Geometric mean

The geometric mean of a series containing n observations is the nth root of the product of the values.
If $x_1, x_2 \ldots, x_n$ are observations then

$$G.M = \sqrt[n]{x_1, x_2 \ldots x_n}$$

$$= (x_1, x_2 \ldots x_n)^{1/n}$$

$$\text{Log GM} = \frac{1}{n}\log(x_1, x_2 \ldots x_n)$$

$$= \frac{1}{n}(\log x_1 + \log x_2 + \ldots + \log x_n)$$

$$= \frac{\sum \log x_i}{n}$$

$$\text{GM} = \text{Antilog } \frac{\sum \log x_i}{n}$$

For grouped data

$$\text{GM} = \text{Antilog } \left[\frac{\sum f \log x_i}{n}\right]$$

GM is used in studies like bacterial growth, cell division, etc.

Example 11

If the weights of sorghum ear heads are 45, 60, 48,100, 65 gms. Find the Geometric mean for the following data

Weight of ear head x (g)	Log x
45	1.653
60	1.778
48	1.681
100	2.000
65	1.813
Total	**8.925**

Solution

Here n = 5

$$\text{GM} = \text{Antilog } \frac{\sum \log x_i}{n}$$

$$= \text{Antilog } \frac{8.925}{5}$$

$$= \text{Antilog } 1.785$$

$= 60.95$

Grouped Data

Example 12

Find the Geometric mean for the following

Weight of sorghum (x)	No. of ear head(f)
50	4
65	6
75	16
80	8
95	7
100	4

Solution

Weight of sorghum (x)	No. of ear head(f)	Log x	f x log x
50	5	1.699	8.495
63	10	10.799	17.99
65	5	1.813	9.065
130	15	2.114	31.71
135	15	2.130	31.95
Total	**50**	**9.555**	**99.21**

Here n= 50

$$GM = \text{Antilog}\left[\frac{\sum f \ \log x_i}{n}\right]$$

$$= \text{Antilog}\left[\frac{99.21}{50}\right]$$

$$= \text{Antilog } 1.9842 = 96.43$$

Continuous distribution

Example 13

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the Geometric mean

Weights of ear heads (in g)	**No of ear heads (f)**
60-80	22
80-100	38
100-120	45

120-140	35
140-160	20
Total	**160**

Solution

Weights of ear heads (in g)	**No of ear heads (f)**	**Mid x**	**Log x**	**f log x**
60-80	22	70	1.845	40 59
80-100	38	90	1.954	74.25
100-120	45	110	2.041	91.85
120-140	35	130	2.114	73.99
140-160	20	150	2.176	43.52
Total	**160**			**324.2**

Here $n = 160$

$$\text{GM} = \text{Antilog}\left[\frac{\sum f \ \log x_i}{n}\right]$$

$$= \text{Antilog}\left[\frac{324.2}{160}\right]$$

$$= \text{Antilog}\left[2.02625\right]$$

$$= 106.23$$

Harmonic mean (H.M)

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If x1, x2…..xn are n observations,

$$\text{H.M} = \frac{n}{\sum_{i=n}^{n}\left(\frac{1}{x_i}\right)}$$

For a frequency distribution

$$\text{H.M} = \frac{n}{\sum_{i=n}^{n} f\left(\frac{1}{x_i}\right)}$$

H.M is used when we are dealing with speed, rates, etc.

Example 13

From the given data 5, 10,17,24,30 calculate H.M.

X	$\frac{1}{x}$
5	0.2000
10	0.1000
17	0.0588
24	0.0417
30	0.4338

$$\text{H.M} = \frac{5}{0.4338} = 11.526$$

Example 14

Number of tomatoes per plant are given below. Calculate the harmonic mean.

Number of tomatoes per plant	20	21	22	23	24	25
Number of plants	4	2	7	1	3	1

Solution

Number of tomatoes per plant (x)	No of plants(f)	$\frac{1}{x}$	$f\left(\frac{1}{x}\right)$
20	4	0.0500	0.2000
21	2	0.0476	0.0952
22	7	0.0454	0.3178
23	1	0.0435	0.0435
24	3	0.0417	0.1251
25	1	0.0400	0.0400
	18		0.8216

$$\text{H.M} = \frac{n}{\sum f\left(\frac{1}{x_i}\right)} = \frac{18}{0.1968} = 21.91$$

Merits of H.M

1. It is rigidly defined.

2. It is defined on all observations.

3. It is amenable to further algebraic treatment.

4. It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.

Demerits of H.M

1. It is not easily understood.
2. It is difficult to compute.
3. It is only a summary figure and may not be the actual item in the series
4. It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.
5. It is rarely used in grouped data.

