

# : Chapter1

## Organizing and Displaying Data

### Introduction:

### Statistics:

Statistics is that area of study which is interested in learning how to collect, organize, and summarize information, and how to answer research questions and draw conclusions.

### Biostatistics:

If the information is obtained from biological and medical sciences, then we use the term biostatistics.

## **Populations:**

A population is the largest group of people or things in which we are interested at a particular time and about which we want to make some statements or conclusions.

## **Samples:**

From the population, we select various elements (or individuals) on which we collect our information. This part of the population on which we collect data is called the sample.

## **Sample Size:**

The number of elements in the sample is called the sample size and is denoted by  $n$ .

## **Variables:**

The characteristics to be measured on the elements of the population or sample are called variables.

## **Variables:**

The characteristics to be measured on the elements of the population or sample are called variables.

## **Example of variables:**

- Height
- no. of cars
- sex
- educational level

# Types of Variables:

## (1) Quantitative Variables:

The values of a quantitative variable are numbers indicating how much or how many of something.

### Examples:

- height
- family size
- age

## **(2) Qualitative Variables:**

The value of a qualitative variable are words or attributes indicating to which category an element of the population belong.

### **Examples:**

- blood type
- educational level
- nationality

## Types of Quantative Variables:

### Discrete Variables:

A discrete variable can have only countable number of values

### Examples:

1. **Family size** ( $x = 0, 1, 2, 3, \dots$  )
2. **Number of patients** ( $x = 0, 1, 2, 3, \dots$  )

## Continuous Variables:

A continuous variable can have any value within a certain interval of values.

**Examples:** - **height** ( $140 < x < 190$ )  
- **blood sugar level** ( $10 < x < 15$ )

**Variable**

**Quantitative**

**Qualitative**

**Discrete**

**Continuous**



## 1.2.Organizing The Data

### Ungrouped (or Simple) frequency distributions :

Used for:

- - qualitative variables
- - discrete quantitative variables with a few different values

### - Grouped frequency distributions :

Used for:

- - continuous quantitative variables
- - discrete quantitative variables with large number of different values

Example: (Simple frequency distribution or ungrouped frequency distribution).

The following data represent the number of children of 16 Saudi women:

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

- Variable =  $X$  = no. of children (discrete, quantitative)
- Sample size =  $n = 16$
- The possible values of the variable are: 0, 1, 2, 3, 4, 5

## Simple frequency distribution of the no. of children

no. of children (variable)	Frequency (no. of women)	Relative Freq. (R. F.) (=Freq /n)	Percentage Freq. (= R.F. * 100%)
0	1	0.0625	6.25%
1	2	0.125	12.5%
2	4	0.25	25%
3	5	0.3125	31.25%
4	2	0.125	12.5%
5	2	0.125	12.5%
<b>Total</b>	<b><i>n</i>=16</b>	<b>1.00</b>	<b>100%</b>

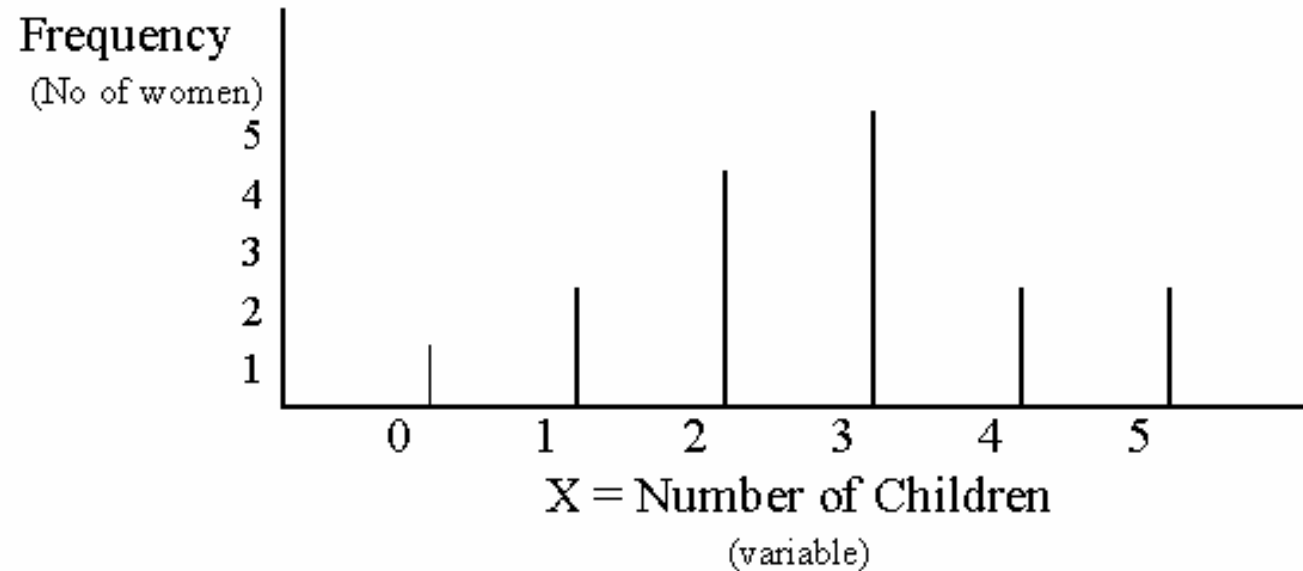
### Note

The sample size =  $n$  = Total of the frequencies

• Relative frequency = frequency/ $n$

Percentage frequency = Relative frequency \* 100%

- **Frequency bar chart** is a graphical representation for the simple frequency distribution.



Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

C3 =

	A	B	C	D	E	F	G	H	I	J
1	no. of children	Frequency								
2	0	1								
3	1	2								
4	2	4								
5	3	5								
6	4	2								
7	5	2								
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										

Sheet1 / Sheet2 / Sheet3

Ready NUM

Start | STAT-106 | Microsoft PowerPoint - ... | A1 - Microsoft Word | Microsoft Excel - B... | En | 10:55 ص

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

C3 =

	A	B	C	D	E	F	G	H	I	J
1	no. of children	Frequency								
2	0	1								
3	1	2								
4	2	4								
5	3	5								
6	4	2								
7	5	2								
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										

Chart Wizard - Step 1 of 4 - Chart Type

Standard Types Custom Types

Chart type:

- Column
- Bar
- Line
- Pie
- XY (Scatter)
- Area
- Doughnut
- Radar
- Surface
- Bubble
- Stock

Chart sub-type:

Clustered Column. Compares values across categories.

Press and Hold to View Sample

Cancel < Back Next > Finish

Sheet1 Sheet2 Sheet3

Ready NUM

Start STAT-106 Microsoft PowerPoint - ... A1 - Microsoft Word Microsoft Excel - B... En 10:56

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

B1 =

	A	B
1	no. of children	Frequency
2	0	1
3	1	2
4	2	4
5	3	5
6	4	2
7	5	2

**Source Data**

Data Range Series

Frequency

Data range:

Series in:  Rows  Columns

Cancel < Back Next > Finish

Sheet1 Sheet2 Sheet3

Point NUM

Start STAT-106 Microsoft PowerPoint - ... A1 - Microsoft Word Microsoft Excel - B... En 10:58

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

	A	B	C	D	E	F	G	H	I	J
1	no. of children	Frequency								
2	0	1								
3	1	2								
4	2									
5	3									
6	4									
7	5									

**Chart Wizard - Step 3 of 4 - Chart Options**

Titles   Axes   Gridlines   Legend   Data Labels   Data Table

Chart title: Frequency

Category (X) axis: No. Children

Value (Y) axis:

Second category (X) axis:

Second value (Y) axis:

