# **INTRODUCTION**

# PREFACE

# W hat is Statistics?

Many if not most, of life's really important question involve partial information. The process of drawing good, reliable conclusions from partial information is central to the study of Statistics. Statistics is described as the study of how to collect, organize, analyze, and interpret information, especially numerical information. A definite advantage of statistical methods is that they can help us make decisions without prejudice. Moreover, Statistics can be used fort making decisions when we are faced with uncertainties. For instance, If we wish to estimate the proportion of people who will have a severe reaction to a knew flu shot without giving the shot to everyone who wants it, Statistics can provide appropriate methods. Before going any further let us define two important terms in statistical studies, namely Population and Sample.

**POPULATION:** this term refers to all measurements or observations of interest. For examples, if there are 600 students in the school that are classified according to blood type, we say we have a population of size 600. The numbers on the cards in a deck, the heights of residents in a certain city, and the lengths of fish in a particular lake, are all examples of populations. Whereas the term **SAMPLE:** is simply a part of the population. But not every sample is useful; the sample must represent the population.

Statistics has two major types that are Descriptive Statistics and Inferential Statistics,

**Descriptive Statistics:** consists of all methods for organizing and summarizing information, while **Inferential Statistics:** consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample from the population.

# **Observational and Experimental Studies**

In many cases the purpose of a statistical study is to investigate whether a relationship exists between two variables or characteristics, such as smoking and lung cancer, height and weight, or educational attainment and annual income. For such studies it is important to distinguish between two types of procedures: observational studies and experimental studies. In an *observational study*, researchers simply observe, in an experimental, researchers design and control. Observational studies can reveal only association, whereas *experimental* studies can establish causation (cause and effect).

# Variables and Data

#### 1. Variables

A characteristic that varies from one person or thing to another is called a variable. A non-numerical-valued variable is called qualitative variable and any numerically valued variable is called quantitative variable. Quantitative variables also can be classified as discrete or continuous. A discrete variable is one whose possible values form a finite (or countable infinite) set of numbers. A continuous variable is a variable whose possible values form some interval of numbers.

#### 2. Data

Observing the values of a variable for one or more people or things yields data. Thus the information collected, organized, and analyzed by statisticians are data so Data can be defined as information obtained by observing values of a variable. Qualitative data, Quantitative data, Discrete data and Continuous data are the data obtained by observing the values of Qualitative, Quantitative, Discrete, and continuous variables, respectively.

**Example 1.** Humans are classified as having one of four blood types: A, B, AB ro O. What kind of data do you receive when you are told your blood type?

**Solution:** Your blood type is Qualitative data obtained by observing your value of the variable "Blood Type".

**Example 2.** The US bureau of the census collects data on household size and publishes the information in current population reports. What kind of data is the number of people in your household?

**Solution:** The number of people in your household is discrete, quantitative data; data obtained by observing the value of the variable "Household Size" for your household.

**Example 3.** In *Information Please Almanac* lists the world's highest waterfalls. The list shows that Angel falls in Venezuela is 3281 feet high, more than twice as high as Ribbon falls in Yosemite, California, which is 1612 feet high. What kind of data are these heights?

**Solution:** The waterfall heights are continuous, quantitative data; data obtained by observing the values of the variable "Height" for the two waterfalls.

# **Levels of Measurements**

When we collect data, it is common to classify the information obtained according to one of the following four levels of measurement: **Nominal Level:** Data at this level of measurement consist of names only or qualities, with no implied criteria by which the data can be identified as greater than or less than other data items. The Faculty of the student at Yarmouk University, Nationality of a person and Blood type are all examples of the Nominal level.

**Ordinal Level:** Data at this level may be arranged in some order, but actual difference between data values either cannot be determined or are meaningless. Faculty member rank (Prof., Assoc., Assis, ... etc.), military rank and student rating at graduation (Excellent, V.Good, ... etc) are all examples of the Ordinal level.

**Interval Level:** It is like the Ordinal level, but it has the additional property that meaningful differences between data values can be computed. Interval-level data may not have an intrinsic zero or starting point. Consequently, differences are meaningful but ratios of data are not. Body temperatures and Intelligence quotient are examples of such a level.

**Ratio Level:** Is similar to the interval level, but it includes an inherent zero as a starting point for all measurements. Consequently, at this level both differences and ratios are meaninggul. Age, Weight and Height of a person are all examples of the Ratio level.

# **Important Notes:**

- 1. You should bring a 3.5-inch floppy disk, submit it to Mr. Baroud to download the projects necessary to your work through the semester, and bring it in each session. ALWAYS keep it clean and in excellent condition.
- 2. You should bring the booklet of STAT 105 in each session, the booklet is available ONLY at Rawa'ah bookshop opposite to the main gate of the university.
- 3. No Homework is valid to be submitted BEYOND the deadline.
- 4. No makeup exams under any circumstances.
- 5. The midterm exam will be in the 9<sup>th</sup> week from the beginning of the semester.

- 6. The final exam will be in the last week of the semester just before the final exams.
- 7. The whole mark will be divided as follows: 40% for HW's 25% for midterm exam 35% for final exam
- 8. All the exams will be theoretical and practical.
- 9. In the beginning of each session you will sit for a short practical QUIZ about the material to be explained, that is you should prepare your self well for the quiz in advance.
- 10. Write your NAME, SECTION NUMBER and SERIAL NUMBER on your disk and on each of the student work sheets in advance.

# **INTRODUCTION TO MINITAB**

# The most commonly used programs for statistical work are obtained from statistical software packages, collections of statistical computer programs written by some organization or individual. We have

computer programs written by some organization or individual. We have chosen Minitab to illustrate the basic ideas of statistical software. You can find the Minitab package is already downloaded on the computer system right in front of you.

Traditionally, Minitab has been applied by typing commands on the MTB> prompt, so-called **session commands** which are available on all platforms. Minitab has recently introduced as menu interface as well, so that commands can also be executed by choosing them from menus and completing dialogue boxes, which are called **menu commands** that may not be available on all platforms. For each application of Minitab, the required session commands and the corresponding menu commands will be separately introduced.

# Startup

To start running the Minitab package, you can find the Minitab icon on the start screen after a little while of switching on the computer in front of you as shown below;



By double clicking on the icon the Minitab program will immediately appear on front of you as shown below;

MINITA	8 Student - U	Intitled				, 			W 🔀 🖭 🛙	🕽 🐼 💶 🛛 🗙
<u>File E</u> dit <u>I</u>	<u>M</u> anip <u>C</u> alc	<u>S</u> tat <u>G</u> raph E	<u>d</u> itor <u>W</u> indov							
<b>2</b>	🞒 X 🖻		I I 🗛 🕯	1 🖪 🗉			?			
E Session	1									
Workshe	et syze:	5000 cell:	3							<b>^</b>
MTB >			-							
Worksh	neet 1 ***	<u> </u>	<b>C</b> 2		05			<u></u>		
Ι.	<u> </u>	2	G	- 4	6	Co	U U	- 28	6	<u>C10</u>
+										
2										
3										
4										
5										
6										
7										
A 8										
Current Work	sheet: Worksh	eet 1						E	ditable	11:24 AM
Start	10 Microsoft	ص ۱۰۵ (Word - 1	ے) 🔯 Explor	ing - Raed	<b>≥</b>	AINITAB Stu	dent - U		En<	ص 11:24 🔬 🎧

In the figure above you can see that there are two secondary windows within the main Minitab window, namely the upper one is **Session** and the lower one is **Work sheet 1** \*\*\* or **Data**. You can see that the active window is the Session, so if you want to work on the Work sheet 1 \*\*\* instead you can simply click anywhere inside the needed window to make it active.

The Session window is that one, which you can type Minitab commands on and get results and graphs on also. Whereas, the Work sheet 1 \*\*\* window is that one, you can enter data only and get data entered from Minitab commands.

# **Storing Data Sets**

To prepare you for Minitab applications, we will know explain how to store one or more data set. As a beginning you need to understand that the data set is stored in a column that is designated by a "C" followed by a number; thus C1 stands for column 1, C2 for column 2, ... etc.

Two commands can be used to store data. One is the **Set** command; the other is the **Read** command.

Set command: is used to store one data set at a time.

Session commands: Follow the syntax;

MTB >	SET	C1	(Enter)
DATA>	# #	#	(Enter)
DATA>	# #	# etc.	(Enter)
DATA>	END		(Enter)
MTB >			

*Menu commands*: We simply enter data in column C1 in the data window directly after Minitab has started.

**Read** command: is used to store two or more data sets simultaneously, provided that the data set are all of the same size.

Session commands: Follow the syntax;

MTB> F	READ	) C1	LC2	2 C	3	•	• •	Cn	(Enter)
DATA>	#1	#2	#3	•	•	•	#n		(Enter)
DATA>	#1	#2	#3	•	•	•	#n		(Enter)
			•						
			•						
			•						
DATA>	#1	#2	#3	•	•	•	#n		(Enter)
DATA>	ENI	)							(Enter)
MTB>									

*Menu commands*: We simply enter data in columns C1, C2, C3 up to Cn in the data window directly after Minitab has started.

There is a more powerful tool for storing data sets especially patterned ones that enables you to enter data with some pattern like arithmetic sequences or constant values. It can be done as follows; *Menu commands*: **Calc > Make Patterned Data...** 

≥м	NITAB - lai	99.MPJ											
File	Edit Manip	Calc Stat	Graph Ed	itor Windo	w Help	<b>+</b>   <b>x</b>   <b>A</b> A		9		- Cle			a real
SUE	ession 3C> Rep	Column Row Sta Standar Extract Extract	Statistics itistics dize from Date/Tin from Date/Tin	ne to Numeri ne to Text	c			8 2-					
We] Ret	.come to M rieving p	Make Pa Make Ma Make In Set Base Random	itterned Data esh Data dicator Variat e	les	•	Simple Set of N Arbitrary Set o Text Values Simple Set of D Arbitrary Set o	umbers f Numbers. ate/Time V. f Date/Time	 alues 9 Values					
A11		Probabil	ity Distributio	ns	+								
III v	Vorksheet 1	***											
÷	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	^
	Weights	Nscor	Po(1)										
1	0.9473		0										
2	0.9655		0										
3	0.9703		0										
4	0.9757		1										- 1
5	0.9770		0										- 1
6	0.9775		2										~
<													>
	6	arasz	2							Edita	ble	2:25 PM	
-	start	ا 🕲 🥥	🕑 » 🔳	2 Microsof		🔁 MINITAB - I	. 🛛	Search Res	Mi	rosoft E		21 2:25	PM

**Example 1**. If you want to store the multiples of 4 between 20 and 50 where each value appears twice and the whole sequence repeats 3 times in column C1, i.e. C1 will contain

20	20	24	24	28	28	32	32	36	36	40	40	44	44	48	48
20	20	24	24	28	28	32	32	36	36	40	40	44	44	48	48
20	20	24	24	28	28	32	32	36	36	40	40	44	44	48	48
1	1		1			.1	C		1 1						

this can be done as shown in the figure below;

Simple Set of Numbers					X
	Store patterned d	ata in:	c1		
	From first value:	20			
	To last value:	50			
	In steps of:	4			
	List each value		2		times
	List the whole se	quence	3		times
Help		ОК		Ca	ncel

# Naming a Column

Naming a column is often useful because it allow us to refer to the column by its name instead of trying to remember its number. This is can be done easily either by using the command Name C3 `Exam' while you are in the session window, or simply write the column name directly above the first entry in the column while you are in the data window.

<u>Note</u>: Actually for the ease of dealing with the data, that is entering, deleting or modifying data, it's much better to use the data window instead of the session window.

# **Copying from Session to Data**

You can copy any number of columns from the Session window to the Data window by using the following procedure,

- 1. Shade the numbers, in the columns, needed only using the mouse or the *shift* and *arrows* keys on the keyboard.
- 2. Press ctrl+c or click on the copy button to copy.

- 3. Click on the first cell in the first empty column.
- 4. Press ctrl+v or click on the paste button to paste.
- 5. Press OK.

