

statistics: It's a branch of mathematics dealing with the collecting, summarizing, presenting and analyzing data which measure numerically, this analysis may lead to conclusions and subsequent decisions.

There are two types of statistics: $\rightarrow$

1) Descriptive: It's branch of statistics which interest to collecting data, organizing data, classifying, tabulating, diagramming, then calculate the various numerical measurements.

2) Inferential: It's branch of statistics which interest to reach to conclusions about the properties of population by using available information from sample.

Descriptive statistics: In statistics we depend upon data, then we can define data as the numbers or values which represent any phenomena (variable) in life, the data may be divided into two parts, the first call ungrouped data, the second call grouped data.

Descriptive Statistics

ungrouped data

the values or numbers which getting it for any phenomena from resources, population].

grouped data

[the data putting in a table and the table contains two columns, classes and frequency].

Construction of frequency distribution table-

To transform the ungrouped data to grouped data, we apply the following steps:-

1. Find the highest value and lowest value.
2. Find the range of data, which is a difference between largest and smallest number in set of data.
3. Suppose the number of class (usually between 5 and 20) which is given in question.
4. Find the class width by using the following

$$\text{Class width} = \frac{\text{Range}}{\text{no. of classes}}$$

5. Put the lowest value in the first class which represent the lower limit of first class interval, then we add class width to it and put the answer in lower limit in the second class interval and so on until we complete the number of classes.
6. Put the upper limit of first class interval which we select by the lower limit of second class interval minus (1) if the data is integer number (and 0.1 if the data is decimal number) and so on.
7. Find the frequency for each class interval.

Example: The following data represent the record high temperatures for each of the (50) states. Construct a frequency distribution table for the data.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Sol: The procedure for constructing a grouped frequency distribution for numerical data follows.

1. $H = 134$ and $L = 100$

2. $R = \text{highest value} - \text{lowest value}$

$$R = 134 - 100 = 34$$

3. here (7) is arbitrary chosen.

4. width = $\frac{R}{\text{no. of classes}} = \frac{34}{7} = 4.9$

5. the 7 classes, as shown.

100 105 110 115 120 125 130

6. $105 - 1 = 104$, so the first class is 100-104

class limits

100 - 104

105 - 109

110 - 114

115 - 119

120 - 124

125 - 129

130 - 134

7- write the numerical values for frequency column, and find the cumulative frequencies, the completed table is:

class limits	frequency	cumulative frequency
100 - 104	2	2
105 - 109	8	10
110 - 114	18	28
115 - 119	13	41
120 - 124	7	48
125 - 129	1	49
130 - 134	1	50

Measures of Central tendency

مقاييس النزعة المركزية

one of the most important aspects of describing a distribution is the Central value around which the observations are distributed. Any mathematical measure which is intended to represent the Center or Central value of a set of observations is known as measure of Central tendency or the Single value, which represents the group of values, is termed as a measure of Central tendency or a measure of location or an average.

Types of average:

- 1- Arithmetic Mean
- 2- Median
- 3- Mode
- 4- Geometric Mean
- 5- Harmonic Mean

Relationship between these types:

the relationship that exists among them is

$H.M \leq G.M \leq A.M$ this inequality could be verified by set of data.

Average

ungrouped data

grouped data

1) Arithmetic Mean

$$A.M = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

$$A.M = \frac{\sum_{i=1}^N f_i X_i}{\sum_{i=1}^N f_i}$$

2) Median

* If (N) is odd

$$Me = \frac{Y_{(N+1)}}{2}$$

* If (N) is even

$$Me = \frac{Y_{N/2} + Y_{N/2+1}}{2}$$

$$Me = L_i + \left[\frac{\left(\sum_{j=1}^N f_j \right) / 2 + F_i}{f_i} \right] \cdot C$$

3) Mode

The Mode here is the value repeated more than the other values

$$Mo = L_i + \left[\frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \right] \cdot C$$

4) Geometric Mode

5) Harmonic Mode

$$H.M = \frac{N}{\sum_{i=1}^N \frac{1}{X_i}}$$

$$H.M = \frac{N}{\sum_{i=1}^N \frac{f_i}{X_i}}$$

Examples

1) Find the Mean for the following

~~7, 13, 16, 20, 23, 27, 29~~

Sol $\bar{x} = \frac{\sum x_i}{n} = \frac{7+13+16+20+23+27+29}{7} = \frac{135}{7} = 19.2$