Legend: Frequency

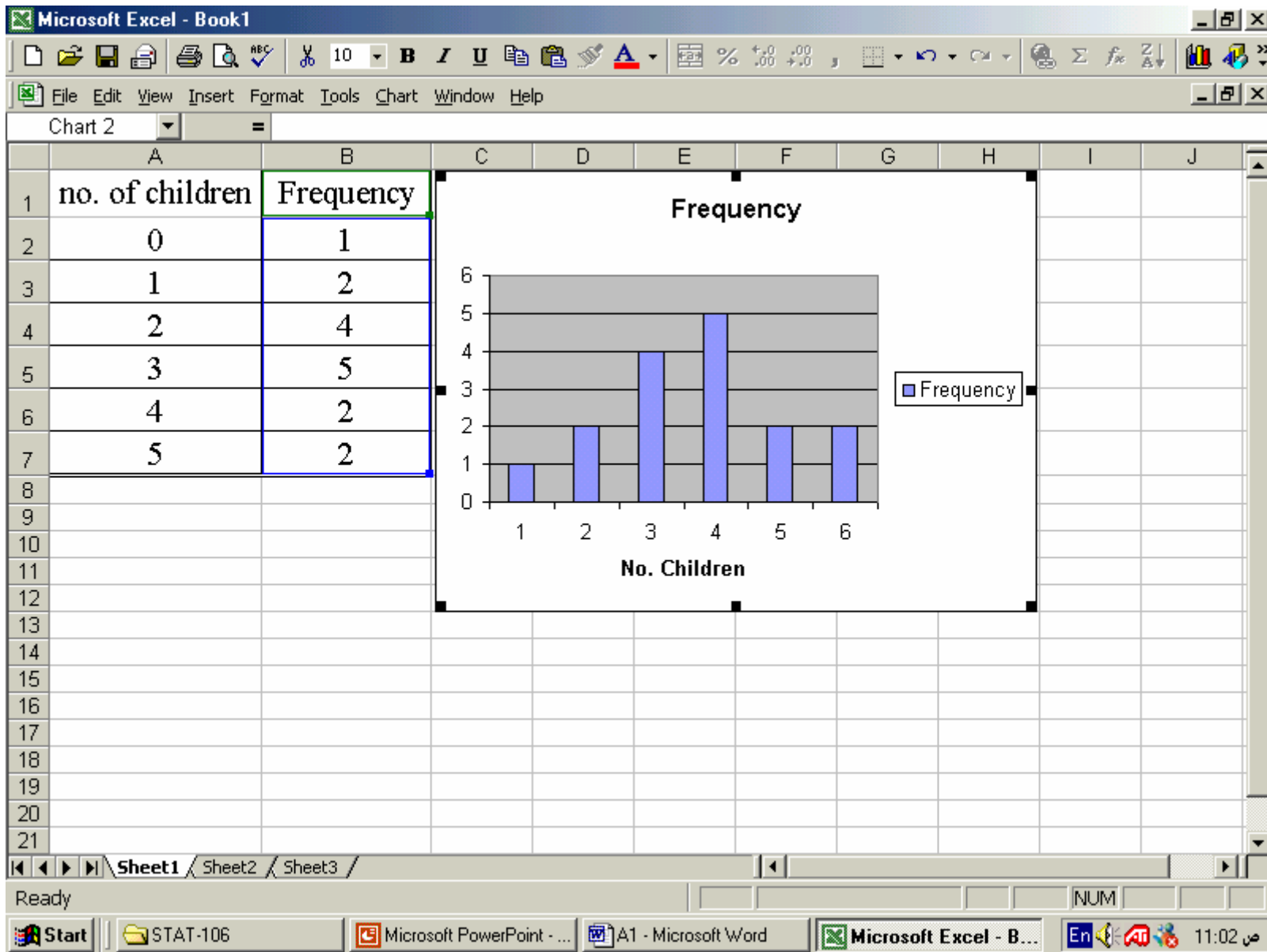
Buttons: ? Cancel < Back Next > Finish

Sheet1 / Sheet2 / Sheet3

Ready NUM

Start | STAT-106 | Microsoft PowerPoint - ... | A1 - Microsoft Word | Microsoft Excel - B... | En 11:01





## Example 1.2:

### (grouped frequency distribution)

The following table gives the hemoglobin level (g/dl) of a sample of 50 men.

17.0	17.7	15.9	15.2	16.2	17.1	15.7	17.3	<u>13.5</u>	16.3
14.4	15.8	15.3	16.4	13.7	16.2	16.4	16.1	17.0	15.9
14.0	16.2	16.4	14.9	17.8	16.1	15.5	<u>18.3</u>	15.8	16.7
15.9	15.3	13.9	16.8	15.9	16.3	17.4	15.0	17.5	16.1
14.2	16.1	15.7	15.1	17.4	16.5	14.4	16.3	17.3	15.8

- Variable =  $X$  = hemoglobin level (continuous, quantitative)
- Sample size =  $n = 50$
- Max = 18.3
- Min = 13.5

## Grouped frequency distribution for the hemoglobin level of the 50 men

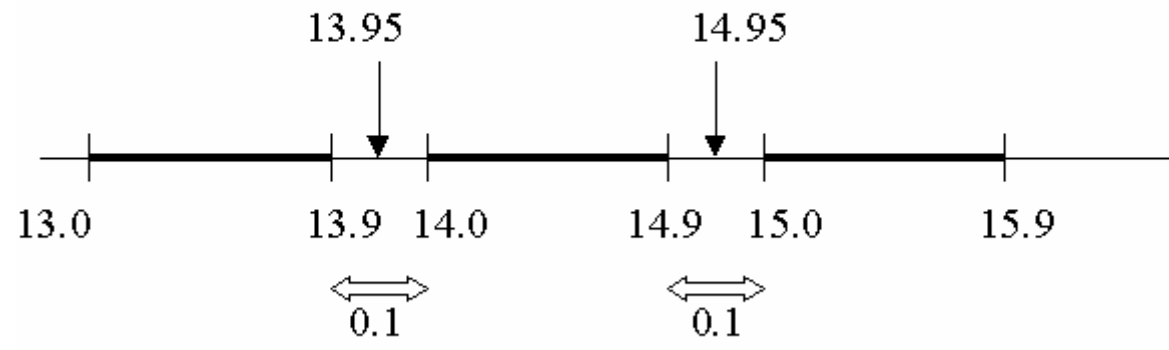
Class Interval (Hemoglobin level)	Frequency (no. of men)	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
13.0 - 13.9	3	0.06	3	0.06
14.0 - 14.9	5	0.10	8	0.16
15.0 - 15.9	15	0.30	23	0.46
16.0 - 16.9	16	0.32	39	0.78
17.0 - 17.9	10	0.20	49	0.98
18.0 - 18.9	1	0.02	50 = $n$	1.00
Total	$n=50$	1.00		

### Notes

- class interval = C. I.
- Cumulative frequency of a class interval = no. of values (frequency) obtained in that class interval or before.

