Geometric Mean

Geometric Mean

- The **Geometric mean** of $x_1, x_2, \ldots, \ldots, \ldots, x_n$ - n positive quantities is the n^{th} root of their product

- *Geometric Mean,* $\boldsymbol{GM} = \sqrt[n]{x_1 * x_2 * \ldots * x_n} = (x_1 * x_2 * \ldots * x_n)^{\frac{1}{n}}$

➢ $\log GM = \log (x_1 * x_2 * \ldots * x_n)^{\frac{1}{n}}$

➢ $\log GM = \frac{1}{n}(\log x_1 + \log x_2 + \cdots + \log x_n)$ [as, $\log x^n = n \log x$;

$\log(AB) = \log A + \log B$]

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

Geometric Mean: Grouped Data

- If $x_1, x_2, \ldots, \ldots, x_k$ - k non zero positive quantities with corresponding frequencies $f_1, f_2, \ldots, \ldots, f_k$ (where $\sum_{i=1}^{k} f_i = n$), then

$$GM = Antilog \frac{\sum_{i=1}^{k} f_i \log x_i}{n}$$

Test Yourself

The Population of Saint Martin’s Island changed by the following percentages each year for the last decade (“+” means increase, “-” means decrease):

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Change	+2%	+3%	-1%	+5%	+7%	+5%	+2%	-1%	+2%	+3%

What is the average population growth per year for the last decade?

Solution

Number of observations, n = 10

Change in Year 2010 = $(1 + \frac{2}{100}) = 1.02$

Change in Year 2011 = $(1 + \frac{3}{100}) = 1.03$

Change in Year 2012 = $(1 - \frac{1}{100}) = 0.99$

Change in Year 2013 = $(1 + \frac{5}{100}) = 1.05$

Change in Year 2014 = $(1 + \frac{7}{100}) = 1.07$

Change in Year 2015 = $(1 + \frac{5}{100}) = 1.05$

Change in Year 2016 = $(1 + \frac{2}{100}) = 1.02$

[*continued in next page*]

Change in Year 2017 = $(1 - \frac{1}{100}) = 0.99$

Change in Year 2018 = $(1 + \frac{2}{100}) = 1.02$

Change in Year 2019 = $(1 + \frac{3}{100}) = 1.03$

Geometric Mean = $\sqrt[10]{1.02 * 1.03 * 0.99 * 1.05 * 1.07 * 1.05 * 1.02 * 0.99 * 1.02 * 1.03}$

$= 1.0267$

Therefore, the average population growth per year for the last decade is 1.0267

Test Yourself

The price of Toyota Camry has fluctuated for last few years. Given below is the change in value for each year. (“+” means increase, “-” means decrease.)

Year	2012	2013	2014	2015	2016	2017	2018	2019
Change	+6%	+3%	+5%	+1%	-2%	-3%	+1%	-4%

What is the average annual change in the price of Toyota Camry within this period? Answer in percentage.

Solution

Number of observations, n = 8

Change in Year 2012 = $(1 + \frac{6}{100}) = 1.06$

Change in Year 2013 = $(1 + \frac{3}{100}) = 1.03$

Change in Year 2014 = $(1 + \frac{5}{100}) = 1.05$

Change in Year 2015 = $(1 + \frac{1}{100}) = 1.01$

Change in Year 2016 = $(1 - \frac{2}{100}) = 0.98$

Change in Year 2017 = $(1 - \frac{3}{100}) = 0.97$

[*continued in next page*]

Change in Year 2018 = $(1 + \frac{1}{100}) = 1.01$

Change in Year 2019 = $(1 - \frac{4}{100}) = 0.96$

Geometric Mean = $\sqrt[8]{1.06 * 1.03 * 1.05 * 1.01 * 0.98 * 0.97 * 1.01 * 0.96}$

$= 1.0082$

Therefore, the average annual change in the price of Toyota Camry in percentage is 0.82%

Significance of Geometric Mean

Geometric mean is usually used for dealing with data related to **growth rates** (like population growth etc.) or interest rates, index number etc.

In many cases, the geometric mean is the best measure to determine the average growth rate of some quantity.

Some uses of Geometric Mean:

1. Proportional growth in share prices, interest
2. Determining aspect ratios
3. UN Human Development Index
4. Average of Price fluctuation

Geometric Mean

Merits	Demerits
1. It is rigidly defined. 2. It is based on all the observations of the series. 3. It is suitable for measuring the relative changes. 4. It gives more weights to the small values and less weights to the large values. 5. It is used in averaging the ratios, percentages and in determining the rate gradual increase and decrease. 6. It is capable of further algebraic treatment.	1. It is not easy to understand by a man of ordinary prudence as it involves logarithmic operations. As such it is not popular like that of arithmetic average. 2. It is difficult to calculate as it involves finding out of the root of the products of certain values either directly, or through logarithmic operations. 3. It cannot be calculated, if the number of negative values is odd. 4. It cannot be calculated, if any value of a series is zero.

