

Methods of Collecting Statistical Data

Statistical data are the first materials for statistics and the statistical analysis cannot be completed without any data. The scientists should collect their data as a sample from the population randomly.

There are three main methods to collect the data:

1. Survey Methods: divided into two types:
 - a. Descriptive Survey.
 - b. Analytical Survey.
2. Personal Methods
3. Logical and Scientific Methods

Data Presentation and Summarization:

There are many ways of presenting data including use of **tables, charts, pictures and statistical measures.**

First: Tabular Presentation: In this method (case), data are presented in frequency tables. Generally, there is more information for the serious reader but at the expense of a loss in number of readers.

For **qualitative** data; using pictures and charts is a common way to summarize and present the results, while for **quantitative** data; using tables and statistical measures is a common way

For **quantitative trait:** the following table includes the body weight of Karadi lambs at weaning (kg).

Weaning weight (kg)	18	19	20	21	22	23	24	25	26	27	28	29	30
Frequency	5	8	11	13	26	33	41	37	25	18	16	12	6

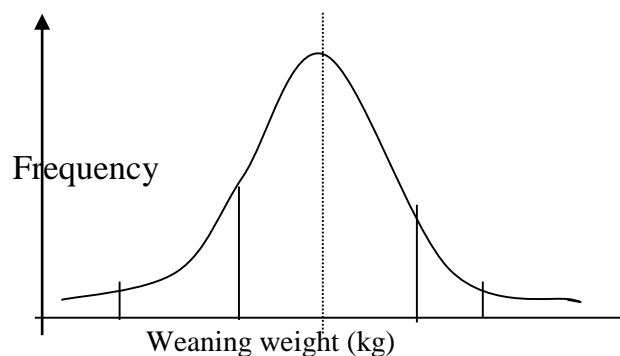


Table Method

1. Simple table
2. Complex table
3. Frequency table

1-Simple table: it is a table which contains one character only, like...

Gender	No. of students
Male	43
Female	22

2-Complex table: It is a table which contains more than one property or character for forest trees as follow:

Tree Species	Height (m)	DBH (cm)	Age (Years)	Leaf Type	Growth Rate (cm/year)
Oak	20	30	100	Deciduous	30
Pine	25	35	80	Evergreen	50
Redwood	90	120	1000	Evergreen	75
Birch	12	20	40	Deciduous	20
Teak	30	60	200	Deciduous	40

3-Frequency table: It is a most important type of table due to its widely use in different sciences.

General rules for constructing frequency table:

Steps for preparing frequency table:

1. Calculate the range.
2. Choose and determine number of classes.
3. Construct the length of width class.
4. Writing the class limits.
5. Calculate the frequency of class.

Class Length

Type of Classes:

1.Real classes: In case of real classes the upper limit of the 1st class is equal to the lower limit of the 2nd class and so on.

2.Unreal classes: In case of unreal classes the upper limit of the 1st class is smaller than the lower limit of the 2nd class.

<u>Real class</u>	<u>Unreal class</u>	<u>Unreal class</u>
5-10	0-10	5-10
10-15	11-21	11-15
15-20	22-32	16-20

Class length = upper limit – lower limit **(For real classes)**

Class length = (upper limit – lower limit) + 1 **(For unreal classes)**

Class length = the different between the value of two class mean

Class length = Second class mean – First class mean

Real class limit or Class boundary for each classes

Lower real class limit = Class mean - 1/2 class length

Upper real class limit = Class mean + 1/2 class length

Lower Real class limit = (lower limit for this class + upper limit for the previous class)/2

Upper Real class limit = (upper limit for this class + lower limit for the next class)/2

Class mean = (Upper class limit + lower class limit)/2

Class mean = (real upper limit + real lower limit)/2

$$\text{class midpoint} = \frac{\text{upper limit} + \text{lower limit}}{2}$$

Note:

1- The first class maybe open if we have a very small value.

2- The last class maybe open a large number.

3- Both, first and last classes may be open.