Mid-Point (Class Mark) of C. I = 
$$\frac{\text{upper limit} + \text{lower limit}}{2}$$

○ True class intervals:



Class Interval	True C. I.	Class mid-point	frequency
13.0 - 13.9	12.95 - 13.95	$(13.0+13.9)/2 = 13.45$	3
<b>14.0 - 14.9</b>	<b>13.95 - 14.95</b>	$(14.9+14.9)/2 = 14.45$	5
15.0 - 15.9	14.95 - 15.95	15.45	15
16.0 - 16.9	15.95 - 16.95	16.45	16
17.0 - 17.9	16.95 - 17.95	17.45	10
18.0 - 18.9	17.95 - 18.95	18.45	1

↑    ↑                    ↑    ↑  
 lower    upper            True    True  
 limits    limits            lower    upper  
 (L.L.)    (U.L.)            limits    limits

$$a = \text{Jump} = 0.1$$

$$\text{True U. L.} = \text{U. L.} + (a/2) = \text{U. L.} + 0.05$$

$$\text{True L. L.} = \text{L. L.} - (a/2) = \text{L. L.} - 0.05$$

$$\text{Width of a class interval (W)} = \text{True U. L.} - \text{True L. L.}$$

$$\text{In the previous example, } W = 13.95 - 12.95 = 1.0$$

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

SUMIF X ✓ = =C4+1.0

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	17.0											
3	17.7											
4	15.9		13.0	13.9								
5	15.2		=C4+1.0									
6	16.2											
7	17.1											
8	15.7											
9	17.3											
10	13.5											
11	16.3											
12	15.8											
13	15.3											
14	16.4											
15	13.7											
16	16.2											
17	16.4											
18	16.1											

Sheet1 Sheet2 Sheet3

Enter NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:15 ص

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

C5 =C4+1

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	17.0											
3	17.7											
4	15.9		13.0	13.9								
5	15.2		14.0									
6	16.2											
7	17.1											
8	15.7											
9	17.3											
10	13.5											
11	16.3											
12	15.8											
13	15.3											
14	16.4											
15	13.7											
16	16.2											
17	16.4											
18	16.1											

Sheet1 Sheet2 Sheet3

Ready NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:17 ص

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

C5 = =C4+1

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	17.0											
3	17.7											
4	15.9		13.0	14								
5	15.2		14.0	14.9								
6	16.2		15.0	15.9								
7	17.1		16.0	16.9								
8	15.7		17.0	17.9								
9	17.3		18.0	18.9								
10	13.5											
11	16.3											
12	15.8											
13	15.3											
14	16.4											
15	13.7											

Sheet1 Sheet2 Sheet3

Ready Sum=164.5 NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:20



Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

F4

	A	B	C	D	E	F	G	H	I	J	K
1											
2	17.0										
3	17.7		L	U	FR	CUM FR					
4	15.9		13.0	13.9		=					
5	15.2		14.0	14.9							
6	16.2		15.0	15.9							
7	17.1		16.0	16.9							
8	15.7		17.0	17.9							
9	17.3		18.0	18.9							
10	13.5										
11	16.3										
12	15.8										
13	15.3										
14	16.4										
15	13.7										

**Paste Function**

Function category: Most Recently Used, All, Financial, Date & Time, Math & Trig, **Statistical**, Lookup & Reference, Database, Text, Logical, Information

Function name: COVAR, CRITBINOM, DEVSQ, EXPONDIST, FDIST, FINV, FISHER, FISHERINV, FORECAST, **FREQUENCY**, FTEST

**FREQUENCY(data\_array;bins\_array)**

Calculates how often values occur within a range of values and then returns a vertical array of numbers having one more element than Bins\_array.

OK Cancel

Sheet1 Sheet2 Sheet3

Edit NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:35

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Formula bar:  $=\text{FREQUENCY}(\$A\$2:\$A\$51;D4)$

	A	B	C	D	E	F	G	H	I	J	K
1											
2	17.0										
3	17.7		L	U	FR	CUM FR					
4	15.9		13.0	13.9		\$A\$51;D4)					
5	15.2		14.0	14.9							
6	16.2		15.0	15.9							
7	17.1		16.0	16.9							
8	15.7		17.0	17.9							
9	17.3		18.0	18.9							
10	13.5										
11	16.3										
12	15.8										
13	15.3										
14	16.4										
15	13.7										

FREQUENCY

Data\_array:  $\$A\$2:\$A\$51$  = {17;17.7;15.9;15.2;

Bins\_array:  $D4$  = 13.9

= {3;47}

Calculates how often values occur within a range of values and then returns a vertical array of numbers having one more element than Bins\_array.

Bins\_array is an array of or reference to intervals into which you want to group the values in data\_array.

