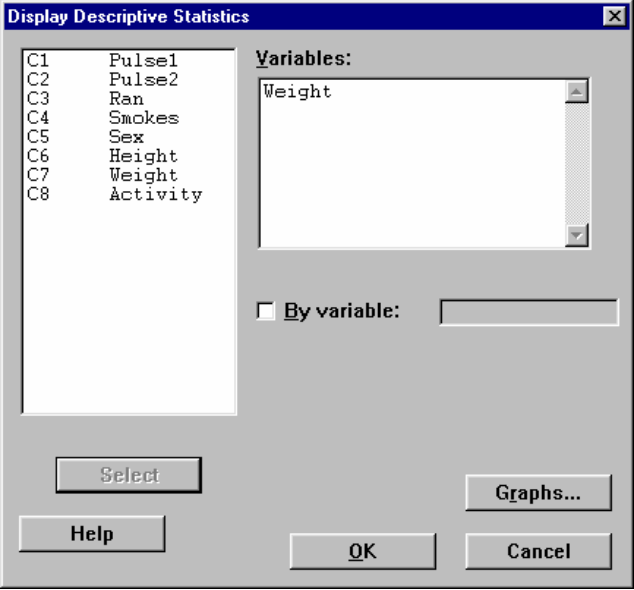
PATH: Stat => Basic Statistics => Display Descriptive Statistics

The following dialog box will appear.

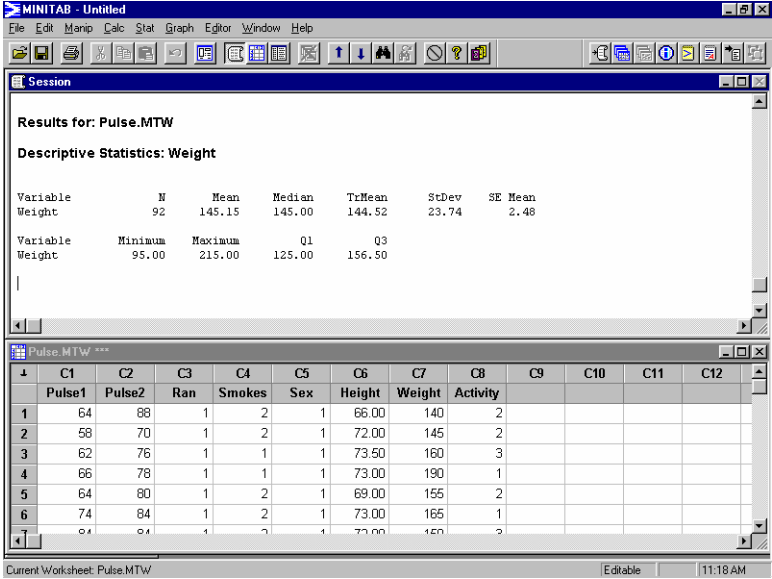
OPTIONS: In the dialog box you may also select a graph to represent your variable. The graph will appear along with the **descriptive statistics** output. See EXAMPLE 2.

EXAMPLE 1: Obtain **Descriptive Statistics** for the variable WEIGHT. Select WEIGHT by *clicking once* on it in the list of variables. Now *click Select*. The variable WEIGHT will be put into the **Variables** box. Now *click OK*. The **descriptive statistics** for the variable WEIGHT will appear in the **Session Window** as shown right.

Dialog Box Input: The variable WEIGHT is selected.



Minitab Output: Descriptive Statistics for the variable WEIGHT are displayed.



Variable	N	Mean	Median	TrMean	StDev	SE Mean
Weight	92	145.15	145.00	144.52	23.74	2.48

Variable	Minimum	Maximum	Q1	Q3
Weight	95.00	215.00	125.00	156.50

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity				
1	64	88	1	2	1	66.00	140	2				
2	58	70	1	2	1	72.00	145	2				
3	62	76	1	1	1	73.50	160	3				
4	66	78	1	1	1	73.00	190	1				
5	64	80	1	2	1	69.00	155	2				
6	74	84	1	2	1	73.00	165	1				

Now, just what do all of these statistics mean?

N: The number of cases in the worksheet for which the **descriptive statistics** were calculated.

Mean: The sum of all the weights divided by the total number of weights (92).

Median: This is the **middle** score of all the weights when they are put in order from smallest to largest.

Tr Mean (Trimmed Mean): The smallest 5% and the largest 5% (the extremes) of the weights are removed, and the **mean** is calculated on the remaining weights.

St Dev (Standard Deviation): A measure of how “spread out” the data are from one another. The larger this value, the more “spread out” are the data.

SE Mean (Standard Error of the Mean): Beyond the scope here.

Minimum: The smallest of the values.

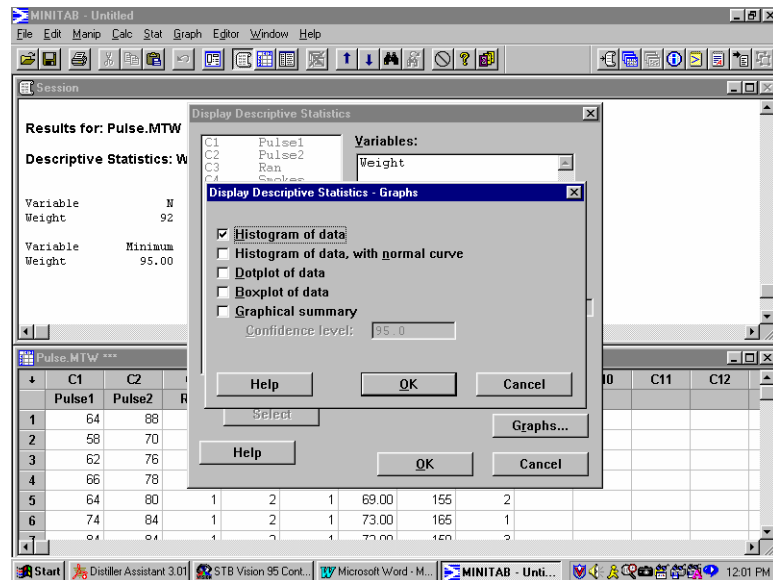
Maximum: The largest of the values.

Q1 (The first quartile or 25th percentile): The value such that 25% of the data fall below it, when the data are ordered from smallest to largest.

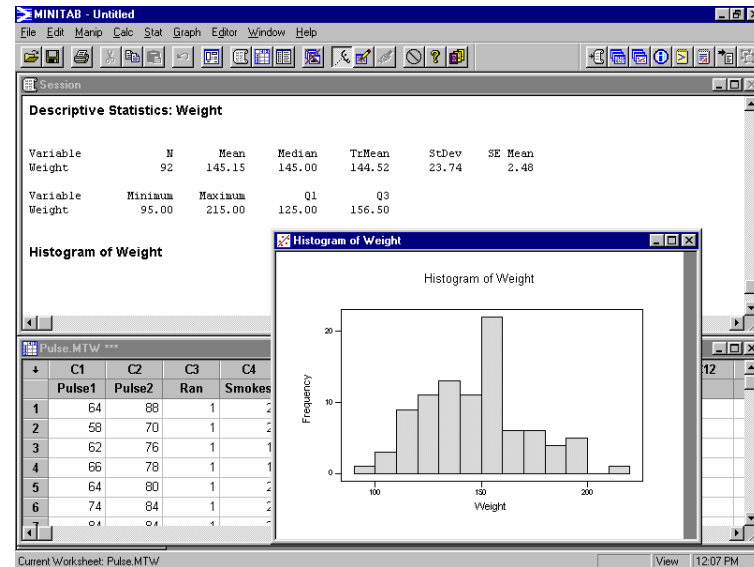
Q3 (The third quartile or 75th percentile): The value such that 75% of the data fall below it, when the data are ordered from smallest to largest.

Example 2: In the dialog box, select your variable, then **click** on the **Graphs** button. In the dialog box that appears, select the type of graph you wish to obtain, by **clicking** in the **box** to the left of it. **Click OK.** The dialog box will close. **Click OK** in the original dialog box. The graph you selected, along with the **descriptive statistics** for your variable, will be displayed.

Dialog Box Input: Histogram is selected



Minitab Output: Histogram, along with **Descriptive Statistics** is displayed.



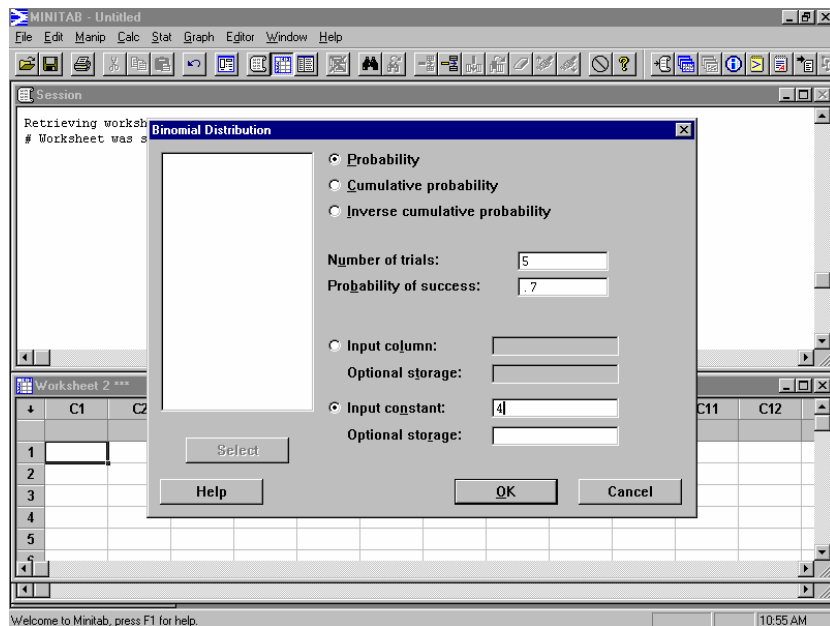
BINOMIAL RANDOM VARIABLES (Calculating Probabilities)

The procedure used to obtain **Binomial Probabilities** in Minitab, is dependent on which probabilities are to be determined. There are 2 choices. 1.) $P(X = x)$. 2.) $P(X \leq x)$. Procedures used to determine the probabilities for #1 and #2 are illustrated here. Probabilities such as $P(X \geq x)$ or $P(a \leq x \leq b)$ can be obtained by using one or both of these procedures in conjunction with basic mathematical calculations

PATH 1 ($P(X = x)$): Calc => Probability Distributions => Binomial => select the “Probability” button => type the number of trials in the “Number of trials:” cell => type the probability of success in the “Probability of success:” cell => select the “Input constant” button => type the value of X in the “Input Constant” cell.

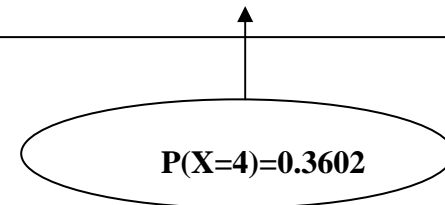
EXAMPLE: According to a survey, 70% of adults say that “staying home with family” is their favorite way of spending an evening. Let X be the number of adults in a sample of five who say staying home with family is their favorite way of spending an evening. Find $P(X = 4)$.

Dialog Box Input: “Probability” button is selected. Number of trials and probability of success are typed in their respective cells. “**Input constant**” button is selected. Value, for which the probability is sought, is typed in the “**Input constant**” cell.



Minitab Output: $P(X=4)$ is calculated.

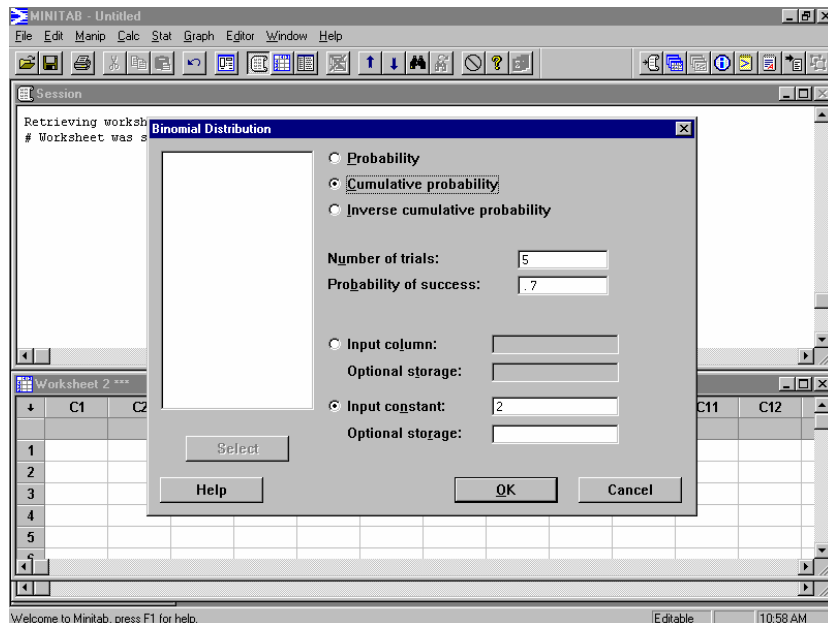
Probability Density Function	
Binomial with $n = 5$ and $p = 0.700000$	
x	$P(X = x)$
4.00	0.3602



PATH 2 ($P(X \leq x)$): Calc => Probability Distributions => Binomial => select the “Cumulative probability” button => type the number of trials in the “Number of trials:” cell => type the probability of success in the “Probability of success:” cell => select the “Input constant:” button => type the value of X in the “Input Constant” cell.

EXAMPLE: According to a survey, 70% of adults say that “staying home with family” is their favorite way of spending an evening. Let X be the number of adults in a sample of five who say staying home with family is their favorite way of spending an evening. Find $P(X \leq 2)$.

Dialog Box Input: “Cumulative probability” button is selected. Number of trials and probability of success are typed in their respective cells. “Input constant:” cell is selected. Value for x is typed in “Input constant” cell.



Minitab Output: $P(X \leq 2)$ is calculated.

Cumulative Distribution Function

Binomial with $n = 5$ and $p = 0.700000$

x	$P(X \leq x)$
2.00	0.1631

$P(X \leq 2) = 0.1631$

BINOMIAL RANDOM VARIABLES (Obtaining a Probability Distribution)

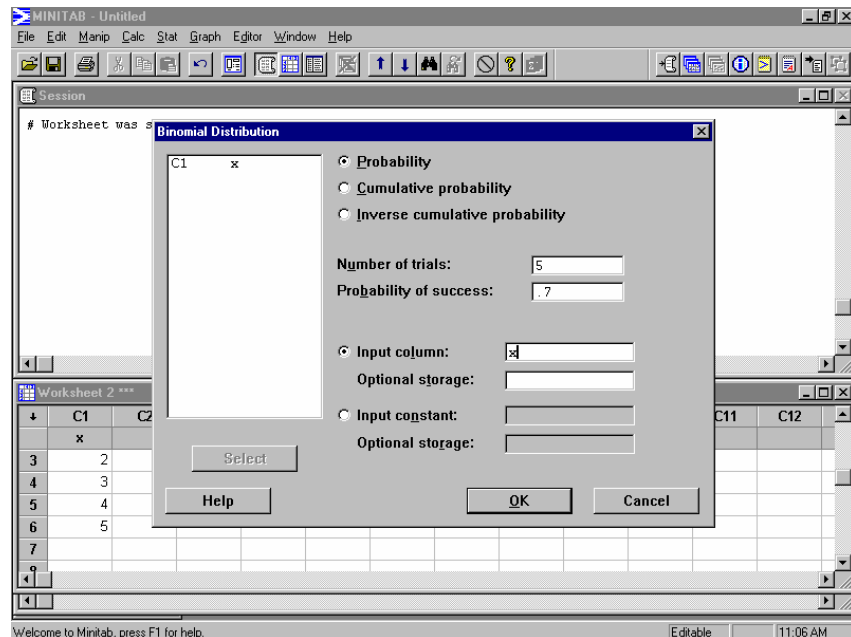
Given all possible values of a Binomial Random Variable X , and the probability of success for one trial, Minitab can be used to construct the corresponding probability distribution.

PATH: (Note: Before beginning, all possible values for X must be stored in a worksheet, and the probability of success must be known.)

Calc => Probability Distributions => Binomial => select the “Probability” button => type the number of trials in the “Number of trials:” cell => type the probability of success in the “Probability of success:” cell => select the “Input column:” button => in the “Input column:” cell, select the name of the column in the worksheet where possible values have been stored.

EXAMPLE: According to a survey, 70% of adults say that “staying home with family” is their favorite way of spending an evening. Let X be the number of adults in a sample of five who say staying home with family is their favorite way of spending an evening. Obtain the probability distribution for this variable.

Dialog Box Input: “Probability” button is selected. Number of trials and probability of success are typed in their respective cells. Input column is selected.



Minitab Output: For each value of x , the corresponding probability is calculated.

Probability Density Function

Binomial with $n = 5$ and $p = 0.700000$

x	$P(X = x)$
0.00	0.0024
1.00	0.0284
2.00	0.1323
3.00	0.3087
4.00	0.3602
5.00	0.1681

BOXPLOT (BOX-AND-WHISKER PLOT)

The **Boxplot** is used to graphically present measures of position for a given data set. Based upon the **Five-Number Summary**, the Boxplot presents a data set's Minimum, First Quartile (Q_1), Median (Q_2), Third Quartile (Q_3), and Maximum values. Minitab uses a convention wherein the whiskers for the boxplot will extend to Adjacent Points, but no further than $1.5 \cdot (IQR)$ below Q_1 or above Q_3 (where $IQR = Q_3 - Q_1$). Beyond that distance, any points will be noted by asterisks and should be examined as potential data outliers.

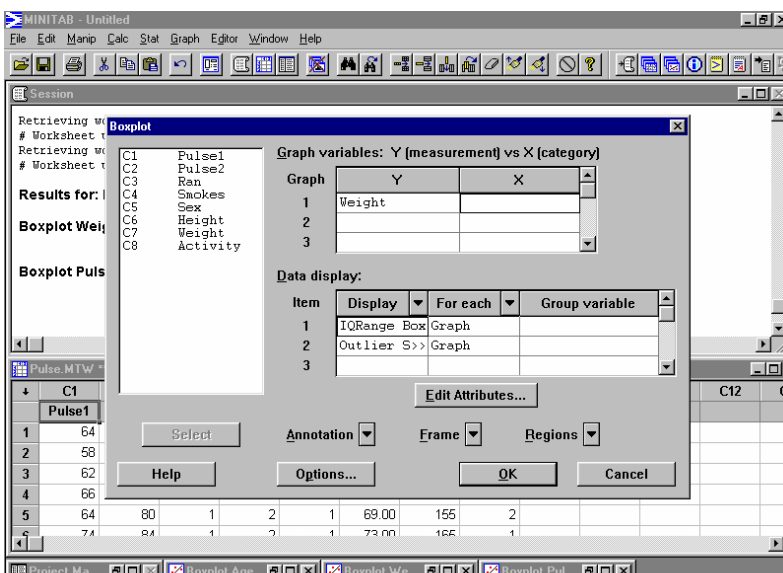
DATA: Pulse.MTW

PATH: Graphs => Boxplot => move variable to the "Y variable" (measurement) cell; insert an "X variable" if Grouped Boxplots are desired (see example 2) => OK

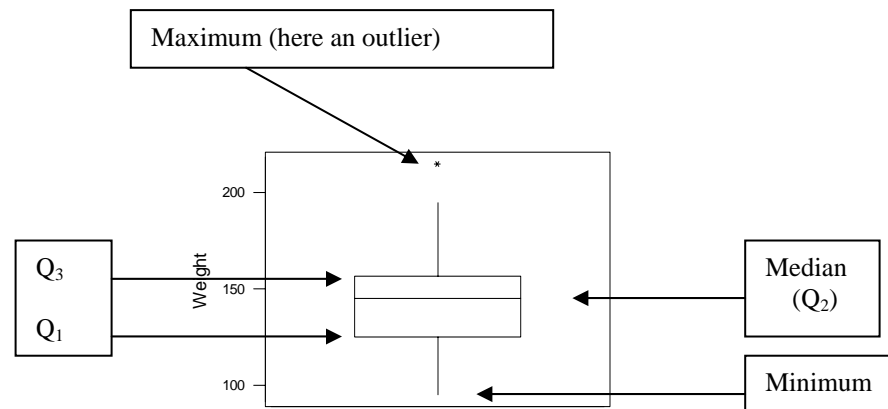
OPTIONS: Note: The default is to produce vertical boxplots. Should you prefer a horizontal Boxplot, select the "Options" button on the dialog box and choose "Transpose X and Y".)

EXAMPLE 1: Obtain a Boxplot of the 'WEIGHT' of the students.

Dialog Box Input: Move 'WEIGHT' into the "Y variable" list.

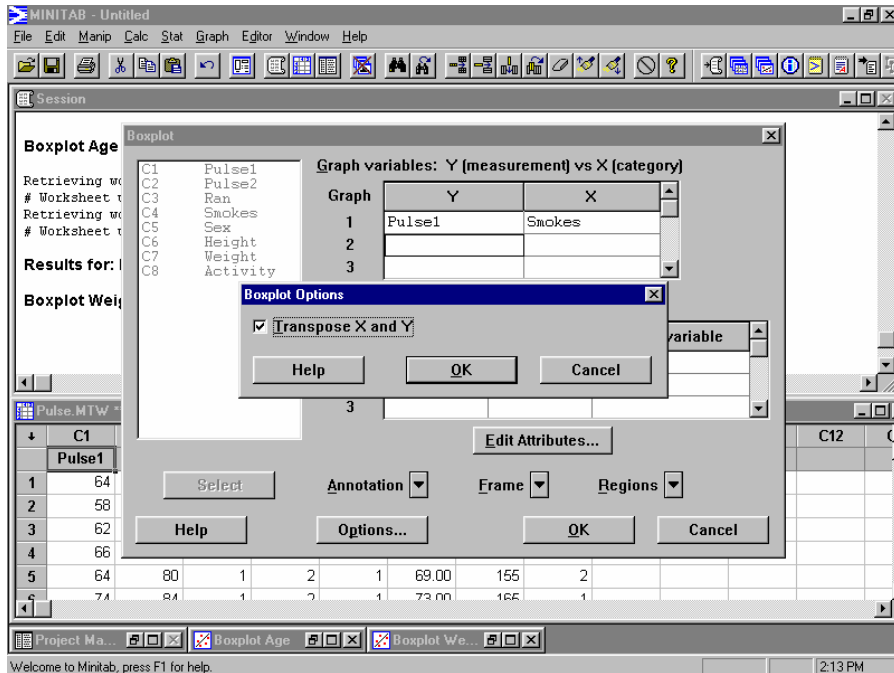


Minitab Output: Boxplot of 'WEIGHT'

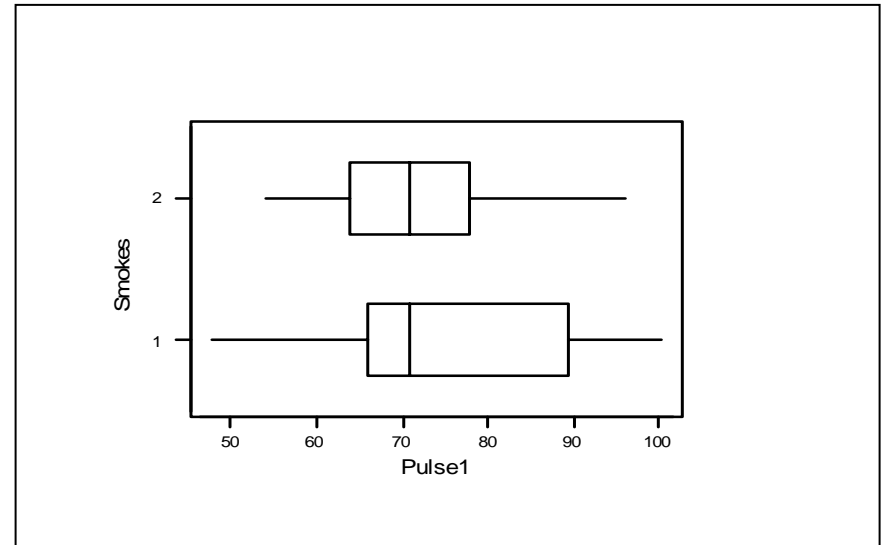


EXAMPLE 2: Obtain a Grouped Boxplot of 'PULSE1' by 'SMOKES'. Transpose the Boxplot orientation from vertical to horizontal.

Dialog Box: Move 'PULSE1' into the "Y variable" cell and 'SMOKES' into the "X variable" cell. Select the "Options" button and "Transpose X and Y" by checking the cell.



Minitab Output: Boxplot of 'PULSE1' by 'SMOKES'
(Grouped by 1 = Smokes, 2 = Not Smoke)



CHI-SQUARE (χ^2): Test of Independence

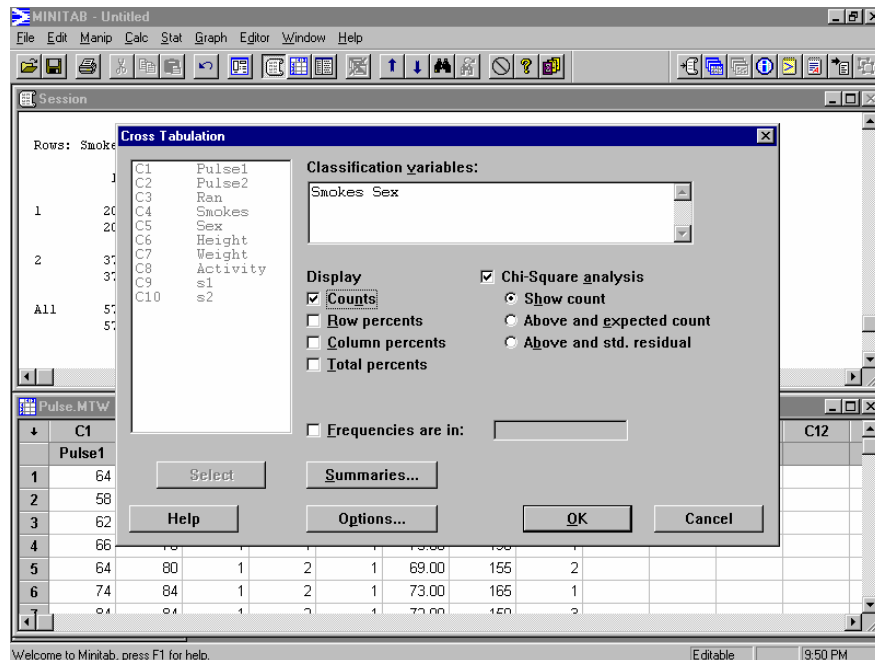
The Chi-Square procedures are used to examine whether there exists an association between two population characteristics. Minitab does not have a direct means by which the Chi-Square Goodness of Fit test may be calculated. However, the Chi-Square Test of Independence can be conducted using Minitab. There is a Chi-Square Test option, but it requires that the data be entered into two or more columns in the form of a contingency table. Therefore, given the nature of most data, it is easier to obtain the Chi-Square statistic through the Cross Tabulation procedure.

DATA: Pulse.MTW

PATH: Stat => Tables => Cross Tabulation => identify the two variables of interest, select the cell data (counts and various percents) and select the Chi-Square box.

EXAMPLE: Is there an association between gender and smoking among the study participants? To examine this, create a cross tabulation of 'SMOKE' (dependent variable) and 'SEX' (independent variable).

Dialog Box Input: Variables 'SMOKE' and 'SEX' entered into the "Classification variables:" cell. Check the Cross Tabulation display items and the Chi-Square analysis box.



Minitab Output: Cross Tabulation table and Chi-Square analysis.

Tabulated Statistics: Smokes, Sex

Rows: Smokes	Columns: Sex		
	1	2	All
1	20	8	28
	20	8	28
2	37	27	64
	37	27	64
All	57	35	92
	57	35	92

Chi-Square = 1.532, DF = 1, P-Value = 0.216

COEFFICIENT OF DETERMINATION (r^2) and the SUMS OF SQUARES IN LINEAR REGRESSION

The *Coefficient of Determination*, represented by r^2 , is produced as part of a regression analysis or may be calculated from a correlation coefficient. The Coefficient of Determination represents the proportion of variation in the observed values of the response variable that may be explained by the predictor variable via regression. Alternatively, it may be obtained by squaring the correlation coefficient between two variables. Mathematically, the coefficient of determination may be obtained by dividing the Regression Sum of Squares (SSR) by the Total Sum of Squares (SST).

The *Sums of Squares*:

Total Sum of Squares (SST): The variation in the observed values of the response variable.

Regression Sum of Squares (SSR): The variation in the observed values of the response variable that is explained by the regression.

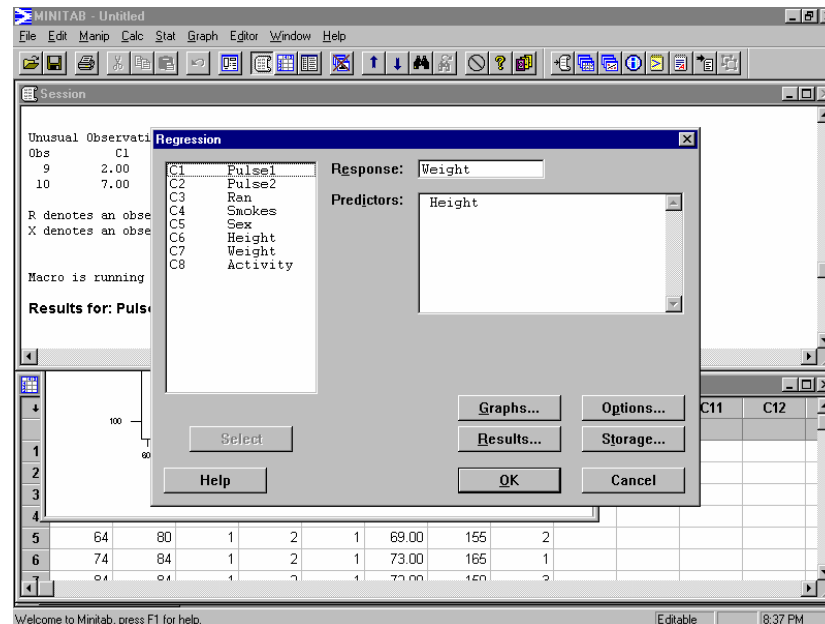
Error Sum of Squares (SSE): The variation in the observed values of the response variable that is not explained by the regression.

DATA: Pulse.MTW

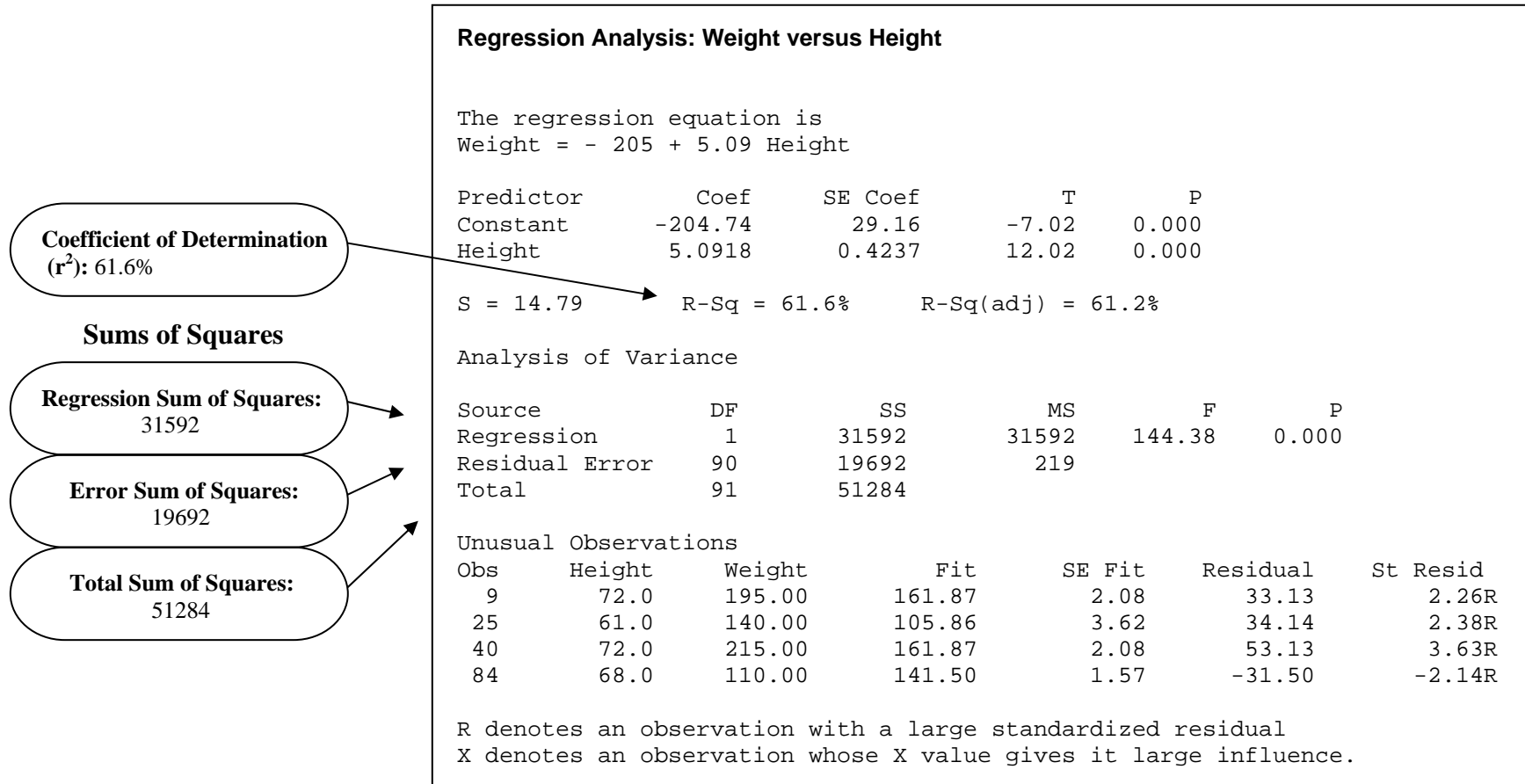
PATH: Stat => Regression => Regression => insert a variable into the "Response:" cell and a variable into the "Predictor:" cell.