# **Opening Files**

1. To Open an Old Worksheet

If you want to open an already existing worksheet from the hard disk you can do the following,

Session commands:

MTB > Retr 'C:\Program Files\MTBWINST\Data\Acid.mtw'

#### Menu commands: File>Open Worksheet...



, press \Progi	Fl for belg Open Worksh	eet				?	2 X
et from ved on	Look in: 🔂	Mtbwin		- 🗈	<u></u>	🔺 🔳	
	Student9 Studnt12 Student1 student8 Data Macros						
C	File <u>n</u> ame:	Minitab (*.mtw)			- -	<u>O</u> pen Cancel	10
)9  1  0  0	Description	Options	Preview		C Merg	<u>H</u> elp	
05							

Open Worksh	neet		<u>?×</u>
Look <u>i</u> n: 🔂	Data	· 🖻 🖻	* 📰
Acid 🔤	🚾 Beetle	🚾 Cartoon	Copper
Acid1	📶 Billiard	🚾 Cod_ex1	🚾 Cranksh
Acid2	🚾 Вохсох	🚾 Cereal	🚾 Crankshd
🕂 Alfalfa	🚾 Врсара	🚾 Cholest	🚾 Defects
Auto 🔤	🚾 Brake	🚾 Cholestc	🚾 Deodoriz
🚾 Azalea	🚾 Camshaft	🚾 Choleste	🚾 Diameter
🚾 Bears	🚾 Cap	🚾 Cities	🚾 Docs 👘
			F
File <u>n</u> ame:	Beetle		<u>O</u> pen
Files of <u>type</u> :	Minitab (*.mtw)	<u> </u>	Cancel
			<u>H</u> elp
Description	. Ogtions Previe	w C Merge	e 🖲 Open

# 2. To Open an Old Project

If you want to open an already existing project from your own floppy drive you can do the following, *Menu commands*: **File>Open Project...** 

$\mathbb{X}$	MINITAB Stud	ent - <mark>Untitled</mark>							- W 🔀	🔳 🔕 🐼	_ 8 ×
Eile	e <u>E</u> dit <u>M</u> anip	<u>Calo Stat G</u> r	aph E <u>d</u> itor <u>W</u>	(indow <u>H</u> el	p						
	<u>N</u> ew		Ctrl+N		× .			102			
	Open Project		Ctrl+O								
	Save Project	43	Ctrl+S	CONTRO	orem File	- WTBUINS	ST\Dete\W	rid mtw			
	Project Descripti	 on		03:41:	49		JI, DUCU, A				
	Open <u>W</u> orkshee	t									
	Save <u>Current</u> W	orksheet									
	Close Workshee	orksheet As t									
	Open Graph			- BWINST	'\Data\Aci αrem File	d1.mtw'	T\ Doto\ 4	ridi mtu			
	Save Workshee	tAs		03:41:	91au 1116 49	Sinipana.	JI (Data) A	JIGI.MCW			
	Other <u>F</u> iles		•								
	Print Worksheet.		Ctrl+P	-							ك
	Prin <u>t</u> Setup										
	E <u>x</u> it					_	_	_	_	_	
	1 C:\Program File	es\\Data\Acid	1.mtw	<b>C</b> 3	C4	C5	C6	C7	C8	C9	C10
	2 C:\Program File	es\\Data\Acid	.mtw					0,	00	00	010
	3 D:\Raed\Lab2	2.MPJ		<u> </u>							
	4 D:\Raed\Laba	I.MPJ									
	2	0.109									
	3	0.110									
	4	0.109									
	5	0.112									
	6	0.109									
	an a Minitab crois	ot filo									
jυρ	en a minitati proje		15			1.500		1			
	Start 🛛 🔯 Ex	ploring - Raed	<b>I</b> INITAE	Stu	🕑 Microsoft W	ord • 🔀 M	icrosoft Excel -			<b>En</b> &E@\}	ص 10:28 ۽

OR



Open Project				?	×
Look jn: 🔂	Data	• 🗈	<u></u>	📸 🛅	
File <u>n</u> ame:	a:try			<u>O</u> pen	
Files of <u>type</u> :	Minitab Project (*.MPJ)		•	Cancel	
				<u>H</u> elp	
Description	. Options Preview	l			

# **Saving Files**

1. Saving Your Session

It's very important to save your commands and the results obtained from it in the end of the class in order to be able to print it on papers using the printer in the lab. This can be done as the following,

Menu commands: File>Save Session Window As...

Save As						?	×
Save jn: 🔂	Mtbwinst	-	<b>E</b>	<u></u>	<b>e</b> *	6-6-	
🛄 Data							_
Macros							
Student							
1							
File <u>n</u> ame:	a:LAB0					<u>S</u> ave	
Save as tupe:	Minitab Session (* TXT)			-		Cancel	1
5575 Bo ()po.	phinical descion((.1771)				-		
						<u>H</u> elp	
							11.

# 2. Saving Your Project

It's very important to save your whole work in the end of the class in case you need to modify on your project later. This can be done as the following,

Session commands:

```
MTB > Save `a:lab0';
SUBC> Project.
```

Menu commands: File>Save Project...

, press \\Progn et from 7ed on	Fl for helm Save Project As Save jn: 🔄 Data 💽 主 🙋	?× 
	≥⊇ try	
(	File name: a:Lab0	<u>S</u> ave 11
	Save as type: Minitab Project (*.MPJ)	Cancel
1		<u>H</u> elp
0	Description Options	
15		

# **Exiting Minitab**

When we are finished with Minitab, we must exit the software. This is accomplished in the following manner:

Session commands: Type Stop command then press (enter).

*Menu commands*: **File** > **Exit** then click on **NO**.

ays	40	68.45	•	67.50	68.50	J 16.	.77	2.65
'ariable	Minimum	Maxim	AINITAE	3				
'ield 'ays	55.000 36.00	67.0 99.	⚠	Save chan	iges to the	project lab11	MPJ before	closing?
MB > Stop.				Yes		lo	Cancel	
Worksheet '	***							
↓ C1	C2	C3	C4	C5	C6	C7	C8	C9

# **Printing Your Work**

After exiting Minitab you need to print your work on the printer, this can be achieved by the following commands, after you insert your floppy in the floppy drive in the case of the printer's computer,

c:\> a: (Moves the reader head to drive a) a:\> dir (Exhibits your files) a:\> copy filename.extension prn

C:\WINDOW	l\$≻cd\					
C:\≻a:						
A:\>dir						
Volume i Volume S Director	n drive A Gerial Numb y of A:\	has no la per is 320	abel 51-19FA			
WORKSHOP	DOC	23.552	07/05/02	11:09a	Workshop.doc	
FALL2002	XLS	111.616	15/01/03	4:45p	Fall2002.xls	
LABØ	MPJ	8.192	23/01/03	1:09p	Lab0.MPJ	
PAPER2	ТХТ	16.332	29/07/02	12:15p	Paper2.txt	
	4 file(s)	15	59.692 byt	tes		
	0 dir(s)	1.29	96.896 by1	tes free	2	
А:\≻сору	paper2.tx1	t prn				

From know on you are able to start Minitab, deal with data and exit easily. The details of other commands will be given at the suitable location of this booklet when it is necessary.

# UNIT I DESCRIPTIVE STUDY OF DATA

Main aspects of describing a data set:

- Summarization and description of the overall pattern of the data by:
- 1. Presentation of tables and graphs
- 2. Examination of the overall shape of the graphed data for important features, including symmetry or departures from it.
- 3. Scanning the graphed data for any unusual observation that seems to stick far out from the major mass of the data.
- Computation of numerical measures for
- 1. A typical or representative value of the center of the data.
- 2. The amount of spread or variation present in the data.

# LAB 1 REPRESENTATION OF DATA BY CHARACTER GRAPHS

# **D**ot-plot

The dot-plot is a graphical display for quantitative data. Dot-plots are particularly useful for showing the relative positions of the data in a data set or for comparing two or more data sets, when the data consist of a small set of numbers (say, less than 20 or 25).

Dot-plot can be performed by Minitab using command Dotplot.

Open the Minitab project Lab1.mpj from your floppy disk.

**Example 1.** Construct a dot-plot for the data Yield stored in C1.

Session commands:

MTB > Dotp 'Yield'



Menu Commands: Graph> Character Graphs> Dotplot...

	- Stat105.MF	ึง	
<u>Eile E</u> dit <u>M</u>	anip <u>C</u> alc <u>S</u> I	tat <u>G</u> raph E <u>d</u> itor <u>W</u> indo	
🗡 🖬 🧉	🗃 👗 🖻 I	Layout	2 🖉 🖾 🖽 📾 📾 📾 🚫 🧣
Session		<u>Plot</u>	
4	9 5899	Lime Series Plot	
-	2 0022	Histogram	
MTB > se	et c6	Boxplot	
DATA> 24	4(1)	Matrix Plot	
MTB > Sa	na ave "C:\li	Draftsman Plot	05.MPJ":
SUBC>	Project;	Contour Plot	
SUBC>	Replace.	3D Plot	
Saving 1	file as: C	3D Wireframe Plot.	at105.MPJ
IIID >		3D Surrace Plot	
		Dotplot	
		Ple Chart Marginal Plot	
Workshe	et 1 ***	Probability Plot	
	<u> </u>	Stem-and-Leaf	
+	Popul	Stein-and-Lea <u>r</u>	Pays Value Ones
1	147	44 Character <u>G</u> raphs	Set Options 1
<u> </u>	509	558 60	Histogram
3	318	211 64	Boxplot 1
4	252	423 61	
5	630	461 55	
6	423	60	Scatter Plot 1
7	678	62	Multiple Scatter Plot 1
4	976	F1	Time Series Plot 1
Create a chara	acter-stule dotolo	t that prints in the Session w	inde Brid
grastart		Exploring	م Seudo <u>3</u> D Plot جمن ۲۰۱۵ (۲۱ 🔤
MINITAB 🔁	- Stat105.MF	Ŋ	
<u>File E</u> dit <u>M</u>	lanip <u>C</u> alc <u>S</u> I	tat <u>G</u> raph E <u>d</u> itor <u>W</u> indo	uw <u>H</u> elp
	<u>s</u> 7		
E Session			
E ocsaion			
4	9 5899		
		otplot	X
MTB > se	et c6 🛛 🎽	ocproc	
DATA> 24	4(1)		Variables:
DATA> er	nd		
MTB > Se	ave "C:		Yield
SUBC>	Projec		
SUBC>	Renlac		
Saving 1	file age		
MTP S	ciic ab.		
HID >			
			By variable:
- Workshe	eet 1 ***		<u>Same scale for all variables</u>
	C1		N C10
			First midpoint:
<u> </u>	Popul		
1	147		Last midpoint:
	500		The large set I
2	505	P = I = = t	
3	318	select	
4	25:		
	202	Help	OK Cancel
5	630		
6	423	60	<u> </u>
7	670	ch	
ļ	0/0		
8	376	61	
Current Works	heet: Workshee	et 1	3:44 PM
Mar III	NAR	A will Commune 1	
start	162 🧶 🕼	2 Exploring	م 03:44 ج MINITAB 🗠 المحمد من المحمد 🗠 🖓 🖓 التقادية المحمد المحم

<u>Note</u>: The *Tick increment* option is used to control the width of the plot since the width of the plot is inversely proportional to the tick increment given that the width of the plot <u>should not</u> exceed the screen width.

<u>Comment</u>: from the previous dot plot it is clear that it has 3 modes (61, 62 and 67) and it is right skewed (or +ve skewed).

#### **Box-Plot**

Box- plot is a graphical technique to display the structure of the data set by using the quartiles and the extreme values of a sample.

To construct a Box plot we need to know the following definitions: The Inter-Quartile Range (IQR) =  $O_3$ - $Q_1$ 

Where  $Q_1$ ,  $Q_2$  and  $Q_3$  are the first, the second and the third quartiles respectively.

*Inner fences:* (Lower)  $LIF = Q_1 - 1.5IQR$ , (Upper)  $UIF = Q_3 + 1.5IQR$ *Outer fences:* (Lower)  $LOF = Q_1 - 3IQR$ , (Upper)  $UOF = Q_3 + 3IQR$ 

In Box plot, data values that lie between the inner and the outer fences are considered possible, or mild, outliers, whereas those lie outside outer fences are considered probable, or extreme, outliers. Box-plot can be obtained using Minitab by the command **Box-Plot**.

The advantages of the box-plot are the following:

- a. It gives the five-number summary.
- b. It gives the outliers and distinguishes between the possible and the probable ones.
- c. It determines, more accurately, the symmetry and skewness of the graph by the following:
  - If  $Q_3 Q_2 > Q_2 Q_1$  then it is skewed to the right (+ve skewed).
  - If  $Q_3 Q_2 < Q_2 Q_1$  then it is skewed to the left ( -ve skewed).

- If  $Q_3 - Q_2 = Q_2 - Q_1$  then it is symmetric.

Considering that the tails are approximately of equal length.

**Example 2.** For the previous sample, construct a box-plot. *Session commands*:

MTB > GStd. MTB > Boxp C1.

# Boxplot



Menu commands: Graph> Character Graphs>Boxplot...