Formula result = 3

OK Cancel

Sheet1 Sheet2 Sheet3

Point NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:39

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

F4 = =FREQUENCY(\$A\$2:\$A\$51;D4)

	A	B	C	D	E	F	G	H	I	J	K
1											
2	17.0										
3	17.7		L	U	FR	CUM FR					
4	15.9		13.0	13.9		3					
5	15.2		14.0	14.9							
6	16.2		15.0	15.9							
7	17.1		16.0	16.9							
8	15.7		17.0	17.9							
9	17.3		18.0	18.9							
10	13.5										
11	16.3										
12	15.8										
13	15.3										
14	16.4										
15	13.7										

Sheet1 Sheet2 Sheet3

Select destination and press ENTER or choose Paste

Start | STAT-106 | Microsoft Power... | A1 - Microsoft W... | Microsoft Exc... | MINITAB - Untit... | 11:39 ص

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window fx Σ Help

F5 = =FREQUENCY(\$A\$2:\$A\$51;[Paste]

	A	B	C	D	E	F	G	H	I	J	K
1											
2	17.0										
3	17.7		L	U	FR	CUM FR					
4	15.9		13.0	13.9		3					
5	15.2		14.0	14.9		8					
6	16.2		15.0	15.9		23					
7	17.1		16.0	16.9		39					
8	15.7		17.0	17.9		49					
9	17.3		18.0	18.9		50					
10	13.5										
11	16.3										
12	15.8										
13	15.3										
14	16.4										
15	13.7										

Sheet1 Sheet2 Sheet3

Select destination and press ENTER or choose Paste

Sum=169 NUM

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untit... 11:40

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window f Σ Help

10 B I U % .00 +.00

FREQUENCY X ✓ =F5-F4

	A	B	C	D	E	F	G	H
1								
2	17.0							
3	17.7	L	U	FR	CUM FR			
4	15.9	13.0	13.9	3	3			
5	15.2	14.0	14.9	=F5-F4	8			
6	16.2	15.0	15.9		23			
7	17.1	16.0	16.9		39			
8	15.7	17.0	17.9		49			
9	17.3	18.0	18.9		50			
10	13.5							
11	16.3							
12	15.8							
13	15.3							
14	16.4							
15	13.7							

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window f Σ Help

20 B I U % .00 +.00

E9 =F9-F8

	A	B	C	D	E	F	G	H	I	J	K
1											
2	17.0										
3	17.7	L	U	FR	CUM FR						
4	15.9	13.0	13.9	3	3						
5	15.2	14.0	14.9	5	8						
6	16.2	15.0	15.9	15	23						
7	17.1	16.0	16.9	16	39						
8	15.7	17.0	17.9	10	49						
9	17.3	18.0	18.9	1	50						
10	13.5										
11	16.3										
12	15.8										
13	15.3										
14	16.4										
15	13.7										