EXAMPLE: Identify the Coefficient of Determination and the Sums of Squares using 'HEIGHT' as the predictor variable and 'WEIGHT' as the response variable.

Dialog Box Input: Variable 'WEIGHT' is entered into the "Response" cell and the variable 'HEIGHT' is entered into the "Predictor" cell.



Minitab Output: Coefficient of Determination and Sums of Squares



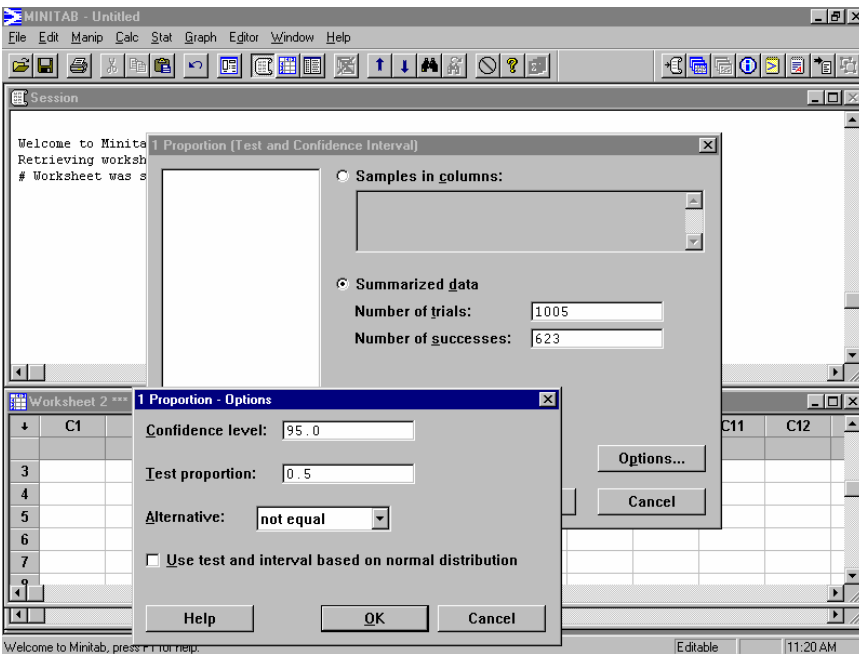
CONFIDENCE INTERVAL FOR ONE POPULATION PROPORTION

In some situations it is desirable to construct a **Confidence Interval** for the proportion (or percentage) of a population possessing a certain characteristic.

PATH: Stat => Basic Statistics => 1 Proportion => select the “Summarized data” option button => enter sample size in the “Number of trials” cell => enter the number of units possessing the characteristic of interest in the “Number of Successes” cell => select the “Options” button => type the desired confidence level in the “Confidence level” cell.

EXAMPLE: Based on a recent poll of 1005 individuals, it was reported that 623 individuals took a daily multi-vitamin. Construct a 95% Confidence Interval for p .

Dialog Box Input: Sample size is entered into the “Number of trials” cell. Number of units possessing the characteristic of interest is entered into the “Number of successes” cell. Confidence level is entered into the “Confidence level” cell.

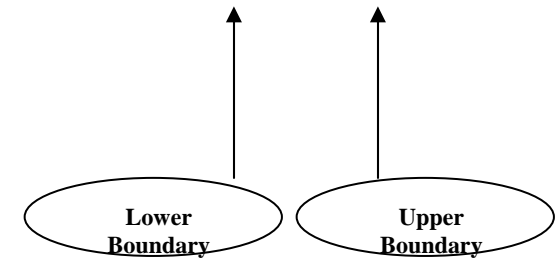


Minitab Output: 95% Confidence Interval for p .

Test and CI for One Proportion

Test of $p = 0.5$ vs $p \text{ not } = 0.5$

Sample	X	N	Sample p	95.0% CI	Exact P-Value
1	630	1005	0.626866	(0.596130, 0.656855)	0.000



CONFIDENCE INTERVAL FOR μ (σ UNKNOWN): ONE-SAMPLE t

Determination of the approach to obtaining a *Confidence Interval* (C.I.) for the mean of a set of data is dependent upon knowledge of the *standard deviation*. Where the standard deviation is **unknown**, the One-Sample t procedure may be used. This procedure establishes a confidence interval for the mean at a given confidence level.

DATA: Pulse.MTW

PATH: Stat => Basic Statistics => 1-Sample t => move variable to the "Variable" cell; enter the sample mean around which the confidence interval will be constructed into the "Test Mean:" cell (see Additional Requirements). => OK

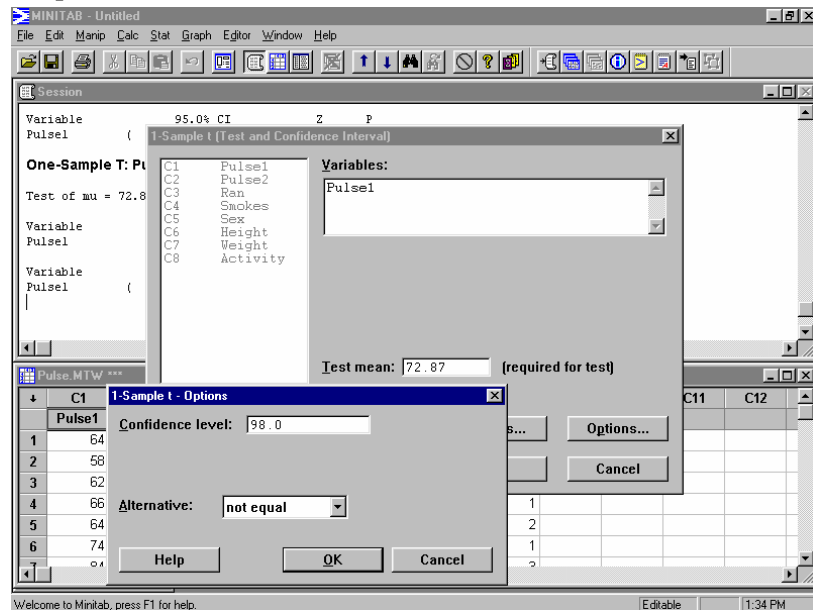
ADDITIONAL REQUIREMENTS:

Note1: Set the "Alternative:" cell to **not equal** by selecting the "Options".

Note2: Enter the desired Confidence Level into "Confidence Level:" cell.

EXAMPLE: The sample's initial mean pulse rate ('PULSE1') is 72.87 beats per minute. Establish a confidence interval for 'PULSE1'.

Dialog Box Input: Move 'PULSE1' into the "Variables" cell; enter 72.87 into the "Test Mean" cell; check the test "Alternative" and "Confidence Level" via the "Options" button.



Minitab Output: Confidence Interval where σ is unknown

One-Sample T: Pulse1

Test of $\mu = 72.87$ vs $\mu \text{ not } = 72.87$

Variable	N	Mean	StDev	SE Mean
Pulse1	92	72.87	11.01	1.15

Variable	98.0% CI	T	P
Pulse1	(70.15, 75.59)	-0.00	1.000

Lower boundary of the C.I.

Upper boundary of the C.I.

Confidence Level

CONFIDENCE INTERVAL FOR μ (σ KNOWN): ONE-SAMPLE Z

Determination of the approach to obtaining a *Confidence Interval* (C.I.) for the mean of a set of data is dependent upon knowledge of the *standard deviation*. Where the standard deviation is **known**, the One-Sample Z procedure may be used. This procedure establishes a confidence interval for the mean at a given confidence level.

DATA: Pulse.MTW

PATH: Stat => Basic Statistics => 1-Sample Z => move variable to the "Variable" cell; enter sigma (standard deviation) into the "Sigma:" cell; enter the sample mean around which the confidence interval will be constructed into the "Test Mean:" cell; see Additional Requirements. => OK

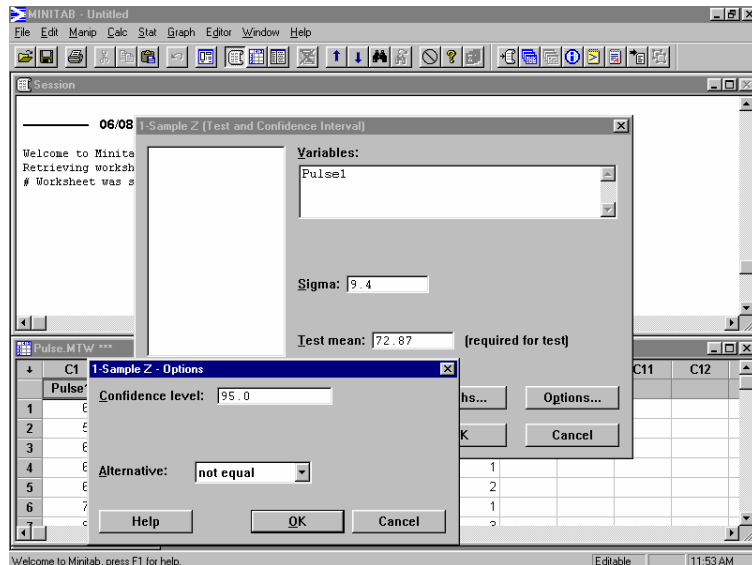
ADDITIONAL REQUIREMENTS:

Note1: Set the "Alternative:" cell to **not equal** by selecting the "Options".

Note2: Enter the desired Confidence Level into the "Confidence Level:" cell.

EXAMPLE: The sample's initial mean pulse rate ('PULSE1') is 72.87. The population standard deviation for adults is known to be 9.4 beats per minute.. Establish a Confidence Interval for the initial sample mean using a 95% Confidence Level.

Dialog Box Input: Move 'PULSE1' into the "Variables" cell; enter 9.4 into the "Sigma:" cell, enter 72.87 into the "Test Mean" cell; check the test "Alternative" and "Confidence Level" via the "Options" button.



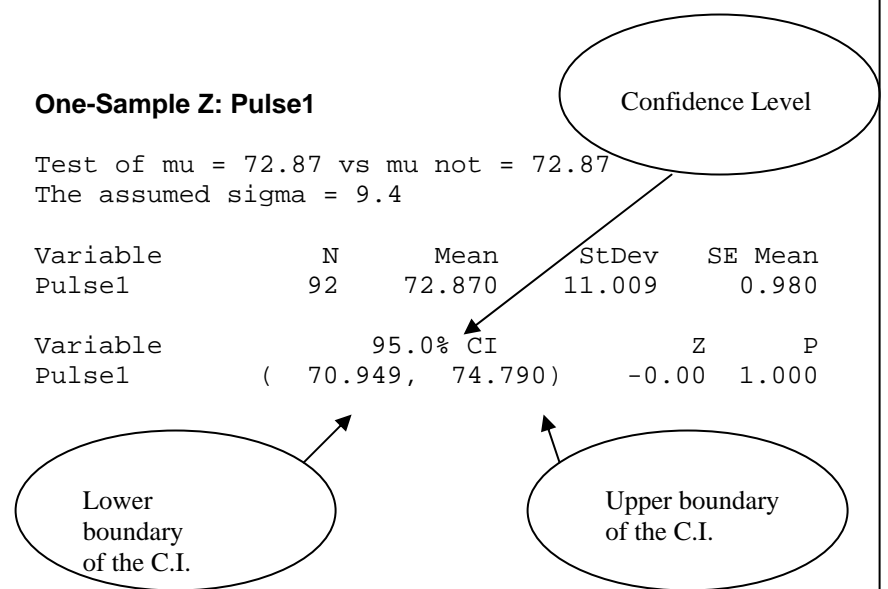
Minitab Output: Confidence Interval where σ is known

One-Sample Z: Pulse1

Test of $\mu = 72.87$ vs $\mu \text{ not } = 72.87$
The assumed sigma = 9.4

Variable	N	Mean	StDev	SE Mean
Pulse1	92	72.870	11.009	0.980

Variable	95.0% CI	Z	P
Pulse1	(70.949, 74.790)	-0.00	1.000



CROSS TABULATION TABLE (Contingency Table)

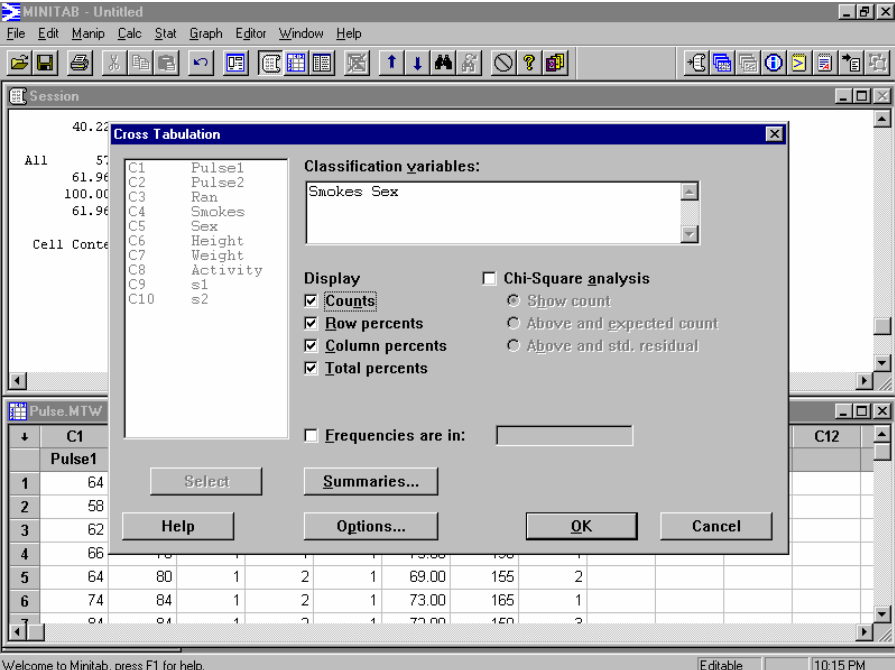
A **Tally** (frequency table) presents the values of a single variable. It is common, however, to want to compare the responses of one variable in relation to a second variable. So, as in the example below, one may wish to examine whether one smokes or not by the sex of the respondent. In that way the data are broken down into male smokers, male non-smokers, female smokers, and female non-smokers. The table that presents these breakdowns in Minitab is called a **Cross Tabulation Table**. It may also be termed a Crosstabs table or a **Contingency Table**. In the case noted here, the table would be a 2-by-2 table as each variable has two possible values. Where a distinction between an independent and a dependent variable exists, the independent variable is normally assigned to the columns and the dependent variable is assigned to the rows of the table. In Minitab the first variable listed in the dialog box will be assigned to the rows. The output may contain cell counts and a number of percentages, depending upon the information desired.

DATA: Pulse.MTW

PATH: Stat => Tables => Cross Tabulation => identify the two variables of interest, select the cell data (counts and various percents) .

EXAMPLE: Is there an association between gender and smoking among the study participants? To examine this, create a Cross Tabulation of 'SMOKE' (dependent variable) and 'SEX' (independent variable). Present the count for each "cell," as well as the column, row and total percentages.

Dialog Box Input: Variables 'SMOKE' and 'SEX' entered into the "Classification variables:" Cell. Check the desired "Display" items.



Minitab Output:

Tabulated Statistics: Smokes, Sex

Rows: Smokes	Columns: Sex		
	1	2	All
1	20	8	28
	71.43	28.57	100.00
	35.09	22.86	30.43
	21.74	8.70	30.43
2	37	27	64
	57.81	42.19	100.00
	64.91	77.14	69.57
	40.22	29.35	69.57
All	57	35	92
	61.96	38.04	100.00
	100.00	100.00	100.00
	61.96	38.04	100.00

Cell Contents --

- Count
- % of Row
- % of Col
- % of Tbl

Interpreting a Cross Tabulation table:

Tabulated Statistics: Smokes, Sex

Rows: Smokes Columns: Sex Where:

SEX: 1 = Male; 2 = Female
Smokes: 1 = Smokes regularly
2 = Does not smoke

	1	2	All
1	20	8	28
	71.43	28.57	100.00
	35.09	22.86	30.43
	21.74	8.70	30.43
2	37	27	64
	57.81	42.19	100.00
	64.91	77.14	69.57
	40.22	29.35	69.57
All	57	35	92
	61.96	38.04	100.00
	100.00	100.00	100.00
	61.96	38.04	100.00

Cell Contents --
Count
% of Row
% of Col

Current Worksheet: Pulse.MTW Editable 9:52 AM

To explain how to interpret the results, let's work with the first data grouping, which appears as follows:

1 (Males)

	1	20	(count)
(Smoke)	71.43		(row %)
	35.09		(column %)
	21.74		(% of total)

Recalling that the variable SMOKES is the row variable, and SEX is the column variable, the grouping shown above are “males who smoke regularly.” Now think, “count, row, column, total.” Of the 92 respondents, 20 are males who smoke regularly (count). They constitute 71.43% of those respondents who smoke regularly (row), 35.09% of males who smoke regularly (column), and these 20 males (out of 92) represent 21.74% of all respondents (total).

Now you try it! (Answers to far right.)

- How many females responded “do not smoke regularly?”
- What % of all respondents are “females who do not smoke regularly?”
- What % of all females are “females who do not smoke regularly?”
- What % of those respondents who “do not smoke regularly” are females?

- (count: 27)
- (% of total: 29.35%)
- (column %: 77.14%)
- (row %: 42.19%)

DISCRETE RANDOM VARIABLES (Obtaining a Probability Histogram)

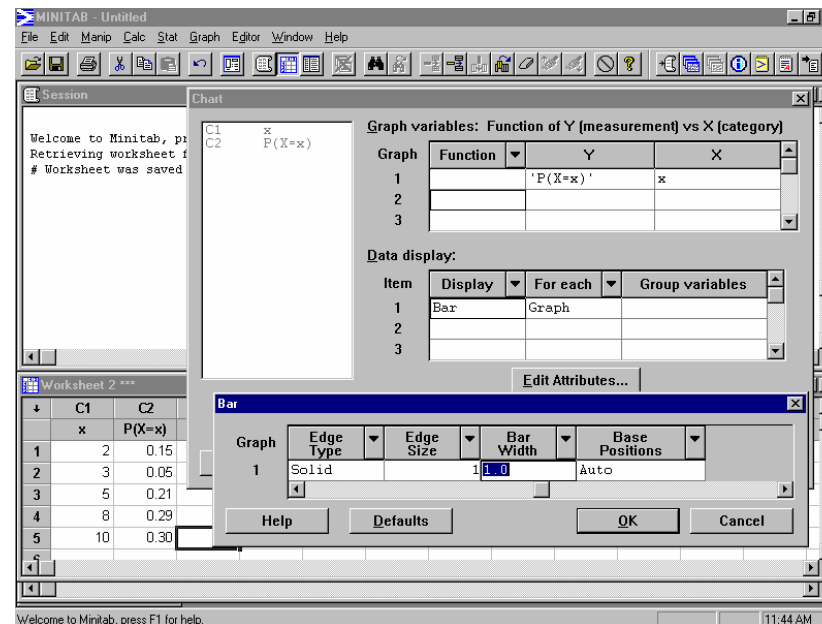
Minitab can be used to obtain a probability histogram for the distribution of a **Discrete Random Variable**.

PATH (Part 1 of 3): (Note: Before beginning, the probability distribution must be stored in a worksheet. The columns in which it is stored must be named. Here, x and $P(X=x)$ were used.) **Graph => Chart => click in the “Y” cell, and select the column of the worksheet that contains the probabilities => click in the “X” cell, and select the column of the worksheet that contains the possible values of the variable => select “Edit Attributes,” and use the scroll bar to locate “Bar Width.” => from the drop-down box located next to “Bar Width,” select 1.0 => click OK.**

EXAMPLE: A discrete random variable X , can assume five possible values: 2, 3, 5, 8, and 10. Its probability distribution is shown below. Construct a **Probability Histogram** for this distribution.

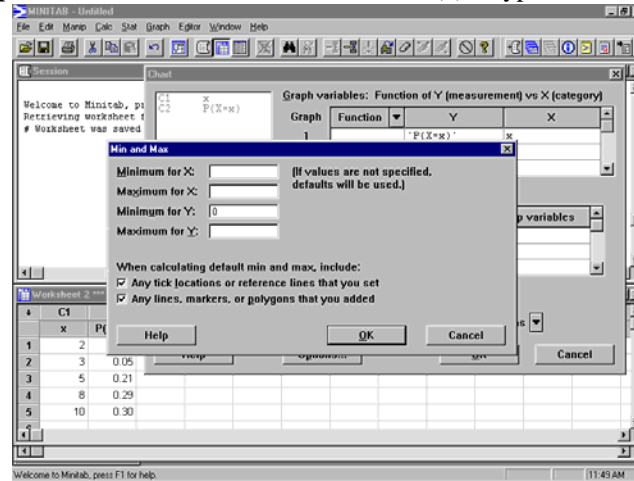
x	2	3	5	8	10
$p(x)$.15	.05	.21	.29	.30

Dialog Box Input: Probabilities and possible values of the variable are put into the “Y” and “X” cells, respectively. “Bar Width” is located, and 1.0 is selected



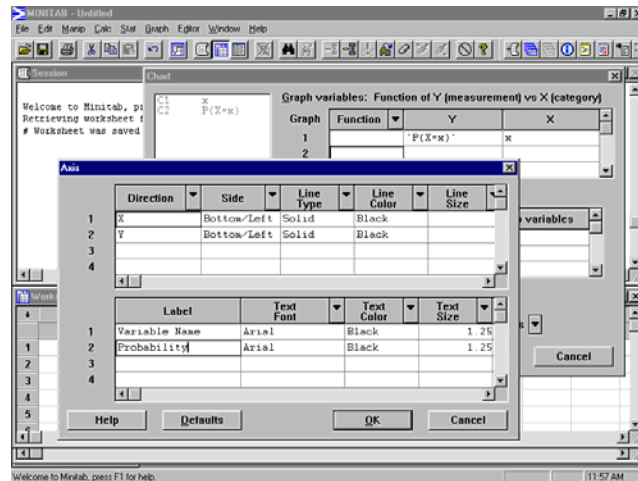
PATH (Part 2 of 3): From the “Frame” drop-down box, select “Min and Max” => in the “Minimum for Y” cell, type 0 => click OK

Dialog Box Input: “Min and Max” is selected. Zero (0) is typed in the “Minimum for Y” cell.

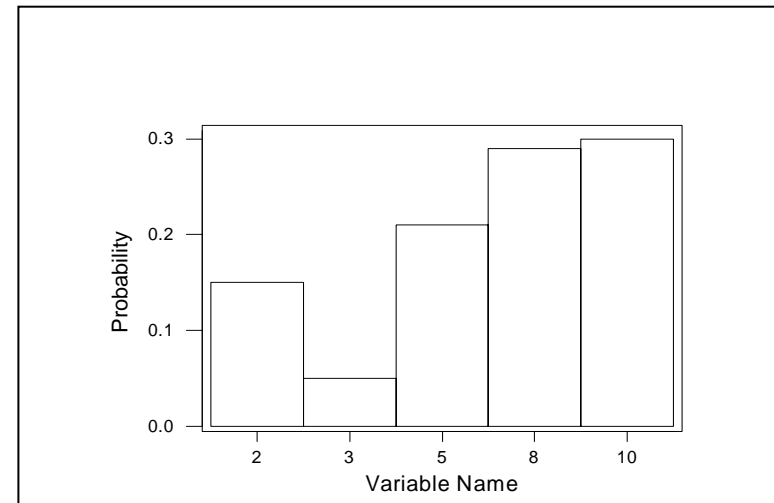


PATH (Part 3 of 3): From the “Frame” drop-down box, select “Axis” => in the “Label” cell for 1, type the variable name => in the “Label” cell for 2, type Probability => click OK to close this dialog box => click OK to construct the probability histogram.

Dialog Box Input: Variable name is typed in the “Label” cell for 1. Probability is typed in the “Label” cell for 2.



Minitab Output: Probability Histogram for a **Discrete Random Variable**



DISCRETE RANDOM VARIABLES (Simulation)

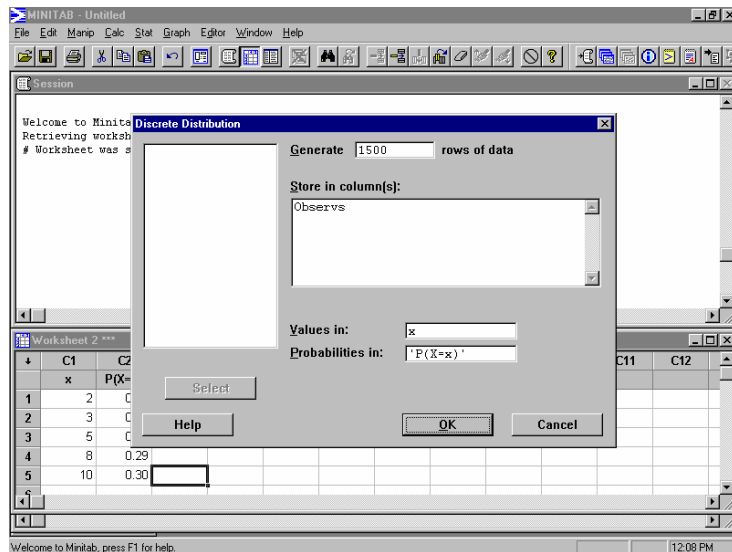
Given the probability distribution of a **Discrete Random Variable**, Minitab can be used to simulate a large number of observations of the variable. The Minitab command *Tally* can then be used to determine counts and percents for the simulated observations.

PATH: (Note: Before beginning, the probability distribution must be stored in a worksheet. The columns in which it is stored must be named; here, x and $P(X=x)$ were used.)
Calc => Random Data => Discrete => type the number of observations to be generated, in the “Generate rows of data” cell => type the name to be used for the column of the worksheet in which the observations will be stored in the “Store in column(s)” cell => type the name of the column in the worksheet containing the possible values of the variable in the “Values in” cell => type the name of the column in the worksheet containing the probabilities for each value of the variable in the “Probabilities in” cell.

EXAMPLE: A discrete random variable X , can assume five possible values: 2, 3, 5, 8, and 10. Its probability distribution is shown below. Use Minitab to simulate 1500 observations of this variable.

x	2	3	5	8	10
$p(x)$.15	.05	.21	.29	.30

Dialog Box Input: Number of observations to be generated is typed in the “Generate rows of data” cell. Name of column to be used to store observations is typed in “Store in column(s)” cell. Names of the columns in the worksheet containing the possible values of the variable, and their corresponding probabilities are typed in the “Values in” and “Probabilities in” cells.



Minitab Output: Partial list of the 1500 observations generated.

Observs5
5
10
5
8
10

DOTPLOT

Dotplots are typically used for small sets of data. A Dotplot uses one dot for each measurement.

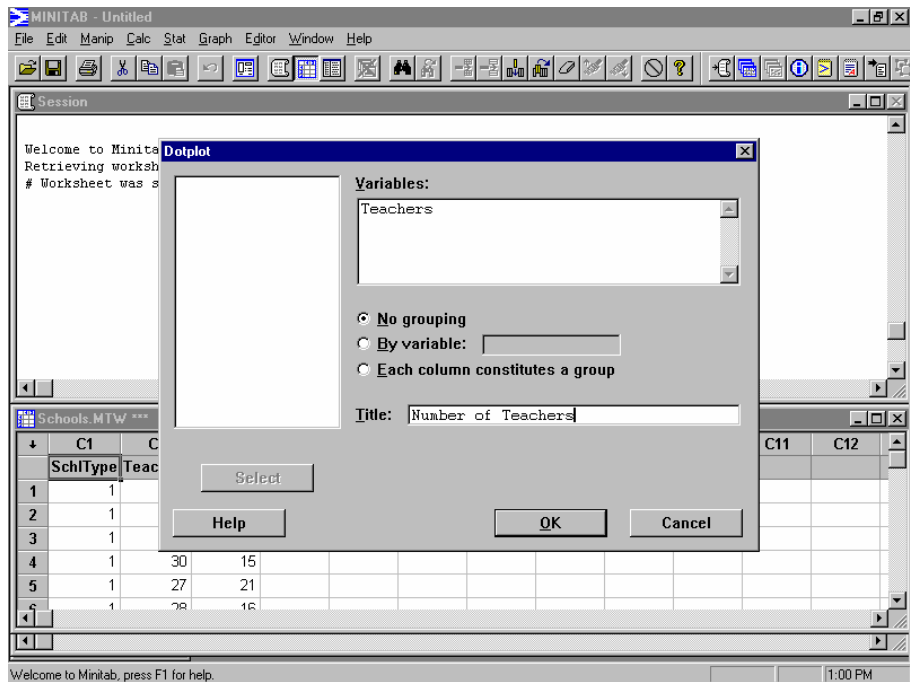
DATA: Schools.MTW

PATH: Graph => Character Graphs => Dotplot => move variable to the “Variables” cell.

OPTIONS: Include a title for the plot by typing it in the “Title” cell.

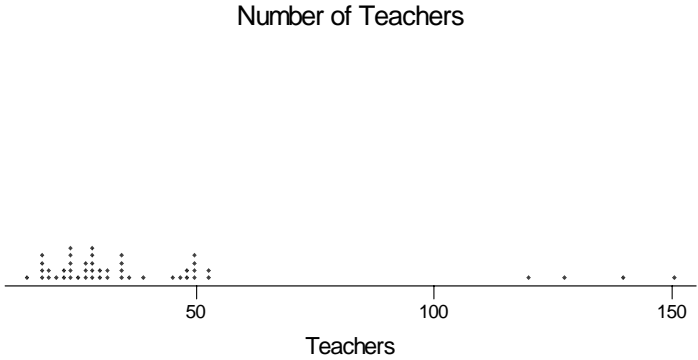
EXAMPLE: Obtain a Dotplot of ‘TEACHERS.’

Dialog Box Input: Variable ‘TEACHERS’ is moved into the “Variables” cell.



Minitab Output: Dotplot of ‘TEACHERS’

Number of Teachers



School	Teachers
1	1
2	1
3	1
4	1, 30, 15
5	1, 27, 21
6	1, 28, 16

FITTED LINE PLOT(Linear Regression)

A *Fitted Line Plot* presents the data points with a regression line drawn through them as determined by the least-squares criterion. The regression equation and related data are output to the Minitab Session Window.