🚬 MINITAB	- STAT105.	MPJ										. 8 X
<u>File E</u> dit <u>M</u>	anip <u>C</u> alc <u>S</u>	itat <u>G</u> raph B	<u>d</u> itor <u>W</u> indow	<u>H</u> elp								
	<b>3</b> X B	R Layou	ıt			<u>ا</u> يد		?				
<b>Session</b> DATA> 24 DATA> e1 MTB > 58	4(1) nd ave "C:\	Plot <u>I</u> ime <u>C</u> hart <u>H</u> istoy	Series Plot  gram ot	L05.MP	J";						-	
SUBC> SUBC>	Project Replace	Matrix Matrix Drafts Conto	: Plot man Plot ur Plot									
Workshee Retriev: MTB >	et size: ing proje	100 <u>3</u> D Pl ct 3D W 3D <u>S</u> i	ot ireframe Plot urface Plot	) ows/d:	ESKTOP\ST.	AT1(	05.MPJ					
<b>↓</b> ₩orkshe	eet 1 ***	Dotpl Pie C Margi	ot hart nal Plot								_	• // 
	C1	Proba	idility Plot	_ C4	CS		C6	C7	C8	C9	C10	
Ļ	Popul	Stem-	and-Lea <u>f</u>	Days	Value	(	Ones					
1	147	44 Chara	icter <u>G</u> raphs 🕠	Set O	ptions		1					
2	509	558	60	– Histor	gram		1					
3	318	211	64	<u>B</u> oxpl	ot		1					
4	252	423	61	<u>D</u> otple	ot <sup>1</sup> 13		1					
5	630	461	55	<u>S</u> tem-	and-Leat		1					
6	423		60	Scatte	er <u>P</u> lot		1					
7	678		62	<u>M</u> ultip	ile Scatter Plot		1					
۹ •	1 376		R1	<u>T</u> ime :	Series Plot		1					
Create a chara	icter-style boxp	lot that prints in	the Session wind	le <u>G</u> rid	u II							
🛃 Start	📓 🤞 🕼 (	🛃 » 🛛 📷	Mi	Pseud	do <u>3</u> D Plot		INITAB	- S 🕅	Document1 - M	i   🔡 🐻	1 1 1	J3:08 e

MINITAB	- STAT105.	MPJ							_ & ×
<u>File E</u> dit <u>M</u>	lanip <u>C</u> alc §	<u>S</u> tat <u>G</u> raph I	E <u>d</u> itor <u>W</u> indo	w <u>H</u> elp					
🖻 🖬 é	<b>3</b> X Pa		1 🚺 🏘 🕯	7 🖪 🗉		<u>s</u> 🛛 🔋			
🗮 Session									_ 🗆 🗵
DATA> 2	4(1)								<u> </u>
DATA> er	nd [	Boxplot						< l	
MTB > St	ave "C:		-						
SUBC>	Projec Replac	C1 Pc C2 SE	opul RS	Variable	: [Yield				
		C3 Yi	ield	E By y	ariable:		1		
		C5 Va	ays alue	, . 	lovolo:				
Retriev	et size: ing proj	C6 Or	nes	<u>0</u> 56	IGVEIS,				
MTB >	5 F5			□ Note	h, confidenc	e level: 90.0			
						,			_
•				Axis					
Workshe	eet 1 ***			<u>M</u> inimu	m position:		-		- 🗆 🖂
	C1			M <u>a</u> xim	Im position:		-	29	C10 🔺
L I	Popul			Tick inc	rement:	5	_		
1	14			<u>_</u>		10			
2	509								
3	318	S	elect	∏ <u>C</u> one	lensed displ	ay			
4	25.				г				
5	63(	Help				<u><u> </u></u>	Cancel		
6	423		60	b/	2	1			
7	678		62	60	1	1			
4	376		R1	78	A	1			
Lurrent Works	heet: Workshe	et I					Editab		3:09 PM
🛃 Start	😹 🦽 😹	🌌 » 📃 💌	(حص 1۰۵ (1	🔍 Exploring	📔 🔁 MINITA	.B 💌 Document1		🕂 🔁 En 🍕 🕻	ە 03:09 🧞

In Minitab the asterisk (\*) denotes a possible, or mild, outlier where the circle  $(\circ)$  denotes a probable, or extreme, outlier.

# **Stem-and-leaf diagram**

This method was developed in the late 1960's by Professor John Tukey. It is often easier to construct than either frequency distribution or histogram and generally displays more information. This can be performed in Minitab using the command **Stem**.

**Example 3.** Construct an <u>ordered</u> stem-and-leaf diagram for the data Days, <u>one value for</u> (ten leaves in) each stem, we let the increment to be  $\underline{10}$ ;

Session commands:

MTB > stem c2; SUBC> incr 10.

**Character Stem-and-Leaf Display** 



Menu commands: Graph>Character Graphs>Stem-and-Leaf...

MINITAB	- STAT105.M	IPJ									l.	- 8 ×
<u>Eile E</u> dit <u>M</u>	anip <u>C</u> alc <u>S</u>	at <u>G</u> raph E	ditor <u>W</u> indow	<u>H</u> elp								
	3 🕹 🕹	E Layou	at	2 30	» II (	3			8			
Cession		<u>P</u> lot	Casian Diak								_	
DATA> 24	ŧ(1)	<u></u> hart.										_
DATA> er	nd We "Coll	Histog	gram	IOS MP.	T							
SUBC>	Project;	Matrix	ot Plot	105.111								
SUBC>	Replace.	<u>D</u> rafts	man Plot									
		Co <u>n</u> to	ur Plot									
Workshee	et size: :	.oc <u>3</u> D PI	ot									
Retrievi	ing projec	t 3DW	ireframe Plot	pows/DI	ESKTOP\ST	AT1	05.MPJ	r				
nib >		3D <u>S</u> t	urface Plot	-								
		Dotple Dia Cl	ot									
		Margi	nal Plot									
Workshe		Proba	bility Plot	C4			CE	67	68	60		
	Popul	Stem-	and-Lea <u>f</u>	Davs	Value	ſ	Ines			0	0.0	
1	147	44 Chara	cter Graphs	Set O	ntions		1					
2	509	558	60	Histor			1					
3	318	211	64	Boxple	pan		1					
4	252	423	61	Dotple	ot		1					
5	630	461	55	Stem-	and-Leaf		1					
6	423		60	Scatte	er <u>P</u> lot		1					
7	678		62	<u>M</u> ultip	le Scatter Plot.	•	1					
	376		<b>E</b> 1	<u>T</u> ime S	Series Plot		1					
Draw a charac	ter based stem.	and leaf plot		<u>G</u> rid								
				<u>C</u> onto	ur			[				
	XXX 44 Mail 10	1 "	xoloring - raid	Pseuc	to 3D Plot		isoft	I > MINITA	B - STAT10	The second		01.14

<u>Comment</u>: from the previous stem-and-leaf diagram it is clear that it is right skewed, since the number of leaves right to the median is greater than that left to the median. The number of modes will be the same regardless of the type of the stem-and-leaf diagram (ordered or modified).

MINITAB	- STAT105.	MPJ								_ & ×
<u>File E</u> dit <u>M</u>	anip <u>C</u> alc <u>S</u>	<u>S</u> tat <u>G</u> raph I	E <u>d</u> itor <u>W</u> indow	<u>H</u> elp						
🖻 🖬 🖉	3 🔏 🖻			0	» 🖪 🖲	3 🔳 🖻		?		
E Session										
DATA> 24 DATA> er MTB > Sa SUBC> SUBC>	4(1) nd ave "C: Projec Replac	C1 Po	af opul RS	 ⊻ariabl	es:				×	<u> </u>
Workshee Retrievi MTB >	et size: ing proj	C3 Y1 C4 Da C5 Va C6 Or	ield ays alue nes	Days						
<b>▲</b>	oot 1 ***			⊏ву∨	variable:					- - 
He WORKSHE	C1								<b>~</b> 0	
+	Popul			□ <u>T</u> rim	outliers				~	
1	14;	1								
2	50									
3	318	S	elect	Increm	ent: 10					
4	25:									
5	63(	Help				<u>0</u> K		Cancel		
6	423		60	67	2	1				
7	678		62	60	1	1				
• • • • •	976	1	E1	79	1	1	I			
Current Works	heet: Workshe	et 1								1:14 PM
Start	💹 🚜 🖄 !	🧭 » 🕅 🕅	Exploring - raid	<b>1</b>	- (حص 10 11 1	Microsoft 1		B-STATIO	Pa En d	1 <b>1</b> 15 • 01 • 15 •

**Example 4.** To obtain a <u>modified</u> stem-and leaf diagram, for the data Days, with <u>two values for</u> (five leaves in) each stem let the Increment be 5,

Session commands:

MTB > stem 'days'; SUBC> incr 5.

# **Character Stem-and-Leaf Display**



Menu commands: Type 5 in Increment.

	- STAT105.	MPJ Stat. Graab. J								
	amp gaid g	en el el						Jal		
	3 <u>a</u> 43		`ā  <u>°</u> ā  o¦a  A⊡					<u> </u>		
E Session										- 🗆 🖂
DATA> 24	4(1)									<u> </u>
DATA> en	nd 🔤	Stem-and-Lea	af						×	
SUBC>	Projec	IC1 Pe		Variable	e.					
SUBC>	Replac	C2 SF	Ś	Daved	5.				1	
		C3 Vi C4 Da	.eld .vs	Daya				_	1	
Norkshee	et size:	C5 Va	lue							
Retriev:	ing proj	C6 01	les					<b>T</b>		
MTB >				· ·						
										-
				□ <u>B</u> y va	ariable:					
Workshe	et 1 ***									<u>- 🗆 ×</u>
	C1								29	C10 🔺
↓ U	Popul			□ <u>T</u> rim	outliers					
1	14;									
2	505		1							
3	318	51	elect	Increme	nτ: [5					
4	25:	Hala				OK		Canaal	1	
5	63(	rielp					<u> </u>	Cancer		
6	423		60	67	2	1				
7	678		62	60	1	1				
	1 376	I	L 61	78	1	1		1	I	
Current Works	heet: Workshe	et 1								1:15 PM
🚮 Start	🐺 🈂 😭	🧭 » 🛛 🔯 E	Exploring - raid	<b>100</b>	- (حص ١٠٥ (1	Microsoft	NINITA	B - STAT10	💾 🔁 En 🍕	D1:15 e

Example 5. To obtain a modified stem-and leaf diagram, for the data Yield, with five values for (two leaves in) each stem let the Increment be <u>2;</u> *Menu commands*: Type 2 in **Increment**.

MINITAB	- STAT105.	MPJ								_ 8 ×
<u>File E</u> dit <u>M</u>	anip <u>C</u> alc <u>(</u>	<u>S</u> tat <u>G</u> raph B	E <u>d</u> itor <u>W</u> indow							
	3 X Pa		1 I M A				ę			
E Session										
8	85									<u> </u>
7	8 67	Stem-and-Lea	əf						×	
5	89	Lot D	1	Variable	o.'					
4	9	C2 SR	pul S		5.				n l	
4	95	C3 Yi C4 Da	eld.	Days				<u> </u>		
3	9 899	C5 Va	lue							
ľ	5 055	Le Un	ies					<b>*</b>		
MTB >				,						
										-
				Г <u>В</u> у ∨а	riable:					<u>▶</u> //:
Workshe	et 1 ***									
	C1								29	C10 🔺
↓	Popul			□ <u>T</u> rim o	outliers					
1	147									
2	509	0.	-1	Increme	-+• [2					
3	318	30	elect	Increme	n. j <i>2</i>				_	
4	25.	Heln	1			OK		Cancel	1	
5	63(	пстр					₽	Cancer		
6	423		60	b/	2	1				
7	678		62	60	1	1				
<b>▲</b>	1 376	1	L 61	79	4	1	1	1		
Current Works	heet: Workshe	et 1						Edita	ble	1:15 PM
🚮 Start	🖉 🙈 🔛	🧭 » 🕅 🔯 E	Exploring - raid	<b>B</b> 1	- (حص 10 (1	Microsoft		AB - STAT10	"aEn<	o1:16 و 🗐

Session commands:

MTB > stem 'Days'; SUBC> incre 2.

#### **Character Stem-and-Leaf Display**



# LAB 2

# **RERESENTATION OF DATA BY HIGH RESOLUTION GRAPHS**

# **Random Sample**

If we are to use information obtained from a sample to draw conclusions about a population, the sample must be representative of the entire population. A type of sample that is representative of the entire population is a random sample so the random sample is defined as follows,

*Definition:* A simple random sample (SRS) of measurements from a population is a one selected in such a manner that every sample of size n from the population has equal chance, probability, of being selected, and every member of the population has equal chance of being included in the sample.

The traditional way to select a SRS is to use the random-number table, but here we will not use such a table because the Minitab software can be employed to obtain a SRS from a population.

**Open** the Minitab project *Lab2.mpj* from your floppy disk.

**Example 1.** Generate (draw or simulate) a SRS of size 5 from the population in C1 and store it in C2.

Session commands:

MTB	$\geq$	base 0	┢	<b>Compulsory</b> in sessions and
MTB	>	Sample 5 'Popul' 'SRS'		exams
MTB	>	Print 'SRS'		Forbidden in HW's

# **Data Display**



Menu commands: Calc> Random Data> Sample From Columns...

MINITAB	- Untitled									8 ×
<u>File E</u> dit <u>M</u>	anip <u>C</u> alc <u>S</u>	itat <u>G</u> raph E	<u>d</u> itor <u>W</u> indo	w <u>H</u> e	lp					
Session	Calc Colu Bow	ulator mn Statistics I Statistics				2				
	Stan	idardize								
Data Dis	spla <u>E</u> xtra	act from Date/T act from Date/T	ime to Numer ime to <u>T</u> ext	ic						
	Mak	e <u>P</u> atterned Da	ita	•						
SRS 422	Mak	e Mes <u>h</u> Data								
125	Mak	e Indicator Vari	ables							
MTB > ba	ase Setj	<u>B</u> ase								
MTB > Se	ampl <u>R</u> an	dom Data		•	Sample From Columns					
MTB >	Prob	ability <u>D</u> istributi	ons		Chi-Square					
	<u>M</u> atr	ices		+	<u>N</u> ormal					_
					<u>E</u>					<u> </u>
<mark>s W</mark> orkshe	et 1 ***				<u>c</u> Uniform					JX
	C1	C2	C3	(		C7	C8	C9	C10	<b>_</b>
↓	Popul	SRS			B <u>e</u> rnoulli Diversiel					
1	147	447			<u>b</u> inomia Discrete					
2	509	558			Integer					
3	318	211			<u>P</u> oisson					
4	252	423			Beta					
5	630	461			– Cauch <u>y</u> …					
6	423				E <u>x</u> ponential					
7	678				<u>G</u> amma					
<b>▲</b>	376				Laplace Logistic					Ţ▼] ▶
Sample randor	n rows from one	or more colum	ns		Lognormal					
					<u>W</u> eibull			Do a f		
Start	20 🧶 🧶 10	2 " ] <u>Q</u> e	.xploring	<b>€</b> ]Ya	hoo! - Mi	[حص ١٠٥ [1 ]		Ľ <mark>∂En</mark> ,	<b>_⊍{{</b> E\}}_⊂02	:16 e
MINITAB									-	a ×

<u>File E</u> dit <u>M</u>	<u>(anip C</u> alc <u>)</u>	<u>S</u> tat <u>G</u> rap	oh E <u>d</u> itor <u>M</u>	/indow <u>H</u> elp						
🖻 🖬 🤞	3 X 🖻		11	M # 🛄			?			
E Session										
MTB > Pr	rint 'SRS	1								<b>_</b>
Data Dis	splay	C1	rom Column Popul	s <u>S</u> ar	nple 5	rows fr	rom <u>c</u> olum	n(s):	×	
SRS 423	147		585	c1				à		
MTB > ba MTB >	ase O			Sto	re samples in'			Y		
				c2	re sampres <u>i</u> n.			<u>_</u>		
Workshe	eet 1 *** C1								29	×
¥	Popul							7		
1	147									
2	509		Select		Sample with re	alacement				
3	318	_	ouren		Sampie <u>w</u> ith re	pracement				
4	252	Н	elp			ОК		Cancel		
5	63		· · ·				━╠┛└		J	
b 7	423									
	0/8									
		·						· · ·		<u> </u>
Current Works	heet: Workshe	et 1						Edita	ble	2:07 PM
🎇 Start	S 🔔 📓	🌌 »]]	🔍 Exploring	· 🖉 Yahoo	o! - Mi 🔁 MINI	TAB 💇 1)	(حص 1+0		BEn _	🍂 Ng 02:07 e

<u>Note</u>: Each time you need to generate a random sample in the class you MUST execute the command (MTB > base 0) first.