Sheet1 Sheet2 Sheet3

Ready NUM

Sheet1 Sheet2 Sheet3

Point

Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untitl... 11:42

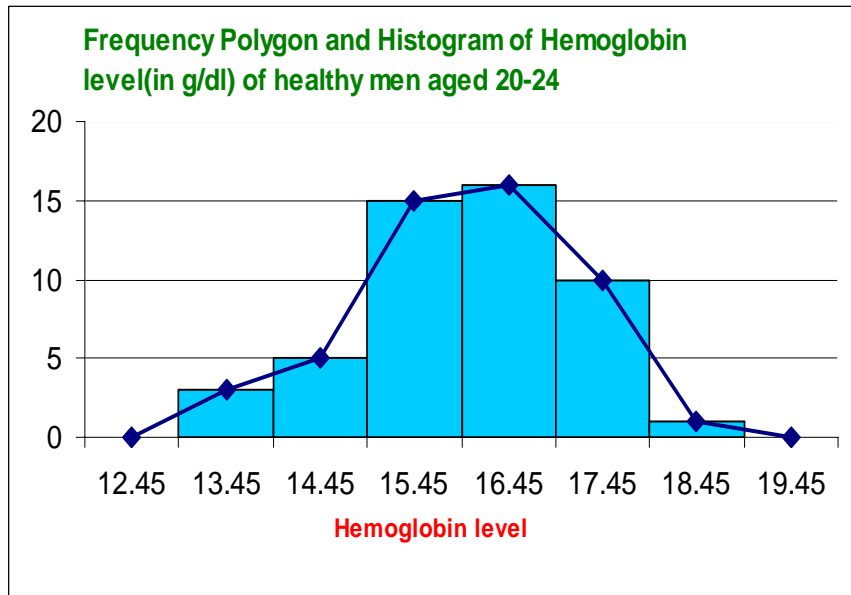
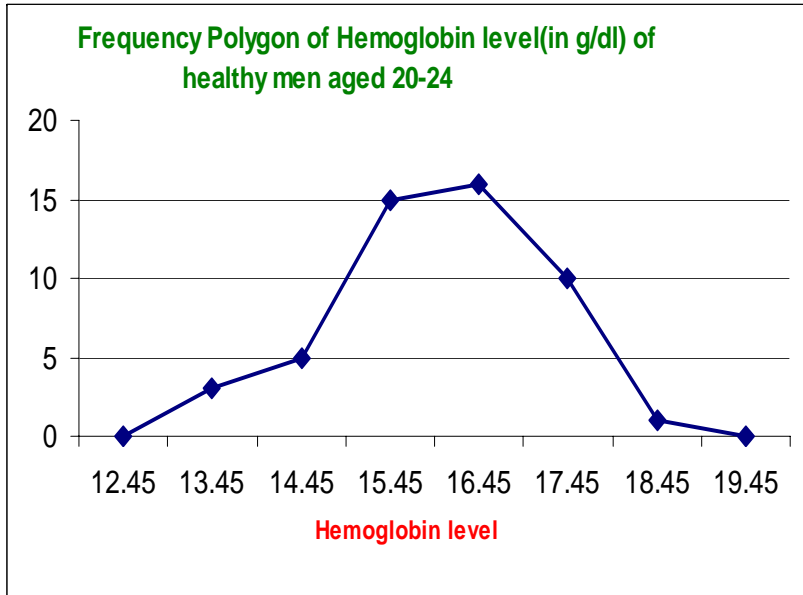
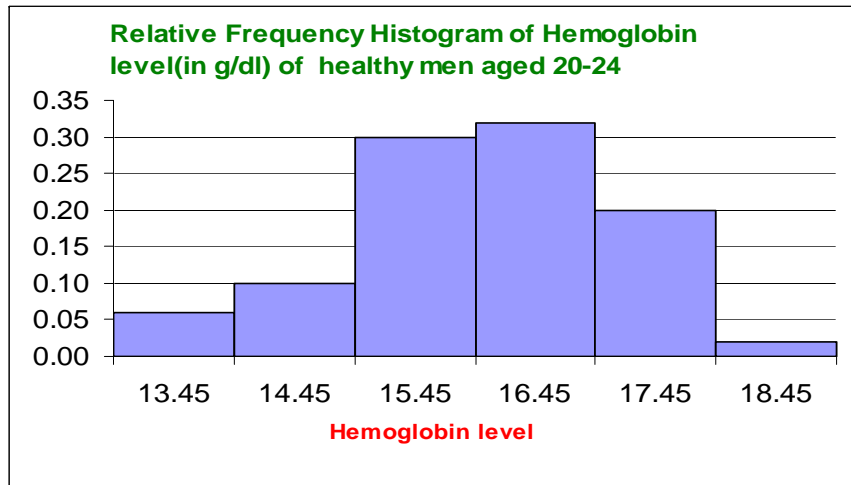
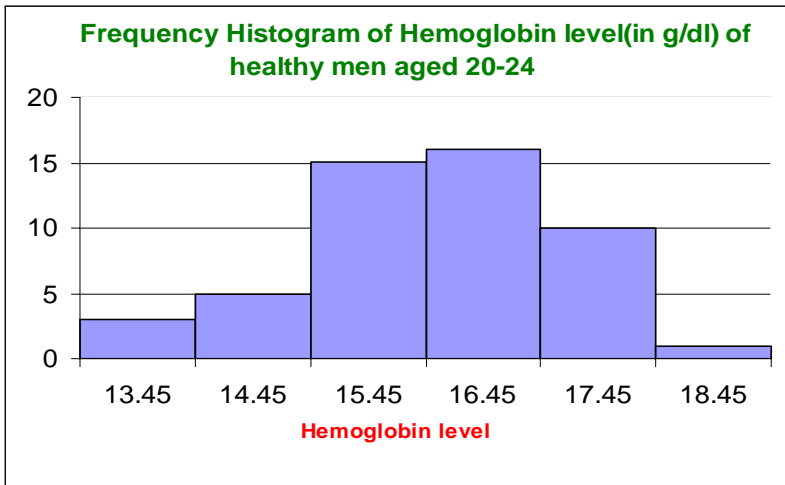
Start STAT-106 Microsoft Power... A1 - Microsoft W... Microsoft Exc... MINITAB - Untitl... 11:41

## Displaying grouped frequency distributions:

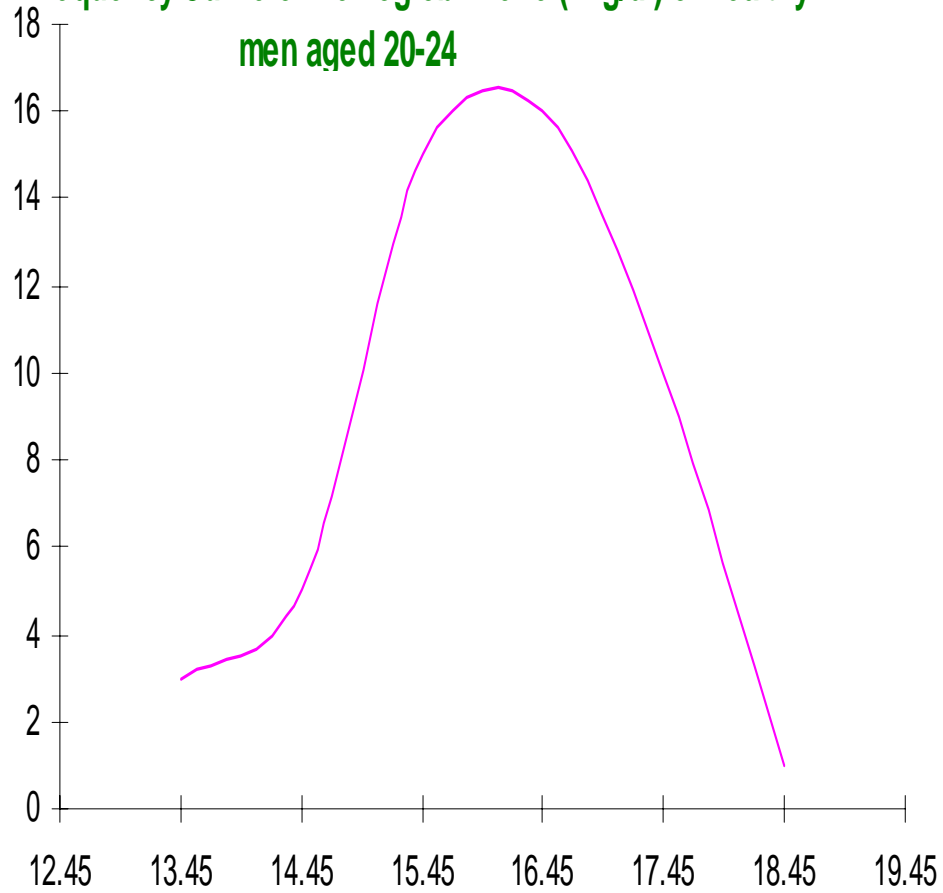
For representing frequency or relative frequency distributions, we have