<u>Class</u>	<u>class midpoint</u>
50-100	(50+100)/2= 75
100-150	125
150-200	175
200-250	225
250-300	275

Steps for preparing frequency table:

Limiting the number of classes: For limiting the number of classes are three methods:

First: Sturge's method: No. of classes=1+3.322log^N
N= no. of observations (individuals of data)

Second method: Yule's method:

No. of classes= 2.5* ⁴√ N

Third method: Approximate method: In this method the no. of classes must be (5) or more.

Limiting classes length (width):

Class length = (Range /no of classes)

Range =Max. value - Min. value

Limiting the frequency for each class.

Frequency: The number of times which a certain category is repeats.

Example: suppose we have a flock of 150 Karadi ewes with different ages (2, 3, 4, 5 and 6) years old.

Age (year)	Frequency	Percentage	
2	30	20 %	0.20
3	30	20 %	0.20
4	60	40 %	0.40
5	15	10 %	0.10
6	15	10 %	0.10
Total	150	$\Sigma = 100 \%$	1

Example: suppose we have the following data include carcass weight of Karadi lambs (kg):

14, 20, 6.5, 19, 11, 9, 8, 10, 8.5, 6, 5, 7, 8, 8, 12

Length = Range = Max. – Min. = 20 – 5 = 15

No. of classes (groups) = $m = 1 + 3.3 \log (n)$

= $1 + 3.3 \log (15)$

= $1 + 3.3 \times 1.176$

= $4.88 \cong 5$

Length of class (group) = R / m (no of classes) = $15 / 5 = 3$

Class	Normal Frequency	Relative Frequency	Frequency %	Class Mean	Real Mean	Accumulative Frequency
5-7	4	0.266	26.6	6	4.5-7.5	4
8-10	6	0.401	40.1	9	7.5-10.5	10
11-13	2	0.133	13.3	12	10.5-13.5	12
14-16	1	0.067	6.7	15	13.5-16.5	13
17-20	2	0.133	13.3	18	16.5-20.5	15
Total	$\Sigma = 15$	$\Sigma = 1$	$\Sigma = 100$			

Relative Frequency = $f_i / \Sigma f_i = 4/15 = 0.266$

Frequency % = $(f_i / \Sigma f_i) * 100 = (4/15) * 100 = 26.6$

Class Mean $X_1 = (5 + 7) / 2 = 6$

$X_2 = (8 + 10) / 2 = 9$

$X_3 = (11 + 13) / 2 = 12$

$X_4 = (14 + 16) / 2 = 15$

$X_5 = (17 + 20) / 2 = 18$

Accumulative frequency: $F_1 = f_1$

$F_2 = f_1 + f_2$

$F_3 = f_1 + f_2 + f_3$

$F_4 = f_1 + f_2 + f_3 + f_4$

$F_5 = f_1 + f_2 + f_3 + f_4 + f_5$

Example: The following data represent the growth radius of fungi (mm) for 40 samples: Prepare cumulative frequency table, then draw histogram, bar chart and pie chart:

60, 30, 70, 80, 71, 84, 50, 66, 25, 57, 70, 60, 50, 53, 70, 61, 52, 79, 39, 50, 72, 58, 64, 68, 43, 70, 96, 73, 46, 41, 45, 50, 60, 52, 80, 86, 48, 41, 39, 67.

$$\begin{aligned} \text{No. of classes} &= 1 + 3.322 \log N \\ &= 1 + 3.322 \log 40 \\ &= 1 + 4.9 = 5.9 = 6 \end{aligned}$$

Range = max. value - min. value = 96 - 25 = 71, Class length = $71/6 \approx 12$

Then organizing the frequency table as follows:

Classes	Frequency(F)	Relative frequency	%Frequency
25-37	2	2/40	(2/40)*100
37-49	8	8/40	(8/40)*100
49-61	12	12/40	(12/40)*100
61-73	11	11/40	(11/40)*100
73-85	5	5/40	(5/40)*100
85-97	2	2/40	(2/40)*100
	$\sum F = 40$	$\sum R F = 40/40$	$\sum \% F = 100\% = 1$

ACumulative frequency:

There are two types of cumulative frequency:

- (1) Less than frequency, it start with zero and it's end with $\sum f = N$.
- (2) More than frequency, it start with $\sum f$ and the end point (class) is zero.