# **Bar Chart and Pie chart**

Both charts, bar and pie, are used to represent discrete and qualitative data in high resolution graphs.

#### 1. Bar Chart

To make a bar chart, the class values are marked a long the horizontal axis and a vertical line (or bar) of height equal to the class frequency is erected over the respective class values. This can be performed using Minitab, for the following data set, as follows:

**Example 2.** Construct a bar chart for the data Value and Coded.

Before constructing a bar chart for any set of data you must create a new column filled with a number of a constant digits equal to the number of observations in the column to be charted.

<u>Note</u>: Bar chart and Pie chart are the only graphs that can be used to represent qualitative, nonnumeric, data. So we will construct the bar chart and the pie chart for the variable Coded which is a coded column of the column Value as shown below;

Code - Nu	meric to Text		X
C1 C3 C4	Popul Value Aux	Code data from columns: C3 Into columns: C5	
		Original values (eg, 1:4-12): New:	
		1	
		2 B	
		3 C	
		4 D	
	-	ОК	Cancel



<u>Comment</u>: from the previous bar-chart it is clear that the chart is right skewed and it has one mode at the value A.

Menu commands: Graph>Chart...

-														
2	MIN	ITAB - la	b22.MPJ											
File	Ed	lit Manip	Calc Stat	Graph E	litor Windo	w Help								
2	🖵	1 3 .	ኤ 🖻 📾	Layout.		1	<b>A A</b>		🖬 🖉 ն		8 .	: 📾 🔂 🖸		*e E
				Plot										
	, Se	ssion		Time Se	ries Plot									
Retrieving project fr MTB > GPro. MTB > Chart Count( Au SUBC> Bar; SUBC> SoFrame; SUBC> SoFane;			Chart. Histora Boxplot Matrix F Draftsm Contou	am Mot Not Nan Plot r Plot	ltak	o\lab22.MP	ı						~	
С	Chart Count(Aux)*( MTB > GStd.			c 3D Plot. 3D Wire 3D Surf	 frame Plot ace Plot									
* M	* NOTE * Character o			Dotplot Pie Cha Margina Probabi	 rt il Plot lity Plot									× ::
111	We	orksheet	1 ***	Stem-ar	nd-Leaf									
	<b>۲</b>	C1	C2	Charact	er Graphs	. ► 5-Т	C6	C7	C8	C9	C10	C11	C12	
		Popul	SRS	Value	Aux	Codeo	1							
1	1	147		1	0	А								
2	2	509		3	0	С								
3	3	318		1	0	А								
4	1	252		1	0	A								
1	5	630		3	0	С								~
<	1000													>
Drav	Draw bar charts, line charts, and area charts for categorical data					al data:				1			5:05 PM	1
	y s	tart 🔰	😑 🕑 !	🖻 🔪 📀	Stat105Te		2 Releas	🕞 🔁 Se	arch Res	🔀 Micro	soft E	EN 🔇 🍡 🛛	📕 <mark>16</mark> 5:09	5 PM

Į	hart							
	C1	Popul	Graph va	riables: Fu	nct	ion of Y (me	easu	rement) vs X (category)
	C3 C4	Value Aux	Graph	Function	-	Y		X A
	C5	Coded	1	Count		Àux		Coded
			2					
			3					-
			Data disp	olay:				
			ltem	Display	-	For each	•	Group variables 🔺
			1	Bar		Graph		
			2			Edit Attribut	es	
			3		_			
			<u>A</u> nnotat	ion 💌	Ē	rame 💌		Regions 💌
	H	lelp	Option	ıs			Ok	Cancel

#### 2. Pie chart

**Example 3.** For the previous example, the pie chart can be drawn using Minitab as follows:

Session commands:

MTB > %Pie 'Value'; SUBC> orde 2; SUBC> Expl 1.

Pie Chart of Coded



<u>Comment</u>: From the previous pie-chart it is clear that the LARGEST slice is for the value A and the SMALLEST slice is for the value D.

-														
2	MINIT	AB - la	ID22.MPJ											킨즈
File	Edit	Manip	Calc Stat	Graph E	ditor Windo	w Help								
B		8	እ 🖻 🛍	Layout		<u> </u>	M 🔏 🕒		h 🖉 🕼		? +€	66	) 💈 🗐	1
Canif				Plot		_								n
	Sess	10 N		Time Se	ries Plot									
R	etrie	eving p	project f	r Histoar	am	itab\	lab22.MPJ							<u>^</u>
M	TB>	GPro. Chart	Count( )	, Boxplot										
s	UBC>	Bar;	;	Matrix I	Plot									
S	UBC>	ScFi	came;	Draftsn	nan Plot									
s	OBC>	SCAI	notation	. Contou	r Piot									
c	hart	Coun	t(Aux)*	c 3D Plot										
W	TB \	CS±4		3D Win 3D Surf	are Plot									
*	NOTE	: * Cł	naracter	g ou lu	0.001100111									
Ι.,	<b>TD</b> ~			Dotplot Pie Cha	art									-
ll "	10 >			Margin	al Plot	2								
<				Probab	ility Plot	2								>
	Worl	(a boot	1 888	Stem-a	nd-Leaf									
Ē		Aaneet		Charac	ter Graphs	)					040			
	·		C2	Malua		1-C	6		- 18		C10	C11	C1Z	-8
	P	opui 147	585	Value	Aux	Coded								_=
		147 500		1	0	A C								
	-	318		1	0	Δ								
	<u> </u>	252		1	0	^ ^								
-	-	630		3	0	<u>с</u>								+
		0.50		J	0	~			-					
Drav	v a pie	chart											5:10 PN	1
7	sta	art	۲ ک	📀 🔌 🐻	Stat105Te		2 Releas	🗸 [ 🧟 Se	arch Res	Micro	soft E		<b>14</b> 5:11	0 PM

Menu commands: Choose Graph>Pie Chart...

> WIN	NITAB - lai	22.MPJ	
File E	dit Manip	Calc Stat Graph Editor Window Help	
<b>e</b>	3 6 3		
E Se	ssion		
		Pie Chart 🛛	~
MTB	> GPro.	Chart data in: Coded	
MTB	> %Pie `> orde	C Chart table	
SUBC	> Expl		
Exec	uting fr	Categories III.	
Macı Sett	ting up p	Frequencies in:	
Pie	Chart: Co	Urder of categories: Increasing 💌	
MTB	> GStd.	Explode slice number Worksheet	
MTB	>	Combine categories   Decreasing >	
	anna)	Start angle of first cat Value ordering	~
			2
W 🖿	orksheet 1	Title:	
÷	C1	C12	
	Popul	Colort	
1	147	5000	_
2	509	Help Ontions OK Cancel	_
3	318		_
4	252		_
5	630		~
Welcom	o to Minitab	wars Et far belo	
weitom	e to minitab,	arcost 1 for help	e pa
s	start J	C C C C C C C C C C C C C C C C C C C	O PIM

Pie Chart	4	×
C1 Popul C3 Value C4 Aux C5 Coded	<ul> <li>Chart data in: Coded</li> <li>Chart table</li> <li>Categories in: Frequencies in: Order of categories: Increasing</li> </ul>	
Help	Combine categories less than: <sup>7</sup> into one group Start angle of first category: 0 Title: 0 Options/ 0K Cancel	>
	In <b>this</b> case the <b>smallest</b> slice will be exploded. But, in the case of <b>Decreasing</b> the <b>largest</b> slice will be exploded.	

<u>Note</u>: In the cases of increasing and decreasing the two extreme slices (the largest and the smallest) will always be adjacent since the pie chart is circular.

# LAB 3 REPRESENTATION OF DATA BY FREQUENCY TABLES

# **G**rouping Data

By suitably organizing data, we can often make a large and complicated set of data more compact and easier to understand. We will discuss grouping, which involves, as the term implies, putting data into groups rather than treating each piece of data individually. Grouping is one of the most common methods for organizing data.

**Open** the Minitab project *Lab3.mpj* from your floppy disk.

# 1. Single-value (Simple) Frequency Table (SFT)

In some cases it is more appropriate to use classes that each represent a single possible numerical value. This is often true for discrete data or qualitative data. Consider the following example;

**Example 1.** To construct an **SFT** for the data Yield from Lab1 we use the Minitab command **Tally**,

Session commands: MTB > tall c1

# Tally for Discrete Variables: Yield



Menu commands: Stat>Tables>Tally...

1	ally	
	C1 Yield C2 Days	Variables:          Vield         Display         Counts         Percents         Cumulative counts         Cumulative percents
	Help	OK Cancel

**Exercise**. (IMPORTANT) Construct an SFT for the data Days and copy the resulting columns onto the data window in the columns C2 and C3 respectively. (Hint: use the steps in the bottom of **p.8** in the introduction to the Minitab package to copy columns)

# 2. Grouped Frequency Table (GFT)

When summarizing a large set of data it is often useful to classify the data into classes or categories and to determine the number of individuals belonging to each class, called the class frequency. A tabular arrangement of data by classes together with the corresponding frequencies is called a frequency distribution or simply a frequency table. Consider the following terminology,

Classes: Categories for grouping data.

Lower class limit: The smallest value that can go into a class.

Upper class limit: The largest value that can go into a class.

Class mark: The midpoint of a class.

**Class width:** The difference between the lower class limit of the given class and the lower class linit of the next higher class.

Frequency: The number of pieces of data in a class.

Frequency Distribution: A listing of classes and their frequencies.

**Relative frequency:** The ratio of the frequency of a class to the total number of pieces of data.

**Relative Frequency Distribution:** A listing of classes and their relative frequencies.

**Cumulative Frequency:** The total frequency of all values less than the upper class limit.

**Cumulative Frequency distribution:** A listing of the upper class limits and their cumulative frequencies.

**Relative Cumulative Frequency:** Is the cumulative frequency divided by the total frequency.

**Relative Cumulative Frequency Distribution:** A listing of the upper class limits and their realtive cumulative frequencies.

**Frequency Histogram:** Consists of a set of rectangles having bases on a horizontal axis with centers at the class marks and lengths equal to the class widths. Areas are proportional to class frequencies.

**Example 3.** To construct a **GFT** for the data Days we can use the command **Histogram**.

Session commands:

MTB > gstd

```
* NOTE * Character graphs are obsolete.
```

```
* NOTE * Standard Graphics are enabled.
```

Professional Graphics are disabled.

```
MTB > hist cl
```

#### Histogram



Histogram of Days N = 40

Note: we can comment in terms of modality and skewness as before.

Histogram	
C1 Days	Variables: Days By variable: Same scale for all variables First midpoint: Last midpoint: Interval width:
Help	OK Cancel

Menu commands: Graph>Character Graphs>Histogram...

You can use the subcommand **Increment** to control, indirectly, the number of classes that you wish to use.

Session commands:

```
MTB > gstd
* NOTE * Character graphs are obsolete.
MTB > hist c1;
Subc> incr 10.
```

#### Histogram

Histogram of	Days N	= 40
Midpoint	Count	
40.0	3	* * *
50.0	5	* * * * *
60.0	10	* * * * * * * * * *
70.0	7	* * * * * * *
80.0	7	* * * * * * *
90.0	4	* * * *
100.0	4	* * * *

There is another type of histograms called High Resolution Histogram, *Session commands*:

```
MTB > GPro.
MTB > Histogram Cl;
SUBC> MidPoint;
SUBC> Bar;
```



Menu commands: Graph> Histogram...

Histogram								×
C1 Days	Graph va	riables:						
	Graph	×		-				
	1	Days						
	2							
	3			<b>_</b>				
	Data disp	ilay:						
	ltem	Display	•	For each	• 0	iroup varia	ables	
	1	Bar		Graph				
	2			Edit Attribute	s			
	3		_					-
	<u>A</u> nnotat	ion 💌	F	rame 💌	Re	gions 💌		
Help	Option	IS		(	ж		Cancel	

Different types of frequency distributions can be obtained from the high resolution Minitab graphs as follows,

Menu commands: Graph>Histogram...