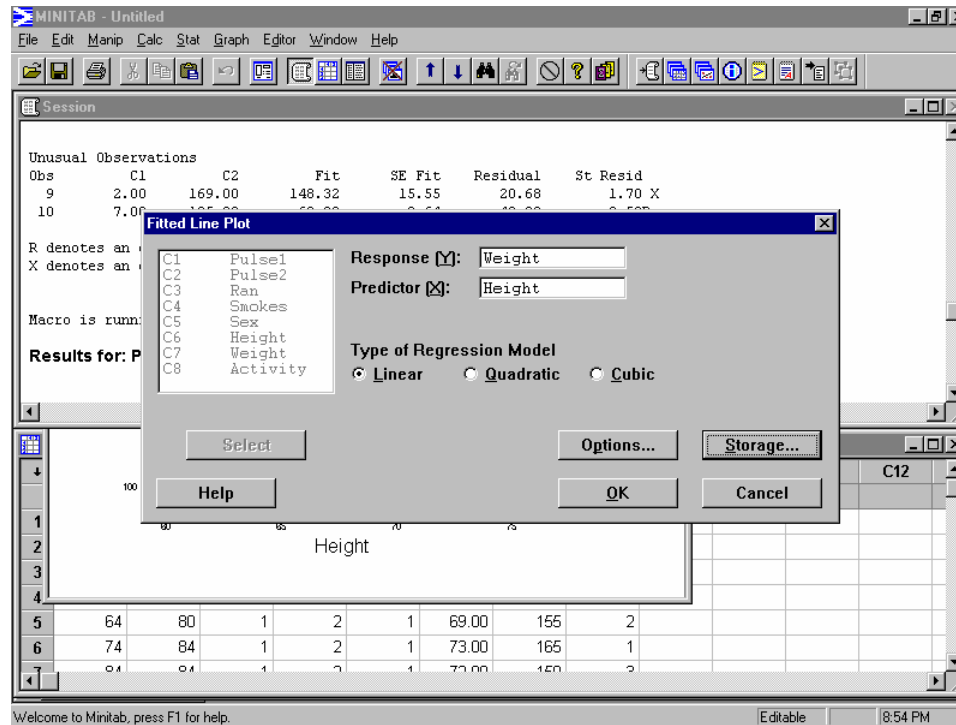
DATA: Pulse.MTW

PATH: Stat => Regression => Fitted Line Plot => insert a variable into the “Response (Y):” cell and a variable into the “Predictor (X):” cell.

RELATED STATISTICAL PROCEDURE: Linear Regression

EXAMPLE: Obtain a Fitted Line Plot for the Regression Equation that predicts ‘WEIGHT’ using ‘HEIGHT’ as the predictor variable.

Dialog Box Input: Variable ‘WEIGHT’ entered into the "Response (Y)" cell and the variable ‘HEIGHT’ entered into the "Predictor (X)" cell.

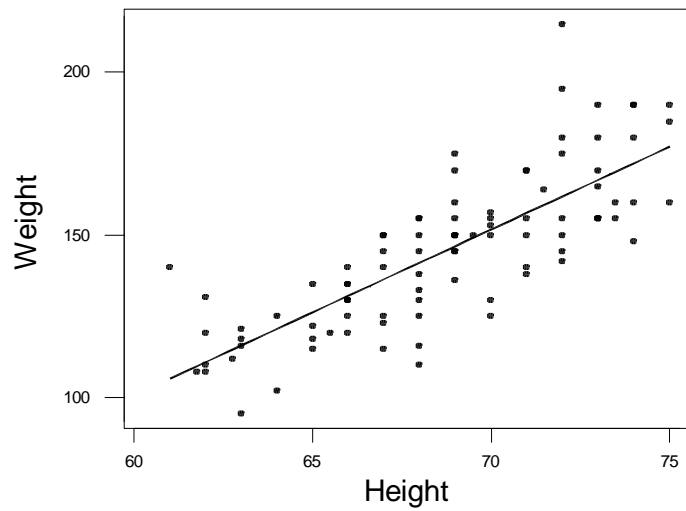


Minitab Output: Fitted Line Plot of Weight versus Height

Regression Plot

$$\text{Weight} = -204.741 + 5.09177 \text{ Height}$$

S = 14.7920 R-Sq = 61.6 % R-Sq(adj) = 61.2 %



Minitab Session Window Output (see Linear Regression):

Regression Analysis: Weight versus Height

The regression equation is
 $\text{Weight} = -204.741 + 5.09177 \text{ Height}$

S = 14.7920 R-Sq = 61.6 % R-Sq(adj) = 61.2 %

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	31591.6	31591.6	144.384	0.000
Error	90	19692.2	218.8		
Total	91	51283.9			

FIVE-NUMBER SUMMARY

The *Five-Number Summary* for a data set, consists of the following values: Minimum, Q_1 , Median (Q_2), Q_3 , and Maximum.

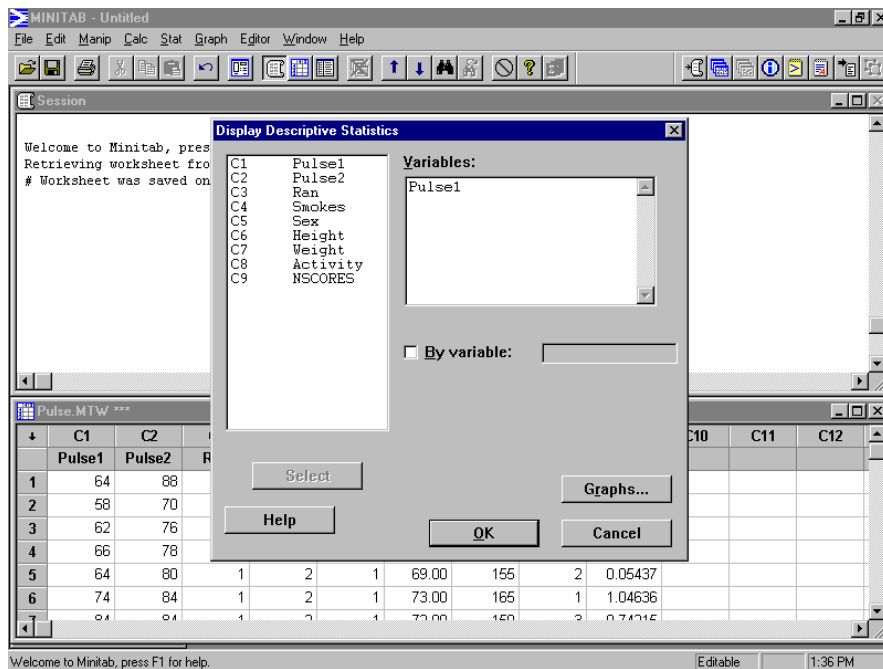
DATA: Pulse.MTW

PATH: Stat => Basic Statistics => Display Descriptive Statistics => move the variable into the “Variables” cell.

OPTIONS: In addition to the Five-Number Summary, you can obtain a histogram, dotplot, or boxplot of the variable by selecting “Graphs” from the “Display Descriptive Statistics” dialog box, and then selecting the type of graph(s) to be displayed.

EXAMPLE: Obtain the Five-Number Summary for the variable ‘PULSE1.’

Dialog Box Input: Variable ‘PULSE1’ is moved into the “Variables” cell.



Minitab Output: Five-Number Summary of ‘PULSE1’

Descriptive Statistics: Pulse1

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Pulse1	92	2.87	71.00	72.61	11.01	1.15

Variable	Minimum	Maximum	Q1	Q3
Pulse1	48.00	100.00	64.00	80.00

FREQUENCY POLYGON/AREA GRAPH

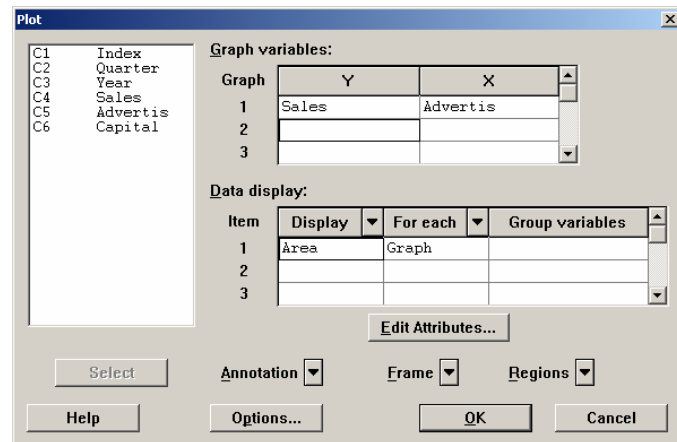
A **Frequency Polygon** graphs data by using lines that connect the midpoints of frequency classes. The height of each midpoint on the graph represents the frequency of occurrence within that class. The line connecting midpoints may begin and end touching the x-axis. Frequency Polygons and Histograms represent two ways to present continuous data. An Area Graph is a Frequency Polygon for which the area under the line has been shaded.

DATA: Market.mtw

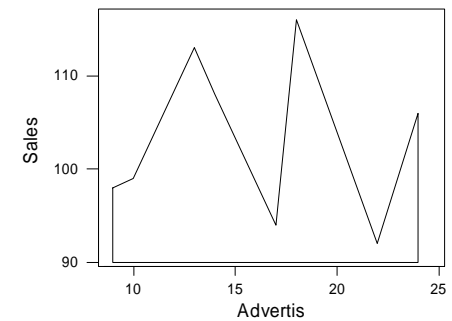
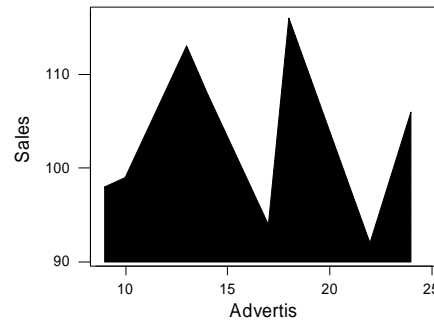
PATH: Graph => Plot => Insert variables into the y-axis and x-axis cells => Change the Data Display to 'Area.'

EXAMPLE: In the y-axis cell insert the variable 'SALES' (sales income) and in the x-axis cell insert the variable "Advertis" (advertising costs). Change the Data Display in Item #1 to 'Area.' An Area Graph is created. To change this to a Frequency Polygon, edit the graph by deleting the shading below the line that connects values.

Dialog Box: Variables are selected and Data Display is changed to 'Area.'



Minitab Output: Area Graph (left) appears and Frequency Polygon (right, after editing of Area Graph).



HISTOGRAM

A *Histogram* is a graph of a distribution in which classes of equal width are placed on the horizontal axis, and either frequencies, relative frequencies, or percentages are placed on the vertical axis. The frequencies, relative frequencies, or percentages are represented by the heights of the bars. The bars are adjacent to each other.

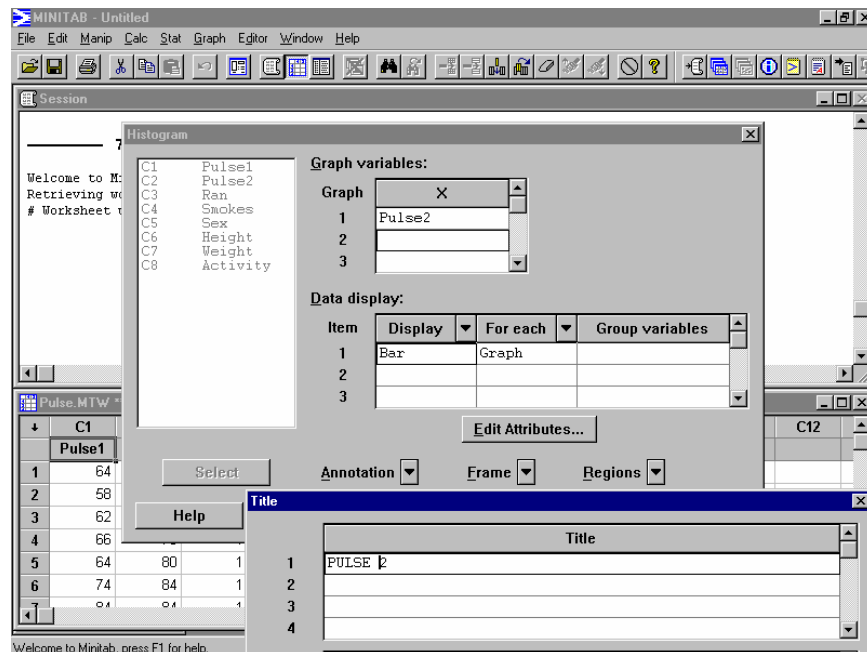
DATA: Pulse.MTW

PATH: Graph => Histogram => move the variable to the “X-variable” cell.

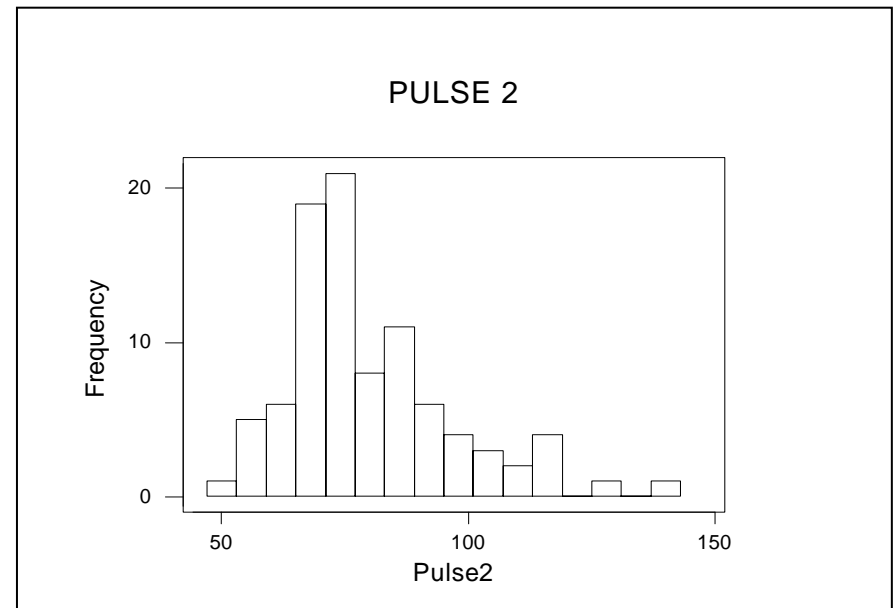
OPTIONS: To add a title, use the “Annotation” drop-down box, and select “Title.” Type in a title.

EXAMPLE: Obtain a Histogram of ‘PULSE2.’

Dialog Box Input: Variable ‘PULSE2’ is moved into the “X-variable” cell.
A title is added.



Minitab Output: Histogram of ‘PULSE 2’



HYPOTHESIS TEST FOR ONE POPULATION PROPORTION

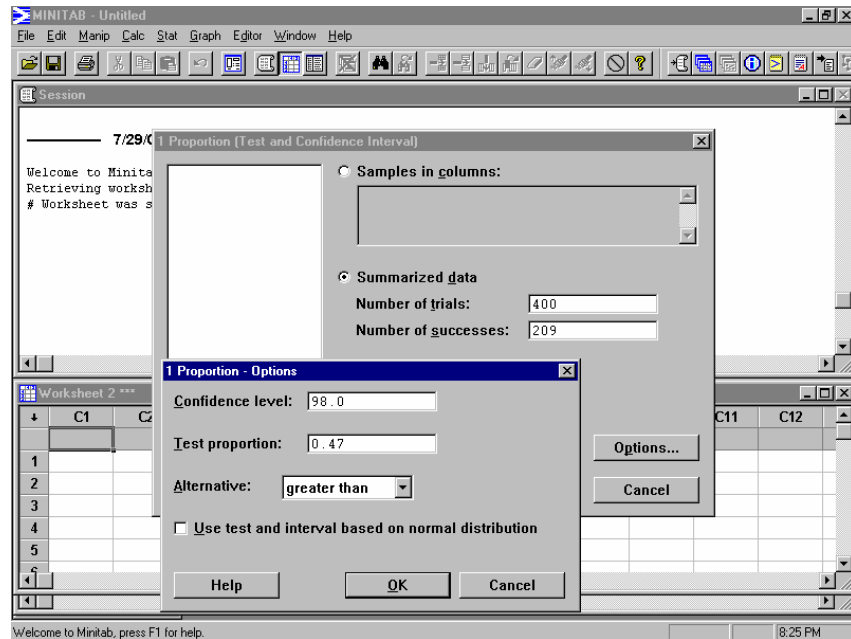
The steps taken in Minitab to conduct a **Hypothesis Test** for one population proportion are very similar to those steps used to obtain a Confidence Interval for one population proportion. The resulting output is also very similar.

PATH: Stat => Basic Statistics => 1 Proportion => select the “Summarized data” option button => enter sample size in the “Number of trials” cell => enter the number of units possessing the characteristic of interest in the “Number of Successes” cell => select the “Options” button => identify significance level by entering the appropriate confidence level in the “Confidence Level” cell (see Additional Note) => enter the test proportion in the “Test proportion” cell => select “greater than,” “less than,” or “not equal “ from the “Alternative” drop-down box.

ADDITIONAL NOTE: Recall the confidence level is not α , but rather $(1 - \alpha)100 = \mathbf{x\%}$. (Example: where $\alpha = .02$, enter a confidence level of $(1 - .02)100 = 98\%$.)

EXAMPLE: A study conducted 1 year ago by a national fitness organization showed that 47% of all males aged 25-45 engage in some type of fitness activity at least 3 times per week. In a random sample taken this year, 209 out of 400 males aged 25-45 stated that they engage in some type of fitness activity at least 3 times per week. At the 2% significance level, test the claim that the percentage of men in this age group who engage in some type of fitness activity at least 3 times per week is greater than 47%.

Dialog Box Input: “Summarized data” is selected. “Number of trials” and “Number of successes” are entered into their respective cells. “Confidence level,” “Test proportion,” and “Alternative” are entered into their respective cells.

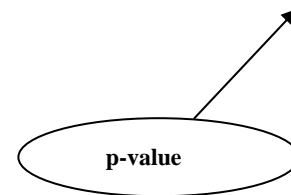


Minitab Output: Hypothesis Test for one Population Proportion

Test and CI for One Proportion

Test of $p = 0.47$ vs $p > 0.47$

Sample	X	N	Sample p	98.0%	Lower Bound	Z-Value	P-Value
1	209	400	0.522500		0.471208	2.10	0.018



HYPOTHESIS TEST FOR μ (σ UNKNOWN): ONE-SAMPLE t

Selection of the statistical approach for a *Hypothesis Test* for the mean of a set of data is dependent upon knowledge of the *standard deviation*. Where the standard deviation is **unknown**, the One-Sample t test may be used. This test compares the mean of a sample with the stated population mean. The test outputs both a value for t and a p -value, thereby providing the results for both approaches to hypothesis testing.

DATA: Pulse.MTW

PATH: Stat => Basic Statistics => 1-Sample t => move variable to the "Variable" cell; enter the mean against which you are testing the data into the "Test Mean:" cell; see **Additional Requirements**. => OK

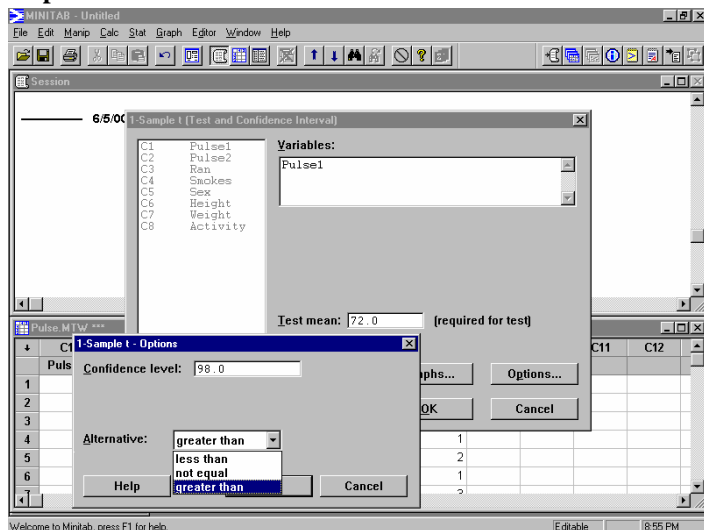
ADDITIONAL REQUIREMENTS:

Note1: Identify the type of hypothesis test (where means are =, <, or >) by selecting the "Options" button and entering the type test into the "Alternative:" cell.

Note2: Identify level of significance of the test by entering the confidence level of the test into the "Confidence Level:" cell.. Recall: the Confidence Level is not α , but rather $(1 - \alpha)100 = x\%$. (Example: where $\alpha = .05$, enter a confidence level of $(1 - .05)100 = 95.0\%$)

EXAMPLE: The mean pulse rate for adults is known to be 72 beats per minute. Determine if the initial pulse rate ('PULSE1') of study participants is significantly higher than that of the population. Test the hypothesis at the $\alpha = .02$ significance level.

Dialog Box Input: Move 'PULSE1' into the "Variables:" cell; enter 72.0 into the "Test Mean:" cell; check the test "Alternative:" and "Confidence Level:" via the "Options" button.



Minitab Output:

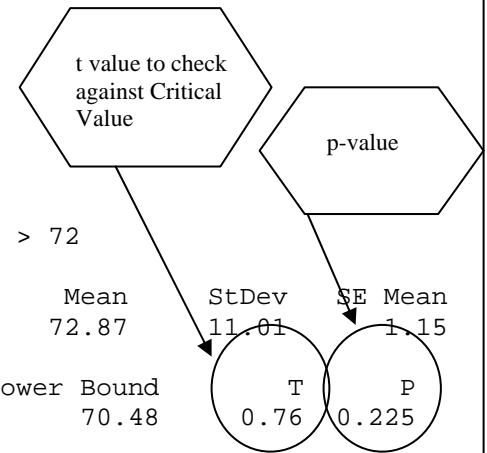
One-Sample T: Pulse1

Test of $\mu = 72$ vs $\mu > 72$

Variable	N	Mean	StDev	SE Mean
Pulse1	92	72.87	11.01	1.15

Variable	98.0% Lower Bound
Pulse1	70.48

T	P
0.76	0.225



HYPOTHESIS TEST OF MEAN (σ KNOWN): ONE-SAMPLE Z

Selection of the statistical approach for a *Hypothesis Test* for the mean of a set of data is dependent upon knowledge of the *standard deviation*. Where the standard deviation is **known**, the One-Sample Z test may be used. This test compares the mean of a sample with the stated population mean, having a known standard deviation. The test outputs both a z-score and a p-value, thereby providing the results for both approaches to hypothesis testing.

DATA: Pulse.MTW

PATH: Stat => Basic Statistics => 1-Sample Z => move variable to the "Variable" cell; enter sigma (standard deviation) into the "Sigma:" cell; enter the mean against which you are testing the data into the "Test Mean:" cell; see Additional Requirements. => OK

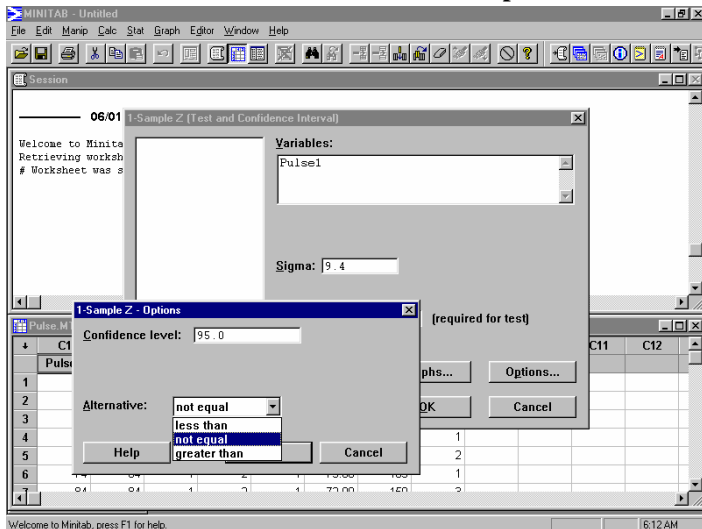
ADDITIONAL REQUIREMENTS:

Note1: Identify the type of hypothesis test (where means are =, <, or >) by selecting the "Options" button and entering the type test into the "Alternative:" cell.

Note2: Identify level of significance of the test by entering the confidence level of the test into the "Confidence Level:" cell. Recall: the Confidence Level is not α , but rather $(1 - \alpha)100 = x\%$. (Example: where $\alpha = .05$, enter a confidence level of $(1 - .05)100 = 95.0\%$)

EXAMPLE: The mean pulse rate for adults is known to be 72 beats per minute with a standard deviation of 9.4 beats per minute. Determine if the initial pulse rate of study participants differs from the population mean. Test the hypothesis at the $\alpha = .05$ significance level.

Dialog Box Input: Move 'PULSE1' into the "Variables:" cell; enter 9.4 into the "Sigma:" cell; enter 72.0 into the "Test Mean:" cell; check the test "Alternative:" and "Confidence Level:" via the "Options" button.



Minitab Output:

MINITAB OUTPUT

Results for: Pulse.MTW

One-Sample Z: Pulse1

Test of $\mu = 72$ vs $\mu \text{ not } = 72$
The assumed sigma = 9.4

Variable	N	Mean	StDev	SE Mean
Pulse1	92	72.870	11.009	0.980

Variable	95.0% CI	Z	P
Pulse1	(70.949, 74.790)	0.89	0.375

Z-score to check against Critical Value

p-value

Z: 0.89
P: 0.375

LINE GRAPH

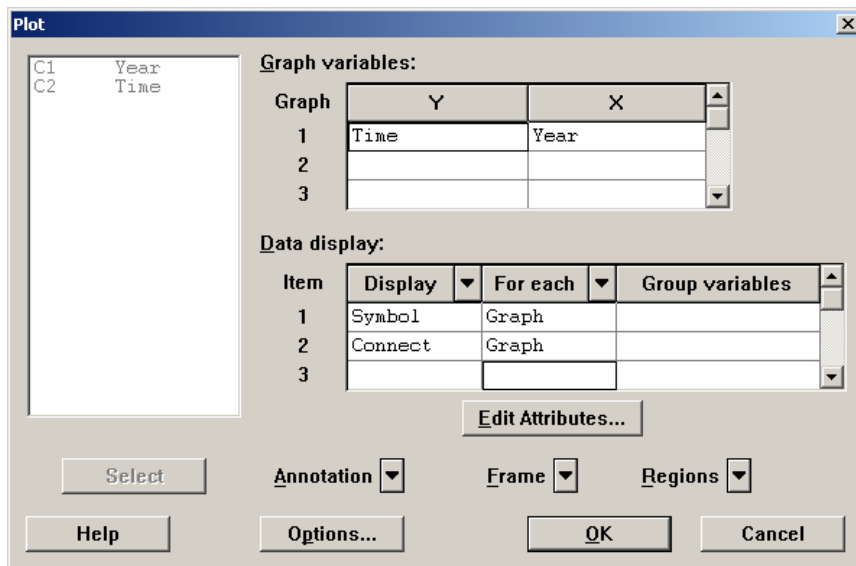
A **Line Graph** summarizes values of a single variable.

DATA: Track15.mtw

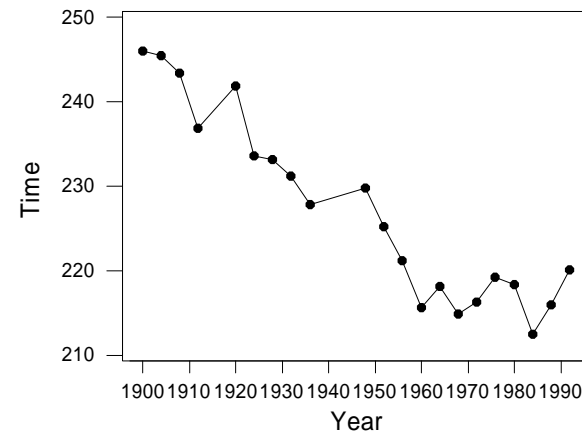
PATH: Graph => Plot => Select variables to plot on the y-axis and x-axis => Change the 'Data Display' Item 2 display type to 'Connect' and select Graph for the 'For Each' cell => OK

EXAMPLE: Create a Line Graph that presents the time to run the 1500 meter run at each of the Olympics from 1900-1992. Enter the variable "TIME" into the y-axis cell and the variable "YEAR" into the x-axis cell. In the second Data Display cell insert the option 'Connect.' ('Symbol' in item #1 will insert a point for each data value. The 'Connect' command inserts the line between the points.)

Dialog Box: Variables are selected and the Data Display option is modified.



Minitab Output: Line Graph of TIME vs. YEAR.



LINEAR CORRELATION

Linear Correlation is used to examine whether or not a relationship exists between two variables. A **Correlation Coefficient** will indicate a direction and degree of relationship between two variables. Correlation does not imply causation. That is, one can not infer from the correlation that one variable has an affect upon the other. Correlation Coefficients range from -1 to $+1$, where -1 = perfect negative correlation, 0 = no correlation, and $+1$ = perfect positive correlation.

DATA: Pulse.MTW

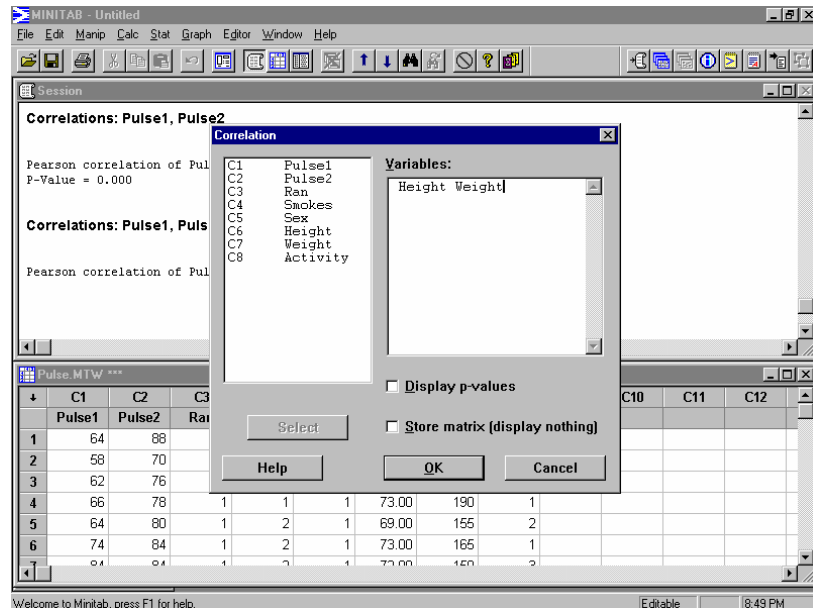
PATH: Stat => Basic Statistics => Correlation => click in the “Variables:” cell, select at minimum two variables from the list of variables in the left cell, toggle off the p-value option, if checked => Ok.

OPTIONS: Selecting more than two variables will result in each variable being correlated with all other variables individually.

RELATED GRAPH: A **Scatterplot** of the two variables of interest is often produced before conducting a correlation. The scatterplot will provide an initial suggestion of whether or not there is a linear relationship between the variables.

EXAMPLE: Is there a relationship between height and weight among study participants? To examine this, create a scatterplot using 'HEIGHT' as the "X variable" and "WEIGHT" as the "Y variable". (see Scatterplot). Run a correlation on the two variables and examine the value of the correlation coefficient.

Dialog Box Input: Variables ‘HEIGHT’ and ‘WEIGHT’ entered into the “Variables:” cell.

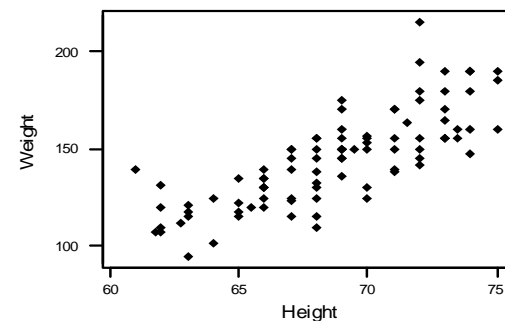


Minitab Output: Correlation

Correlations: Height, Weight

Pearson correlation of Height and Weight = 0.785

Minitab Graph: Scatterplot of ‘HEIGHT’ and ‘WEIGHT’ (done separately)



LINEAR REGRESSION

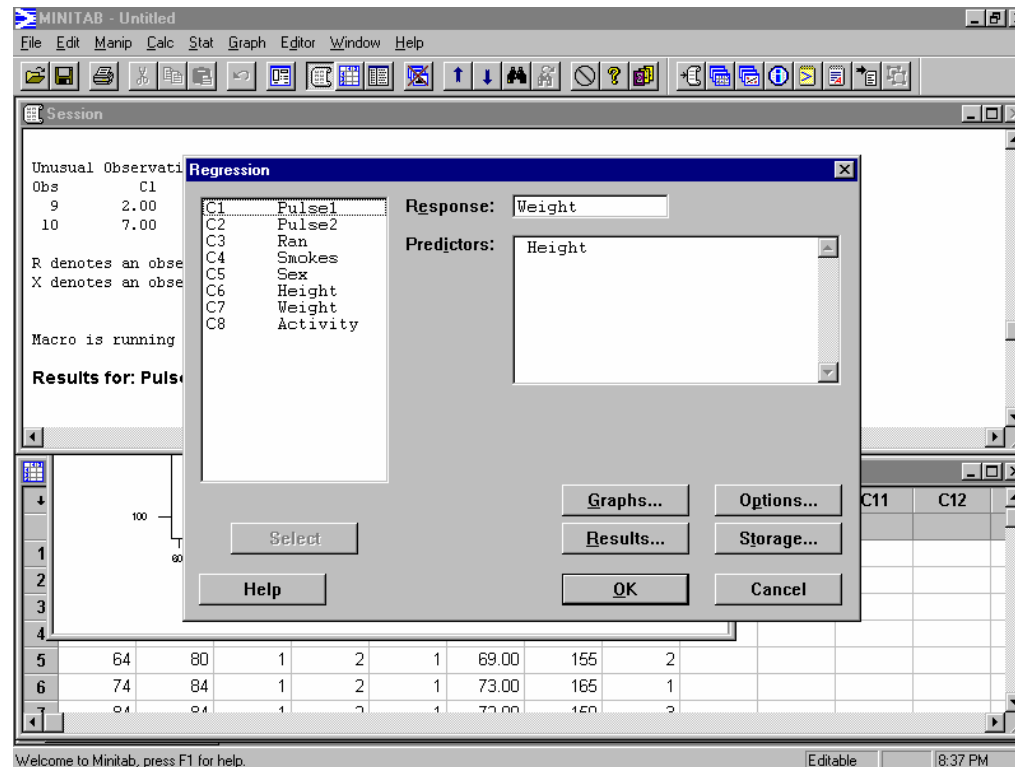
Linear Regression is used to examine the relationship between two or more variables. The method used employs the least-squares criterion to establish a regression line that best fits the data points. The Response variable represents the “Y” variable and the Predictor variable represents the “X” variable in the equation. The results are output to the Minitab Session Window.