Steps for calculating:

- Preparing frequency table from the raw data.
- Convert normal frequency to cumulative frequency.

Less than frequency class More than frequency class

Less than	20		More than	20
=	=	30	=	=
=	=	40	=	=
=	=	50	=	=
=	=	60	=	=
70 or less			more than	70

Second: Chart Presentation: In this method data are presented in different ways as charts: we can use the above data to describe the following main types of charts (note that there are a lot of types could be used now in Excel):

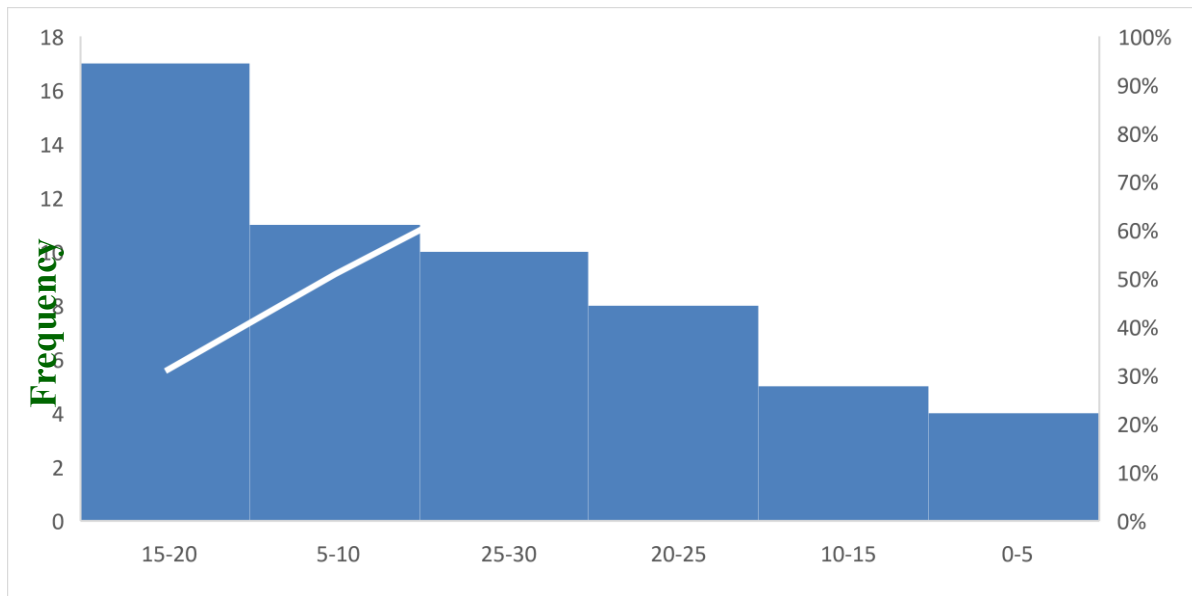
a. Histogram

The length or height of rectangles represent the class frequency, the base (width) represent class length.

The number of class= the number of rectangles (bars).

Example: Prepare a histogram for the following information.