Select : from the dialogue box **Options** Check: **Type of Histogram** required.

ð		Histogram Options		
Po ; ra	listogram C1 Day:		Type of Histogram C Frequency C Percent C Density	<ul> <li>Cumulative Frequency</li> <li>Cumulative Percent</li> <li>Cumulative Density</li> </ul>
2			Type of Intervals ⓒ MidPoint	⊂ CutPoint
1			Definition of Intervals • Automatic • Number of intervals: • Midpoint/cutpoint positiv	ons:
+		Select	Transpose X and Y	OK Cancel
	Help	Options	ОК	Cancel

Executing the previous boxes gives;



# UNIT II DESCRIPTIVE MEASURES

# LAB 4

# SAMPLE MEAN AND STANDARD DEVIATION

 $\mathbf{S}$  ometimes we are interested in one number which represents the typical number of the data set. In unit I we defined one such number viz.,  $Q_2$  which is called also the median of the data set. Another number frequently used as a location measure is the sample mean. Similarly, we defined the range of the sample which gives some ideas about the variation in the sample data. Another measure for dispersion is the sample standard deviation.

It is useful to know that the sample mean and the sample standard deviation are denoted by  $\bar{x}$  and s respectively for the sample  $x_1, x_2, ..., x_n$ .

# **1.** Calculation of $\bar{x}$ and s using the sample (raw) data:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
,  $s = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$ ,  $M = Medain = Q_2$ 

where  $S^2$  is called the sample variance.

The sample mean, standard deviation and median are calculated by Minitab using the commands Mean, Stan and Medi respectively.

**Open** the Minitab project *Lab4.mpj* from your floppy disk.

**Example 1.** Calculate the sample mean, standard deviation and median for the data stored in C1 named Days:

which are stored in C1,

Session commands:

```
MTB > name k1 'RawMean =' k2 'RawStDev ='
MTB > name k5 'GFTMean =' k6 'GFTStDev ='
MTB >
MTB > Mean 'Days' 'RawMean ='.
```

# Mean of Days

Mean of Days = 68.300
```
MTB > StDev 'Days' 'RawStDev ='.
```

#### **Standard Deviation of Days**

```
Standard deviation of Days = 16.786
MTB > Median 'Days' k7.
```

#### **Median of Days**

```
Median of Days = 66.500
MTB > name k7 'Median ='
```

Menu commands:	<u>Calc&gt;Column</u>	Statistics
----------------	-----------------------	------------

Column	Statistics			
K1 K2 K3 K4 K5 K6 K7	RawMean = RawStDev = SFTMean = SFTStDev = GFTMean = GFTStDev = Median =	Statistic ○ Sum ● Mean ○ Standard devia ○ Minimum ○ Maximum ○ Range	CM CSu ation CN CN CN	edian Im of squares total nonmissing missing
		Input variable:	Days	-
		Store result in:	'RawMean ='	(Optional)
- F	lelp	]	ОК	Cancel

**2.** Calculation of  $\bar{x}$  and s for SFT and GFT: In case of grouped data,

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} x_i f_i \quad , s = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{k} (x_i - \overline{x})^2 f_i} \quad , n = \sum_{i=1}^{k} f_i , k = \text{number of classes}$$

were xi's and fi's are the classmarks and their corresponding frequencies respectively.

**Example 2.** For the SFT of the data in C1, which is stored in the columns C2 and C3, calculate the mean and the standard deviation. <u>Note</u>: It is much easier to achieve that using the session commands.

Session commands:

```
MTB > name k3 'SFTMean =' k4 'SFTStDev ='
MTB > let k3=sum(c2*c3)/sum(f)
MTB > let k4=sQrT(sum((x-k3)**2*f)/(sum(f)-1))
```

**Example 3.** For the GFT of the data in C1, which is stored in the columns C3 and C5, calculate the mean and the standard deviation.

```
Session commands:
    MTB > name k5 'GFTMean =' k6 'GFTStDev ='
    MTB > let k5=sum(c4*c5)/sum(fi)
    MTB > let k6=sqrt(sum((xi-k5)**2*fi)/(sum(fi)-1))
    MTB >
    MTB > print k1-k6
```

#### **Data Display**

RawMean =	68.3000
RawStDev =	16.7855
SFTMean =	68.3000
SFTStDev =	16.7855
GFTMean =	69.5000
GFTStDev =	17.2389

<u>Note</u>: The raw measures is identical to the SFT ones, where the GFT measures differ slightly but obvious.

#### **3. Descriptive Statistics:**

To give more than one statistic for one or more columns at a time we can use the command **Desc**, and to store them in a column for each measurement we can use the command **Stat**;

Session commands:

```
MTB > Desc 'Days'-'fi'.
```

#### Descriptive Statistics: Days, x, f, xi, fi

Variable	Ν	Mean	Median	TrMean	StDev	SE Mean
Days	40	68.30	66.50	68.33	16.79	2.65
x	35	68.54	68.00	68.65	16.74	2.83
f	35	1.1429	1.0000	1.0645	0.4300	0.0727
xi	7	70.00	70.00	70.00	21.60	8.16
fi	7	5.714	5.000	5.714	2.430	0.918
Variable	Minimum	Maximum	01	03		
Days	36.00	99.00	55.25	80.75		
x	36.00	99.00	56.00	81.00		
f	1.0000	3.0000	1.0000	1.0000		
xi	40.00	100.00	50.00	90.00		
fi	3.000	10.000	4.000	7.000		

<u>Note</u>: TrMean means Trimmed Mean where Minitab removes the smallest 5% and the largest 5% of the values (rounded to the nearest integer), and then averages the remaining values. The SE Mean will be explained later, in Lab 6.

Menu commands: Stat>Basic Statistics>Display Descriptive Statistics...

<u>Note</u>: If you want to use the command **Stat** you should select columns of equal length ONLY and it is much easier to use the session commands to execute the **Stat** command than the menu ones;

Menu commands: Stat>Basic Statistics>Store Descriptive Statistics...

Store De	scriptive Statistics		- 🔀
C1 C2 C3 C4 C5	Days x f xi fi	<b>/ariables:</b> Days-fi	
	MINITAB Invalid var Column ler		
	Use 1 or m	OK	gth
а н	elp	Statistics OK	Options
7	Store Descriptive	Statistics	
Minimum Store I	Descriptive Statistics	- Statistics	
I I M I I SE I I St I I Va	ean E of mean andard deviation iriance	<ul> <li>✓ Sum</li> <li>✓ Minimum</li> <li>✓ Maximum</li> <li>✓ Range</li> </ul>	N nonmissing N missing N total Cumulative N Percent
	rst quartile edian hird quartile terquartile range	☐ Sum of squares ☐ Skewness ☐ Kurtosis ☐ MSSD	☐ Cumulative percent
	Help		OK Cancel
	Help	ОК	Cancel

## LAB 5 PERCENTILES

## efinitions: *Minimum (min.):* Is the smallest sample observation.

Maximum (max.): Is the largest observation of the sample.

Range (R): Is the difference between max. and min.

 $P^{th}$  percentile: Is that value of the variable, below which p% of the sample lies.

*Quartiles:* 3 values (denoted by  $Q_1$ ,  $Q_2$  and  $Q_3$ ) within the data divides the data into 4 quarters. That is  $Q_1$  is the value that exceeds 25% of the data and  $Q_2$  is the value exceeds 50% of the data ... etc.

*Deciles:* 9 values (denoted by  $D_1, D_2 \dots D_9$ ) within the data divides the data into 10 tenths. That is  $D_1$  is the value that exceeds 10% of the data and  $D_2$  is the value exceeds 20% of the data ... etc.

*Percentiles:* 99 values (denoted by  $P_1$ ,  $P_2$  ...  $P_{99}$ ) within the data divides the data into 100 hundredths. That is  $P_1$  is the value that exceeds 1% of the data and  $P_{34}$  is the value exceeds 34% of the data ... etc.

Note that there is a one-to-one correspondence among the quartiles, the deciles and the percentiles;

 $Q_1=P_{25}$ ,  $Q_3=P_{75}$ ,  $D_1=P_{10}$ ,  $D_9=P_{90}$ ,  $Q_2=D_5=P_{50}=The Median$ .

Assume that the sample has observations  $x_1, x_2, ..., x_n$ . The increasingly (ascending) ordered sample is denoted by  $x_{(1)}, x_{(2)}, ..., x_{(n)}$ . The P<sup>th</sup> percentile is denoted by P<sub>p</sub> where the subscript p denotes the *number* of the percentile.

#### **Procedure**:

- 1. <u>Sort</u>: the sample in ascending order.
- 2. <u>Locate</u>: the p<sup>th</sup> percentile that is the *rank* of the percentile using the equation  $q = \frac{p}{100}(n+1)$ , where n > 10, n is the sample size and m is

the greatest integer less than or equal to q (i.e.,  $m \le q \le m+1$ ).

3. <u>Calculate</u>: the *value* of the p<sup>th</sup> percentile can be obtained by interpolating between the values  $x_{(m)}$  and  $x_{(m+1)}$  according to the formula:  $P_p = x_{(m)} + (q-m) [x_{(m+1)} - x_{(m)}]$ . If p = 25, 50 or 75 then the 25<sup>th</sup>, 50<sup>th</sup> or 75<sup>th</sup> percentiles are called

If p = 25, 50 or 75 then the  $25^{th}$ ,  $50^{th}$  or  $75^{th}$  percentiles are called the first, the second and the third quartiles, denoted by  $Q_1$ ,  $Q_2$ , and  $Q_3$  respectively.

**Open** the Minitab project *Lab5.mpj* from your floppy disk.

Sorting Data We can sort the data in C1 increasingly in C2 as follows: Session commands: MTB > sort c1 c2

MTB > print c2

#### **Data Display**

Sorted							
55	57	58	60	60	61	61	61
62	62	62	64	65	67	67	

Menu commands: Manip > Sort ...

Sort		
C1 Yield	Sort column(s):	
C2 Sorted	Yield	
		~
	Store sorted column(s) in:	
	Sorted	
	Sort by column: Yield	Descending
	Sort by column:	🗖 Descending
	Sort by column:	🗖 Descending
	Sort by column:	🗖 Descending
Help	ок	Cancel

**Example 1.** To compute the 35<sup>th</sup> percentile of the data Yield which are stored in C1 and sorted in C2, using Minitab, use only Session commands as follows:

```
MTB > sort c1 c2
MTB > \# Calculation of the 35<sup>th</sup> percentile
MTB > name k1 'q35 =' k2 'P35 ='
```

```
MTB > let k1=35/100*(15+1)
MTB > prin k1
```

#### **Data Display**

q35 = 5.60000

then (the rank) q = 5.6, m = 5 and m+1 = 6MTB > let k2=c2(5)+(k1-5)\*(c2(6)-c2(5))MTB > print k2

#### **Data Display**

P35 = 60.6000

then (the value) P35 = 60.6

**Example 2.** To compute the  $7^{th}$  decile =  $70^{th}$  percentile of the previous data, do as follows:

```
MTB > # Calculating the 70th percentile
MTB > name k3 'q70 =' k4 'P70 ='
MTB > let k3=70/100*16
MTB > print k3
```

#### **Data Display**

q70 = 11.2000 MTB > let k4=c2(11)+(k3-11)\*(c2(12)-c2(11)) MTB > prin k4

#### **Data Display**

P70 = 62.4000

 $Q_1$ ,  $Q_2$ ,  $Q_3$ , min and max are the main measures of the Five-number summary of the data that can be obtained from the command **Desc**: *Session commands*:

MTB > desc cl

#### **Descriptive Statistics: Yield**

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Yield	15	61.467	61.000	61.538	3.378	0.872
Variable Yield	Minimum 55.000	Maximum 67.000	Q1 60.000	Q3 64.000		

## UNIT III PROBABILITY DISTRIBUIONS

## LAB 6

## **DISCRETE AND BINOMIAL DISTRIBUTIONS**

et X be a discrete random variable assumes values  $x_1, x_2, ..., x_n$  with

probabilities  $p_1, p_2, ..., p_n$  respectively. Any table, formula or graph gives the different values of the r.v. with their corresponding probabilities is called a <u>probability distribution</u>.

#### **Probability Plot**

To construct a probability plot we plot the <u>probabilities of X versus</u> the corresponding values using the command **Plot**.

Open the Minitab project Lab6.mpj from your floppy disk.

**Example 1.** Construct a probability plot for the probability distribution of the r.v. X stored in C1 and C2.

Session commands:

```
MTB > # Probability plot of discrete r.v.
MTB > gstd
* NOTE * Character graphs are obsolete.
* NOTE * Standard Graphics are enabled.
        Professional Graphics are disabled.
        Use the GPRO command to enable
        Professional Graphics.
MTB > plot c2 c1
```



*Menu commands*: **Graph>Character Graph>Scatter Plot...** and select the option **Scale...** and fill in the *Minimum position* of *X Axis* by -0.5.

Scatter Plot		
C1 x C2 P(x)	Y variable: <sup>1</sup> P (x) <sup>1</sup> X variable: x	_
	<ul> <li>Use symbol: *</li> <li>Use labels:</li> </ul>	<u> </u>
	Annotate	. Scale
Help	ОК	Cancel

	icatter P	Plot Scale			
	C1	Y Axis			
2	C2	Minimum position:			
		Maximum position:			
2		Tick increment:			
3		X Axis			
1		Minimum position:	5		
		Maximum position:			
2		Tick increment:			
2					e
	He	Неір	ОК	Cancel	cel

The previous probability plot can be changed into a probability histogram by erecting adjacent bars one over each value of the r.v. with a height equal to its probability as follows;



Also the probability plot can be changed into a probability polygon simply by connecting the asterisks with line segments as follows;



#### **Discrete Random Samples**

Minitab has the command **Random** that will simulate observations of a specified random variable. In particular, by employing the **Discrete** subcommand of the Random command, we can generate random observations from any (finite-valued) discrete random variable.