Percentiles

The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The x^{th} percentile is that value below which x percent of values in the distribution fall. It may be noted that the median is the 50^{th} percentile.

For raw data, first arrange the n observations in increasing order. Then the x^{th} percentile is given by

$$P_x = \left(\frac{x(n+1)}{100}\right)^{th} \text{ item}$$

For a frequency distribution the x^{th} percentile is given by

$$P_x = l + \left(\frac{(x.n/100) - cf}{f} \times C\right)$$

Where

l = lower limit of the percentile calss which contains the x^{th} percentile value (x. n /100)

cf = cumulative frequency uotp l

f = frequency of the percentile class

C= class interval

N= total number of observations

Percentile for Raw Data or Ungrouped Data

Example 15

The following are the paddy yields (kg/plot) from 14 plots:

30,32,35,38,40.42,48,49,52,55,58,60,62,and 65 (after arranging in ascending order). The computation of 25^{th} percentile (Q_1) and 75^{th} percentile (Q_3) are given below:

$$P_{25}(or\ Q_1) = \left(\frac{25(14+1)}{100}\right)^{th} \text{ item}$$

$$= \left(3\frac{3}{4}\right)^{th} \text{ item}$$

$$= 3^{rd} \text{ item} + (4^{th} \text{ item} - 3^{rd} \text{ item}) \left(\frac{3}{4}\right)$$

$$= 35 + (38\text{-}35) \left(\frac{3}{4}\right)$$

$$= 35 + 3 \left(\frac{3}{4}\right) = 37.25 \text{ kg}$$

$$P_{75}(or\ Q_3) = \left(\frac{75(14+1)}{100}\right)^{th} \text{ item}$$

$$= \left(11\frac{1}{4}\right)^{th} \text{ item}$$

$$= 11^{th} \text{ item} + (12^{th} \text{ item} - 11^{th} \text{ item}) \left(\frac{1}{4}\right)$$

$$= 55 + (58\text{-}55) \left(\frac{1}{4}\right)$$

$$= 55 + 3 \left(\frac{1}{4}\right) = 55.75 \text{ kg}$$

Example 16

The frequency distribution of weights of 190 sorghum ear-heads are given below. Compute 25th percentile and 75th percentile.

Weight of ear-heads (in g)	**No of ear heads**
40-60	6
60-80	28
80-100	35
100-120	55
120-140	30
140-160	15
160-180	12
180-200	9
Total	**190**

Solution

Weight of ear-heads (in g)	**No of ear heads**	**Less than class**	**Cumulative frequency**
40-60	6	< 60	6
60-80	28	< 80	34
80-100	35	<100	69
100-120	55	<120	124
120-140	30	<140	154
140-160	15	<160	169
160-180	12	<180	181
180-200	9	<200	190
Total	**190**		

47.5

142.5

or P_{25}, first find out $\left(\frac{25(190)}{100}\right)$, and for $P_{75}, \left(\frac{75(190)}{100}\right)$, and proceed as in the case of median.

For P_{25}, we have $\left(\frac{25(190)}{100}\right) = 47.5$.

The value 47.5 lies between 34 and 69. Therefore, the percentile class is 80-100. Hence,

$$P_{25} = Q_1 = l + \left(\frac{(25.n/100) - cf}{f} \times C\right)$$

$$= 80 + \left(\frac{(47.5) - 34}{35} \times 20\right)$$

$$= 80 + \left(\frac{(13.5)}{35} \times 20\right)$$

$= 80 + 7.71$ or 87.71 g.

Similarly,

$$P_{75} = l + \left(\frac{(75.n/100) - cf}{f} \times C\right) \text{ Class}$$

$$= 120 + \left(\frac{(142.5) - 121}{30} \times 20\right)$$

$$= 120 + \left(\frac{(21.5)}{30} \times 20\right)$$

$= 120 + 14.33 = 134.33$ g.

Quartiles

The quartiles divide the distribution in four parts. There are three quartiles. The second quartile divides the distribution into two halves and therefore is the same as the median. The first (lower).quartile (Q1) marks off the first one-fourth, the third (upper) quartile (Q3) marks off the three-fourth. It may be noted that the second quartile is the value of the median and 50^{th} percentile.