DATA: Pulse.MTW

PATH: Stat => Regression => Regression => insert a variable into the “Response:” cell and a variable into the “Predictor:” cell.

EXAMPLE: Determine a Regression Equation that predicts ‘WEIGHT’ using ‘HEIGHT’ as the predictor variable.

Dialog Box Input: Variable ‘WEIGHT’ entered into the "Response" cell and the variable ‘HEIGHT’ entered into the "Predictor" cell.



Minitab Output: Regression

Response Variable: (Weight)

y-intercept: (-205)

Direction of Slope and orientation of Correlation Coefficient (+/-): here a “-” meaning a negative Slope and negative Correlation

Slope of Line: (5.09)

Predictor Variable: (Height)

Coefficient of Determination (r^2): 61.6%

Outliers: Noted by an “R”
Influential Observations: Noted by an “X”

Regression Analysis: Weight versus Height

The regression equation is
Weight = - 205 + 5.09 Height

Predictor	Coef	SE Coef	T	P
Constant	-204.74	29.16	-7.02	0.000
Height	5.0918	0.4237	12.02	0.000

S = 14.79 R-Sq = 61.6% R-Sq(adj) = 61.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	31592	31592	144.38	0.000
Residual Error	90	19692	219		
Total	91	51284			

Unusual Observations

Obs	Height	Weight	Fit	SE Fit	Residual	St Resid
9	72.0	195.00	161.87	2.08	33.13	2.26R
25	61.0	140.00	105.86	3.62	34.14	2.38R
40	72.0	215.00	161.87	2.08	53.13	3.63R
84	68.0	110.00	141.50	1.57	-31.50	-2.14R

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.

MEASURES OF CENTER

Measures of Center are values that tend to cluster around the “middle” of a set of data values. The mean, median, and *mode* are Measures of Center. To determine the mode of a data set using Minitab, the *Tally* procedure is used.

DATA: Pulse.MTW

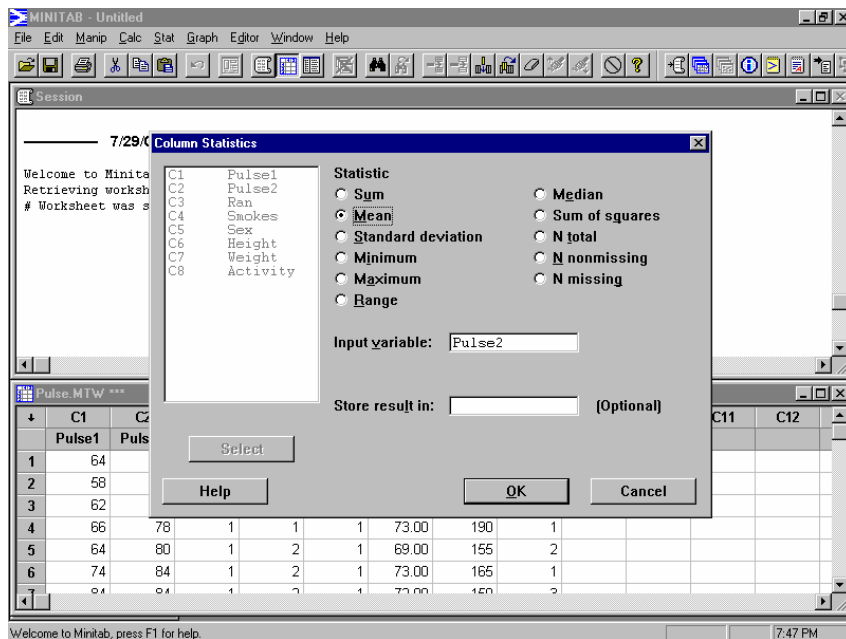
PATH: Calc => Column Statistics => select “Mean” => move the variable to the “Input Variable” cell.

OPTIONS: To obtain the median, select “Median” instead of “Mean.”

RELATED STATISTICAL PROCEDURE: The “Mean” and “Median” can also be obtained using **PATH:** Stat => Basic Statistics => Descriptive Statistics.

EXAMPLE: Obtain the mean of ‘PULSE2.’

Dialog Box Input: “Mean” is selected. Variable ‘PULSE2’ is moved into the “Input Variable” cell.



Minitab Output: Mean of ‘PULSE2’

Mean of Pulse2

Mean of Pulse2 = 80.000

MEASURES OF VARIATION

Measures of Variation are values that indicate how much variation exists within a data set. That is, how “spread out” the data values are. The range and the standard deviation are Measures of Variation.

DATA: Pulse.MTW

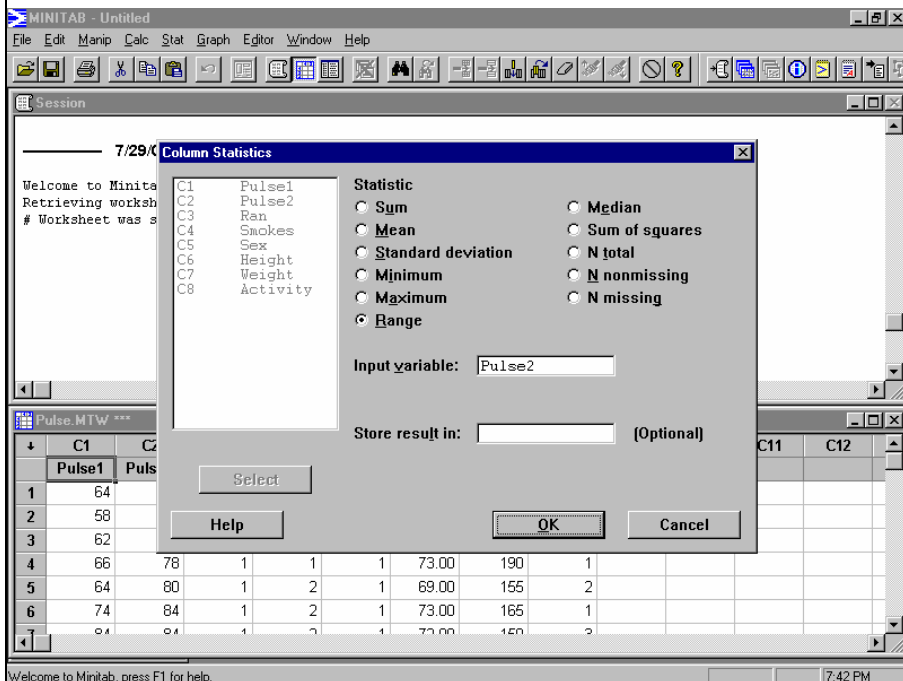
PATH: Calc => Column Statistics => select “Range” => move the variable to the “Input Variable” cell.

OPTIONS: To obtain the standard deviation, select “Standard Deviation” instead of “Range.”

RELATED STATISTICAL PROCEDURE: Measures of Variation can also be obtained using **PATH:** Stat => Basic Statistics => Descriptive Statistics.

EXAMPLE: Obtain the range for ‘PULSE2.’

Dialog Box Input: “Range” is selected. Variable ‘PULSE2’ is moved into the “Input Variable” cell.



Minitab Output: Range for ‘PULSE2’

Range of Pulse2

Range of Pulse2 = 90.000

NORMAL DISTRIBUTION (Probabilities for a Normally Distributed Variable)

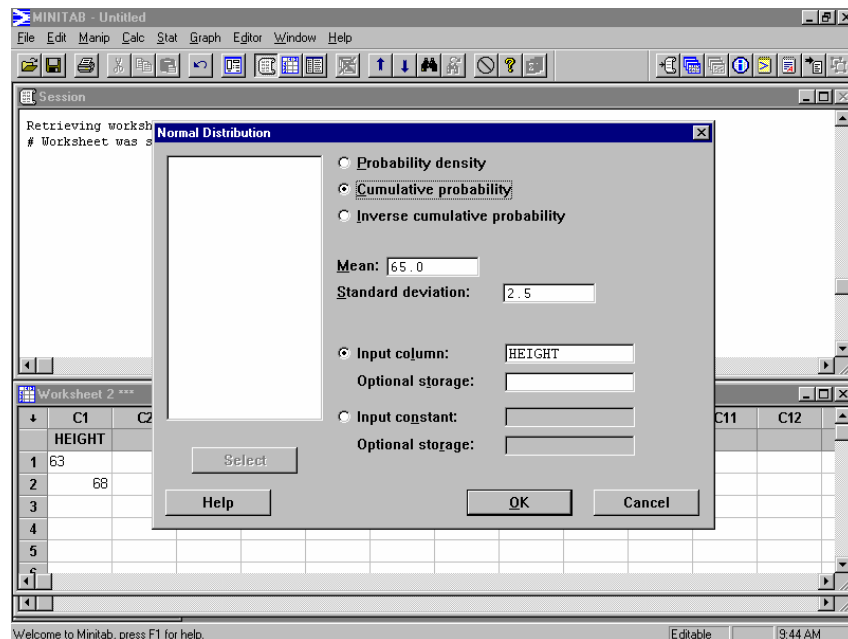
Minitab can be used in determining areas (probabilities) under the normal curve between 2 specified values, less than a specified value, or greater than a specified value, for a given *mean* and *standard deviation*. Note that in order to determine the actual probabilities, some basic arithmetic operations will need to be performed by the user.

Before beginning the process, the value or values must be stored in a column of the Minitab worksheet.

PATH: Calc => Probability Distributions => Normal => select the “Cumulative probability” button => type the mean in the “Mean” cell => type the standard deviation in the “Standard deviation” cell => select the “Input column” button => type the name of the column (variable) in which the values were stored, in the “Input column” cell.

EXAMPLE: The heights of females, aged 18 – 45 are normally distributed with a mean of 65 inches, and a standard deviation of 2.5 inches. Minitab will be used to find the probability that a randomly selected female between 18 and 45 years of age will be between 63 and 68 inches tall ($P(63 \leq x \leq 68)$).

Dialog Box Input: The mean and standard deviation are typed in the corresponding cells. “Input column” button is selected, and the name of the column (HEIGHT) is typed in the “Input column” cell.



Minitab Output: Probabilities can now be determined using basic arithmetic operations.

Cumulative Distribution Function

Normal with mean = 65.0000 and standard deviation = 2.50000

x	P(X <= x)
63.0000	0.2119
68.0000	0.8849

.2119, is the probability that $X \leq 63$.

.8849 is the probability that $X \leq 68$.

The probability that X is between 63 and 68 is $.8849 - .2119 = .673$

NORMAL PROBABILITY PLOT

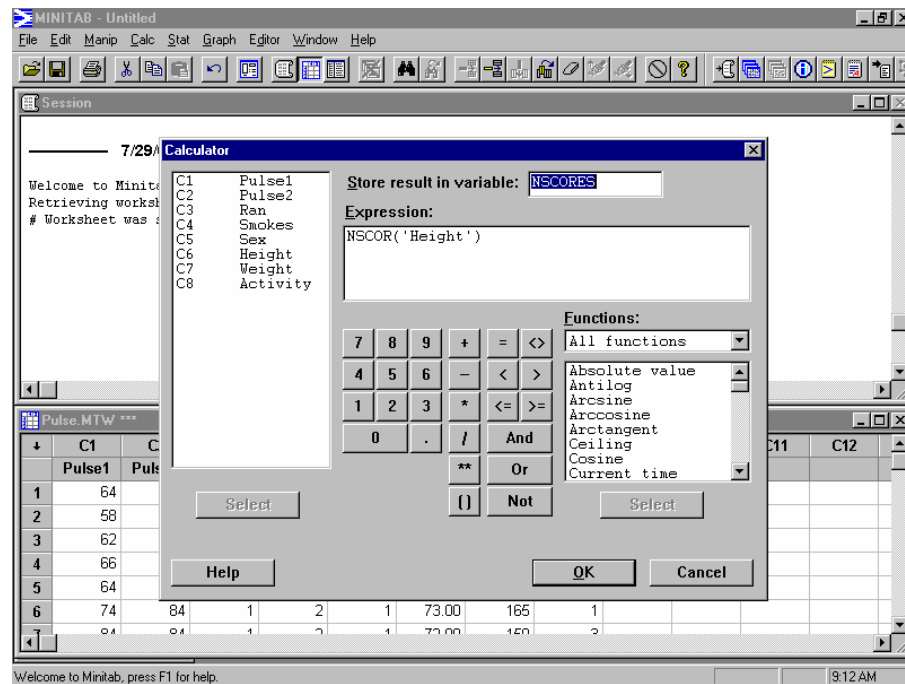
Detecting normality from a *histogram* is difficult when data sets are not large. A *Normal Probability Plot* compares the data with what would be expected of data that is normally distributed. If the plot exhibits a linear pattern, then the data is probably normally distributed. If the plot exhibits significant curvature, then the data is not normally distributed.

DATA: Pulse.MTW

PATH (Part 1 of 2): Calc => Calculator => type NSCORES in the “Store result in variable” cell. From the “Functions” list box, select “Normal Scores.” Move variable to the “Expression” cell (entered inside the parentheses by selecting a variable) => Click OK.

EXAMPLE: Obtain a Normal Probability Plot for the variable ‘HEIGHT.’

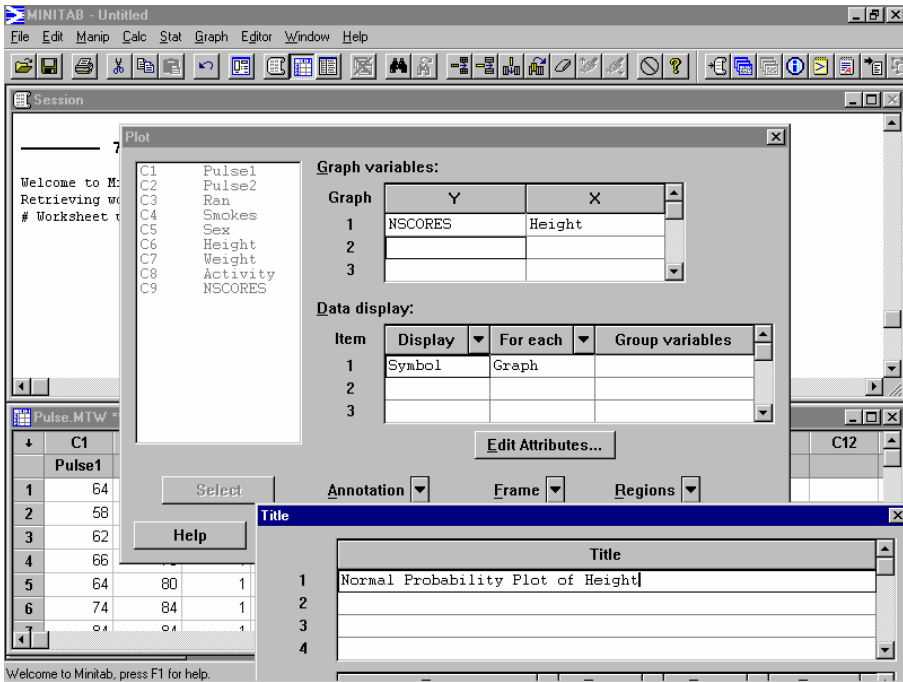
Dialog Box Input: This stores the normal scores in a column of the worksheet named ‘NSCORES’.



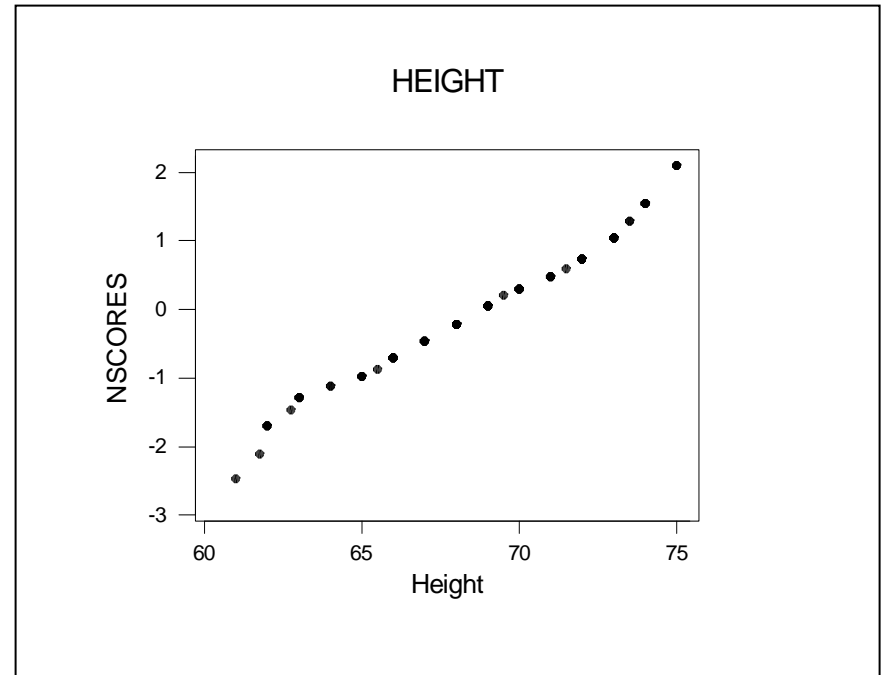
PATH (Part 2 of 2): Graph => Plot => move “NSCORES” to the “Y-variable” cell. Move the variable of interest to the “X-variable” cell. Click OK.

OPTIONS: To add a title, use the “Annotation” drop-down box, and select “Title.” Type in a title.

Dialog Box Input: ‘NSCORES’ is moved into the “Y-variable” cell, and variable ‘HEIGHT’ is moved into the “X-variable” cell. A title is added.



Minitab Output: Normal Probability Plot of ‘HEIGHT’



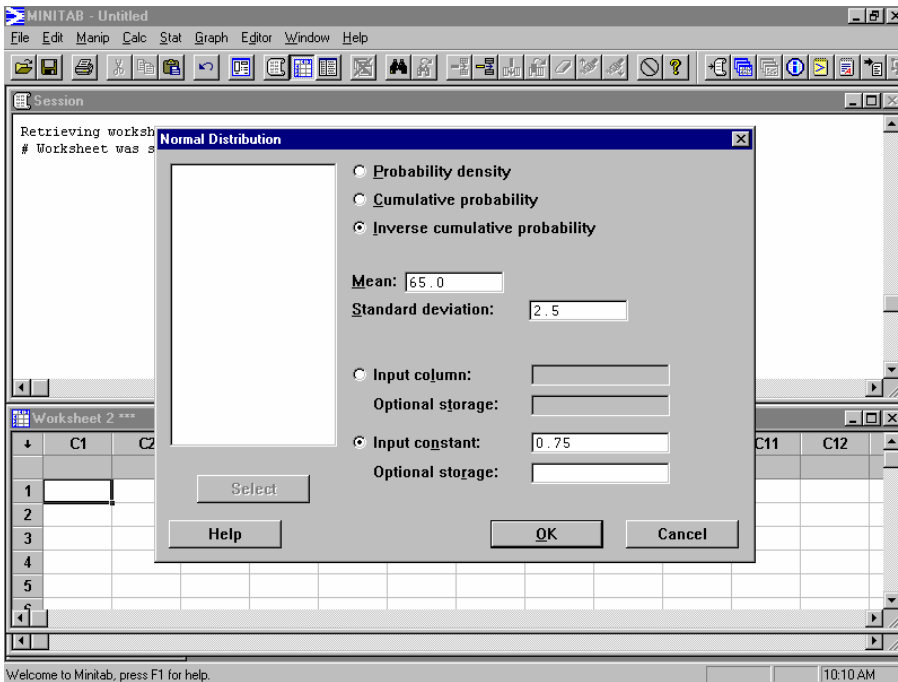
NORMALLY DISTRIBUTED VARIABLE (Determining an Observation Corresponding to a Specified Percentile)

Given a normally distributed variable, with a specified *mean* and *standard deviation*, Minitab can be used to determine an observation corresponding to a designated percentile.

PATH: Calc => Probability Distributions => Normal =>select the “Inverse cumulative probability” button => type the mean in the “Mean” cell => type the standard deviation in the “Standard deviation” cell => select the “Input constant” button => type the desired percentile (in decimal form) in the “Input constant” cell.

EXAMPLE: Heights of females aged 18 – 45 years, are normally distributed with a mean of 65 inches, and a standard deviation of 2.5 inches. Determine the height that corresponds to the 75th percentile.

Dialog Box Input: “Inverse cumulative probability” is selected. The mean and standard deviation are typed in the corresponding cells. “Input constant” button is selected, and the desired percentile is typed in the “Input constant” cell.



Minitab Output: Value corresponding to the 75th percentile is given.

Inverse Cumulative Distribution Function

Normal with mean = 65.0000 and standard deviation = 2.50000

P(X <= x)	x
0.7500	66.6862

66.6862 is the value that corresponds to the 75th percentile.

NORMALLY DISTRIBUTED VARIABLE (Simulating a Normal Distribution)

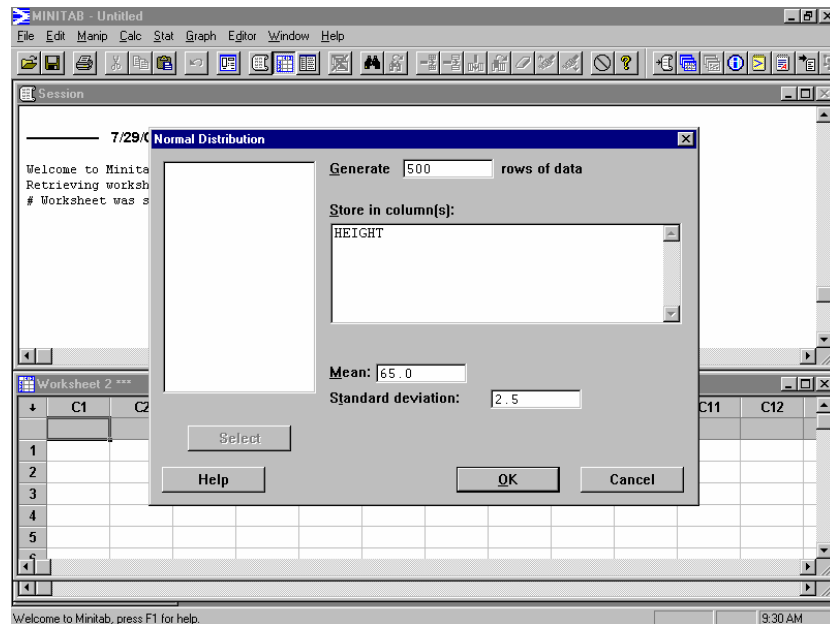
There are times when simulating a variable is useful in helping to visualize the distribution of the variable. Minitab can be used to simulate a specified number of observations for a normally distributed variable with a given *mean* and *standard deviation*.

PATH: Calc => Random Data => Normal => type the number of values to be simulated, in the “Generate rows of data” cell => type the name of the variable in the “Store in column(s)” cell => type the mean in the “Mean” cell => type the standard deviation in the “Standard deviation” cell.

OPTIONS: To view the observations in the form of a histogram, use Minitab’s *Histogram* procedure after running the simulation.

EXAMPLE: It is known that the heights of females aged 18 – 45 are normally distributed with a mean of 65 inches, and a standard deviation of 2.5 inches. Minitab will be used to simulate 500 values of a normally distributed variable, given this mean and standard deviation.

Dialog Box Input: 500 is typed in the “Generate rows of data” cell. HEIGHT is typed in the “Store in column(s)” cell. The values of the mean and standard deviation are typed in their corresponding cells.



Minitab Output: Simulated observations are placed in a column of the worksheet. Note that what is shown below is a partial list.

HEIGHT
63.0303
67.0651
69.1892
67.6786
65.5887
64.9849
66.2807
63.9197
62.2256

PARETO CHART

A **Pareto Chart** is a special type of bar chart. Frequencies displayed in a Pareto Chart are ordered from highest to lowest. A cumulative relative frequency curve is also included. The purpose of a Pareto Chart is to call attention to the most frequently occurring values of a nominal variable. Pareto Charts are often associated with quality control situations to present production defects in order of frequency, but it can be used to present any form of qualitative data.

DATA: Pulse.mtw

PATH: Stats => Quality Tools => Pareto Chart

EXAMPLE: Obtain a **Pareto Chart** for the variable “ACTIVITY.” In the **Pareto Charts** dialog box insert the variable of interest, here “ACTIVITY,” into the “**Chart Defects Data In**’ cell and select the **OK** button.

Dialog Box: Insert variable into the top right cell (‘Chart defects data in’)

Pareto Chart

Chart defects data in:

BY variable in: [optional]

Default [all on one page, same ordering of bars]

One chart per page, same ordering of bars

One chart per page, independent ordering of bars

Chart defects table

Labels in:

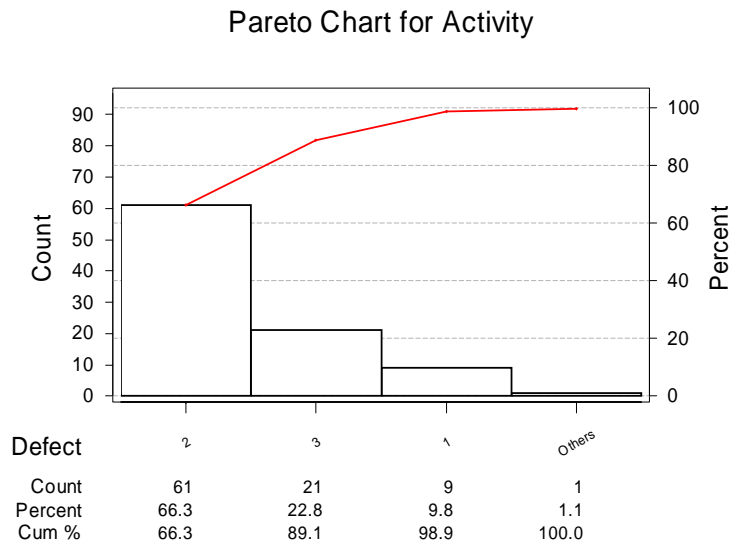
Frequencies in:

Combine defects after the first % into one

Title:

Select Help OK Cancel

Minitab Output: The number of occurrences by activity level is listed in order of decreasing frequency.



PIE CHART

A *Pie Chart* is used for presenting categorical data. The chart consists of a circle subdivided into sections. The size of each section is proportional to the quantity it represents.

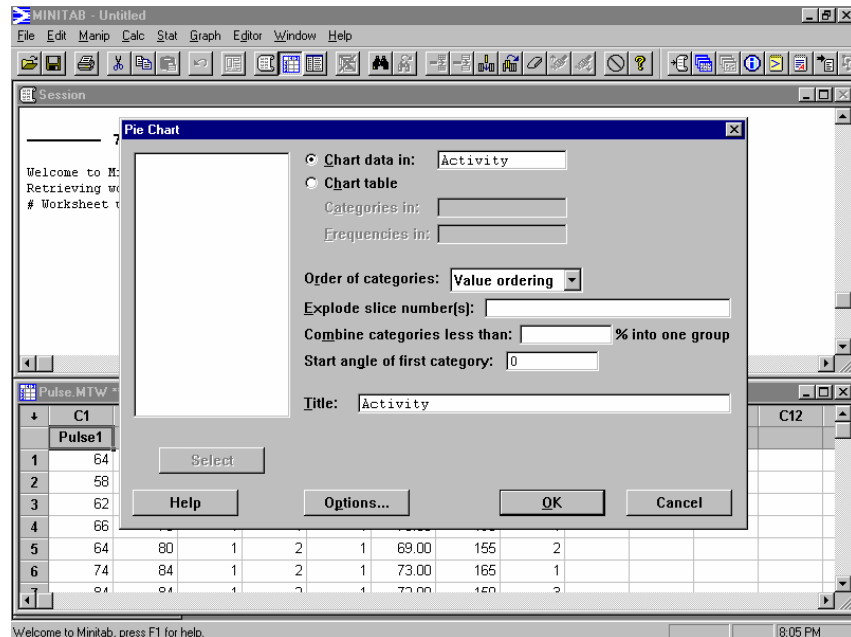
DATA: Pulse.MTW

PATH: Graph => Pie Chart => move variable to the “Chart data in:” cell.

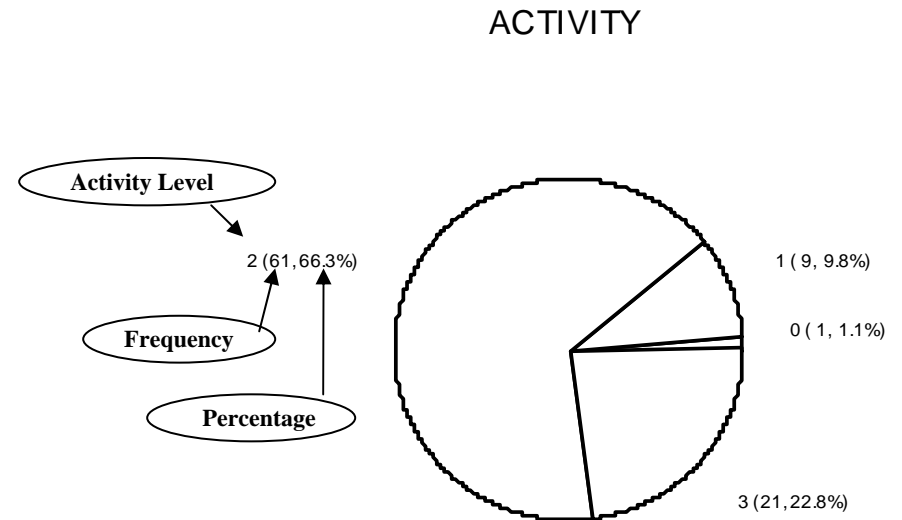
OPTIONS: Include a title for the chart by typing it in the “Title” cell. Slices may be exploded, rotated, combined, etc. using options in the dialog box.

EXAMPLE: Obtain a Pie Chart of the ‘ACTIVITY’ of the students.

Dialog Box Input: Variable ‘ACTIVITY’ moved into the “Chart data in” cell and a title is added.



Minitab Output: Pie Chart of ‘ACTIVITY’



Selected Options:

First, it is important to note that by default, Minitab starts slices at the 3 o'clock position and goes in a counter-clockwise direction from there. In other words, the slice in the 3 o'clock position is the first slice, and proceeding counter-clockwise, the next slice is the second slice, etc. In the case of the Pie Chart on the previous page, the first slice is the small one with the label "0 (1, 1.1%)." The second slice is the one counter-clockwise above it (1 (9, 9.8%)), etc.

Order of categories: This option orders the slices of the pie.

Value ordering is the default. Using **Value ordering** places the slices in alphabetic, or in numeric order.

Worksheet is used when dealing with text-type data.

Increasing places the slices in increasing order by slice size.

Decreasing places the slices in decreasing order by slice size.