**Example 2.** For the previous example, generate a r.s. of size 1000 observations from the random variable X, store it in C3 and check weather the generated sample is representative for the r.v. X.

Session commands:

```
MTB > base 0
MTB > rand 1000 c3;
SUBC> disc c1 c2.
MTB > tall c3
```

#### Tally for Discrete Variables: Discr

<u>Note</u>: constructing an SFT for the sample is a useful tool to check, roughly, the goodness of fit of the sample to the probability distribution. This can be achieved by simply comparing the relative frequencies in the sample by the probabilities in the distribution.

Discrete Distribution			×
$\begin{array}{ccc} C1 & \mathbf{x} \\ C2 & P(\mathbf{x}) \\ \end{array}$	<u>G</u> enerate 1000	rows of data	
C3 Discr C4 P(X=x) C5 P(X=x)	<u>S</u> tore in column(s	):	
C6 X=x C7 B(5,.2)	Discr		×
	,		_
	<u>V</u> alues in:	c1	1
1	<u>P</u> robabilities in:	c2	ĺ
Select			
Help		<u>0</u> K	Cancel

Menu commands: Calc>Random Data>Discrete...

<u>Note</u>: by comparing the relative frequencies in the SFT by the probabilities in the distribution it seems that the generated r.s. is a representative (good) one.



#### The Binomial Distribution

The random variable X is said to have a binomial distribution with parameters n, p, denoted by X:B(n, p), if its probability law is given by

$$P(x) = f(x) = P(X = x) = {\binom{n}{x}} p^{x} (1-p)^{n-x} , x = 0, 1, 2, ..., n$$

where the integer n denotes the number of trials and the fraction p denotes the probability of success, are both positive.

#### Calculating the Probabilities

To calculate the Bionmial probabilities using Minitab, we use the command **PDF** with the subcommand **Binomial**.

**Example 1.** For a binomial distribution with n = 5 and p=0.2 calculate p(x), where  $x=0,1,2,\ldots,5$ .

Session commands:

```
MTB > pdf c1 c4;
SUBC> bino 5 .2.
```

#### Menu commands: Calc>Probability Distribution>Binomial...

Binomial Distribution	x
C1 x C2 P(x) C3 Discr C4 P(X=x) C5 P(X<=x) C6 X=x C7 B(5,.2)	<ul> <li>Probability</li> <li>Cumulative probability</li> <li>Inverse cumulative probability</li> </ul>
	Number of trials:     5       Probability of success:     .2
	© Input column: x Optional storage: c4
Select	C Input co <u>n</u> stant: Optional sto <u>r</u> age:
Help	<u>O</u> K Cancel

Select: **Probability** 

#### Calculating the Areas

To calculate the areas under the Binomial curve to the left of a point x using Minitab, we use the command **CDF** with the subcommand **Binomial**.

**Example 2.** From the last example calculate  $F(x)=P(X \le x)$ , the area to the left (under) x, where x=0,1,2,...,5. Session commands: MTB > cdf c1 c5;

SUBC> bino 5 .2.

#### Menu commands: Calc>Probability Distribution>Binomial... Select: Cumulative probability

Binomial Distribution		×
C1 x C2 P(x) C3 Disor C4 P(X=x) C5 P(X<=x) C6 X=x C7 B(5,.2)	<ul> <li>Probability</li> <li><u>C</u>umulative probability</li> <li><u>Inverse cumulative probability</u></li> </ul>	
	Number of trials: 5	
	Pro <u>b</u> ability of success: . 2	
	© Input column: x	
	Optional storage: c5	
1	C Input co <u>n</u> stant:	
Select	Optional sto <u>r</u> age:	
Help	<u>O</u> K Cancel	

Finding Points of Given Probabilities

To find value of the Binomial r.v. with given probability to its left using Minitab, we use the command **InvCDF** with the subcommand **Binomial**.

**Example 3.** From the last example find the values of X that lie areas in column c5 to its left.

```
Session commands:

MTB > invcdf c5 c6;

SUBC> bino 5 .2.
```

```
Menu commands: Calc>Probability Distribution>Binomial...
Select: Inverse cumulative probability
```

Binomial Distribution	x
C1 x C2 P(x) C3 Discr C4 P(X=x) C5 P(X<=x) C6 X=x C7 B(5,.2)	<ul> <li>Probability</li> <li>Cumulative probability</li> <li>Inverse cumulative probability</li> <li>Number of trials: 5</li> <li>Probability of success: .2</li> </ul>
Select	<ul> <li>Input column: 'P(X&lt;=x)'</li> <li>Optional storage: 'X=x'</li> <li>C Input constant: Optional storage:</li> </ul>
Help	<u>O</u> K Cancel

## Generating a Binomial r.s.

**Example 4.** Generate 1000 random observations from the random variable X, where X:B(5,2), store it in c7 and check the goodness of fit of the sample to the distribution.

Session commands:

```
MTB > base 0
MTB > rand 1000 c7;
SUBC> bino 5 .2.
MTB > tall c7
```

#### Tally for Discrete Variables: B(5,.2)

B(5,.2)	Count
0	325
1	408
2	205
3	55
4	7
N=	1000

MTB > print c1-c2 c4-c6

#### **Data Display**

Row	х	P(x)	P(X=x)	P(X<=x)	X=x
1	0	0.004	0.32768	0.32768	0
2	1	0.219	0.40960	0.73728	1
3	2	0.442	0.20480	0.94208	2
4	3	0.277	0.05120	0.99328	3
5	4	0.050	0.00640	0.99968	4
б	5	0.008	0.00032	1.00000	5

## Menu commands: Calc>Random Data>Binomial...

Binomial Distribution		×
$\begin{bmatrix} C1 & \mathbf{x} \\ C2 & P(\mathbf{x}) \end{bmatrix}$	Generate 1000	rows of data
C3 Discr C4 P(X=x) C5 P(X<=x) C6 X=x C7 B(5,.2)	<u>Store in column(s):</u> 'B(5,.2)'	<u> </u>
	Number of trials:	
	Probability of success:	.2
Select		
Help		<u>O</u> K Cancel

## LAB 7 NORMAL DISTRIBUTION

In the world around us, we observe a variety of populations and random variables. Many are intrinsically different. But some such as aptitude test scores, heights of women and yield weight share an important characteristic: the probabilities associated with them are equal, at least approximately, to areas under a normal curve, that is, a bellshaped curve like the one shown below.



The normal distribution has two parameters,  $\mu$  and  $\sigma$ , for the mean and standard deviation of the normal distribution, denoted by X:N( $\mu$ ,  $\sigma^2$ ). The normal curve has the equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^3}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

Basic properties of the normal curve;

- 1. The curve is symmetric about the vertical line  $x=\mu$ .
- 2. The area under the curve is unity.
- 3. Small values of  $\sigma$  make the curve peaked, where large values make it flat. The curve approaches the horizontal axis as x approaches  $\pm \infty$ .
- 4. Most of the area under the curve lies between  $\mu$ -3 $\sigma$ ,  $\mu$ +3 $\sigma$ .

**Open** the Minitab project *Lab7.mpj* from your floppy disk.

Calculating Densities Under the Normal Curve

The densities under the normal curve can be computed using the following Minitab command **PDF**.

**Example 1.** To calculate the density under a normal curve with  $\mu=5$ ,  $\sigma=3$  at each of the points in C1.

Session commands:

```
MTB > # Calculation of f(x).
MTB > pdf c1 c2;
SUBC> norm 5 3.
```

Menu commands: Calc>Probability Distribution>Normal... Select: Probability density

Normal Distribution		x
C1 x C2 f(x) C3 P(X<=x) C4 X=x C5 Test C6 Nscores C7 Sample C8 Nscore	<ul> <li>Probability density</li> <li>Cumulative probability</li> <li>Inverse cumulative probability</li> <li>Mean: 5</li> <li>Standard deviation: 3</li> </ul>	
Select	<ul> <li>Input column: x</li> <li>Optional storage: 'f (x)'</li> <li>Input constant: Optional storage:</li> </ul>	
Help	<u>O</u> K Canc	el

Calculating Areas Under the Normal Curve

The areas under the normal curve can be computed using the following Minitab command **CDF**.

**Example 2.** To calculate the area under a normal curve with  $\mu=5$ ,  $\sigma=3$  that lies to the left of each of the points in C1.

Session command:

```
MTB > # Calculation of F(x).
MTB > cdf cl c3;
```

SUBC> norm 5 3.

#### Menu commands: Calc>Probability Distributions>Noraml... Select: Cumulative probability

Normal Distribution		x
C1 x C2 f(x) C3 P(X<=x) C4 X=x C5 Test C6 Nscores C7 Sample C8 Nscore	<ul> <li>Probability density</li> <li><u>C</u>umulative probability</li> <li><u>Inverse cumulative</u></li> <li><u>Mean:</u></li> <li><u>S</u>tandard deviation:</li> </ul>	y bility e probability 3
Select	<ul> <li>Input column:</li> <li>Optional storage:</li> <li>Input constant:</li> <li>Optional storage:</li> </ul>	x 'P(X<=x)'
Help		<u>O</u> K Cancel

Finding Points Under the Normal Curve

The values of the normal r.v. can be found, given the areas under the normal curve and to the left of each, using the following Minitab command **InvCDF**.

**Example 3.** Determine the x-values where X has a normal distribution with  $\mu$ =5 and  $\sigma$ =3, that has areas in C3 to its left.

Session commands:

```
MTB > InvCDF c3 c4;
SUBC> Noraml 5 3.
MTB >
MTB > print c1-c4
```

**Data Display** 

Row	x	f(x)	P(X<=x)	X=X
1	-4.0	0.001477	0.001350	-4.0000
2	-1.0	0.017997	0.022750	-1.0000
3	0.5	0.043173	0.066807	0.5000
4	2.0	0.080657	0.158655	2.0000
5	5.0	0.132981	0.500000	5.0000
б	8.0	0.080657	0.841345	8.0000
7	9.5	0.043173	0.933193	9.5000
8	11.0	0.017997	0.977250	11.0000
9	14.0	0.001477	0.998650	14.0000

#### Menu commands: Calc>Probability Distribution>Normal... Select: Inverse cumulative probability

Normal Distribution		×
C1 x C2 f(x) C3 P(X<=x) C4 X=x C5 Test C6 Nscores C7 Sample C8 Nscore	<ul> <li><u>P</u>robability density</li> <li><u>C</u>umulative probab</li> <li><u>Inverse cumulative</u></li> <li><u>M</u>ean: <u>5</u></li> <li><u>S</u>tandard deviation:</li> </ul>	, ility e probability 3
	<ul> <li>Input column:</li> <li>Optional storage:</li> <li>Input constant:</li> </ul>	'P(X<=x)'
Select Help	Optional sto <u>r</u> age:	<u>O</u> K Cancel

Generating a Normal r.s.

**Example 4.** Generate a sample of size 2000 from the r.v. X:N(5,9), store it in C7 and check the normality assumption of the sample.

Session commands:

```
MTB > base 0
MTB > rand 1000 c7;
SUBC> norm 5 3.
```

Menu commands: Calc>Random Data>Normal...

Normal Distribution	x
C1 x C2 f(x) C3 P(X<=x) C4 X=x C5 Test C6 Nscores C7 Sample C8 Nscore	Generate 2000 rows of data Store in column(s): Sample
	Mean: 5 Standard deviation: 3
Select	
Help	<u>O</u> K Cancel

But to check the normality assumption is the question of the next section and will be revisited after introducing that section.

#### Normal Probability Plot (NPP)

Normal Probability Plot can be used, roughly, as an aid for deciding whether a sample is drawn from a population whose distribution is (approximately) normal. Minitab has a command called **Nscores** that can be used to determine the normal scores for a data set. We can employ the Minitab's **Plot** command to obtain a Normal Probability Plot.

Procedure for Character NPP:

- 1. Calculate the <u>normal scores</u> (ZScores) of the sample.
- 2. <u>Plot</u> the NScores versus the observations of the sample.

**Example 5.** A sample of 15 cars yields the following data on miles driven 10.2 9.2 11.8 10.3 13.7 6.6 8.9 7.7 8.7 12.7 3.3 5.7 8.3 10.6 12.0. Do the data imply that the miles driven by <u>all</u> the cars (population) are normally distributed?

This question can be answered by constructing an NPP for the sample stored in C5 using the following commands;

```
Use the GPRO command to enable
Professional Graphics.
MTB > plot c6 c5
```





<u>Comment</u>: Since the asterisks formed a straight line approximately, we can conclude that the sample is drawn from a normal population. *Menu commands*: **Calc>Calculator...** 

Calculator	x
$\begin{array}{ccc} C1 & \mathbf{x} \\ C2 & f(\mathbf{x}) \\ C3 & P(X \le \mathbf{x}) \\ C4 & X = \mathbf{x} \\ C5 & T = \mathbf{x} \end{array}$	Store result in variable: Nscores
C5 Test C6 Nscores C7 Sample C8 Nscore	
	<u>F</u> unctions:
	7 8 9 + = ↔ All functions ▼
	4 5 6 < > Minimum (rows) A Missing data code
	1 2 3 * <= >= Natural log Normal scores
	0 . / And Now N missing
1	**     Or     N missing (rows)       N nonmissing     •
Select	() Not Select
	NSCOR(number)
Help	<u>O</u> K Cancel

Scatter Plot	×
$ \begin{array}{c ccc} C1 & x \\ C2 & f(x) \\ C3 & P(X < x) \\ C4 & X = x \\ C5 & Test \\ C6 & Nscores \\ C7 & Sample \\ C8 & Nscore \\ \end{array} $	Y variable: Nscores X variable: Test Use symbol: * O Use labels:
Select	<u>A</u> nnotate <u>O</u> K Cancel

Menu commands: Graph>Character Graph>Scatter Plot...