The following graphical presentations:

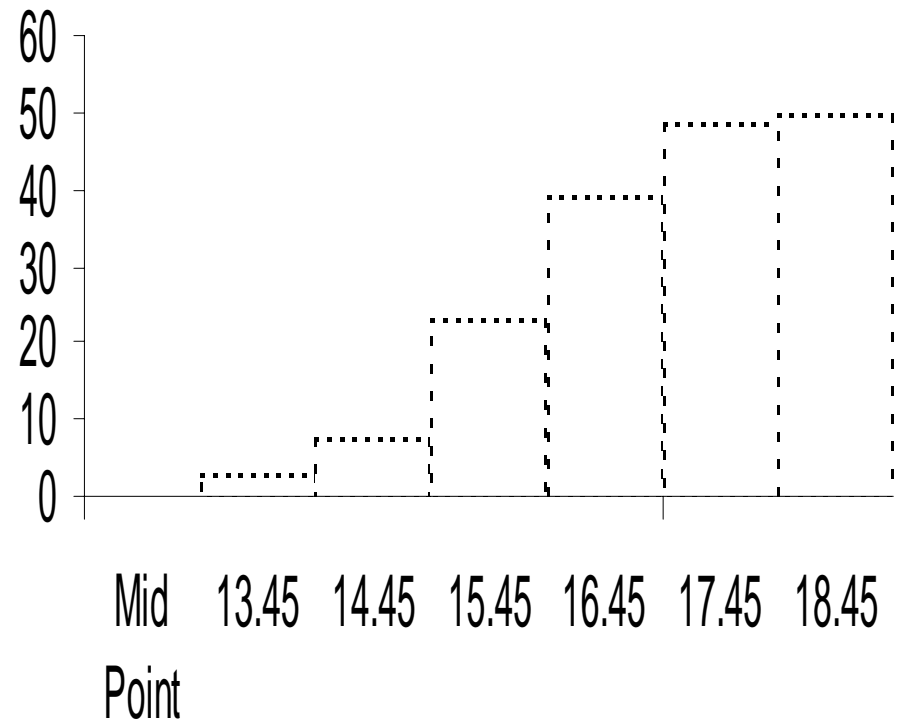
- Histograms
- Polygon
- Curves



Frequency Curve of Hemoglobin level(in g/dl) of healthy men aged 20-24



Cumulative frequency ogive of the Hemoglobine level





For representing cumulative frequency or cumulative relative frequency distributions:

- Cumulative Curves
- Cumulative Polygon

C.I.	True C. I.	Frequency	Cumulative Freq.	mid-points
15 - 19	14.5 - 19.5	8	8	17
20 - 24	19.5 - 24.5	16	24	22
25 - 29	24.5 - 29.5	32	56	27
30 - 34	29.5 - 34.5	28	84	32
35 - 39	34.5 - 39.5	12	96	37
40 - 44	39.5 - 44.5	4	100	42
Total		$n=100$		