Options Button:

"Label slices with category and:" allows frequency and count for the slices to be included in the chart. It is from the *pull-down menu* that you will be able to choose one of four options. Choosing **"nothing,"** will produce a **Pie Chart** with slices labeled by categories, and nothing else. Choosing **"frequency, percent,"** will produce a **Pie Chart** with slices labeled by categories, as well as frequencies (counts) and percents. Choosing **frequency** will produce a **Pie Chart** with slices labeled by categories and frequencies, and finally, choosing **percent** will produce a **Pie Chart** with slices labeled by categories and percents. In this example, we will choose **"frequency, percent."** *Click OK.* You will be returned to the **Pie Chart Dialog Box.**

Exploding a Slice:

When making a **Pie Chart** in **Minitab**, you may choose to **explode** a slice. **Exploding** a slice means that the slice is moved away from the center of the pie. This is useful for emphasis of a particular category. Knowing that the first slice of the Pie Chart is at 3 o'clock, one can "assign" a number to each slice in a counter-clockwise direction. To explode a specific enter the "assigned" slice number into the **"Explode slice number"** cell.

Title:

Add a title to your **Pie Chart** by entering one into the **"Title"** cell.

SAMPLING DISTRIBUTION OF THE MEAN (Obtaining Means for Generated Samples)

Once samples are generated and stored in columns of a worksheet, Minitab can be used to obtain the means of the samples.

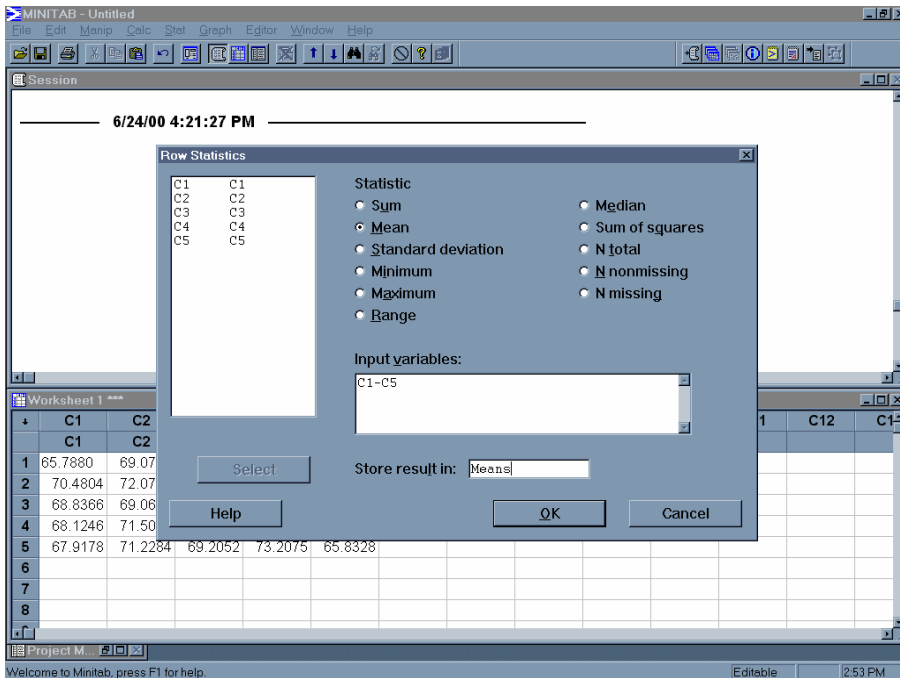
PATH: Calc => Row Statistics => select the “Mean” button => type the names of the columns in which the samples are stored, in the “Input variables:” cell => type the name to be used for the column in which the sample means will be stored.

RELATED STATISTICAL PROCEDURES: For how to generate samples, see *SAMPLING DISTRIBUTION OF THE MEAN (Simulation)*. After the means of the samples have been generated, Minitab’s *HISTOGRAM* procedure can be used to obtain a *Histogram* of the means.

EXAMPLE: The heights of American males aged 18-74 are normally distributed with a mean of 69.1 inches, and a standard deviation of 2.1 inches. 1500 samples of size 5 have been generated by Minitab. Obtain the means for these samples.

Dialog Box Input: “Mean” button is selected. Columns of the worksheet that contain the generated samples are typed in “Input variables:” cell.

Minitab Output: A partial list of the samples and their corresponding means.



C1	C2	C3	C4	C5	Means
65.7880	69.0764	71.8685	72.3236	72.9464	70.4006
70.4804	72.0730	72.8336	71.1516	70.6830	71.4443
68.8366	69.0675	72.7578	72.9450	68.2463	70.3707
68.1246	71.5071	68.5967	69.2977	70.3669	69.5786
67.9178	71.2284	69.2052	73.2075	65.8328	69.4783

SAMPLING DISTRIBUTION OF THE MEAN (Simulation)

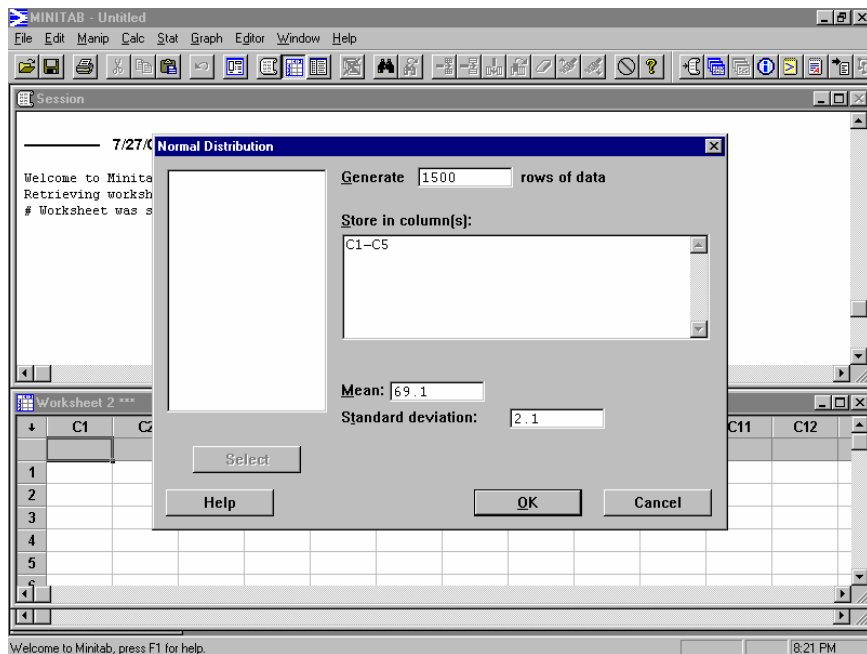
Given a normally distributed variable with **mean** μ , and **standard deviation** σ , Minitab can be used to simulate a large number of samples of a given size for the variable.

PATH: Calc => Random Data => Normal => type the number of samples to be generated in the "Generate rows of data" cell => identify the columns in the worksheet to be used for storing the samples (see NOTE) => type the value of the mean in the "Mean" cell => type the standard deviation in the "Standard deviation" cell.

NOTE: Number of columns to be used will depend on sample size desired. For example, if the sample size is 6, columns will be identified as C1-C6, or any 6 contiguous columns in the worksheet.

EXAMPLE: The heights of American males aged 18-74 are normally distributed with a mean of 69.1 inches, and a standard deviation of 2.1 inches. Simulate 1500 samples of size 5 for this variable.

Dialog Box Input: Number of samples to be generated is typed in the "Generate rows of data" cell. Columns into which the samples will be placed are identified. *Mean* and *standard deviation* are typed in their respective cells.



Minitab Output: A partial list of the 1500 generated samples.

C1	C2	C3	C4	C5
65.7880	69.0764	71.8685	72.3236	72.9464
70.4804	72.0730	72.8336	71.1516	70.6830
68.8366	69.0675	72.7578	72.9450	68.2463
68.1246	71.5071	68.5967	69.2977	70.3669
67.9178	71.2284	69.2052	73.2075	65.8328

SCATTERPLOT

A *Scatterplot* is used to display information about the relationship involving two different variables. One point is plotted for each (x,y) pair in the data set.

DATA: Pulse.MTW

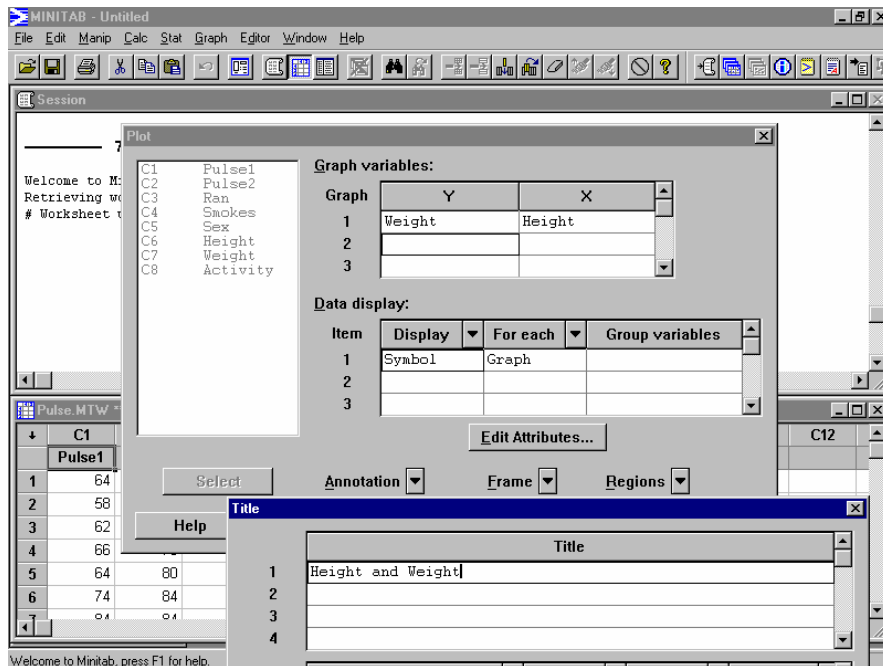
PATH: Graph => Plot => move the response variable to the “Y-variable” cell, and the predictor variable to the “X-variable” cell.

OPTIONS: Include a title for the plot by using the “Annotation” drop-down box, and typing in a title.

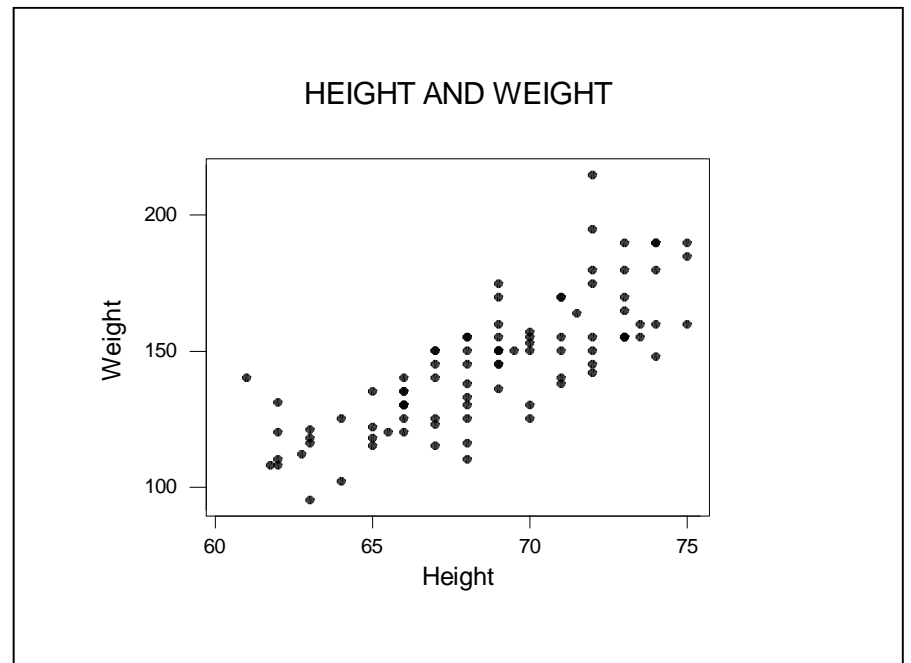
RELATED STATISTICAL PROCEDURES: Correlation, Linear Regression.

EXAMPLE: Obtain a scatterplot of the ‘HEIGHT’ and ‘WEIGHT’ of the students.

Dialog Box Input: Variable ‘WEIGHT’ is moved into the “Y-variable” cell, and ‘HEIGHT’ is moved into the “X-variable” cell. A title is added.



Minitab Output: Scatterplot of ‘HEIGHT’ and ‘WEIGHT’



SIMPLE RANDOM SAMPLE (SRS)

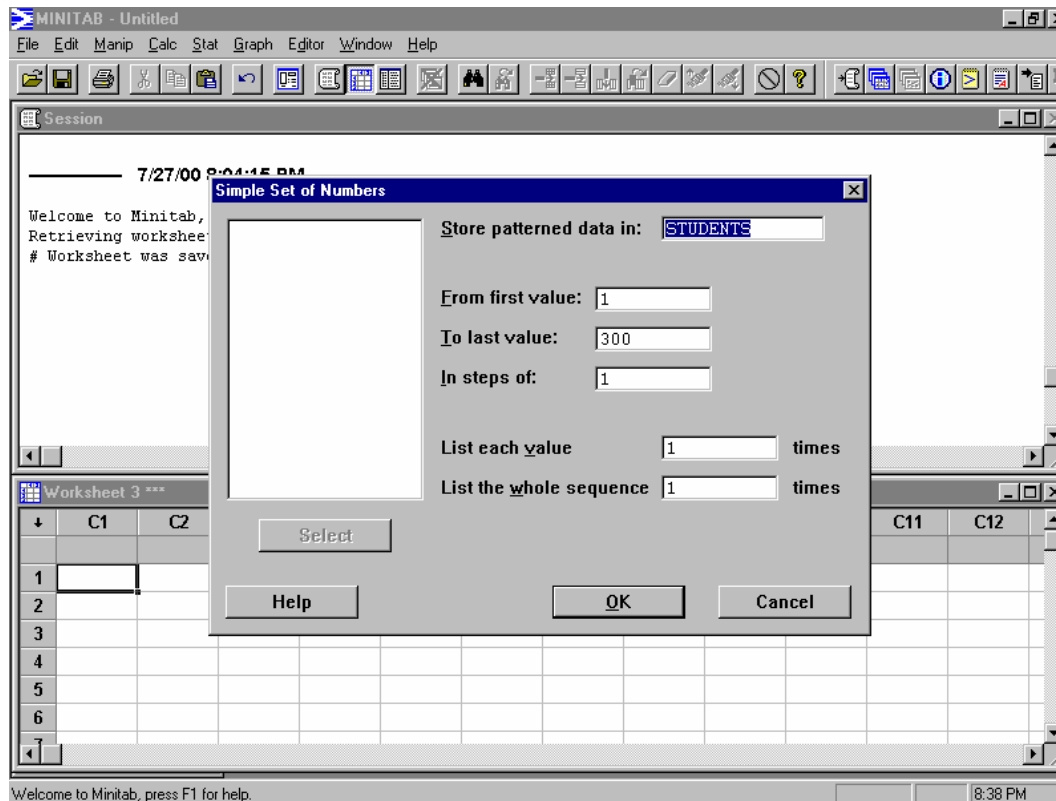
A *Simple Random Sample* is the most basic of sampling methods. The sample is chosen in a way such that each sample of a given size is equally likely to be selected.

PATH (Part 1 of 2): Calc => Make Patterned Data => Simple Set of Numbers => type desired column name into the “Store patterned data in” cell. Type 1 in the “From first value” cell. Type the total number of units in the sampling frame into the “To last value cell” (Note: This is not the same as sample size.)

EXAMPLE: From a group of 300 STAT 101 students, we wish to select 5 to attend a lecture presented by a renowned statistician.

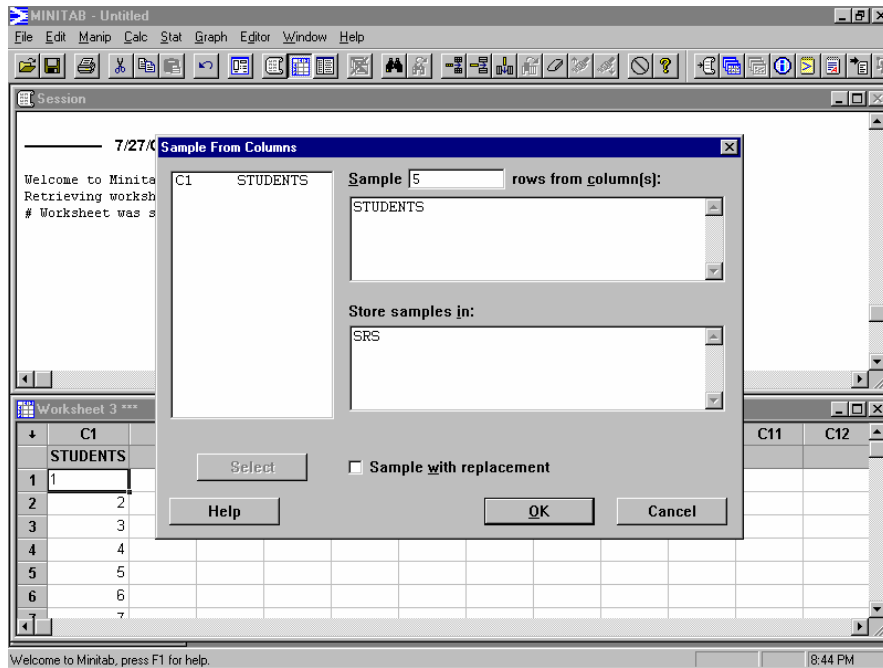
PROCESS: The numbers 1-300 will be placed in a column of the Minitab worksheet. Minitab will then randomly select 5 of these numbers and place them in the first available empty column. Note: The numbers obtained in the SRS column will vary, as this is a random sampling process.

Dialog Box Input: The numbers 1-300 will be placed in a column named ‘STUDENTS.’



PATH (Part 2 of 2): Calc => Random Data => Sample From Columns => type the sample size into the “Sample” cell (here 5)=> move the column ‘STUDENTS’ into the “Sample __ rows from column(s):” cell. Type SRS in the “Store Samples In:” cell.

Dialog Box Input: 5 numbers are randomly selected and placed in the column labeled ‘SRS’



Minitab Worksheet Column: Simple Random Sample of 5 Students

SRS
242
168
180
293
61

STEM-AND-LEAF PLOT

A *Stem-and-Leaf Plot* is similar to a histogram in that it shows where the data are concentrated, the general shape of the distribution, the range of the data and whether or not outliers are present. The Stem-and-Leaf Plot places the last digit of a value in the leaf, and all prior digits in the corresponding stem.

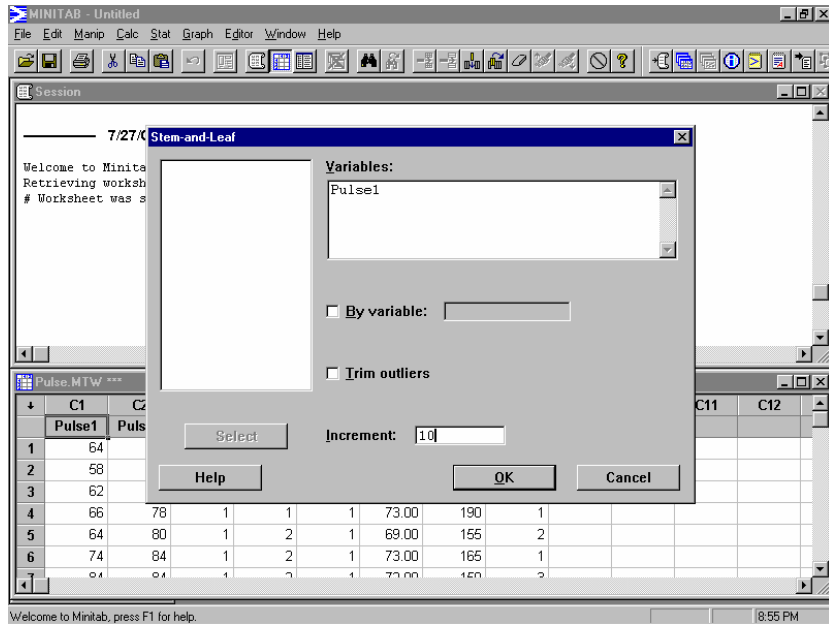
DATA: Pulse.MTW

PATH: Graph => Stem-and-Leaf => move variable to the “Variable” cell. Type the number 10 in the “Increment” cell.

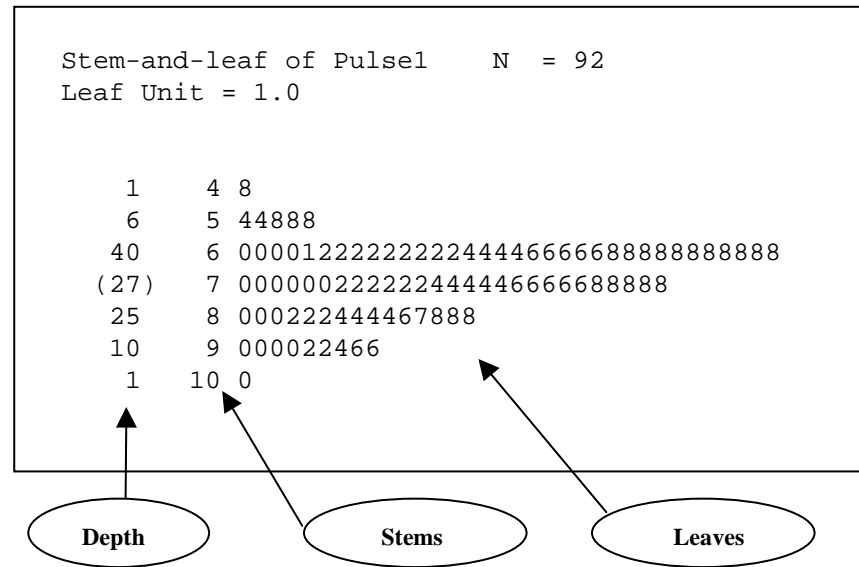
OPTIONS: To obtain a Stem-and-Leaf Plot using double stems, type the number 5 in the “Increment” cell.

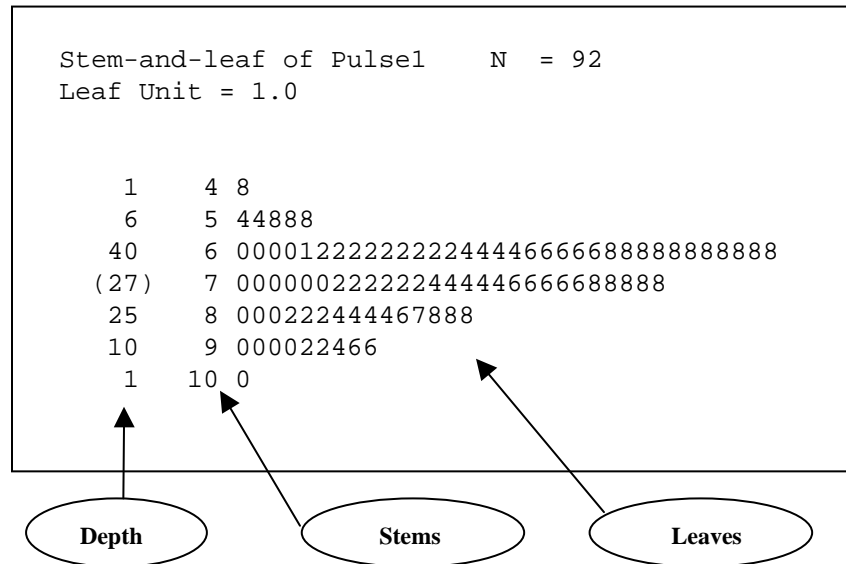
EXAMPLE : Obtain a Stem-and –Leaf Plot for the variable ‘PULSE1.’

Dialog Box Input: Variable ‘PULSE1’ is moved into the “Variable” cell.



Minitab Output: Stem-and-Leaf Plot of ‘PULSE 1’





Your plot displays the pulse rates of 92 people (denoted by $N=92$). There are several parts to the display. The first column (1, 6, 40, etc.) displays the **Depths**. The second column displays the **Stems** (4, 5, 6, etc.), and the remainder of the display shows the **Leaves**. The first **Stem** is 4, and the first **Leaf** is 8. This corresponds to one individual whose reported pulse was 48. The second **Stem** is 5. The first **Leaf** adjacent to this **Stem** is 4 and the second **Leaf** is also 4. This corresponds to two individuals whose reported pulses were both 54.

The **Leaf Unit** tells you where to place the decimal point in each data value. Since the **Leaf Unit** here is equal to 1.0, the decimal point goes at the end of each data value. Therefore the first observation is 48.0. If the **Leaf Unit** were 0.10, the first data value would be 4.8.

In explaining **Depth**, an understanding of the **median** is helpful. Recall from “The Five-Number Summary,” that the **median** is the middle value in a data set that has been ordered from smallest to largest.

The number in the **Depth** column in parentheses (27, in this example), is important. It is the line that contains the **median** (in this data set the **median** is 72). It is also the only line that contains the exact number of data values designated by the **Depth** number (there are 27 data values in this line). Note that in the line above this line, the **Depth** number is 40, but there are not 40 data values in that line (there are 34). However, there are 40 data values, if we add all of the data values on this line, and all of the data values on each line *above* it ($34+5+1 = 40$).

Now let’s look at the line with **Depth** number 25. There are not 25 data values in this line (there are 15) however, there are 25 data values, if we add all of the data values on this line, and each line *below* it ($15+9+1 = 35$).

Depending on your data, and the type of **stem-and-leaf plot** you define (for example, single versus double stems), the depth value of the line containing the median, may not be in parentheses. This is because the **median** is located between two lines.

TALLY

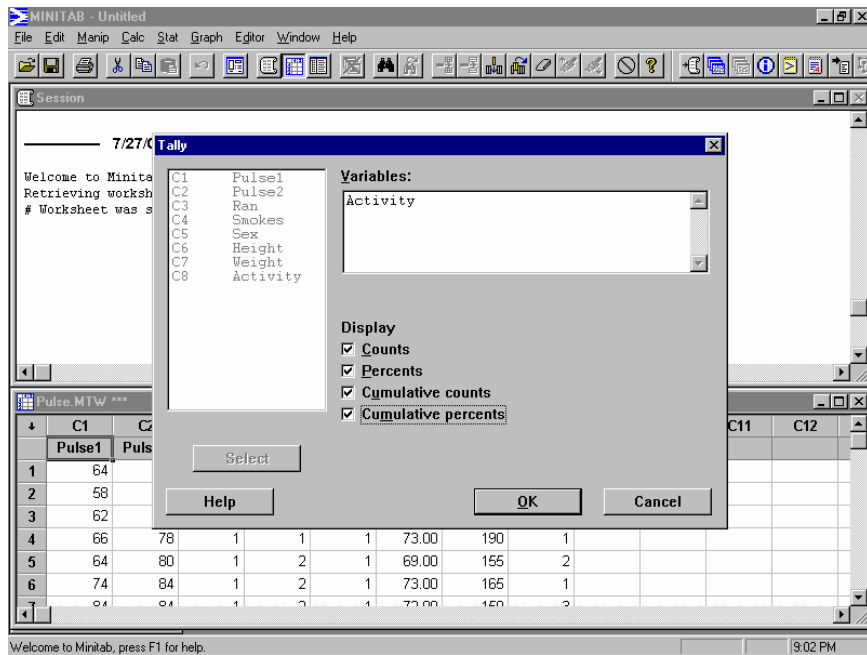
The *Tally* command is used to display any or all of the following, for each value of one or more variables: counts, cumulative counts, percents, cumulative percents. Tally can also be used to obtain a Single-Value Grouped Data Table, as well as the *Mode* of a data set.

DATA: Pulse.MTW

PATH: Stat => Tables => Tally => move variable(s) to the “Variables” cell. Select any or all of, counts, cumulative counts, percents, cumulative percents.

EXAMPLE: Use the Tally command to obtain counts, cumulative counts, percents, and cumulative percents for the variable ‘ACTIVITY.’

Dialog Box Input: Variable ‘ACTIVITY’ is moved to the “Variables” cell. Counts, cumulative counts, percents, and cumulative percents are selected.



Minitab Output: Tally of ‘ACTIVITY’

Tally for Discrete Variables: Activity

Activity	Count	CumCnt	Percent	CumPct
0	1	1	1.09	1.09
1	9	10	9.78	10.87
2	61	71	66.30	77.17
3	21	92	22.83	100.00
N=	92			

MINITAB DATA MANIPULATION: PROCEDURES TO EDIT MINITAB DATA FILES

CALCULATE

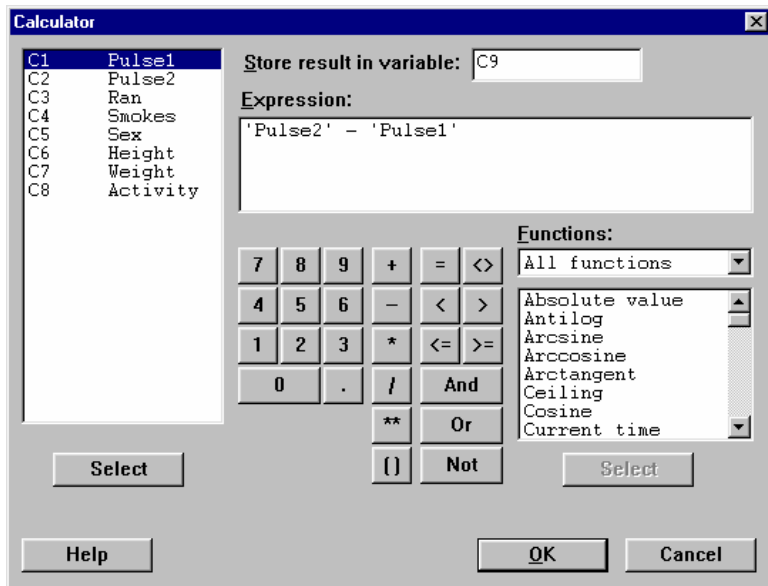
Minitab's calculator allows you to perform basic arithmetic operations, or very complex mathematical functions in the worksheet. Only very basic material will be covered here.

DATA: Pulse.mtw

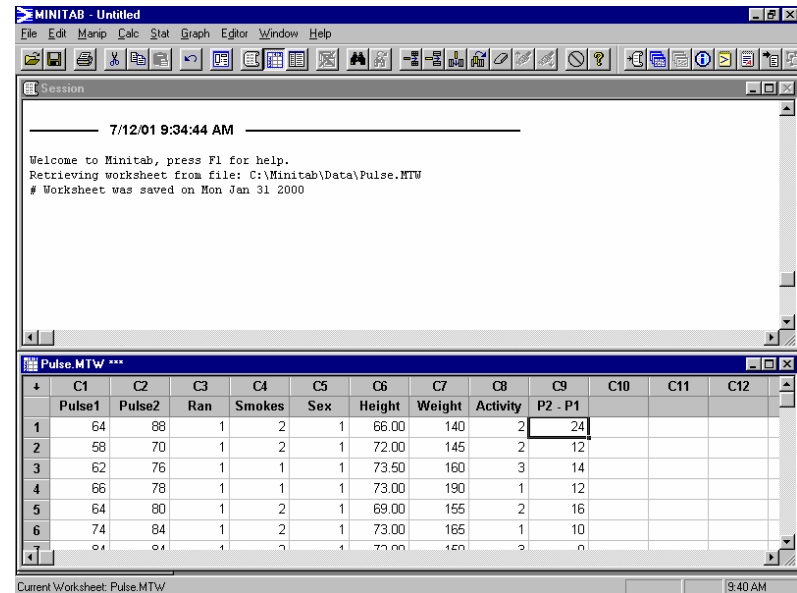
PATH: Calc => Calculator

EXAMPLE 1: Suppose you wish to calculate the difference between PULSE2 and PULSE1, and store this result in column C9.

Dialog Box Input: Enter C9 into the "Store Results in variable" cell and in the "Expression" cell enter the formula that would find the difference in pulses.