2) If you have the frequency dist. table which represents the temperatures of the following:

classes 20-24 25-29 30-34 35-39 40-44

Frequency = 1 2 6 5 6

Classes	(f_i)	x_i	$f_i \cdot x_i$
---------	---------	-------	-----------------

20-24	1	22	22
25-29	2	27	54
30-34	6	32	192
35-39	15	37	185
40-44	6	42	252

$$\sum f_i = 20$$

$$\sum f_i x_i = 705$$

$$x_i = \frac{L_i + U_i}{2}, \text{ for } x_1 = \frac{L_1 + U_1}{2} = \frac{20 + 24}{2} = \frac{44}{2} = 22$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{705}{20} = 35.25, \text{ the average of temperature in Baghdad at 20 days}$$

3) Find the Median for the following -

13, 20, 7, 16, 23, 29, 27

Sol 7, 13, 16, 20, 23, 27, 29 (ترتيب تصاعدي)
 $n = 7$ (odd)

$$Me = \frac{Y_{n+1}}{2} = \frac{Y_{7+1}}{2} = \frac{Y_8}{2} = Y_4 = 20$$

4) Find the Median for the data which represents the speed for many computers

317, 469, 375, 444, 216, 392, 285, 469

Sol 216, 285, 317, 375, 392, 444, 469, 469 (ترتيب تصاعدي)
 $n = 8$ (even)

$$Me = \frac{Y_{(\frac{n}{2})} + Y_{(\frac{n}{2}+1)}}{2} = \frac{Y_{\frac{8}{2}} + Y_{(\frac{8}{2}+1)}}{2} = \frac{Y_4 + Y_5}{2} = \frac{375 + 392}{2} = 383.5 \text{ the median}$$

5) Find the Median for the following results

classes	f_i
40-49	5
50-59	15
60-69	22
70-79	28
80-89	10

L_i = the lower limit for the median class

F_i = cumulative frequency at the start of median class

f_i = is the median class frequency.

c = the median class length.

Sol

classes	f_i	C-F	F_i
40-49	5	less than 49	5
<u>50-59</u>	<u>15</u>	= = 59	<u>20</u>
60-69	22	= = 69	42
70-79	28	= = 79	70
80-89	10	= = 89	80

$$\sum f_i = 80$$

$$\text{Median order} = \frac{\sum f_i}{2} = \frac{80}{2} = 40$$

$$F_i = 20, f_i = 15, C = 10, L_i = 50$$

$$Me = L_i + \left[\frac{(\sum f_i)/2 + F_i}{f_i} \right] \cdot c$$

$$= 50 + \left[\frac{40 - 20}{15} \right] \cdot 10 = 50 + \frac{20}{15} \cdot 10 = 50 + \frac{40}{3}$$

$$= 63.33 \text{ the Median}$$

6) Find the mode for the following samples:

A. $X_i = 2, 11, 4, -3, 7, 13$

B. $X_i = 9, 10, 5, 9, 7, 8, 6, 9, 10, 4$

C. $X_i = 2, 0, 3, 1, 2, 4, 2, 5, 4, 1, 0, 4$

Sol

For A) there no mode exists.

For B) $Mo = 9$

For C) $Mo = 2, 4$

- * L_i = the lower limit for mode class
- * f_i = the frequency for mode class (greatest freq.)
- * C = the length for mode class

7) Consider the following frequency distribution table which represents particles in five levels of energy classes

Classes	40-49	50-59	60-69	70-79	80-89
Frequency	5	15	20	30	10

Find the Mode

Sol, the greatest freq is 30, then

$$f_i = 30, f_{i-1} = 20, f_{i+1} = 10$$

The mode class is 70-79, $L_i = 70$, $C = 10$

$$\begin{aligned}
 Mo &= L_i + \left[\frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \right] \cdot C \\
 &= 70 + \left[\frac{30 - 20}{2(30) - 20 - 10} \right] \cdot 10 = 73.33 \text{ the Mode}
 \end{aligned}$$

8) Find the Geometric mean for the following data 3, 2, 6, 4

$$\text{Sol, } G.M = \sqrt[n]{x_1 \cdot x_2 \cdots x_m} = \sqrt[4]{3 \times 2 \times 6 \times 4} = \sqrt[4]{144} = 3$$

9) Find the harmonic mean of the following data 3, 2, 6, 4

$$\text{Sol, } H.M = \frac{n}{\sum \frac{1}{x_i}} = \frac{4}{\frac{1}{3} + \frac{1}{2} + \frac{1}{6} + \frac{1}{4}} = 3.2$$