There is another effective and numerical method in Minitab which can be used to draw Normal Probability Plot for a given set of data; *Menu commands*: **Graph>Probability Plot...** 

Select: Distribution (Normal)



Normal Probability Plot for Test

There is also another method in Minitab to check the normality assumption of a given data set.

Menu Commands: Stat>Basic Statistics>Display Descriptive Statis...

```
Select Graph option and tick on
```

#### Histogram of data, with normal curve

Display Descriptive Statistics				
C1 x C5 Test	Variables:			
Display Descriptive Stat	tistics - Graphs			
<ul> <li>Histogram of data</li> <li>Histogram of data, with normal curve</li> <li>Dotplot of data</li> <li>Boxplot of data</li> </ul>				
Confidence level: 95.0				
Help	ОК	Cancel		
		Graphs		
Help	ОК	Cancel		

Histogram of Test, with Normal Curve



The last numerical and the multipurpose, uses more than one test statistic, method for checking the normality assumption is the Normality Test given as follows;

Menu commands: Stat>Basic Statistics>Normality Test...

Normality Test	
C1 x C5 Test C9 Try	Variable: Reference probabilities: Tests for Normality • Anderson-Darling • Ryan-Joiner (Similar to Shapiro-Wilk) • Kolmogorov-Smirnov Title:
Help	OK Cancel

Normal Probability Plot



<u>Rule</u>: If P-Value < 0.05 we would Reject that the sample is drawn from a normal population with a confidence level of 95%.

<u>Comment</u>: Since the P-Value = 0.945 >> 0.05 then we conclude that the sample is drawn from a normal population.

Now return back to the generated sample in Example 4 which was named *Sample*, then the normality test gives;



#### Normal Probability Plot

<u>Comment</u>: It is obvious that Sample was drawn from a normal population.

Now as an exercise we tested the normality assumption of the sample in Try and we concluded that;

<u>Comment</u>: Since P-Value = < 0.05 the sample Try was NOT drawn from a normal population.

<u>Remember</u>: in the home work the only method you can use to test the normality assumption is the Character NPP to be printed on the printer.



## Normal Probability Plot

63

## LAB 8 VERIFICATION OF THE CENTRAL LIMIT THEOREM

# **T** he Mean & Standard deviation of $\overline{X}$

Suppose a random sample of size n is to be drawn from a population with mean  $\mu$  and standard deviation  $\sigma$ . Then  $\mu_{\overline{X}} = \mu$ , i.e., for each sample, the mean of the random variable  $\overline{X}$  is equal to mean of the population. The standard deviation of  $\overline{X}$  is

 $\sigma_{\overline{X}} = \begin{cases} \sqrt{\frac{N-n}{N-1}} \frac{\sigma}{\sqrt{n}} & \text{if sampling without replacement} \\ \frac{\sigma}{\sqrt{n}} & \text{if sampling with replacement} \end{cases}$ 

where N is the population size. If the sample size (n) is small (n≤0.05N) relative to the population size then,  $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$  in both cases. In this lab we will deal, only, with infinite populations such as normal or binomial populations where the sampling as always assumed to be with replacement. So we will consider always that  $\mu_{\overline{X}} = \mu$  and  $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ . It is important to notice that drawing with replacement is equivalent to drawing from an infinite population and vice versa.

Open the Minitab project Lab8.mpj from your floppy disk.

**Example 1.** Given the following sample: 1, 10, 100 and 1000 then describe this sample. Session commands:

MTB > desc cl

## **Descriptive Statistics: C1**

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Cl	4	278	55	278	484	242
Variable Cl	Minimum 1	Maximum 1000	Q1 3	Q3 775		

Compare the theoretical result of  $\sigma_{-x}$  with SE Mean in the last printout!

#### **Central Limit Theorem (CLT)**

Suppose a random sample of size n is to be taken from a population with mean  $\mu$  and standard deviation  $\sigma$ . Then the random variable  $\overline{X}$  has approximately the distribution of a normal variable with mean  $\mu_{\overline{X}} = \mu$ 

and standard deviation  $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$  as n increases without limit.

The standardized version of  $\overline{X}$  is  $Z = \frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}}$  which has

approximately a standard normal distribution. Minitab can be employed to verify the CLT by drawing different large samples from some distribution, calculating their means and showing that the distribution of the population of these means are approximately normal. Practically speaking we mean by a large sample is that with a size  $\geq 30$ .

Procedure of verification:

- 1. Generate 50 random samples each of size 5, 15, 30 and 40 from a Binomial distribution with n=10 and p=0.1.
- 2. Calculate the mean of each sample and store the means in a separate column named Means.
- 3. Test the normality assumption of the means column.

Note that we chose B(10, 0.1) because it is not Normal, it is a discrete distribution and it is much skewed to the left. Also, note that we can not place each sample in a separate column because we will face a hard difficulty in step 2 above if so, but instead we will place the samples in separate rows which would make it much easier to achieve step 2.

**Example 2.** The following example verifies the CLT for B(10,0.1) by generating 50 samples each of size n=5.

Session commands:

```
MTB > base 0
MTB > rand 5 cl-c50;
SUBC> bino 10 .1.
MTB > rmean cl-c50 c51
MTB > GPro.
* NOTE * Professional Graphics are enabled.
        Standard Graphics are disabled.
        Use the GSTD command to enable
        Standard Graphics.
MTB > %NormPlot 'Means'.
```

```
Executing from file:
D:\Minitab13\MACROS\NormPlot.MAC
Macro is running ... please wait
```

#### **Normal Prob Plot: Means**

## Normal Probability Plot



<u>Comment</u>: it is obvious that we accept that the means population is Normally distributed.

Menu commands: Calc>Random Data>Binomial...

Binomial Distribution		
	Generate 5	rows of data
	Store in column(s):	
	c1-c50	
	Number of trials:	
	Probability of success:	.9
Help		OK Cancel

Here is the commands for calculating the row mean. Session commands: MTB > RMean C1-C50 C51 Menu commands: Calc>Row Statistics...

Row Statistics		
C39 C40 C41 C42 C43 C44 C45 C46 C47 C48 C49	Statistic Sum Mean Standard deviation Minimum Maximum Range Input variables:	<ul> <li>Median</li> <li>Sum of squares</li> <li>N total</li> <li>N nonmissing</li> <li>N missing</li> </ul>
C50 C51 Means K1 K2 K3 K4 K5	C1-C50 Store result in: c51	
Help	ОК	Cancel

Now repeat Example.2 three times for n=15, 30 and 40. State your conclusion about the three additional results.

**Example 3.** The following example verifies the CLT for B(10,0.1) by generating 50 samples each of size n=15.

#### Session commands:

```
MTB > base 0
MTB > rand 15 cl-c50;
SUBC> bino 10 .1.
MTB > rmean cl-c50 c51
MTB > GPro.
* NOTE * Professional Graphics are enabled.
        Standard Graphics are disabled.
        Use the GSTD command to enable
        Standard Graphics.
MTB > %NormPlot 'Means'.
Executing from file:
D:\Minitab13\MACROS\NormPlot.MAC
Macro is running ... please wait
```

#### **Normal Prob Plot: Means**



Normal Probability Plot

**Example 4.** The following example verifies the CLT for B(10,.1) by generating 50 samples each of size n=30.

```
Session commands:
MTB > base 0
MTB > rand 30 c1-c50;
SUBC> bino 10 .1.
MTB > rmean c1-c50 c51
MTB > GPro.
* NOTE * Professional Graphics are enabled.
Standard Graphics are disabled.
Use the GSTD command to enable
Standard Graphics.
MTB > %NormPlot 'Means'.
Executing from file:
D:\Minitabl3\MACROS\NormPlot.MAC
Macro is running ... please wait
```

#### **Normal Prob Plot: Means**



#### Normal Probability Plot

**Example 5.** The following example verifies the CLT for B(10,.1) by generating 50 samples each of size n=40.

#### Session commands:

```
MTB > base 0
MTB > rand 40 c1-c50;
SUBC> bino 10 .1.
MTB > rmean c1-c50 c51
MTB > GPro.
* NOTE * Professional Graphics are enabled.
        Standard Graphics are disabled.
        Use the GSTD command to enable
        Standard Graphics.
MTB > %NormPlot 'Means'.
Executing from file:
D:\Minitabl3\MACROS\NormPlot.MAC
Macro is running ... please wait
```

#### **Normal Prob Plot: Means**



## Normal Probability Plot

General comment: From all of the above we saw that the result of the normality test was getting better as the number of the samples getting larger which verifies clearly the CLT.

## UNIT IV ESTIMATION AND TESTING

## LAB 9

## **CONFIDENCE INTERVALS ABOUT THE MEAN**

Definitions

Parameters: Are the characteristics of a random variable (or population).

- **Estimation:** Is the approximating value of the population parameter using the sample elements.
- **Statistic:** Is a function based on the sample elements used to estimate the population parameter.

## **Point Estimation**

In most applications, the population under consideration undergoes the Normal Distribution. In addition, the most frequently used measure of central tendency is the population Mean  $\mu$ . Subsequently, one is interested in estimating the mean of the normal population for small samples or estimating the mean of any population for large samples.

**Theorem:** Let  $X_1, X_2, \ldots, X_n$  be a random sample from a normal population with mean  $\mu$  and variance  $\sigma^2$ (i.e.N ( $\mu$ ,  $\sigma^2$ )), then  $\overline{X}$  is an *unbiased*, that is E( $\overline{X}$ ) =  $\mu$ , estimator of  $\mu$ .

So,  $\overline{X}$  is called a *Point Estimate* of  $\mu$ , since it gives a single numerical value.

## **Interval Estimation (Confidence Intervals)**

One can compute a point estimate for a population parameter once a random sample is drawn from the population. However, in many applications it is preferable to obtain an interval of values that has a high probability of containing  $\mu$ . Such intervals are often called Confidence Intervals denoted by C.I.

In general, a C.I., of a <u>confidence level</u>  $(1-\alpha)100\%$ , about a population parameter  $\mu$  with a point estimate  $\hat{\theta}$ , that has a variance  $\sigma_{\hat{\mu}}$ , and a <u>confidence coefficient</u>  $z_{1-\alpha/2}$  is given by the form  $\hat{\mu} \pm Z_{1-\alpha/2} \sigma_{\hat{\mu}}$ . Consequently, if  $\hat{\mu} = \overline{X}$  then  $\sigma_{\hat{\mu}} = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ . Since there are four cases to be studied depending on the sample size, the distribution of the population and the status of  $\sigma$ , it is very useful to consider this flow chart as a brief figure for the procedure of finding the Confidence Interval;



## **1. Small sample C.I. for \mu of a Normal population** <u>Part I: $\sigma^2$ is known</u>

**Theorem:** If X<sub>1</sub>,X<sub>2</sub>,....,X<sub>n</sub> is a random sample from N( $\mu$ ,  $\sigma^2$ ), then a (1- $\alpha$ )100% C.I. for  $\mu$  is given by  $\overline{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ .

A C.I. can be computed in Minitab using the command **OneZ**.

Open the Minitab project Lab9.mpj from your floppy disk.

**Example 1.** Given the sample in C1, if the population is  $N(\mu, 0.0004)$ , find a 95% C.I. for the mean weight.

Here, n=25 (small sample),  $100(1-\alpha) = 95$  and  $\sigma$  is known so we must use 1-Sample Z to find the C.I.

Session commands:
```
MTB > let c2=nsco(c1)
MTB > %Qqplot 'Weights';
                                        It is optional in this case because it
                                        is mentioned in the problem that the
SUBC>
           Normal;
                                        population is Normal.
SUBC>
           MLE;
                                        But, in general, it is compulsory in
           Allpts;
                                        the case of small samples and
SUBC>
                                        unknown populations.
           Table;
SUBC>
           Conf 95;
SUBC>
           Ci.
SUBC>
Executing from file:
D:\Minitab13\MACROS\Qqplot.MAC
```

## **Prob Plot for Weights**



Session commands:

MTB > OneZ 'Weights'; SUBC> Sigma 0.02; SUBC> Confidence 95.

### **One-Sample Z: Weights**

The assur	ned s	igma =	0.02		
Variable	N	Mean	StDev	SE Mean	95.0% CI
Weights	25	0.99826	0.02228	0.00400	( 0.99042, 1.00610)

1-Sample Z (Test and Con	fidence Interval)
КЗ	Variables:
	Weights
	Sigma: 0.02
1-Sample	Z - Options
Confider	
Comuci	
Alternati	ve: not equal
Help	
2 He	lp OK Cancel

#### *Menu commands*: **Stat> Basic Statistics > 1- Sample Z...** Select : **Options...** and fill in **Confidence level:**

## <u>Part II: $\sigma^2$ is unknown</u>

**Theorem:** If  $X_1, X_2, ..., X_n$  is a random sample from  $N(\mu, \sigma^2)$  and  $\sigma^2$  is unknown, then a  $(1-\alpha)100\%$  C.I. for  $\mu$  is given by  $\overline{X} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}$ , Where *s* is the sample standard deviation.