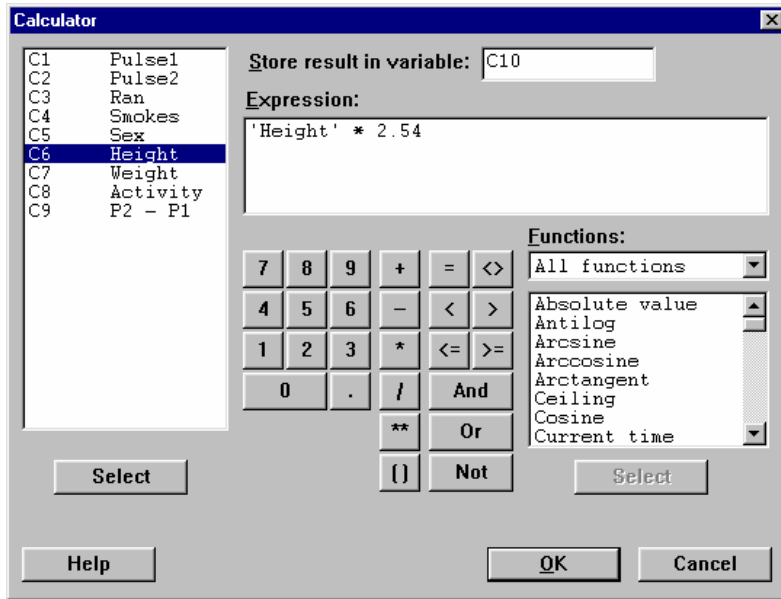


Minitab Output: The difference in pulses is recorded in column C9. (column label added.)

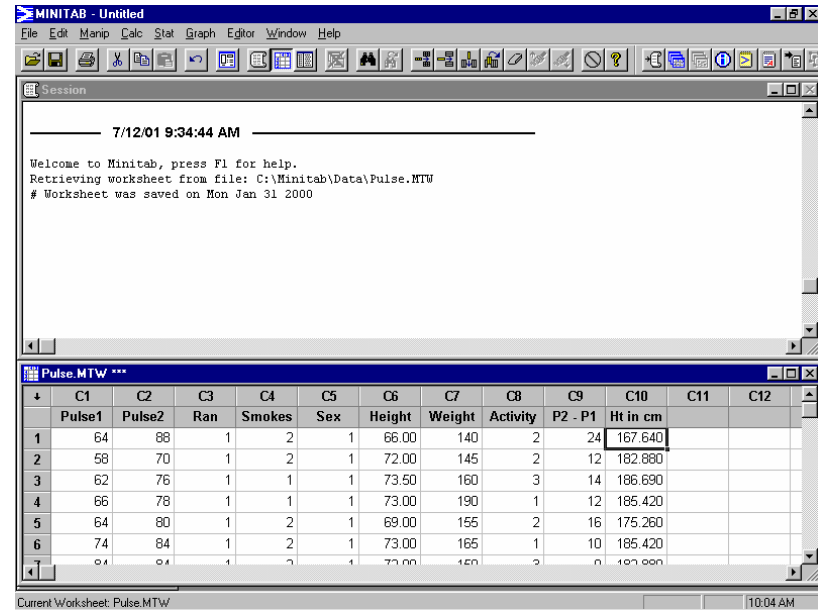


EXAMPLE 2: Suppose you wish to convert HEIGHT (given in inches) to HEIGHT in centimeters, and store this result in column C10.

Dialog Box Input: Enter C10 in the “Store results in” cell. Since 1 inch equals approximately 2.54 centimeters, the variable HEIGHT is multiplied (denoted by *) by 2.54.



Minitab Output: The conversion of inches to centimeters is saved in C10. (column label added.)



Calculating Individual Column or Row Statistics:

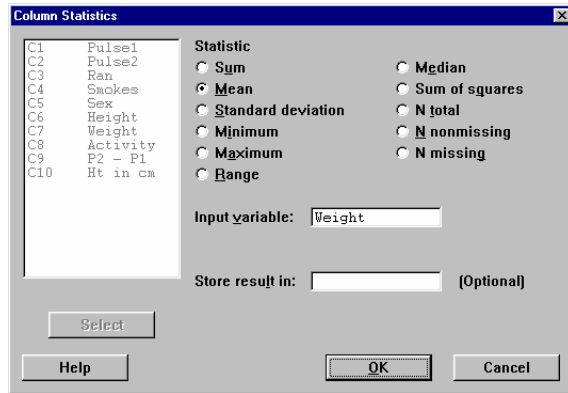
Various statistics can be calculated within the columns and rows of the worksheet. It is important to note that calculated column statistics will be displayed in the **Session Window**, and calculated row statistics are calculated across the rows of the designated columns, and **stored in the corresponding rows of a new column**. (Examples on following page.)

INDIVIDUAL COLUMN STATISTICS:

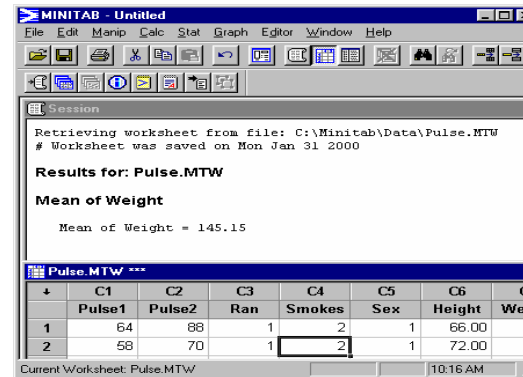
PATH: Calc => Column Statistics

EXAMPLE: Calculate the mean of column C7 WEIGHT.

Dialog Box Input: Select the desired statistic and enter the variable of interest into the “Input Variable” cell.



Minitab Output: Unless otherwise specified, the result is placed into The Session window (here the mean of the weights is 145.15).

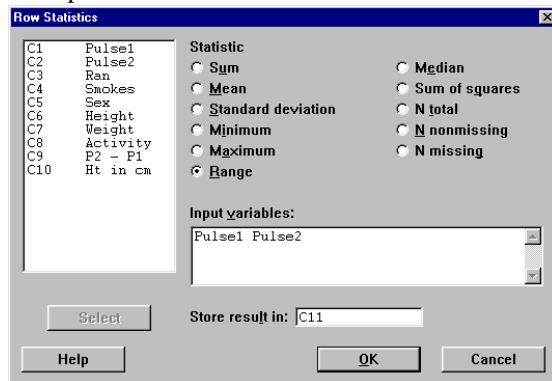


INDIVIDUAL ROW STATISTICS:

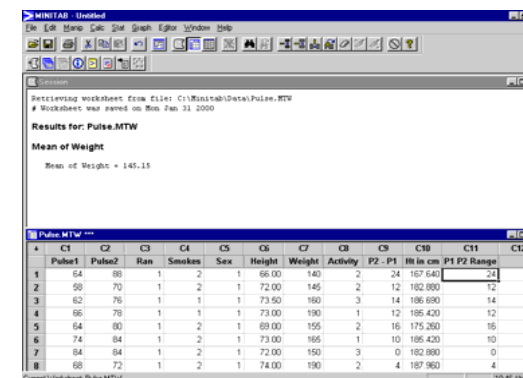
PATH: Calc => Row Statistics

EXAMPLE: Calculate the rowwise range of PULSE1 and PULSE2 and store the result in column C11

Dialog Box Input: Select the desired statistic, enter the variable(s) into the “Input Variables” cell and indicate the column into which the results are to be placed.



Minitab Output: The range of the data are placed in C11. (column label added.)



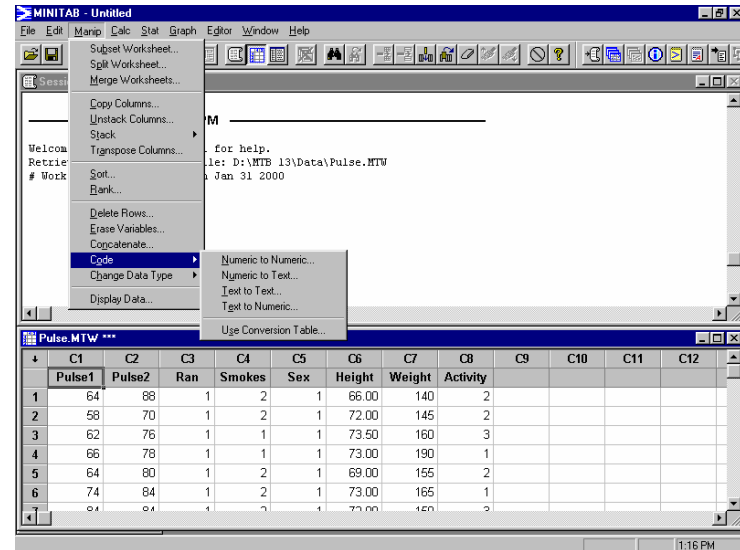
CODE

The **Code** option allows for the conversion of data from one type (numeric, text, or date/time) to another. For numeric data, it can be used to recode data values into groups.

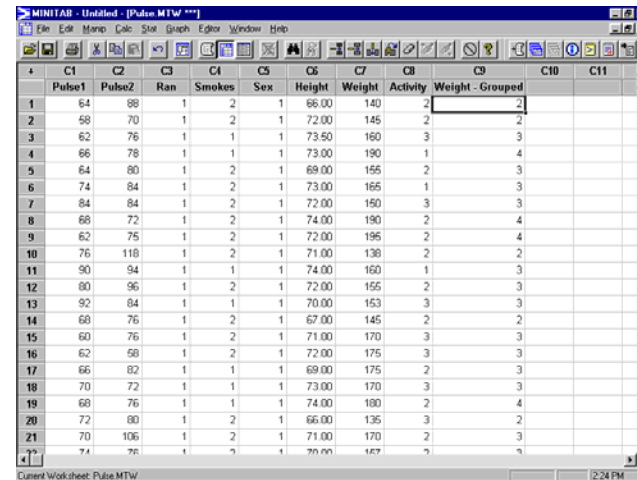
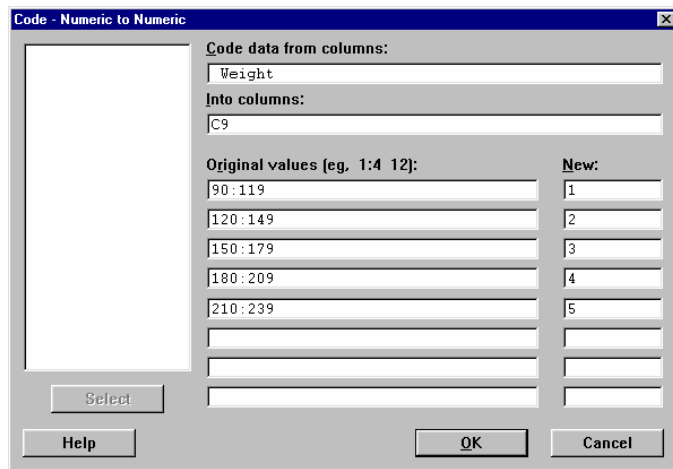
DATA: Pulse.mtw

PATH: Manip => Code => select appropriate option

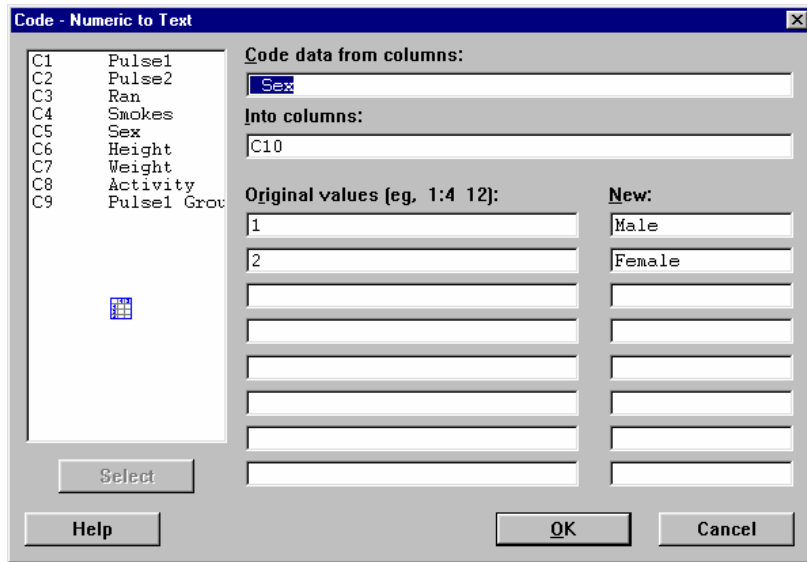
NOTE: As Minitab allows for many, many variables, always code into a new column. In doing so, both the original form of the data and its new format will be available for data analysis.



EXAMPLE 1: Numeric to Numeric. There are 36 different weights reported for this study's participants. Presenting a Tally of the variable 'WEIGHT' would result in a lengthy table. Recode the 'WEIGHT' values into five equal-sized groups beginning at 90 pounds, with each group representing a 30-pound range. **HOW:** Select Numeric to Numeric; identify the "Code from" column and the column into which the data are to be placed (here C9); enter the original value(s) and the new value; select okay



EXAMPLE 2: Numeric to Text. The variable ‘SEX’ has numeric values representing Male (1) and Female (2) participants. As a qualitative variable, these values can be converted to Text equivalents. This will allow for tables as produced via a Tally to have more meaningful labels in the margins. HOW: Select Numeric to Text; identify the “Code from” column and the column into which the data are to be placed (here C10); enter the original value(s) and the new text value; select okay



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10-T	C11
	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity	Weight - Grouped	Sex (text)	
1	64	88	1	2	1	66.00	140	2	2	Male	
2	58	70	1	2	1	72.00	145	2	2	Male	
3	62	76	1	1	1	73.50	160	3	3	Male	
4	66	78	1	1	1	73.00	190	1	4	Male	
5	64	80	1	2	1	69.00	155	2	3	Male	
6	74	84	1	2	1	73.00	165	1	3	Male	
7	84	84	1	2	1	72.00	150	3	3	Male	
8	68	72	1	2	1	74.00	190	2	4	Male	
9	62	75	1	2	1	72.00	195	2	4	Male	
10	76	118	1	2	1	71.00	138	2	2	Male	
11	90	94	1	1	1	74.00	160	1	3	Male	
12	80	96	1	2	1	72.00	155	2	3	Male	
13	92	84	1	1	1	70.00	153	3	3	Male	
14	68	76	1	2	1	67.00	145	2	2	Male	
15	60	76	1	2	1	71.00	170	3	3	Male	
16	62	58	1	2	1	72.00	175	3	3	Male	
17	66	82	1	1	1	69.00	175	2	3	Male	
18	70	72	1	1	1	73.00	170	3	3	Male	
19	68	76	1	1	1	74.00	180	2	4	Male	
20	72	80	1	2	1	66.00	135	3	2	Male	
21	70	106	1	2	1	71.00	170	2	3	Male	
22	74	76	1	2	1	70.00	167	2	3	Male	

SAMPLE OUTPUT:

Tally for Discrete Variables: Weight - Grouped, Sex, Sex (text)

Weight - Grouped	Count	Sex	Count	Sex (text)	Count
1	13	1	57	Female	35
2	35	2	35	Male	57
3	34	N=	92	N=	92
4	9				
5	1				
N=	92				

(Recall that there were 36 different weights and now there are five weight groups.)

Note: The order for the recoding of the ‘SEX’ variable has changed for the new variable ‘SEX (text)’ and is now based upon alphabetic order.

NOTE: Always check a couple of the values that have been converted to check that the procedure has accomplished what you intended it to do.

DATA ENTRY

Before you can analyze your data, it must be entered into a **Worksheet**. The **Worksheet** consists of rows and columns into which you will enter your data. Each column represents a variable – a piece of information common to all data sources, and each row a case – all the different variable responses from one data source. **Minitab** opens to an empty **Worksheet** in the **Data Window** as shown below.

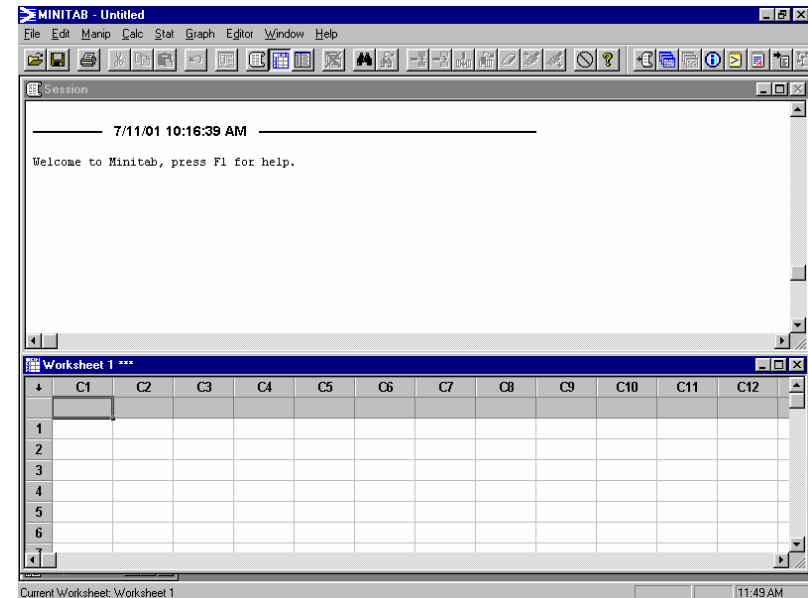
DATA ENTRY

Minitab will accept three types of data:

- **Numeric.** Numbers, and *, the missing value symbol in Minitab. Numbers can be positive or negative and can contain a decimal point. An asterisk (*) is used to represent missing numeric data.
- **Text.** Can consist of letters, numbers, spaces and special characters. Limit for text is 80 characters in length.
- **Date/Time.** Data can be dates such as Sep-22-1938, or 4/17/58. Times such as 08:30:17 AM can be used, as well as a combination of both date and time, such as 9/22/38 08:30:17 AM.***

Data can be entered into a **Worksheet**:

- Directly.
- From existing **Minitab** worksheets.
- From other applications such as EXCEL.
- From a textfile

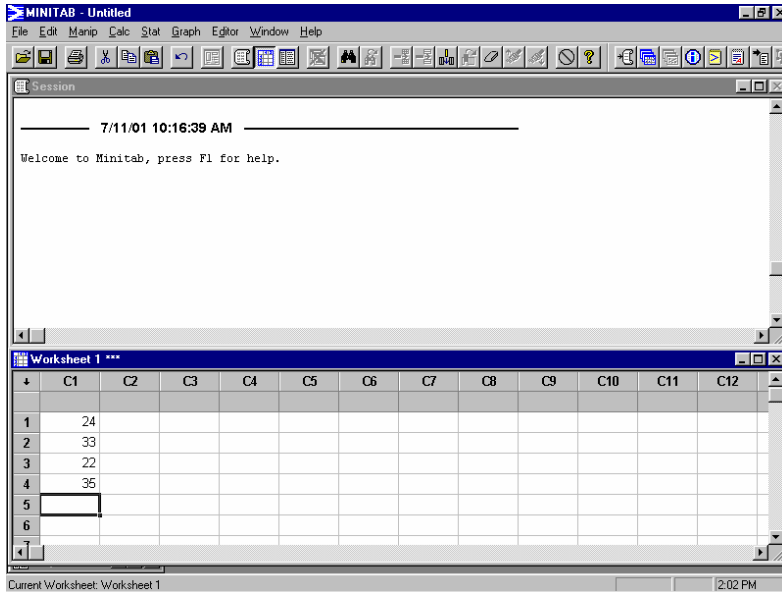


Direct Data Entry.

Data can be entered into the **Worksheet** by rows or columns. To designate the manner in which you wish to enter the data, use the direction arrow in the upper left corner of the **worksheet**. The default for the arrow is pointing downward, as is shown above. This means the data will be entered in columns. To change the direction of the arrow, click once on it. The arrow will now point to the right, and the data will subsequently be entered in rows.(See examples to right.)

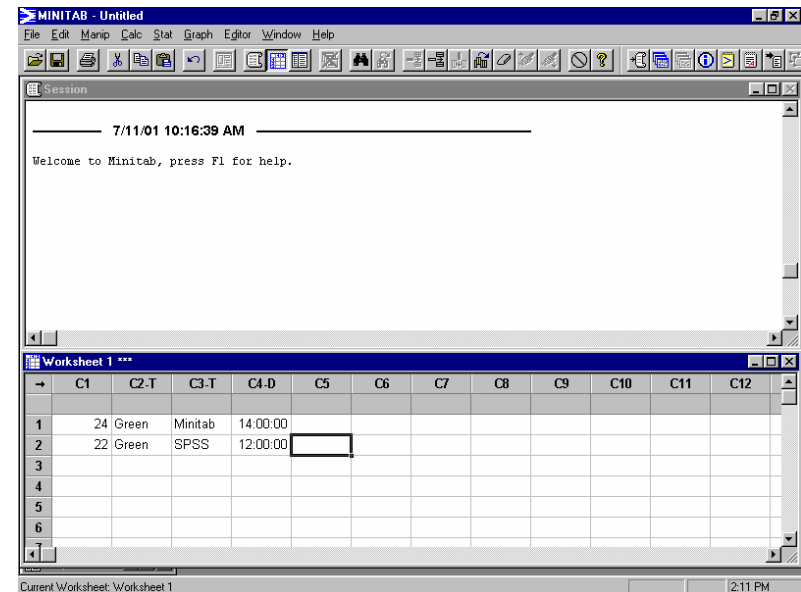
In order to enter a data value into a cell within the worksheet, the cell must be *active*. A cell is made active by clicking once in the cell.

Entering Data Into Columns. Direction arrow pointing downward. Make active the cell in which you wish to enter your first data value. Enter your first data value. Press *ENTER* to move to the next cell. Note that the “down arrow” key on your keyboard will perform the same function. Continue your data entry in this manner. Press *CTRL ENTER* to move to the top of the next column.



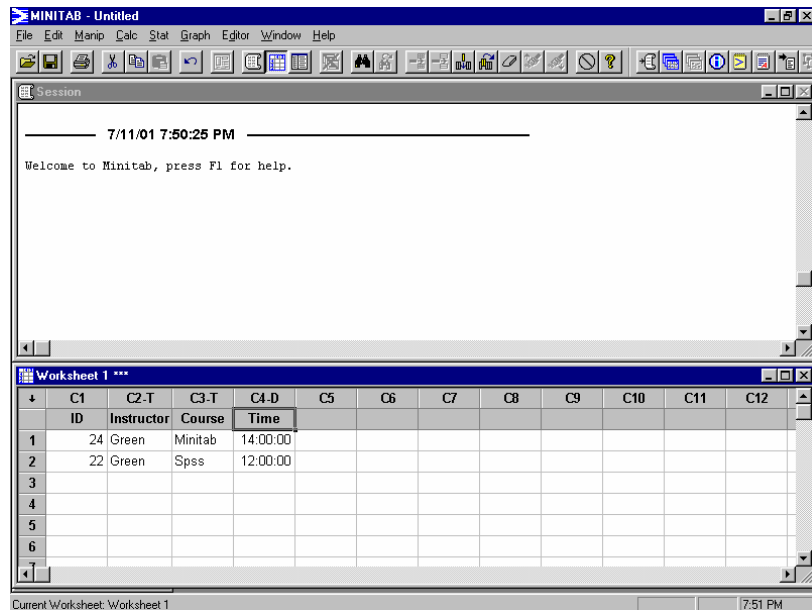
Entering Data Into Rows. Direction arrow pointing to the right. Make active the cell in which you wish to enter your first data value. Enter your first data value. Press *ENTER*, *TAB*, or use the “right arrow” key on your keyboard to move to the next cell. Continue your data entry in this manner. Press *CTRL ENTER* to move to the beginning of the next row.

Note the change in column headings (from C1, C2, C3, C4) to C1, C2-T, C3-D, C4-D. Because text was entered into C2, **Minitab** has now automatically identified this column as a text column. In column C3 dates were entered, therefore **Minitab** has identified this column as a date/time column. Likewise for column C4 where times were entered.



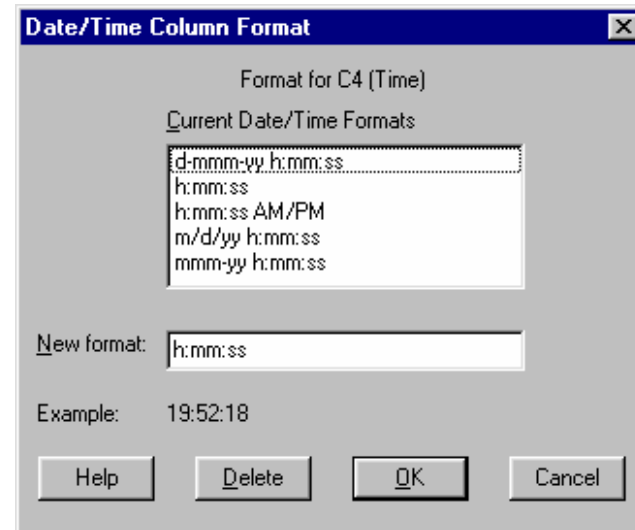
Naming Variables (columns). To name your variables make the **column name cell** the active cell by *clicking once* in it. The **column name cell** is located directly below the alpha-numeric column identifier (C1, C2, etc.). Type in the variable name of your choice. The following restrictions apply:

- name can be no longer than 31 characters.
- name cannot begin or end with a space.
- name cannot include the symbol # or ‘.
- name cannot start with the symbol *.



Entering Date/Time Data. Before entering **Date/Time Data**, you must first format the column into which it will be entered. First select the column in which the **Date/Time** will appear, by *clicking once* on the column heading. Then follow the **Path:** Editor => Format Column => Date/Time. A dialog box will appear as is shown to the right.

Several **Date/Time** formats will be available to you. An example of the format will appear as seen in the above dialog box. Select a format by *clicking once* on it. So, what if I do not like any of the given formats?? Good question! You can create your own format! Suppose you would like your **Date/Time** format to appear as Sep-22-1938. Simply *click once* in the **New Format** box, and then *type* *mmm-d/yyyy* (because you will be using three characters (Sep) for the month, and four characters for the year (1938). The day takes care of itself!). The dialog box will now appear as is shown below:



Click once on **OK**. You may now enter the **Date/Time Data** in this column, in your newly created format. Note that when you exit **Minitab**, your new format will automatically be saved in the list of existing formats, so that you may use it again if you wish.

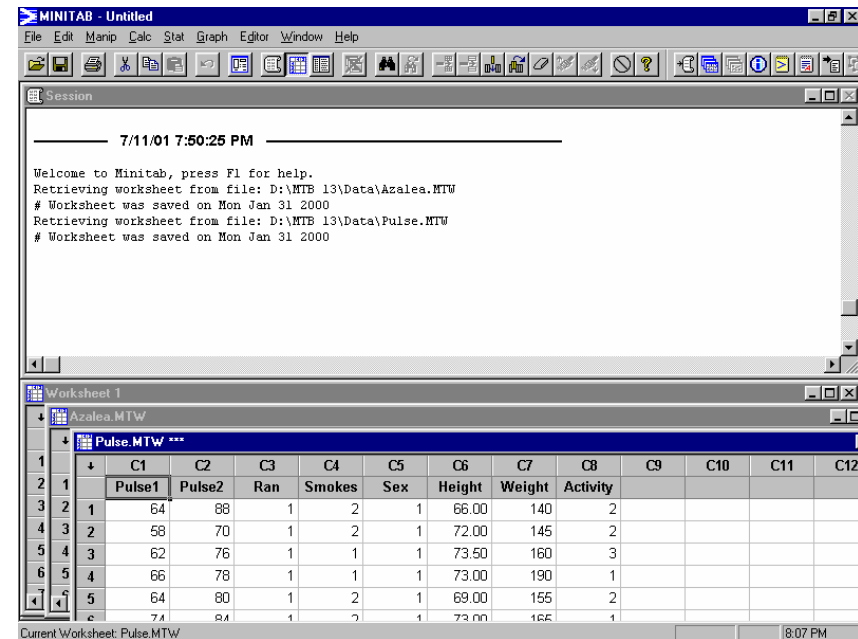
Retrieving an Existing Minitab Worksheet.

Worksheets are saved with the extension **.MTW**. Approximately 109 small, sample **worksheets** come as part of the Minitab software. To open a **Worksheet**, go to the directory in which it has been saved, be sure that the “Files of Type” box indicates “Minitab (*.MTW),” and select the desired file. Multiple worksheets may be opened within a project at any given time. **The active Worksheet (the one one can do operations on) will have three asterisks (***) following the Worksheet’s name.** In the following example there are three worksheets within the project, with the **PULSE.MTW** worksheet being the one currently in use.

Path: File => Open Worksheet



Active Worksheet: Pulse.mtw (***)



Entering Excel Files:

Entering an Excel file is done in the same manner as opening a Minitab worksheet. Select the “Open Worksheet” option off the File menu and identify the file type as an excel file. The information contained in the first row of the Excel file will become the column names in the worksheet.

Entering Text Data:

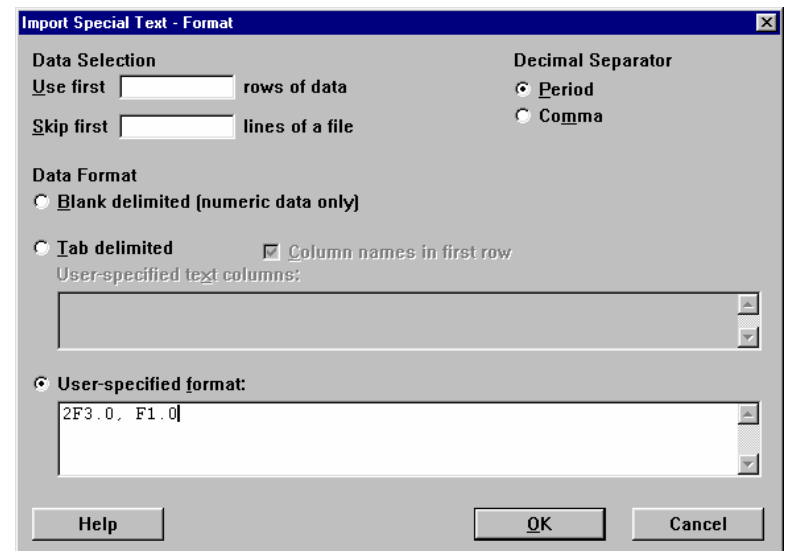
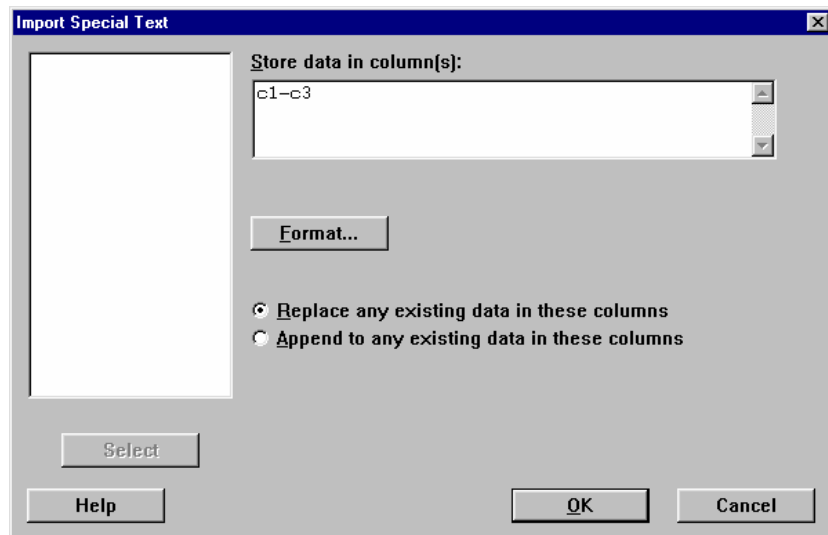
Quite often data are available in a raw data format, such as a text-based file, for which a format must be defined before the data may be entered into Minitab. To accomplish this task, one must be familiar with the structure of the raw data file. That is, one must know where the data for each variable lies within the raw data. An Example of this process is noted below.

PATH: File -> Other Files -> Import Special Text -> Identify the columns into which the data are to be placed and , using the Format button, identify the structure of the raw data file.

EXAMPLE: Assume that a data file contains the three variables ID NUMBER, AGE, and GENDER. The ID variable takes up three columns (so there is somewhere between 0 and 999 cases). AGE also uses three columns (there is the potential for respondents beyond 99 years old). Gender uses one column. The first three cases in the raw data file appear as follows:

```
0010321      (ID = 1; Age = 32; Gender = 1 (Male))
0020481      (ID = 2; Age = 48; Gender = 1 (Male))
003008       (ID = 3; Age = 8; Gender = blank (unknown))
```

- 1) Identify the columns into which the data are to be entered.
- 2) As these data have a fixed size for each variable, in the Format Dialog Box select "User Specified Format". In this case the first two variables, ID and AGE, both use three columns to hold the data. Therefore, the first entry is 2 (for the two variables) followed by an F (symbol indicating Numeric data; an A indicates alpha or text-based data). The 2F is followed by the size of the field, which in this case is 3 columns wide, with no decimals, or 3.0. (If a decimal is involved, then 3.1 would indicate that the variable used three columns and the third column is a decimal value.) A comma is entered, followed by the size and type of the next variable, here F1.0, representing GENDER.
- 3) Once the entire format has been identified, follow the OK buttons and the data will be entered into the worksheet.
- 4) Refer to the Help screens for additional considerations regarding the entry of raw data files.