Elementary Biostatistics



Fig. 1.2 Frequency histogram of ages of women having children

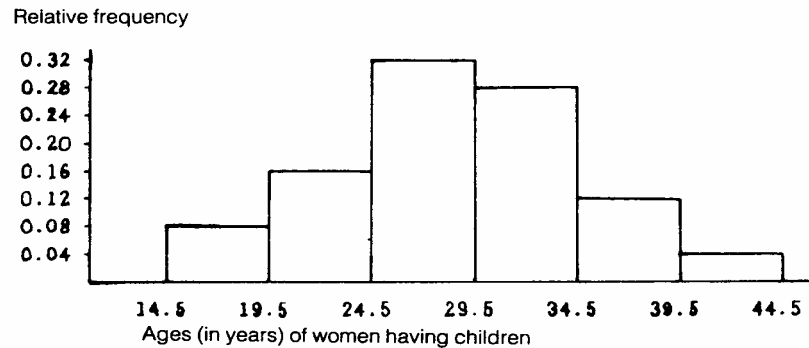


Fig. 1.3 Relative frequency histogram of ages of women having children

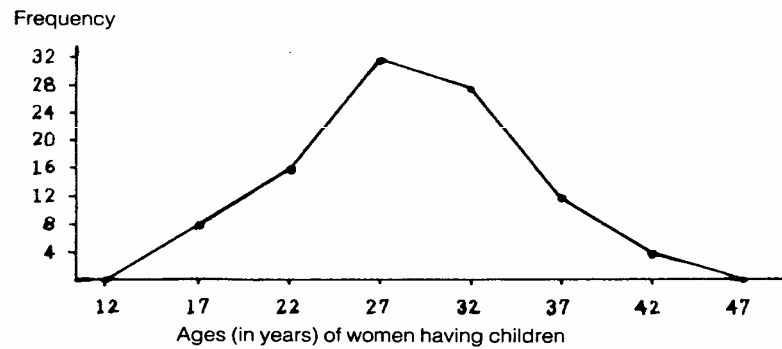


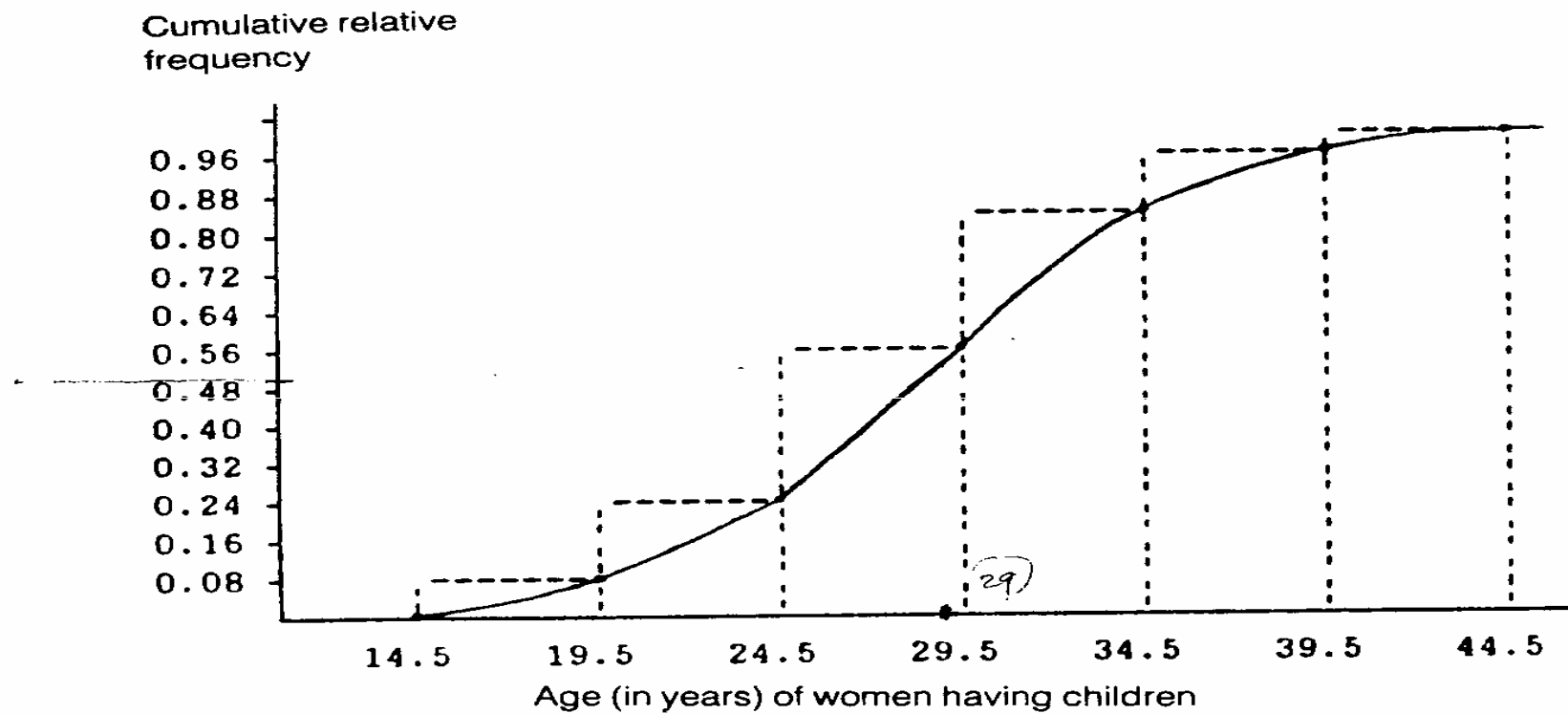
Fig. 1.4 Frequency polygon for ages of women having children

Fr

Cur  
is by a cu  
gon. The  
histogram  
the recta  
that the c

Og  
rectang  
quency  
rectang

is given in Fig. 1.7.



**Fig. 1.7** Cumulative relative frequency ogive of age of women having children

### EXERCISES