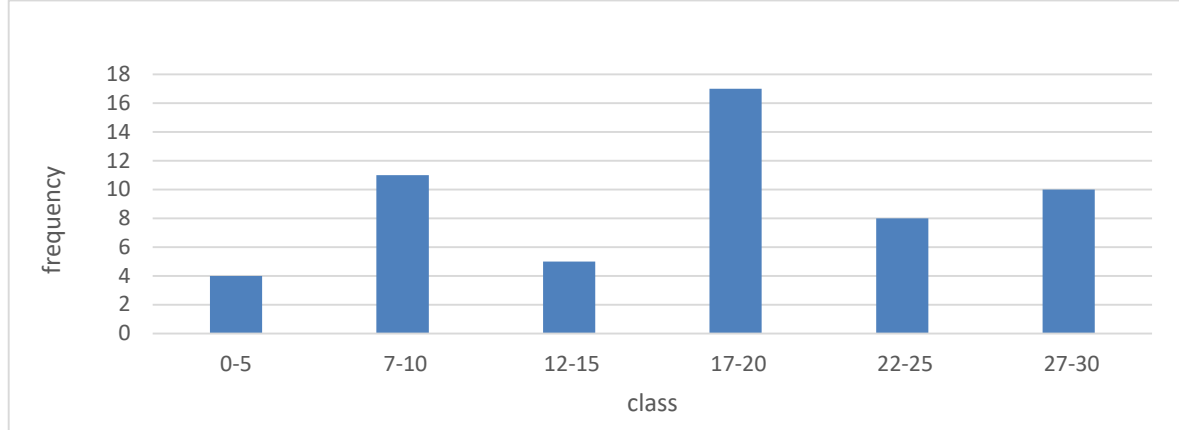
<u>Classes</u>	<u>freq.</u>
0-5	4
5-10	11
10-15	5
15-20	17
20-25	8
25-30	10



b-Bar chart

It is similar to histogram but there are the constant (fixed) spaces between rectangles (bars), the real class cannot use in drawing bar chart because there are space be classes.

Example: if you are given the following information



b. Bars Diagram

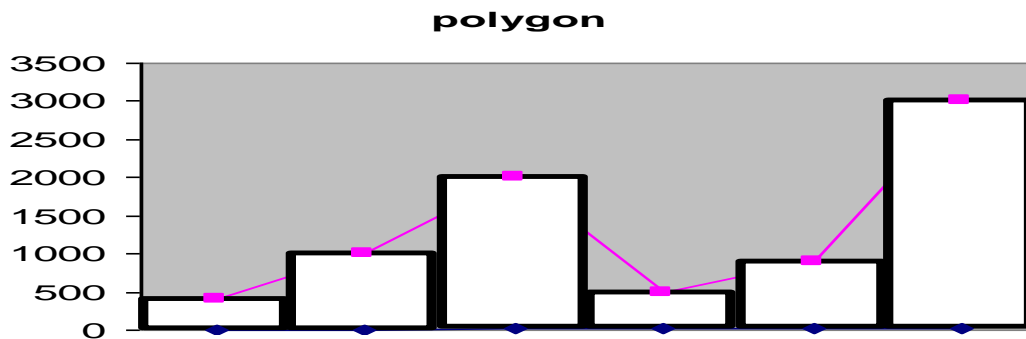
Frequency polygon::

Can be drawn with histogram or without depending on histogram.
If you are given the following information, prepare frequency polygon.

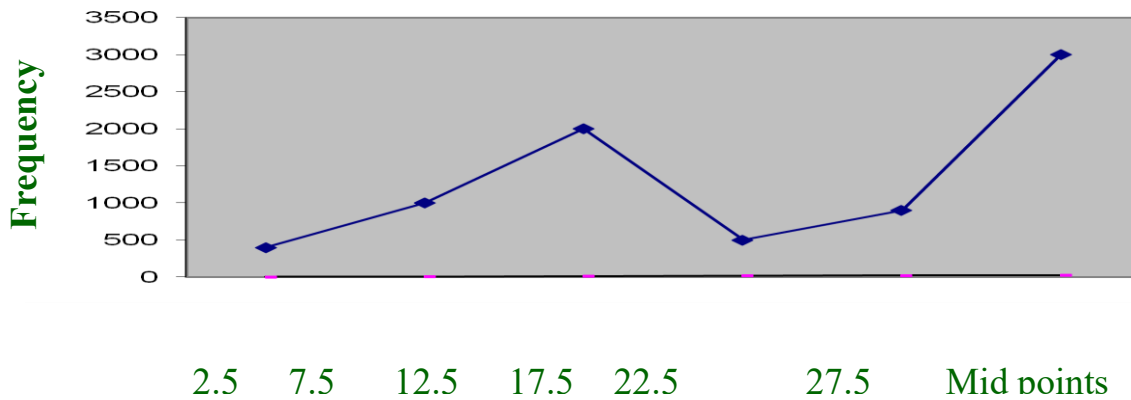
<u>Class</u>	<u>Class midpoint</u>	<u>frequency</u>
0-5	2.5	400
5-10	7.5	1000
10-15	12.5	2000
15-20	17.5	500
20-25	22.5	900
25-30	27.5	3000