DISPLAY DATA

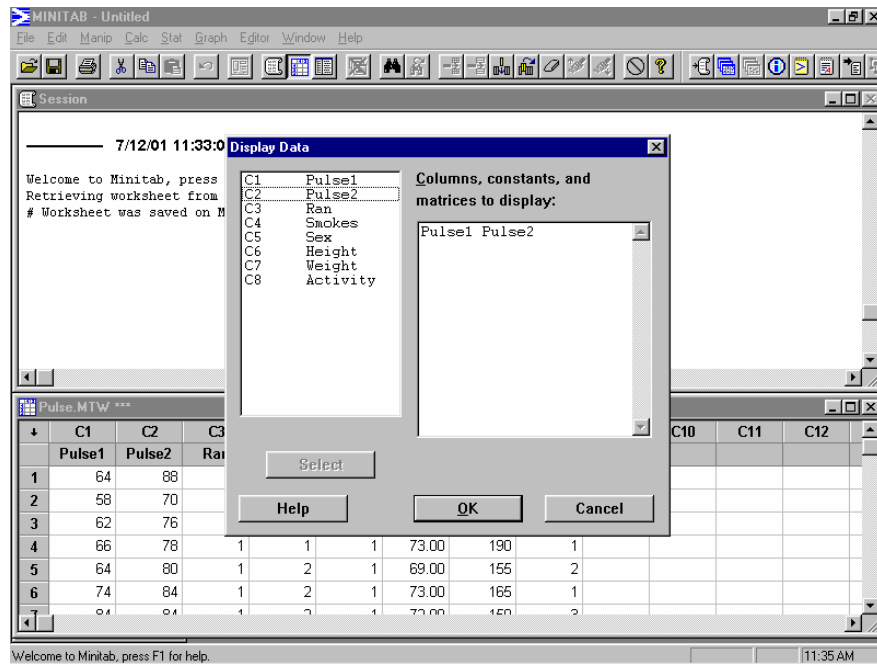
On occasion it may be useful to obtain a list of values for a variable(s). The **Display Data** command will place a listing of a variable's values into the Session Window from which a printout may be obtained.

DATA: Pulse.MTW

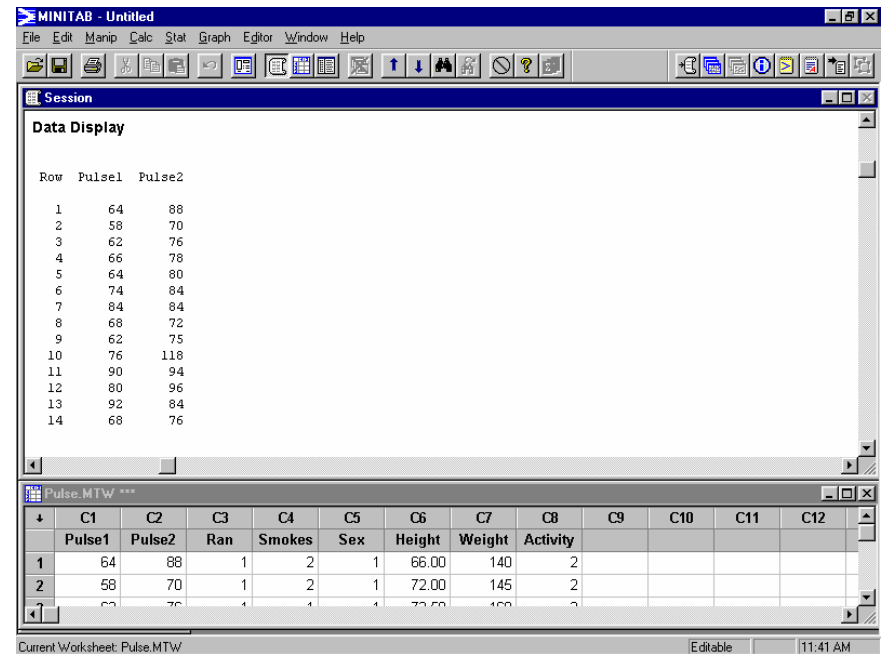
PATH: Manip => Display Data => insert the variable(s) to be displayed into the right box ("Columns, constants, etc.") => OK

EXAMPLE: Display the data associated with the variables 'PULSE1' and 'PULSE2.'

Dialog Box Input: Move 'PULSE1' and "PULSE2" into the "Columns, constants, and matrices to display" box..



Minitab Output: A summary listing of all values of the selected variables displayed in the **Session Window**.



EDITING A WORKSHEET

Once data are in a **worksheet (data window)**, it can be edited. The data can be rearranged, rows and columns can be inserted or deleted, cut or copied, and pasted, and values within the individual cells can be changed. (This is not an exhaustive list).

PATH: from the Main Menu Bar select **EDIT (or MANIP or EDITOR, as noted below)**.

In all cases, the data you wish to edit, must be *selected* first.

- An individual cell is *selected* by *clicking once* in the cell.
- A column is *selected* by clicking once in the alpha-numeric (C1, C2, etc.) cell.
- A row is *selected* by clicking once in the cell containing the row number.

A group of rows or columns is *selected by clicking and dragging*. For example, suppose you wish to *select* columns C1, and C2 from the PULSE worksheet. *Click once and hold down* in the alpha-numeric column heading C1. *Drag* the mouse over to C2. Both columns will now be highlighted (*selected*) as shown below (left). Now suppose you wish to *select* rows 2 and 3 in the PULSE worksheet. *Click once and hold down* on row 2, then *drag* down to row 3. Both rows will now be highlighted (*selected*) as shown below (right)..

The screenshot shows the MINITAB interface with the PULSE.MTW worksheet open. The columns C1 and C2 are highlighted in black, indicating they have been selected. The data in the worksheet is as follows:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity				
1	64	88	1	2	1	66.00	140	2				
2	58	70	1	2	1	72.00	145	2				
3	62	76	1	1	1	73.50	160	3				
4	66	78	1	1	1	73.00	190	1				
5	64	80	1	2	1	69.00	155	2				
6	74	84	1	2	1	73.00	165	1				

The screenshot shows the MINITAB interface with the PULSE.MTW worksheet open. The rows 2 and 3 are highlighted in black, indicating they have been selected. The data in the worksheet is as follows:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity				
1	64	88	1	2	1	66.00	140	2				
2	58	70	1	2	1	72.00	145	2				
3	62	76	1	1	1	73.50	160	3				
4	66	78	1	1	1	73.00	190	1				
5	64	80	1	2	1	69.00	155	2				
6	74	84	1	2	1	73.00	165	1				

Editing Commands:

Undo: Undoes the *last* action taken in the worksheet. Note that **undo** only “remembers” the *last* action taken.

Path: Edit => Undo

CLEAR vs. DELETE: *It is extremely important to understand the difference between the CLEAR command and the DELETE command. CLEAR inserts a “place holder” whereas DELETE removes data and allows the data below the deletion point to move up. For a data file based upon cases (where related responses from one source are contained in one row), the use of the DELETE command results in the data from one case shifting into another case, thereby rendering many analytical techniques useless.*

Clear: Removes the cell contents of the selected data cells. The empty cells will remain with the cleared data values being replaced by the Minitab missing value symbol (*) for numeric data or a blank cell for text data.

Path: Edit => Clear Cells (Pressing the *backspace* key after selecting the data will perform the same function)

Delete: Deletes the selected data from the worksheet. **All remaining column data will move up (row data will move left).**

Path: Edit => Delete Cells (Pressing the *delete* key after selecting the data will perform the same function.)

But what if the rows or columns that I wish to delete are not next to each other (are non-contiguous)??

Deleting Non-Contiguous Columns.

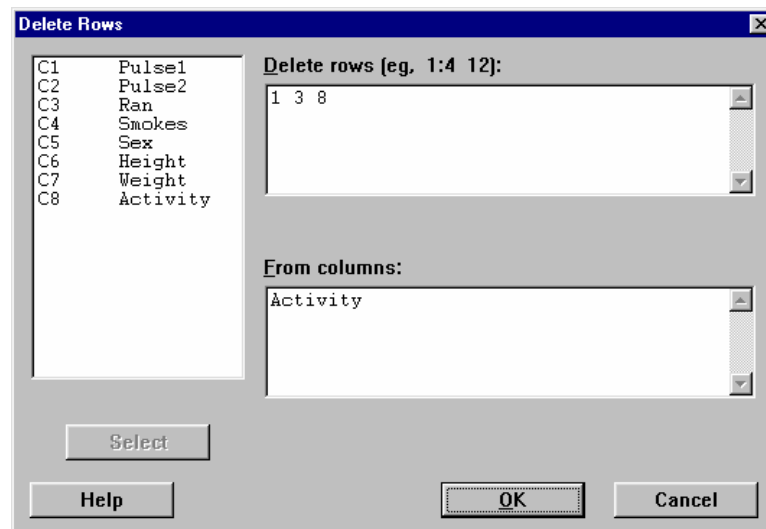
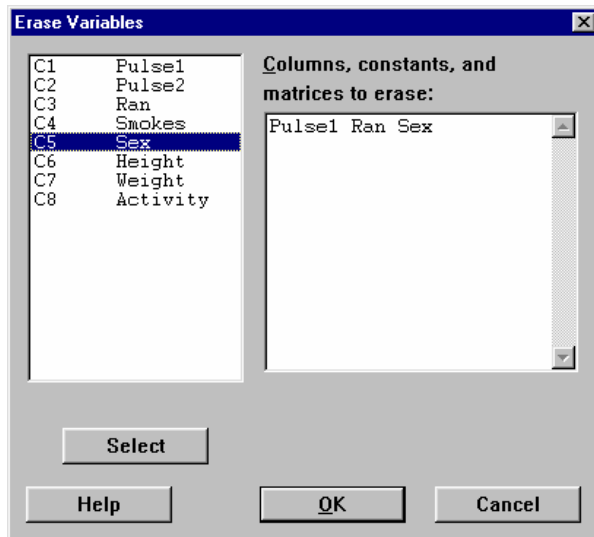
Path: Manip => Erase Variables (Be warned, that when using this command, you cannot **undo** the action.)

A dialog box as shown below left, will appear into which you can enter any combination of columns to be erased. To enter the columns to be erased, select the column(s) in the dialog box on the left by *clicking once* on the alpha-numeric column name. This will highlight the column name. Now press *select*. This will move the column name to the box on the right. When all choices have been made, *click* OK. The contents of columns C1, C3, and C5 will be erased. The empty columns will remain.

Deleting Non-Contiguous Rows.

Path: Manip => Delete Rows (When using this command, you cannot **undo** the action.)

A dialog box will as shown below right, will appear into which you can enter the rows to be erased. Enter the rows to be deleted in the “delete rows” box. In the example shown, rows 1, 3, and 8 will be deleted. You must now designate the columns from which these rows will be deleted. When you *click once* in the “from columns” box, a list of variables (columns) will appear in the box to the left. Select the columns from which the rows will be deleted by *clicking once* on the variable. Now *click* select. When you have completed this process, *click* OK. The chosen rows will be deleted and the remaining rows will move up.



Cut, and Copy:

Path: Edit => Copy Cells OR Edit => Cut Cells

It is important to note the difference between the **cut** and **copy** commands. **Cut** will actually *remove* the selected data from the worksheet and place it on the Clipboard. All remaining rows or columns will move up or left. **Copy** will simply place a copy of the selected data on the Clipboard, leaving the selected data in the worksheet that it was copied from. Copying non-contiguous rows and columns is done in a manner similar to deleting non-contiguous rows and columns.

Paste:

Path: Edit => Paste Cells

Once data is **cut** or **copied** to the Clipboard, it can be pasted within the same or another worksheet. In the worksheet *select* the area where the data will be pasted. This is done by designating an anchor cell. The first cell of the Clipboard contents will be pasted here, followed by the remainder.

Inserting Cells, Rows, and Columns:

Path: Editor => Insert Cells/Insert Rows/Insert Columns

Note here that you are now using the **Editor Menu**, as opposed to the **Edit Menu** used previously. Before inserting cells, rows, or columns, an area of the worksheet must be *selected*** **Cells** and **Rows** will be inserted **within** the area you select. **Columns** will be inserted to the **left** of your selection.**

****Area Selection:** Suppose you want to add two cells (one in row 2 and one in row 3) to the column PULSE1 from the PULSE worksheet. *Select* cell 2 and cell 3 in the column PULSE1. Next use the above **Path** to choose **Insert Cells**. Two empty cells will appear (one in row 2 and one in row 3) in the column PULSE1, and all other cells will move down. The values that previously appeared in cells 2 and 3 in this column, are now in cells 4 and 5.

Now suppose you want to insert 3 rows in positions 7, 8, and 9. *Select* rows 7, 8, and 9 (or simply cells 7, 8, and 9 will do). Next, use the above **Path** to choose **Insert Rows**. Three empty rows will appear in positions 7, 8, and 9, and all other rows will move down. The rows that were previously rows 7, 8, and 9, are now rows 10, 11, and 12.

Finally, suppose you want to insert a column in between PULSE1 and PULSE2. *Select* the column PULSE2 (remember, the new column will be inserted to the left of your selection). Next, use the above **Path** to choose **Insert Columns**. A new column will be inserted between PULSE1 and PULSE2, and all other columns will move to the right. Note that had you wanted to add two columns in between PULSE1 and PULSE2, you would have *selected* PULSE2 **and** RAN. To add three columns you would have *selected* PULSE2, RAN, and SMOKES, etc.

Moving Columns:

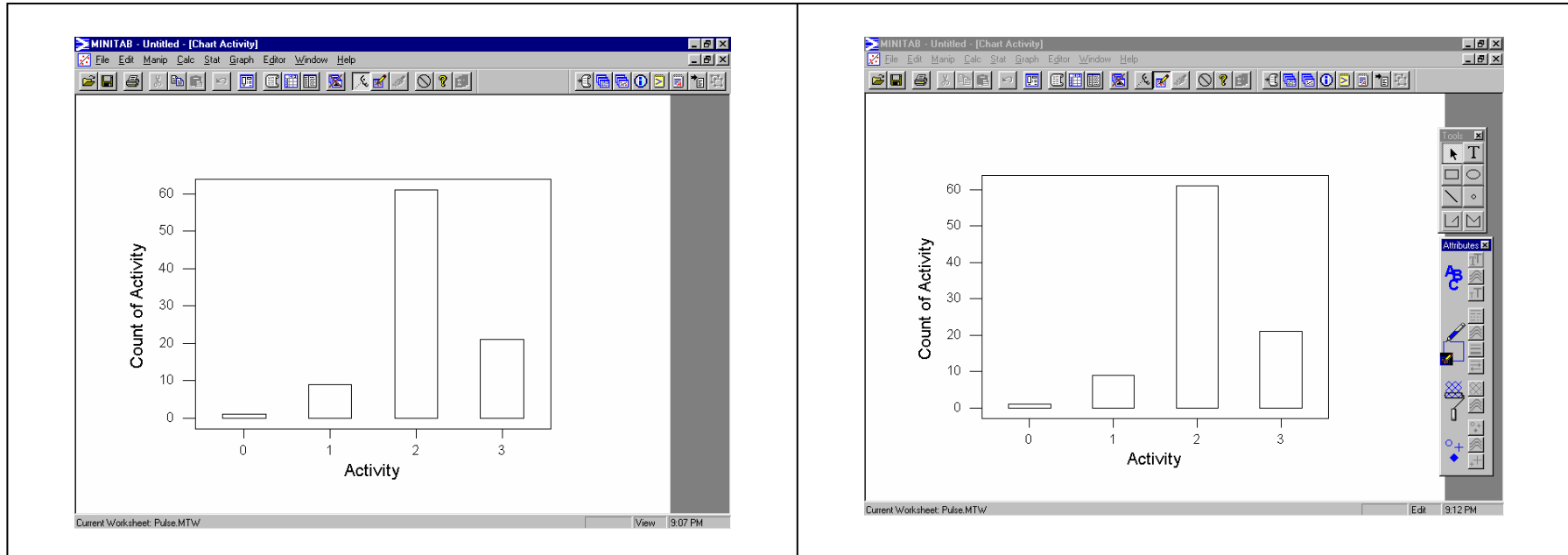
Path: Editor => Move Columns

Suppose you now want to move a column or columns. Begin by *selecting* the column or columns you want to move. Now use the above **Path**. The following dialog box will appear. You will use this dialog box to designate where you would like the *selected* column(s) moved to. Choosing **Before column C1**, will move the *selected* column(s) into the position before C1, and all columns will move to the right. Choosing **After last column in use**, will move the *selected* column(s) to the first position **after** the last column being occupied in the worksheet (in this case the *selected* column(s) would be moved to the first empty column **after** the column ACTIVITY).

Finally, you may designate a position between any of the existing columns, by choosing **Before Column** and then selecting the column(s) of your choice. The column(s) you want to move will be placed **before** the column you designated in the dialog box, and all columns will move to the right. Once you've made your choice, *click* **OK**.

EDITING CHARTS AND GRAPHS: TOOL AND ATTRIBUTE PALETTES

Once you've made your chart or graph, there may be times when you want or need to make some changes to it. Minitab's **Tool and Attribute Palettes** are used to do this. The options and editing features discussed here apply to each of the charts and graphs discussed in this manual. Not all options from each of the palettes will be discussed. In each example, a basic **Bar Chart** will be used.



To begin editing, the graph you want to work on, must be the active window, and must be in the **Edit Mode**. To enter the **Edit Mode**, do one of the following: *Double-click* in the graph window, or use the **Path**: Editor => Edit.

Sometimes, the **Tool and Attribute Palettes** which help you do your editing, do not automatically display when you enter the **Edit Mode**. To make them appear:

Path: Editor => Show Tool Palette and/or Editor => Show Attribute Palette

All editing begins with a *selection*. The **Selection Tool** (diagonally pointing arrow in the upper left box) in the **Tool Palette** is used for this purpose. This arrow must be *selected* before selecting areas or text for editing, within your chart.

Beginning with the **Tool Palette**, you will “draw” an ellipse, and add some text within this ellipse. To “draw” the ellipse, *click once* on the picture of the ellipse (second column, second row of the **Tool Palette**). Now locate the place on your chart where you want to place the ellipse. Holding down the left mouse key, begin to drag the mouse. You will see the ellipse appear as you are dragging the mouse. If you've not done this before in, it will take some practice to get it just the way you want it, so be patient. Once your ellipse looks the way you want it to look, release the left mouse key. The ellipse surrounded by four **selection dots** will now be added to your chart. To remove the **selection dots**, *click once* anywhere outside of the ellipse.

Now, you will add some text to the inside of the ellipse. *Click once* on the “T” (upper right box) in the **Tool Palette**. Now *click once* inside the ellipse. Type in the text that you want to appear inside of the ellipse, inside the **Text Box**. Now *click OK*. The text will now appear inside the ellipse on your chart.

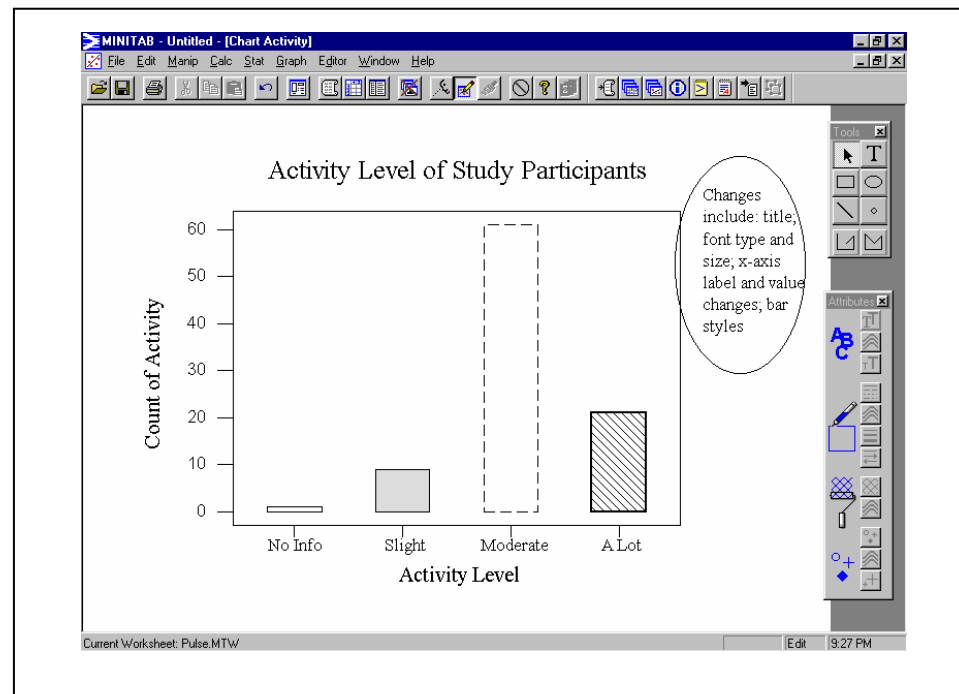
Now that you know how to choose a tool and select a place within your chart or graph to use that tool, do not be afraid to experiment with the remaining tools not discussed here. Anything you do, can always be undone by simply *selecting* what you wish to remove, and then pressing the delete key.

Next you will be working with the **Attribute Palette**. You will use the **Attribute Palette** to change font and text size, the line type surrounding the bars, the foreground fill pattern, the color of the foreground fill pattern, and the background color of selected objects.

To change font styles, *click once* on the text for which you wish to change the font style. **Selection Dots** will now appear around the selected text. Now *click once* in the upper right box (the one containing two different styles of “T’s”) in the **Attribute Palette**. A *drop-down menu* with font choices will appear. *Select* a font. Your *selected* text will now appear in the font style chosen. To change font size, *select* the text you wish to change, and then *click once* in the third small box on the right, in the **Attribute Palette**. A *drop-down menu* with font size choices will appear. *Select* a size. The text will now be resized to the font size you chose.

Now you will change the line type surrounding the bars. *Select* a bar by *clicking once* on the outline of it. The bar will be surrounded by **Selection Dots**. Now *select* the fourth box on the right, in the **Attribute Palette**. From the *drop-down menu* select a line style. The new line style will now appear around the *selected* bar. Next, you will choose a background color for a bar. Once again *select* a bar. Now *select* the ninth box on the right, in the **Attribute Palette**. Choose a color. This color will now appear as the background color of the *selected* bar.

Next, you will select a fill pattern, and fill pattern color for a bar. Once again, first *select* a bar. Now *select* the eighth box on the right in the **Attribute Palette**. *Select* a fill pattern. Now *select* the fifth box on the right in the **Attribute Palette**. *Select* a fill pattern color. Depending on the pattern and colors chosen, your chart should now appear somewhat changed. To close the **Tool** and **Attribute Palettes** *click once* in the small rectangle in the upper left corner of each box.



EXPORTING TABLES AND GRAPHS

Minitab data analysis in the form of tables, graphs, text, etc. may be copied from the Session Window, Graph Windows, etc. to a document created in Microsoft Word or another program.

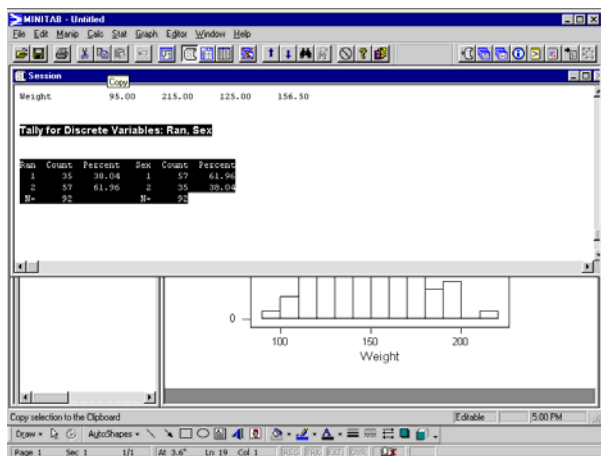
DATA: Pulse.mtw

PATH: From a Minitab window copy the material to be transferred to Word, go to Word and paste the material in the desired location.

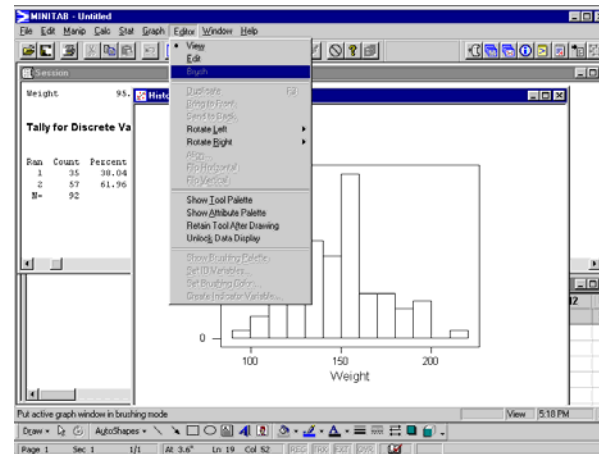
NOTE REGARDING GRAPHS: In most cases transferring information requires that the desired materials be highlighted and copied. However, for graphs the process is slightly different. First select the graph to be copied and then go to the Main Menu Bar 'Edit' menu and select 'Copy Graph.' If the Copy Graph option is not available you, you must go to the "Editor" menu and select the 'View" option at the top of the list, Right below 'View' is the 'Edit' option. When 'View' is selected, one can copy a graph. When 'Edit' is selected one can edit the characteristics of the graph, but not copy it.

EXAMPLES:

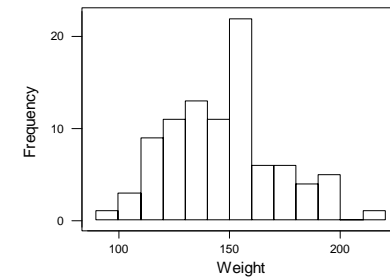
Session Window: Highlight items to copy. Go to Word and select Paste.



Graph Window: See Note above. Editor option to 'View' Edit Menu option 'Copy Graph.'



Graphs to Word: Select Edit, Paste Special, Picture (Metafile) or Bitmap.



PATTERNED DATA

Any column in the **Minitab** work sheet can contain data (numbers, text, or date/time) that follow a pattern. The pattern (numbers, or date/time) need not be an equally spaced pattern. Some examples of **Patterned Data** are:

- The numbers 1 through 10.
- The months January through December.
- Strongly Agree, Agree, No Opinion, Disagree, Strongly Disagree.
- 9:00 AM, 10:00 AM, 1:00 PM, 5:00 PM.

PATH: Calc => Make Patterned Data => Simple Set of Numbers

EXAMPLE 1: Filling a Column with an Equally Spaced Pattern of Numbers. Suppose you wish to enter the numbers 1 through 5, in steps of 1 (that is, 1, 2, 3, 4, 5) into column C2. Enter the column into which you want the **patterned data** to appear (here C2). Choose the first value that you want to appear (here 1). Choose the last value that you want to appear (here 5). Now enter the increments (steps) that you want between each number. Finally, enter the number of times you want each value to appear and the number of times you want the entire sequence to appear. *Click* OK. The numbers 1 through 5 will appear in column C2 as shown right.

Simple Set of Numbers

Store patterned data in: C2

From first value: 1

To last value: 5

In steps of: 1

List each value 1 times

List the whole sequence 1 times

Select

Help OK Cancel

MINITAB - Untitled

File Edit Manip Calc Stat Graph Editor Window Help

Session

07/11/2001 10:58:51 AM

Welcome to Minitab, press F1 for help.

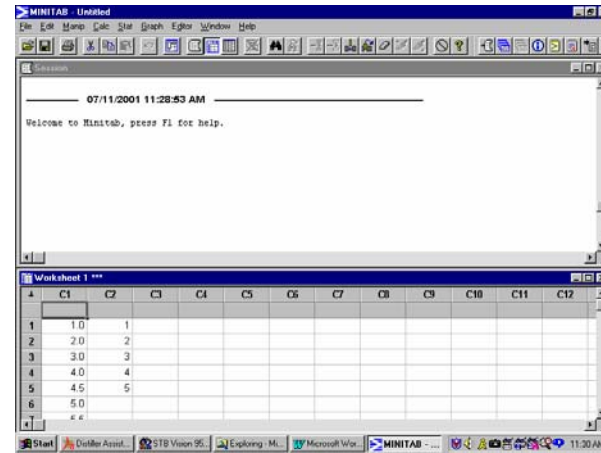
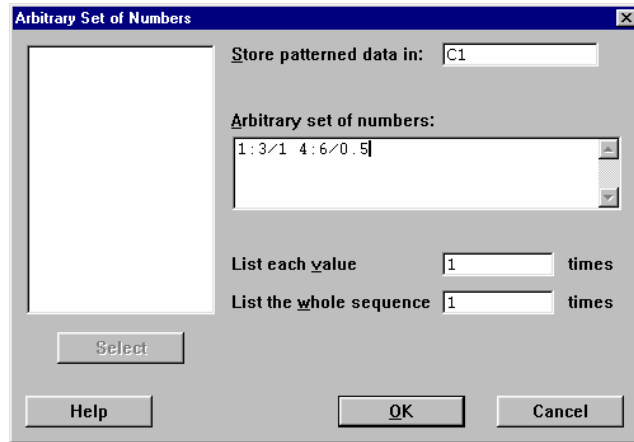
Worksheet 1 ***

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1		1										
2		2										
3		3										
4		4										
5		5										
6												

Current Worksheet: Worksheet 1 11:09 AM

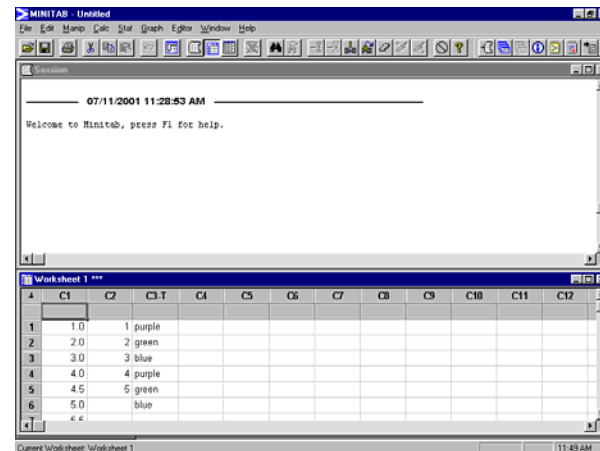
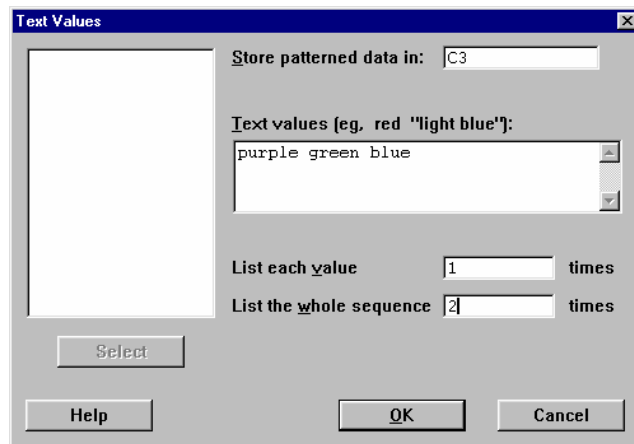
PATH: Calc => Make Patterned Data => Arbitrary Set of Numbers

EXAMPLE 2: Filling a Column with an Unequally Spaced Pattern of Numbers. Suppose you wish to enter the numbers 1, 2, 3, in steps of 1, and the numbers 4 through 6 in steps of 0.5 into column C1. The only difference in this dialog box and the one from the previous example is in the designation of numbers. In the box labeled **arbitrary set of numbers**, you would type **1:3/1** and **4:6/0.5**. This “tells” Minitab that you wish to have the numbers 1 through (:) 3 in steps of 1 (/1), and the numbers 4 through 6 in steps of 0.5 entered into column C1. *Click OK.* The designated **patterned data** is entered into Column C1 as shown next.