Raw or ungrouped data

First arrange the given data in the increasing order and use the formula for Q1 and Q3 then quartile deviation, Q.D is given by

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Where $Q_1. = \left(\frac{n+1}{4}\right)^{th}$ item and $Q_3. = 3\left(\frac{n+1}{4}\right)^{th}$ item

Example 18

Compute quartiles for the data given below (grains/panicles) 25, 18, 30, 8, 15, 5, 10, 35, 40, 45

Solution

5, 8, 10, 15, 18, 25, 30, 35, 40, 45

$$Q_1. = \left(\frac{n+1}{4}\right)^{th}$$

$$= \left(\frac{10+1}{4}\right)^{th}$$

$= (2.75)^{th}$ item

$= 2^{nd}$ item + $\left(\frac{3}{4}\right)(3^{rd}$ item – 2^{nd} item)

$$= 8+\frac{3}{4}(10\text{-}8)$$

$$= 8+\frac{3}{4} \text{ x } 2$$

$= 8+1.5$

$= 9.5$

$$Q_3. = 3\left(\frac{n+1}{4}\right)^{th}$$

$= 3 \text{ x } (2.75)^{\text{th}}$ item

$= (8.75)^{\text{th}}$ item

$= 8^{\text{th}}$ item $+ \left(\frac{1}{4}\right)(9^{\text{th}}$ item $- 8^{\text{th}}$ item)

$= 35+\frac{1}{4}(40\text{-}35)$

$= 35+1.25$

$= 36.25$

Discrete Series

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{n+1}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than $\left(\frac{n+1}{4}\right)$, then the corresponding value of x is Q1

Step4: Find $3\left(\frac{n+1}{4}\right)$

Step5: See in the cumulative frequencies, the value just greater than $3\left(\frac{n+1}{4}\right)$,then the corresponding value of x is Q3

Example 19

Compute quartiles for the data given bellow (insects/plant).

X	5	8	12	15	19	24	30
f	4	3	2	4	5	2	4

Solution

x	f	cf
5	4	4
8	3	7
12	2	9
15	4	13
19	5	18
24	2	20

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} item = \left(\frac{24+1}{4}\right) = \left(\frac{25}{4}\right) = 6.25^{th}\, item$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} item = 3\left(\frac{24+1}{4}\right) = 18.75\text{th item } \therefore Q1 = 8;\ Q3 = 24$$

Continuous series

Step1: Find cumulative frequencies

Step2: Find $\left(\frac{n}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than $\left(\frac{n}{4}\right)$, then the corresponding class interval is called first quartile class.

Step4: Find $3\left(\frac{n}{4}\right)$ See in the cumulative frequencies the value just greater than $3\left(\frac{n}{4}\right)$ then the corresponding class interval is called 3rd quartile class. Then apply the respective formulae

$$Q_1 = l_1 + \frac{\frac{n}{4} - m_1}{f_1} \times c_1$$

$$Q_3 = l_3 + \frac{3\left(\frac{n}{4}\right) - m_3}{f_3} \times c_3$$

Where l_1 = lower limit of the first quartile class

f_1 = frequency of the first quartile class

c_1 = width of the first quartile class

m_1 = c.f. preceding the first quartile class

l_3 = 1ower limit of the 3rd quartile class

f_3 = frequency of the 3rd quartile class

c_3 = width of the 3rd quartile class

m_3 = c.f. preceding the 3rd quartile class

Example 20: The following series relates to the marks secured by students in an examination.

Marks	No. of Students
0-10	11
10-20	18
20-30	25
30-40	28
40-50	30
50-60	33
60-70	22
70-80	15
80-90	12
90-100	10

Find the quartiles

Solution

C.I	f	cf
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	12	194
90-100	10	204
	204	

$$\left(\frac{n}{4}\right) = \left(\frac{204}{4}\right) = 51$$

$$3\left(\frac{n}{4}\right) = 153$$

$$Q_1 = l_1 + \frac{\frac{n}{4} - m_1}{f_1} \times c_1$$

$$= 20 + \frac{51 - 29_1}{25_1} \times 10 = 20 + 8.8 = 28.8$$

$$Q_3 = l_3 + \frac{3\left(\frac{n}{4}\right) - m_3}{f_3} \times c_3$$

$$= 60 + \frac{153 - 145}{22} \times 12 = 60 + 4.36 = 64.36$$

Questions

1. The middle value of an ordered series is called

a)2nd quartile b) 5th decile

c) 50th percentile d) all the above

Ans: all the above

2. For a set of values the model value can be

a) Unimodel b) bimodal

c) Trimodel d) All of these

d) Ans: all the above

3. Mode is suitable for qualitative data.

Ans: True

4. Decile divides the group in to ten equal parts.

Ans: True

5. Mean is affected by extreme values.

Ans: True

6. Geometric mean can be calculated for negative values.

Ans: False

7. Define mean and median

8. For what type of data mode can be calculated.
9. Explain how to calculate the arithmetic mean for raw and grouped data.
10. Explain how to calculate median and mode for grouped data.