The Steps are:

- 1- drawing histogram
- 2- pointing the class midpoints at the upper part of rectangles.
- 3- Joining between class midpoints with straight line
- 4- joining the first class midpoint with its lower limit.
- 5- the last class midpoint with its upper limit



Drawing frequency polygon without using histogram



Pie Chart:

Also called circuit method, Steps for drawing pie chart:

1- Calculating percentage frequency or normal percentage for classes.

2- Convert percentage value to pie chart value by using this equation:

$$\text{Pie chart value} = \text{percentage value} * 3.6$$

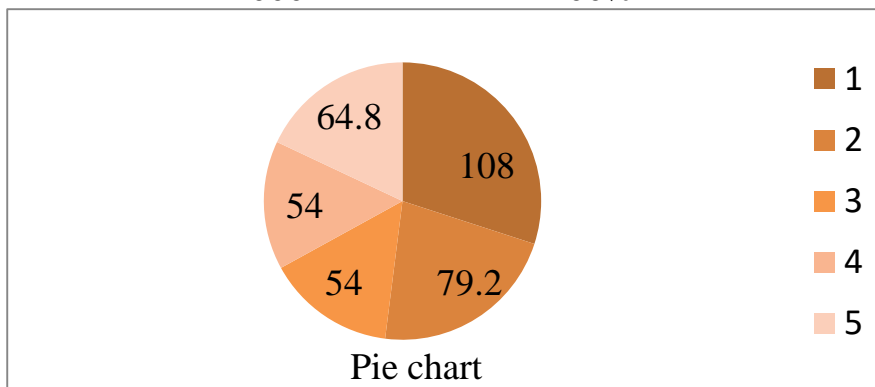
$$= \% \text{Freq.} * 360 / 100$$

3- Preparing pie chart

Example: construct pie chart, if the no. of students at the departments of college of agriculture are:

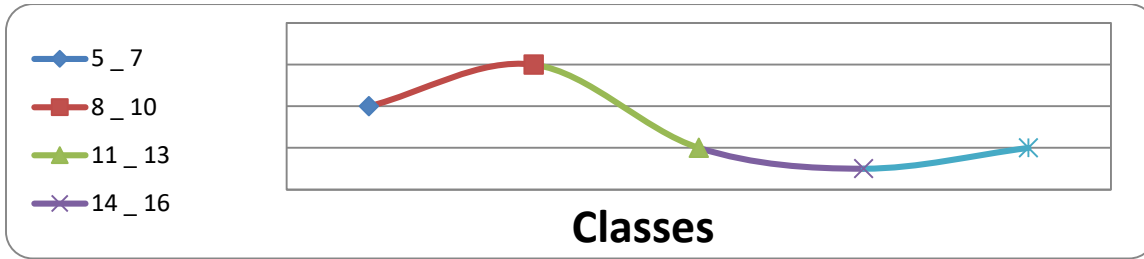
Example the following table represent number of students at 5 scientific departments prepare pie chart :

Departements	no. of students	percentage value	pie chart value
1-Biology	300	=30	30*3.6=108
2-Chemistry	220	=22	22*3.6=79.2
3-Physics	180	=18	18*3.6=64.8
4-Mathematic	150	=15	15*3.6=54
5- Geology	150	=15	15*3.6=54
	1000	100%	360