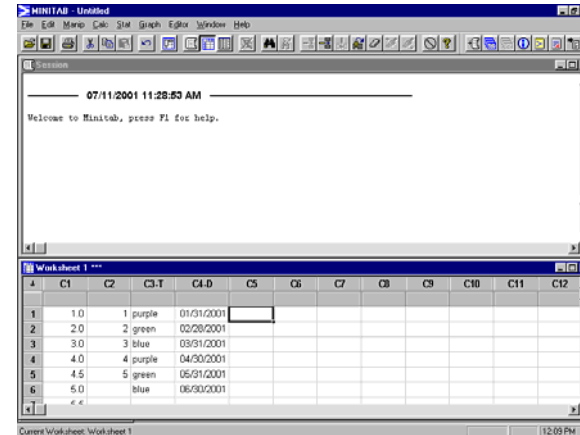
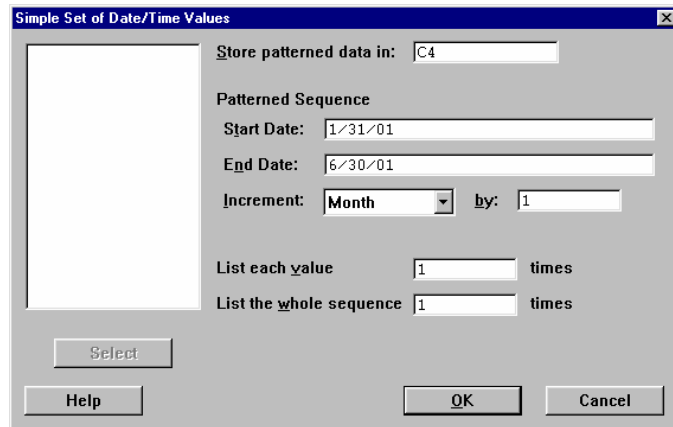
PATH: Calc => Make Patterned Data => Text Values

EXAMPLE 3: Filling a Column With Patterned Text. Now suppose you wish to have the colors purple, green, and blue, appear in two sequences, one time each, in column C3. Choose the column (C3), the text values (purple, green, and blue). Each value is to be listed once, and the sequence is to be listed twice. *Click OK.* Column C3 now contains the **patterned data**.



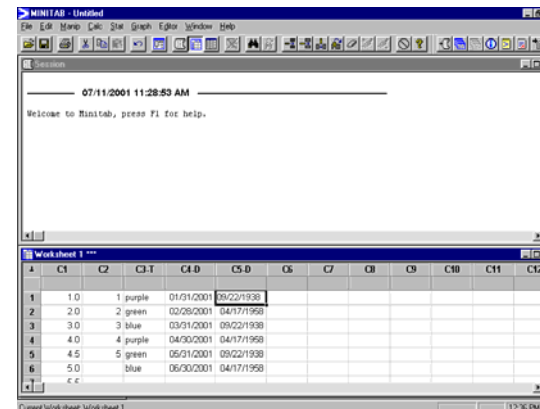
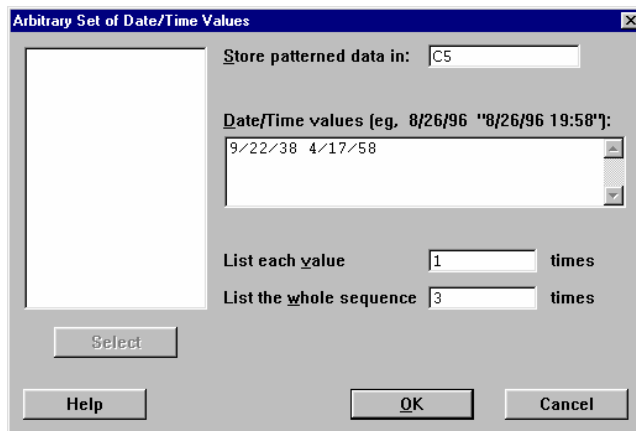
PATH: Calc => Make Patterned Data => Simple Set of Date/Time Values

EXAMPLE 4: Filling a Column with an Equally Spaced Pattern of Dates/Times. Suppose you wish to enter one sequence of the dates 1/31/00 through 6/30/01, each date appearing once, into column C4. Enter the column into which you want the **patterned data** to appear (C4). Enter the start date and end date. Select the increment by using the pull-down menu in the **Increment** box. Enter the number of times each value (date) is to be listed (1) and the number of times the sequence is to be repeated (1). *Click OK*. Column C4 will now contain the patterned data as shown next.



PATH: Calc => Make Patterned Data => Arbitrary Set of Date/Time Values

EXAMPLE 5: Filling a Column with an Unequally Spaced Pattern of Dates/Times. Suppose you wish to enter three sequences of the dates 9/22/38 and 4/17/58, one time each into column C5. As with each previous example, a column is designated, as well as the values, how many times each value is to be listed, and how many times the sequence is to be repeated. The worksheet now appears as shown below.

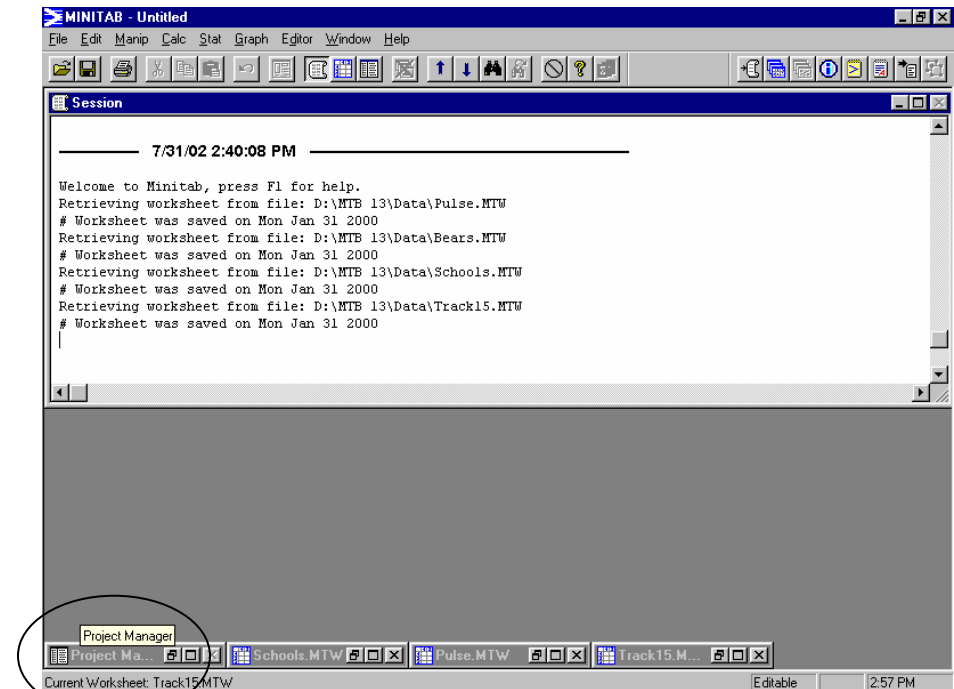
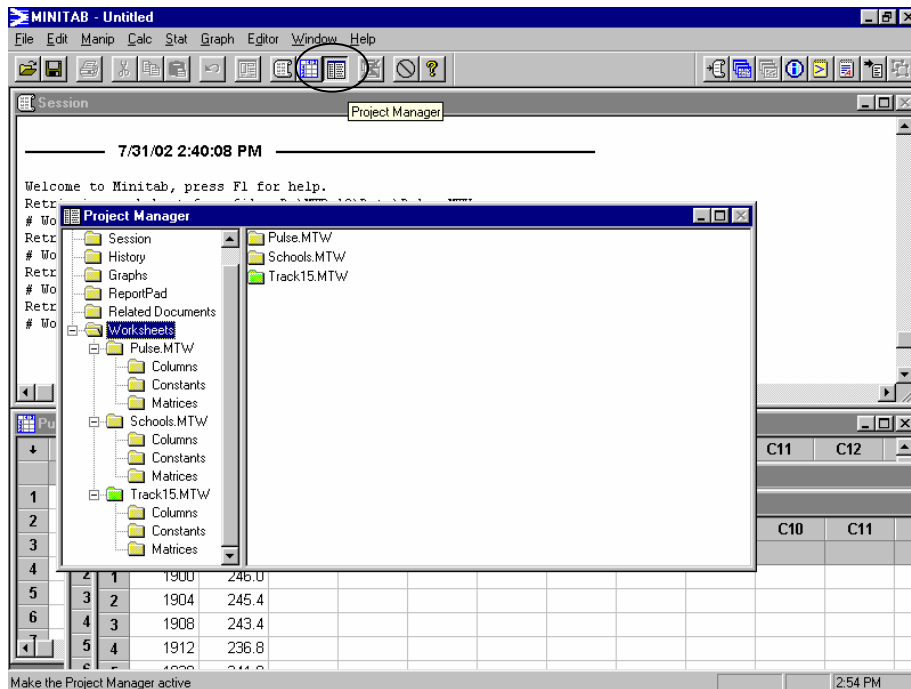


PROJECT MANAGER

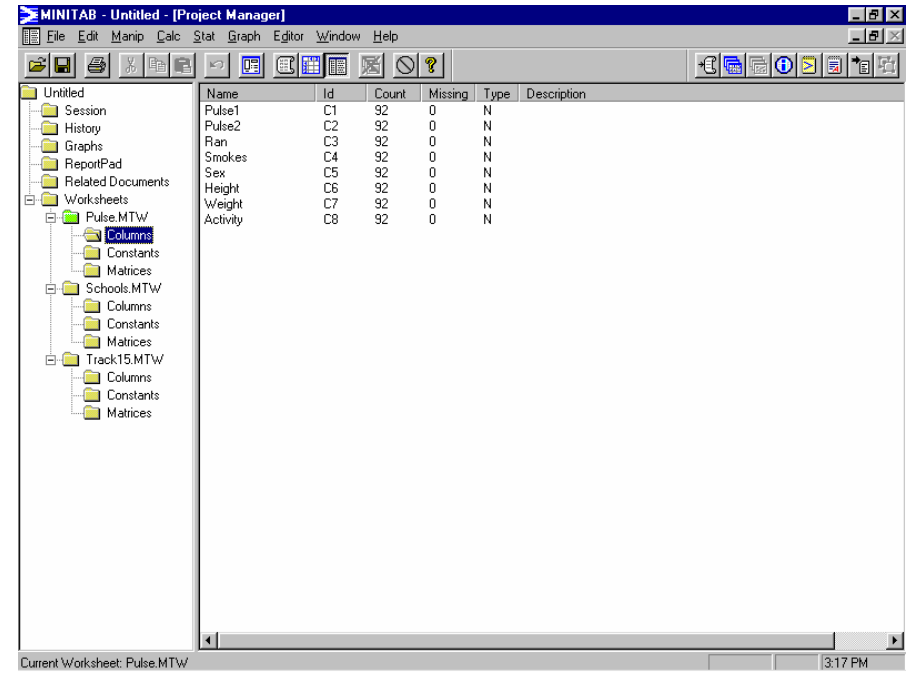
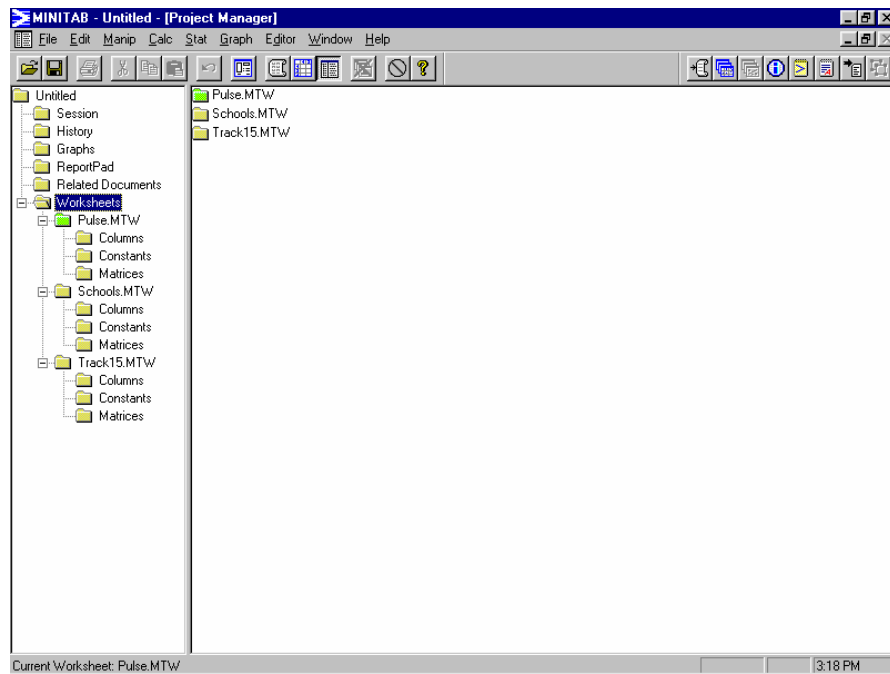
The **Project Manager** resides in the background of every Minitab session. Its primary function is to help one maneuver among the various Minitab windows.

DATA: Pulse.mtw; Schools.mtw; and Track15.mtw

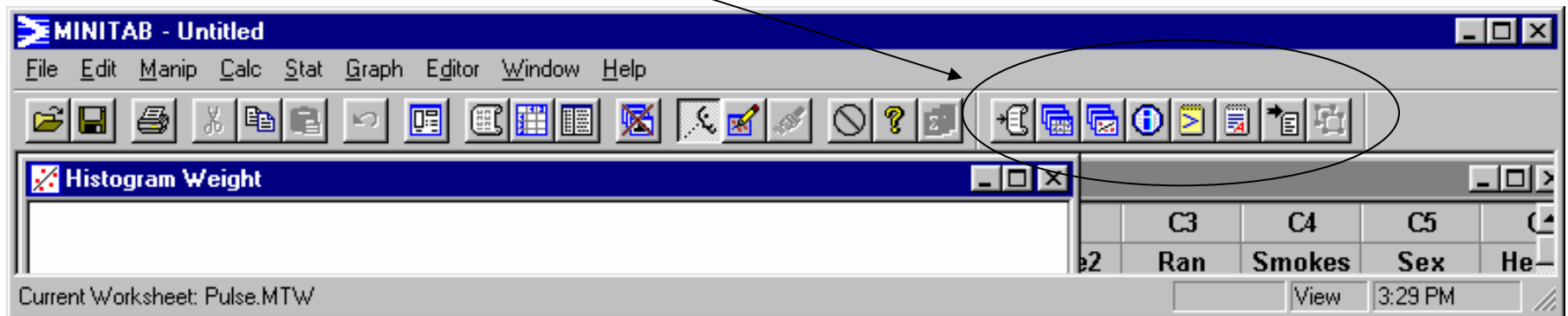
PATH: Select the **Project Manager** icon off of the Standard Toolbar or, if Minitab components (windows) have been downsized, it will be visible within the project window.



The **Project Manager** may be used to recall the history of the project (allows one to review prior analysis); look at the contents of the ReportPad; move between worksheets, edit the name of individual worksheets, examine the characteristics of the worksheet (e.g. the variables in the Pulse.mtw worksheet), as well as access a number of other windows.



Associated with the **Project Manager** is the **Project manager Toolbar**, which can be toggled on/off from the Main Menu 'Windows' dropdown menu.



REPORTPAD

The **ReportPad** serves as a simple word processing window within Minitab. Results of one's data analysis may be copied to this window and prepared for printing as a document or the information may be collected here and transferred to another word processing program.

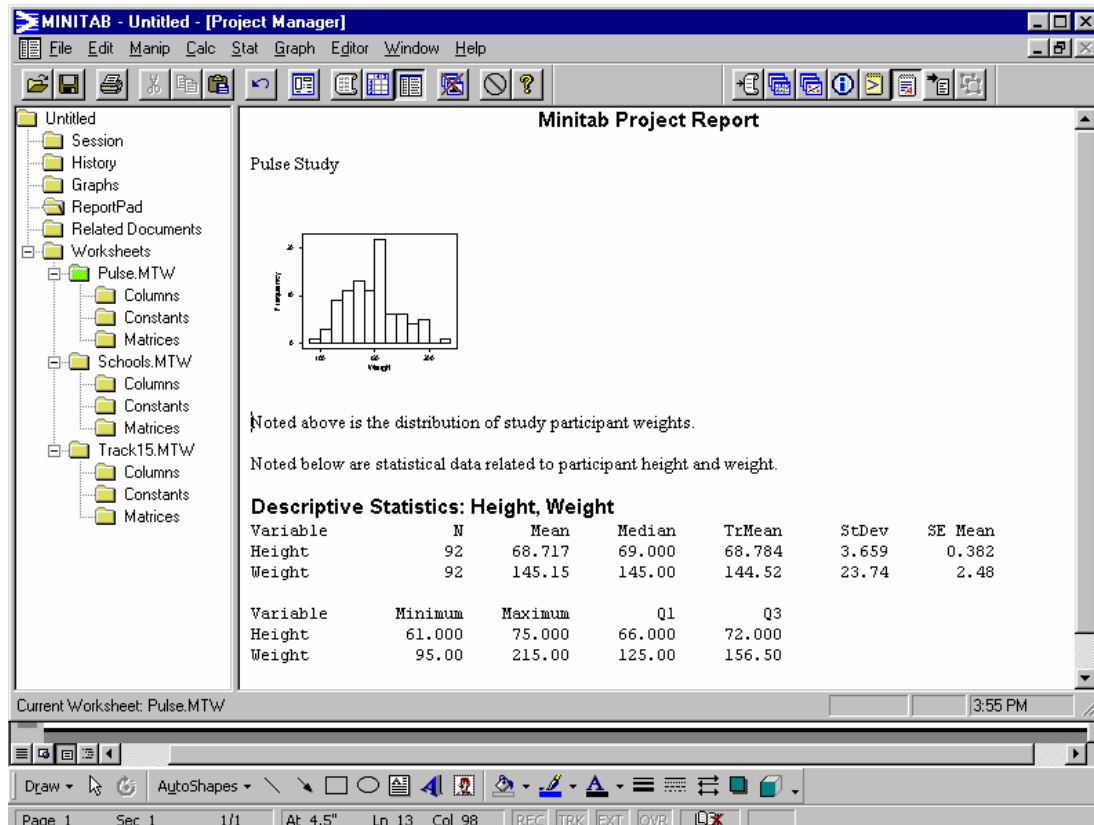
DATA: Pulse.mtw

PATH: Access the ReportPad from the Project Manager or via the Project Manager Toolbar icon, a notepad with a red corner.



ReportPad Icon

EXAMPLE: Shown below is a portion of a ReportPad containing a graph and some statistical data.



SPLIT WORKSHEET

On occasion it may be necessary to use the *Split Worksheet* command so that analysis on a sub-set of the data may be conducted. For example, one might wish to examine the relationship between height and weight for only female study participants. To do this one would first have to split the data set based upon the ‘SEX’ variable. In the case of a dichotomous variable such as sex, the result of splitting the worksheet will be three worksheets – the original, a worksheet containing only male subjects, and a worksheet containing only female participants.

DATA: Pulse.MTW

PATH: Select the Minitab Worksheet to be split => Manip => Split Worksheet => insert the variable by which the file is to be divided into the “By variables:” cell => OK

EXAMPLE: Split the worksheet by the variable ‘SEX’.

Dialog Box Input: Move 'SEX' into the "By variables"

The dialog box 'Split Worksheet' is open over a worksheet. The 'By variables:' list contains 'Sex'. The 'Include missing as a BY level' checkbox is unchecked. The background worksheet shows columns C1 (Pulse1), C2 (Pulse2), C3 (Ran), C4 (Smokes), C5 (Sex), C6 (Height), C7 (Weight), and C8 (Activity). Rows 1-6 are visible.

Result: Three Worksheets – the original, one for male, and one for female.

The Session window shows the following results:
 Pearson correlation of Pulse1 and Pulse2 = 0.616
 Correlations: Height, Weight
 Pearson correlation of Height and Weight = 0.785
 Plot Weight * Height

The worksheet is split into three sheets:
 - Pulse.MTW(Sex = 1)
 - Pulse.MTW(Sex = 2) (Current Worksheet)
 - Pulse.MTW(Sex = 3)

The active sheet (Sex = 2) shows the following data:

Row	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity				
1	96	140	1	2	2	61.00	140	2				
2	62	100	1	2	2	66.00	120	2				
3	78	104	1	1	2	68.00	130	2				
4	82	100	1	2	2	68.00	138	2				
5	100	115	1	1	2	63.00	121	2				
6	68	112	1	2	2	70.00	135	2				

APPENDIX A: Data File Descriptions

PATH to Data Files: File => Open Worksheet => set the "Look in:" cell to Data => select the desired data file => Open

PATH to Data File Descriptions: Help => Search Help => Index => enter data file name => select file => Display

MARKET.MTW

This data set consists of a company's quarterly data on sales revenues, advertising expenses, capital expenses, and the ad agency used from 1991 to 1992.

Column	Name	Count	Description
C1	Index	8	Index
C2	Quarter	8	Quarter
C3	Year	8	Year
C4	Sales	8	Total sales
C5	Advertis	8	Advertising expenses
C6	Capital	8	Capital expenses
C7	AdAgency	8	Advertising agency

PULSE.MTW

Students in an introductory statistics course participated in a simple experiment. Each student recorded his or her height, weight, gender, smoking preference, usual activity level, and resting pulse. Then they all flipped coins, and those whose coins came up heads ran in place for one minute. Then the entire class recorded their pulses once more.

Column	Name	Count	Description
C1	Pulse1	92	First pulse rate
C2	Pulse2	92	Second pulse rate
C3	Ran	92	Running: 1 = ran in place; 2 = did not run in place
C4	Smokes	92	Smoking: 1 = smokes regularly; 2 = does not smoke regularly
C5	Sex	92	Sex: 1 = male; 2 = female
C6	Height	92	Height in inches
C7	Weight	92	Weight in pounds
C8	Activity	92	Usual level of physical activity: 1 = slight; 2 = moderate; 3 = a lot

SCHOOLS.MTW

In December 1975, many public school teachers in Madison, Wisconsin called in sick as a protest against administrators.

Column	Name	Count	Description
C1	SchlType	48	School level: 1 = elementary; 2 = middle; 3 = high
C2	Teachers	48	Number of teachers in the school
C3	Sick	48	Number who called in sick

TRACK.MTW

Winning times in seconds for the men's 1500-meter run in the Olympics from 1900 to 1992.

Column	Name	Count	Missing	Description
C1	Year	21	0	Year
C2	Time	21	0	Time in seconds

APPENDIX B: CONDUCTING A HYPOTHESIS TEST

The steps used to conduct a hypothesis test are noted below, along with an example, which uses the t-test.

Steps in the Test of a Hypothesis:

- A) **State the Given (known) information:** μ , σ , α , etc.
- B) **State the Null Hypothesis and its Alternative:**
 - 1) Identify whether the test is right tailed ($<$), left-tailed ($>$), or not equal.
 - 2) State the hypotheses in the form: $H_0: \mu = \mu_x$ $H_a: \mu = \mu_x$ (where μ_x = population mean)
- C) **Determine the Critical Value:** In the case of the One Sample t: 1) Critical Value = a t table value based upon a given level of significance (α) AND the degrees of Freedom of the statistic (d.f.), or 2) Critical Value if using p-value approach is the stated level of significance.
- D) **Determine the Test and compute the Test Statistic:** Dependant upon what measure is being tested.
- E) **Decision:** Compare the Test Statistic with the Critical Value and decide whether to Retain the Null Hypothesis or Reject it in favor of the Alternative Hypothesis.
- F) **Conclusion:** State a conclusion in terms of the data. The conclusion should go beyond a simple retention or rejection of the null hypothesis to interpret what this decision means in terms of the data studied.

EXAMPLE: The mean pulse rate for adults is known to be 72 beats per minute. Determine if the initial pulse rate ('PULSE1') of study participants is significantly higher than that of the population. Test the hypothesis at the $\alpha = .02$ significance level.

Dialog Box Input: Move 'PULSE1' into the "Variables:" cell; enter 72.0 into the "Test Mean:" cell; check the test "Alternative:" and "Confidence Level:" via the "Options" button.

Test of the Hypothesi

Given: $\mu = 72.0$; $\alpha = .02$

Hypotheses: $H_0: \mu = 72.0$ $H_a: \mu > 72.0$

Critical Value: where $\alpha = .02$; d.f. = 91 (use nearest table value greater than d.f.; used 100); C.V. = +2.081

Test: 1-Sample t

Test Statistic: 0.76

Decision: Fail to Reject H_0

Conclusion: At the $\alpha = .02$ significance level there is not sufficient evidence to reject the null hypothesis that the sample had a mean pulse rate greater than that of the population (72.0 beats per minute) and accept the alternative hypothesis that the pulse rate was greater for the sample.

MINITAB OUTPUT

One-Sample T: Pulse1

Test of mu = 72 vs mu > 72

Variable	N	Mean	StDev	SE Mean
Pulse1	92	72.87	11.01	1.15

Variable	98.0% Lower Bound	T	P
Pulse1	70.48	0.76	0.225

t value to check against Critical Value

p-value

INDEX

A

APPENDIX A: Data File Descriptions, 86
APPENDIX B: CONDUCTING A HYPOTHESIS TEST, 87
AREA GRAPH, 34
Attribute Palette
 Editing Charts and Graphs, 77

B

Bar Chart, 11
BAR CHART, 11
BASIC STATISTICS (Descriptive Statistics), 12
Binomial Probabilities, 14
Binomial Random Variable
 Obtaining Probability Distribution, 16
BINOMIAL RANDOM VARIABLES (Calculating Probabilities), 14
BINOMIAL RANDOM VARIABLES (Obtaining a Probability Distribution), 16
Boxplot, 17
BOXPLOT (BOX-AND-WHISKER PLOT), 17

C

CALCULATE, 62
Calculator, 62
Chi-Square, 19
 Chi-Square Goodness of Fit Test, 19
 Chi-Square Independence Test, 19
CHI-SQUARE (): Test of Independence, 19
CODE
 Recoding data, 65
Coefficient of Determination, 20
COEFFICIENT OF DETERMINATION (r^2), 20
Column Statistics, 63, *See also* Descriptive Statistics
Confidence Interval
 for One Population Proportion, 36
 One Population Proportion, 22
 Sigma Known
 One Sample Z, 24
 Sigma Unknown
 1 Sample t, 23
CONFIDENCE INTERVAL FOR ONE POPULATION PROPORTION, 22
CONFIDENCE INTERVAL FOR μ (σ KNOWN): ONE-SAMPLE Z, 24
CONFIDENCE INTERVAL FOR μ (σ UNKNOWN): ONE-SAMPLE t, 23
Contingency Table, 19, *See* Cross Tabulation Table

CONVENTIONS USED IN THE MANUAL, 3

Conventions used in this Manual
 ADDITIONAL REQUIREMENTS, 3
 OPTIONS, 3
 PATH, 3
 PROCESS, 3
 RELATED GRAPHS, 3
 RELATED STATISTICAL PROCEDURES, 3
Correlation. *See* Linear Correlation
Correlation Coefficient, 20, *See* Linear Correlation
Cross Tabulation, 19
Cross Tabulation Table, 25
 Interpreting, 26
CROSS TABULATION TABLE (Contingency Table), 25

D

DATA, 4
Data Entry
 Direct, 67
 Directly into Columns, 68
 Directly into Rows, 68
 Excel Files, 70
 Retrieving Existing Worksheets, 70
 Text/ASCII type files, 71
DATA ENTRY, 67
Data File Descriptions
 finding, 4
Data Files
 Pulse.MTW, 86
 Schools.MTW, 86
Data Formats
 Date/Time, changing, 69
Data Types
 Date/Time, 67
 Numeric, 67
 Text, 67
Date/Time, 67
Descriptive Statistics, 12, *See also* Column and Row Statistics
Discrete Random Variable
 Obtaining a Probability Histogram, 27
 Simulating, 29
DISCRETE RANDOM VARIABLES (Obtaining a Probability Histogram), 27
DISCRETE RANDOM VARIABLES (Simulation), 29
Display Data, 72
DISPLAY DATA, 72
DOCUMENT OVERVIEW, 3
DOTPLOT, 30
Dotplots, 30

E

Editing Charts and Graphs, 76
EDITING CHARTS AND GRAPHS: TOOL AND ATTRIBUTE PALETTES, 76
EDITING THE WORKSHEET, 73
EXPORTING TABLES AND GRAPHS, 78

F

Fitted Line Plot, 31
FITTED LINE PLOT(Linear Regression), 31
Five-Number Summary, 17, 33
FIVE-NUMBER SUMMARY, 33
frequencies, 11
frequency distributions, 11
FREQUENCY POLYGON, 34

G

GETTING STARTED WITH MINITAB, 6

H

Histogram, 35
HISTOGRAM, 35
History Window, 8
hypothesis test, 87
Hypothesis Test
 One Population Proportion, 36
 One Sample Z test, 38
 Sigma Unknown
 1 Sample t, 37
HYPOTHESIS TEST FOR ONE POPULATION PROPORTION, 36
HYPOTHESIS TEST FOR μ (σ UNKNOWN): ONE-SAMPLE t, 37
HYPOTHESIS TEST OF MEAN (σ KNOWN): ONE-SAMPLE Z, 38

I

Info Window, 8
Inter Quartile Range (IQR), 17
INTRODUCTION, 3

L

LINE GRAPH, 39
Linear Correlation, 40
 LINEAR CORRELATION, 40
 Linear Regression, 41
 Fitted Line Plot, 31
 Least-Squares Criterion, 41
 Scatterplot, 55
 LINEAR REGRESSION, 41

M

Maximum, 13
 Boxplot, 17
 Five Number Summary, 33
 Mean, 13, 43
 Trimmed, 13
Measures of Center, 43
 MEASURES OF CENTER, 43
Measures of Variation, 44
 MEASURES OF VARIATION, 44
 Median, 13, 43
 Boxplot, 17
 Five Number Summary, 33
 Minimum, 13
 Boxplot, 17
 Five Number Summary, 33
 Minitab
 Accessing on campus, 6
 Data Window (Worksheet), 6
 Graph Window, 6
 Opening and Starting, 6
 Session Window, 6
 Mode. *See* Tally
 Tally, 60

N

Naming Variables, 69
 NORMAL DISTRIBUTION, 45
Normal Probability Plot, 46
 NORMAL PROBABILITY PLOT, 46
 Normally Distributed Variable
 Probabilities for, 45
 NORMALLY DISTRIBUTED VARIABLE (Determining an
 Observation Corresponding to a Specified Percentile), 48
 NORMALLY DISTRIBUTED VARIABLE (Simulating a Normal
 Distribution), 49
 Numeric, 67

O

One Sample Z test
 Sigma Known - Hypothesis Test, 38
 One-Sample t test
 Confidence Interval (sigma unknown), 23
 Hypothesis Test, 37
 One-Sample Z test
 Confidence Interval (sigma known), 24

P

Pareto Chart
 Cumulative frequency curve, 50
 PARETO CHART, 50
 Patterned Data, 79
 PATTERNED DATA, 79
Pie Chart, 51
 PIE CHART, 51
 Project
 in Minitab, 9
 Project Manager, 8
 PROJECT MANAGER, 82
 PROJECTS AND WORKSHEETS, 9
 p-value
 for Hypothesis Test where Sigma Unknown, 37
 for Hypothesis where sigma known, 38

Q

Q1, 13
 Boxplot, 17
 Five Number Summary, 33
 Q3, 13
 Boxplot, 17
 Five Number Summary, 33

R

r^2 . *See* Coefficient of Determination
 Range, 44
 Recode. *See* CODE
 REPORTPAD, 84
 ReportPad Window, 8
 Row Statistics, 63, *See also* Descriptive Statistics

S

Sampling Distribution of the Mean

Obtaining Means for Generated Samples, 53
 Simulation of, 54
 SAMPLING DISTRIBUTION OF THE MEAN (Obtaining Means
 for Generated Samples), 53
 SAMPLING DISTRIBUTION OF THE MEAN (Simulation), 54
Scatterplot, 55
 SCATTERPLOT, 55
Simple Random Sample (SRS), 56
 SIMPLE RANDOM SAMPLE (SRS), 56
Single-Value Grouped Data Table
 Tally, 60
Split Worksheet, 85
 SPLIT WORKSHEET, 85
 Standard Deviation, 13, 44
 Standard Error of the Mean, 13
Stem-and-Leaf Plot, 58
 STEM-AND-LEAF PLOT, 58
 Sum of Squares
 Error Sum of (SSE), 20
 Regression Sum of Squares (SSR), 20
 Total Sum of (SST), 20
Sums of Squares, 20

T

t
 One Sample t, 87
Tally, 60
 TALLY, 60
 Text, 67
 Tool Palette, 76
 Editing Charts and Graphs, 76

W

Window
 Data (Worksheet), 7
 Graph, 7
 History, 8
 Info, 8
 ReportPad, 8
 Session, 7
 Worksheet
 in Minitab, 9

Z

z-score
 One Sample Z Test. *See*