A C.I. can be computed in Minitab, when  $\sigma^2$  is unknown, using the command **OneT**.

**Example 2.** For the previous example, suppose that  $\sigma^2$  were unknown. Find a 95% C.I. for the mean weight.

Here, n=25 (small sample),  $100(1-\alpha) = 95$  and  $\sigma$  is unknown, s will be used instead, so we must use 1-Sample t.

Session commands:

MTB > OneT 'Weights'.

## **One-Sample T: Weights**

Variable	N	Mean	StDev	SE Mean	95.0% CI
Weights	25	0.99826	0.02228	0.00446	( 0.98907, 1.00746)

<u>Note</u>: The TC.I. is slightly wider than the ZC.I. because:

- 1. The curve of the t-distribution is shorter but wider than that of the Normal distribution.
- 2. The population StDev. In TC.I. is unknown and substituted by s which is less accurate.

Menu commands: Stat> Basic Statistics> 1-Sample t...

Select: Options... and fill in Confidence level:

1-Sample t (Test and Cont	fidence Interval) 🛛 🔀
C1 Weights	Variables:
C3 Po(1)	Weights 🔗
1-Sample t	- Options 🛛 🛛
Confidenc	e level: 95.0
Alternative	: not equal
Help	OK Cancel
	Graphs Options
Help	OK Cancel

## 2. Large Sample C.I. for the mean of a general population

One can employ the CLT to get the following theorem, when the sample size  $\ge 30$ .

**Theorem:** If  $X_1, X_2, \ldots, X_n$  is a random sample from a population with mean  $\mu$  and variance  $\sigma^2$ , regardless of the distribution of the population, then an approximate  $(1-\alpha)100\%$  C.I. for  $\mu$  is given by  $\overline{X} \pm z_{1-\alpha/2} \frac{s}{\sqrt{n}}$ . Where n is assumed to be large.

A C.I. can be computed in Minitab using the command **OneZ**. **Example 3.** Compute a 90% C.I. for mean  $\lambda$  using the sample in C3. Here, n=100 (large sample),  $100(1-\alpha) = 90$  and  $\sigma$  is unknown so we must calculate s for the given sample before we find the C.I.

```
Session commands:
MTB > stan c3 k3
```

## Standard Deviation of Po(1)

```
Standard deviation of Po(1) = 0.95007
MTB > OneZ 'Po(1)';
SUBC> Sigma K3;
SUBC> Confidence 90.
```

## One-Sample Z: Po(1)

The	assumed	sigma	=	0.950066	

Variable	Ν	Mean	StDev	SE Mean		90.0% CI
Po(1)	100	0.9200	0.9501	0.0950	(	0.7637, 1.0763)

*Menu commands*: Stat> Basic Statistics > 1- Sample Z...

1-Sample Z (Test and Confidence Interval)						
K3	Variables:					
	Sigma: K3 1-Sample Z - Options					
	Confidence level: 90					
Help 2	Alternative: not equal  Help OK Cancel					

<u>Note</u>: If the confidence level is NOT mentioned in the problem, it can be assumed to be 95%.

<u>Note</u>: If the confidence level is 95%, which is the default value in the dialogue box of testing, you may NOT write it in the session commands.

# **LAB 10**

# **TESTING HYPOTHESES ABOUT THE MEAN**

# Definitions

**Statistical Hypothesis:** Is a statement about the value or values of the population parameter(s).

**Null Hypothesis H**<sub>o</sub>: Is that hypothesis that generally declares that no change from the current, or the last, situation exists.

Alternative Hypothesis  $H_1$ : Is that hypothesis indicating that a conjecture or new idea of an investigator is true.

- **Test Statistic:** Is a calculation from the sampled data, the value of which will be used to decide whether to accept or reject the null hypothesis, it is sometimes called the Test Ratio (TR) or the Test Function.
- **Critical Region:** Is the set of values of the test statistic that lead to the rejection of  $H_0$ , it is also known as the rejection region.

**Type I error:** Is made if  $H_0$  is rejected while it is actually true.

Significance Level ( $\alpha$ ): Is the probability that the test statistic will fall within the critical region when H<sub>0</sub> is true; that is, it is equal to the probability of type I error of the test.

**P-value** (**P**): Is the area of the region lies beyond the test function value.

## 1. Small Sample Hypothesis Testing About the Mean $\mu$ :

Two cases will be considered, the first when  $\sigma^2$  is known and the other when  $\sigma^2$  is unknown, under the assumption that the sample is drawn from a normal population.

# <u>Part I: $\sigma^2$ is known:</u>

To test  $H_0$ :  $\mu = \mu_0$  vs.

 $H_1: \mu > \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $Z_0 > z_{1-\alpha}$  (right-tailed test)

H<sub>1</sub> :  $\mu < \mu_0$  Reject H<sub>0</sub>, at level  $\alpha$ , if Z<sub>0</sub> < -z<sub>1- $\alpha$ </sub> (left-tailed test)

 $H_1: \mu \neq \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $|Z_0| > z_{1-\alpha/2}$  (two-tailed test)

Where  $Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$  is the value of the test function and  $z_{1-\alpha}$  is the

value of N(0,1) under which an area  $1-\alpha$  lies.

A simpler approach of testing hypotheses is using the P-value, following the rule; if  $P < \alpha$  then *Reject* H<sub>0</sub>, otherwise *Do not Reject* H<sub>0</sub>.



Again we refer to the flow chart given in the previous lab.

**Example 1.** For the sample of weights stored in C1, <u>do</u> the data provide sufficient evidence to conclude that the mean weights, <u>differs</u> from <u>1</u> lb.? At <u>5%</u> sig. level and the <u>population standard deviation</u> is <u>0.02</u>.

To answer this question is to test the following hypotheses  $H_0$ :  $\mu=1$  vs.  $H_1: \mu\neq 1$  where  $\alpha=0.05$  and  $\sigma=0.02$ , using Minitab;

Session command:

```
MTB > OneZ 'Weights';
SUBC> Sigma 0.02;
SUBC> Test 1;
SUBC> Confidence 95.
```

## **One-Sample Z: Weights**

Test of mu =	1 vs	mu not =	1	
The assumed	sigma	= 0.02		
Variable	N	Mean	StDev	SE Mean
Weights	25	0.99826	0.02228	0.00400

Variable		95.	0%	CI	Z	P
Weights	(	0.99042,	1.	00610)	-0.43	0.664

<u>Note</u>: To distinguish among the three cases of the alternative hypothesis, you may identify the following notation;

- 1. If  $H_1$  has < sign write **alte -1** in SUBC> on the session.
- 2. If  $H_1$  has  $\neq$  sign write **alte 0** in SUBC> on the session.
- 3. If  $H_1$  has > sign write **alte 1** in SUBC> on the session.

Since 0.66 > .05, our decision is NOT to Reject H<sub>0</sub>. Consequently, <u>the</u> <u>data provide sufficient evidence to conclude that</u> *the mean weight, for all of the one-pound containers,* does not *differ from 1 lb. at 5% sig. level.* 

#### Menu commands: Stat> Basic Statistics> 1-Sample Z... Fill in: Test mean: and Select Options... to fill in Alternative:

1-Sample Z (Test and Con	fidence Interval) 🛛 🔀
C1 Weights C2 Nscor C3 Po(1)	Variables: Weights Sigma: 0.02 Test mean: 1 (required for test) Graphs Options
Help	OK Cancel

1-Sample	Z (Test and Con	fidence Interval) 🛛 🔀
C1	Weights	Variables:
C2 C3	Nscor Po(1)	Weights
	1-Sample Z - O	ptions 🔀
	Confidence la	evel: 95
	Alternati∨e:	not equal
	Help	OK Cancel
		Graphs Options
Не	lp	OK Cancel

<u>Note</u>: If the significance level is NOT mentioned in the problem, it can be assumed to be 5%.

<u>Note</u>: If the significance level is 5%, which is the default value in the dialogue box of testing, you may NOT write it in the session commands. <u>Note</u>: If the alternative hypothesis is Not equal,  $\neq$ , you may NOT write it in the session commands.

**Example 2.** For the weights in C1, <u>test</u> that the mean weights <u>exceeds</u> 0.99 lb. Given that the true, or the <u>population</u>, standard deviation is <u>0.02</u>.

To answer this question is to test the following hypotheses  $H_0$ :  $\mu$ =0.99 vs.  $H_1$ :  $\mu$  > 0.99 where  $\alpha$ = 0.05, since it was not mentioned in the question, and  $\sigma$  = 0.02, using Minitab; *Session command*:

```
MTB > OneZ 'Weights';
SUBC> Sigma 0.02;
SUBC> Test 0.99;
SUBC> Confidence 95;
SUBC> Alternative 1.
```

## **One-Sample Z: Weights**

```
Test of mu = 0.99 vs mu > 0.99
The assumed sigma = 0.02
```

Variable	N	Mean	StDev	SE Mean
Weights	25	0.99826	0.02228	0.00400
Variable	95.0%	Lower Bound	Z	P
Weights		0.99168	2.07	0.019

Since 0.02 < .05, our decision is to Reject H<sub>0</sub>. Consequently, <u>the data</u> <u>provide sufficient evidence to conclude that</u> the mean weights, for all of the one-pound containers, exceeds 0.99 lb.

#### Menu commands: Stat> Basic Statistics> 1-Sample Z... Fill in: Test mean: and Select Options... to fill in Alternative:

<u>Part II :  $\sigma^2$  is unknown:</u>

To test  $H_0$  :  $\mu = \mu_0$  vs.

 $H_1: \mu > \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $T_0 > t_{1-\alpha,n-1}$  (right-tailed test)

 $H_1: \mu < \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $T_0 < -t_{1-\alpha, n-1}$  (left-tailed test)

 $H_1: \mu \neq \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $|T_0| > t_{1-\alpha/2,n-1}$  (two-tailed test)

Where  $T_0 = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$  is the value of TR and  $t_{1-\alpha,n-1}$  is the value of t-

distribution, with n-1 degrees of freedom, under which an area  $1-\alpha$  lies.

If  $\sigma$  is assumed to be unknown we use the t-test. From the last printout note that s=0.02228. We will perform t-test for the previous hypotheses using Minitab.

**Example 3.** For the weights in C1, Test the <u>claim</u> that the mean weights, for all of the one-pound containers, <u>differs</u> from <u>1</u> lb.

Session command:

```
MTB > OneT 'Weights';
SUBC> Test 1.
```

## **One-Sample T: Weights**

Test of m	u = 1 vs	mu not =	1	
Variable Weights	N 25	Mean 0.99826	StDev 0.02228	SE Mean 0.00446
Variable Weights	( 0.98	95.0% CI 907, 1.0074	16)	т Р 39 0.700

Menu commands: Stat>Basic Statistics>1-Sample t...

#### **2.** Large Samples Hypothesis Testing about $\mu$ ( $n \ge 30$ ):

For the testing of hypothesis about a population mean  $\mu$ , regardless of the population distribution, the Large Sample Z-test can be used to perform such a test.

To test  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $Z_0 > z_{1-\alpha}$  (right-tailed test)  $H_1 : \mu < \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $Z_0 < -z_{1-\alpha}$  (left-tailed test)  $H_1 : \mu \neq \mu_0$  Reject  $H_0$ , at level  $\alpha$ , if  $|Z_0| > z_{1-\alpha/2}$  (left-tailed test)

Where  $Z_0 = \frac{\overline{X} - \mu_0}{\sigma(s)/\sqrt{n}}$  is the value of the Test Statistic and  $z_{1-\alpha}$  is

the value of N(0,1) under which an area 1- $\alpha$  lies. If  $\sigma$  is unknown it can be replaced by s.

**Example 4.** For the sample in C3, test the <u>claim</u> that the mean  $\lambda$  is <u>less</u> than 2 at <u>10%</u> sig. level.

To answer this question is to test the following hypotheses  $H_0$ :  $\mu=2$  vs.  $H_1$ :  $\mu < 2$  where  $\alpha=0.10$ , and  $\sigma$  is unknown, so we have to calculate the sample StDev before testing;

Session commands:

MTB > stan c3 k3

#### Standard Deviation of Po(1)

```
Standard deviation of Po(1) = 0.95007
MTB > OneZ 'Po(1)';
SUBC> Sigma K3;
SUBC> Test 2;
SUBC> Confidence 90;
SUBC> Alternative -1.
```

#### **One-Sample Z: Po(1)**

```
Test of mu = 2 vs mu < 2
The assumed sigma = 0.950066
Variable
                                StDev
                                        SE Mean
                 Ν
                        Mean
Po(1)
               100
                      0.9200
                               0.9501
                                         0.0950
             90.0% Upper Bound
Variable
                                    Ζ
                                            Ρ
                        1.0418
                               -11.37 0.000
Po(1)
```

## Menu commands: Stat> Basic Statistics> 1-Sample Z... Fill in: Test mean: and Select Options... to fill in Alternative:

1-Sam	nple Z (Test and Con	fidence Interval)	×
C1 C2 C3	Weights Nscor Po(1)	Variables: 'Po(1)'	<ul> <li></li> </ul>
	1 Secola 7 Dette	Sigma: K3	
	Confidence level:	test)	
	Alternative:	ess than  Cancel	
9987	Help	OK Cancel	