Spring semester, 2024-2025

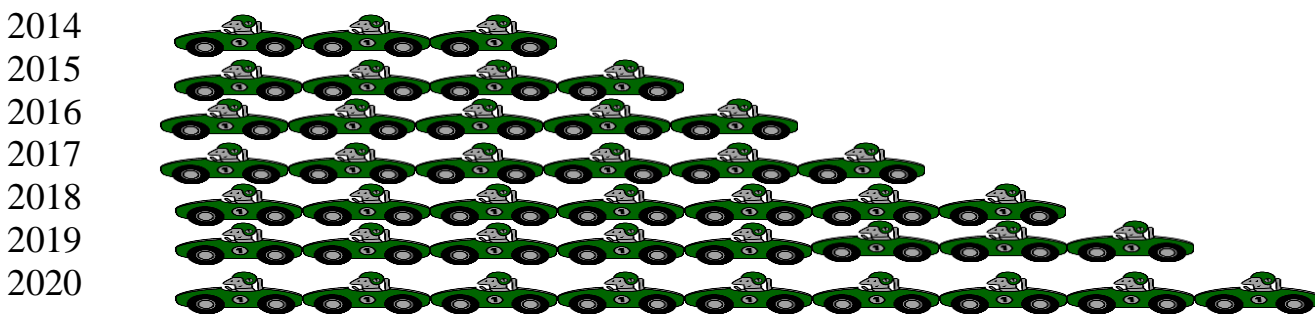
Line Diagram (Frequency Polygon)



Pictorial chart

Drawing the pictures and their values as a symbol for example, the no. of popular or the buildings and construction of houses.

the following graphs represent development of traveling devices (cars) for period (2014-2020) in Erbil city. If = 60,000 cars



Example: The following values represent the amount of yield (ton / hectare) of wheat in forty farms.

2.6, 2.2, 4.1, 3.5, 2.0, 3.2, 3.7, 3.0, 3.7, 3.4,
1.6, 3.1, 3.3, 3.8, 3.1, 2.4, 3.1, 2.5, 4.3, 3.4,
3.6, 2.9, 3.3, 3.9, 3.4, 3.3, 3.1, 3.7, 4.4, 3.2,
4.1, 1.9, 3.5, 2.3, 3.8, 3.2, 2.6, 3.9, 3.0, 4.2

Summarizing data in a frequency table.

Answer

1. Range = Max. value – Min. value

Range = 4.4 – 1.6 = 2.8 kg

2. No. of classes = $2.5 \sqrt[4]{n}$, No. of classes = $2.5 \sqrt[4]{40} = 6$

3. Class length = $\frac{Range}{no. of classes}$ Class length = $\frac{2.8}{6} = 0.467 \approx 0.5$

4. Write the class limit for these six classes start with smaller value.
5. Calculate the no. of frequency for each classes.

Summarize in the frequency table

No. of class	Class limit	Real class limit	Class mean	Frequency
1	1.5 – 1.9	1.45 – 1.95	1.7	2
2	2.0 – 2.4	1.95 – 2.45	2.2	4
3	2.5 – 2.9	2.45 – 2.95	2.7	4
4	3.0 – 3.4	2.95 – 3.45	3.2	15
5	3.5 – 3.9	3.45 – 3.95	3.7	10
6	4.0 – 4.4	3.95 – 4.45	4.2	5
Total				40

Relative frequency distribution

Calculate the Relative frequency for each classes by this equation:

$$\text{Relative frequency for each classes} = \frac{f_i}{\sum f_i}$$

Percentage frequency distribution

$$\% \text{ frequency} = \text{Relative frequency for each classes} \times 100$$

Cumulative frequency distribution

No. of class	Class limit	Normal frequency	Relative frequency	Percentage frequency	Cumulative frequency
1	1.5 – 1.9	2	0.05	5	2
2	2.0 – 2.4	4	0.1	10	6
3	2.5 – 2.9	4	0.1	10	10
4	3.0 – 3.4	15	0.375	37.5	25
5	3.5 – 3.9	10	0.25	25	35
6	4.0 – 4.4	5	0.125	12.5	40
		40	1	100%	