Harmonic Mean

Harmonic Mean

- If $x_1, x_2, \ldots, \ldots, \ldots, x_n$ - n values of a series of data, then the harmonic mean of the given data will be the **reciprocal** of the arithmetic mean of the **reciprocal** of the given value.
- For $x_1, x_2, \ldots, \ldots, \ldots, x_n$ - n nonzero observations the harmonic mean HM will be

$$HM = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\cdots+\frac{1}{x_n}}$$

- Reciprocal of $x = \frac{1}{x}$

Harmonic Mean: Grouped Data

- If $x_1, x_2, \ldots, \ldots, x_k$ - k non zero positive quantities with corresponding frequencies $f_1, f_2, \ldots, \ldots, fk$ (where $\sum_{i=1}^{k} f_i = n$), then the Harmonic mean HM will be calculated as follows

$$HM = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \cdots + \frac{f_k}{x_k}}$$

Test Yourself

Evan was driving up a mountainous road, with some flat stretches among the steep climbs. As a result, his speed varied for each 10 kilometers, as seen below:

Distance Travelled (km)	10	10	10	10	10
Speed (km/h)	10	25	6	15	8

What was his average speed throughout the drive?

Solution

Harmonic Mean = $\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\cdots+\frac{1}{x_n}}$

Number of observations, n = 5

Harmonic Mean = $\frac{5}{\frac{1}{10}+\frac{1}{25}+\frac{1}{6}+\frac{1}{15}+\frac{1}{8}}$ = 10.03

Therefore, the average speed throughout the drive was 10.03 km/h

Test Yourself

Selim rides his bicycle to deliver food. Given below is his speed for each kilometer he travelled today.

Kilometer	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
Speed (km/h)	3	2	6	4	4	5

What was his average speed on the 6km ride?

Solution

$$\text{Harmonic Mean} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\cdots+\frac{1}{x_n}}$$

$$\text{Harmonic Mean} = \frac{6}{\frac{1}{3}+\frac{1}{2}+\frac{1}{6}+\frac{1}{4}+\frac{1}{4}+\frac{1}{5}} = 3.53$$

Therefore, his average speed on the 6km ride is 3.53 km/h

Significance of Harmonic Mean

- Harmonic mean is used for calculating mean data for values obtained by combining two scales, like distance and time for speed.
- Harmonic means are often used in **averaging ratios** (e.g., the average travel speed given a duration of several trips, combined work rates).
- The weighted harmonic mean is used in finance to average multiples like the price-earnings ratio because it gives equal weight to each data point.
- Other uses: Thin lens equation in optics, parallel resistors in electricity etc.

Harmonic Mean

Merits	Demerits
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.	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.

Weighted Mean

Weighted Mean

- The **weighted mean** is a **special case** of the arithmetic mean. It occurs when there are **several observations** of the same value.
- Instead of each data point contributing equally to the final mean, some data points contribute more **"weight"** than others.
- Note: If all the weights are equal, then the weighted mean equals the arithmetic mean.
- Suppose Shumi's Hot Cake offers three different kinds of burger packages- small, medium and large for Tk. 100, Tk. 125 and Tk. 150 respectively.
- Of the last 10 burgers sold 3 were small, 4 were medium and 3 were large.
- Using the usual formula of the arithmetic mean, the average price of burgers sold:

$$\overline{x} = \frac{100 + 100 + 100 + 125 + 125 + 125 + 125 + 150 + 150 + 150}{10} = 125$$

Weighted Mean

- The mean selling price of the last 10 burger packages sold is Tk. 125.
- An easier way to find the mean selling price is to determine the weighted mean.
- In this method we multiply each observation by the number of times it happens as described below –

$$\bar{x}_W = \frac{(3*100)+ (4*125) + (3*150)}{10} = \frac{1250}{10} = 125$$

Weighted Mean: Formula

- In the last problem, the weights were frequency counts
- However, any measure of importance could be used as a weight.
- In general, the weighted mean of a set of numbers designated $X_1, X_2, \ldots., X_n$ with the corresponding weights $W_1, W_2, \ldots., W_n$ is computed by:

$$\bar{x}_W = \frac{\sum(WX)}{\sum W} = \frac{W_1X_1 + W_2X_2 + \cdots + W_nX_n}{W_1 + W_2 + \cdots + W_n}$$

Example

Madina Construction Company pays its part time employees on an hourly basis. For different level of employee, the hourly rates are Tk. 50, Tk. 75 and Tk. 90 respectively. There are 260 hourly employees, 140 of which are paid at Tk. 50 rate, 100 at Tk. 75 and 20 at the Tk. 90 rate. What is the mean hourly rate paid to the employees?

- To find the mean hourly rate, we multiply each of the hourly rates by the number of employees earning that rate as follows -

$$\bar{x}_W = \frac{\sum(WX)}{\sum W} = \frac{140*50 + 100*75 + 20*90}{140+100+20} = \frac{16300}{260} = \text{Tk } 62.69$$

- The weighted mean hourly wage is Tk. 62.69 or Tk. 63.00 (approximately).

Test Yourself

The postal service handles five basic types of letters and cards: 1st class, airmail, special delivery, registered and certified. The mail volume during a given year is given in the following table.

Types of Mailing	gm delivered (in millions)	Price per gm
1st class	77600	0.13
AIR mail	19000	0.17
Special delivery	1300	0.35
Registered mail	750	0.40
Certified mail	800	0.45

What was the average revenue per gm for these services during the year?

Solution

Types of Mailing	gm delivered (in millions)	Price per gm	Total price of each type (price * gm delivered)
1^{st} class	77600	0.13	10088
AIR mail	19000	0.17	3230
Special delivery	1300	0.35	455
Registered mail	750	0.40	300
Certified mail	800	0.45	360

Total gm delivered (in millions) = 77600 + 19000 + 1300 + 750 + 800 = 99450

Total price of all the types of Mailing = 10088 + 3230 + 455 + 300 + 360 = 14433

Average revenue of all types of services per gm = $\frac{\text{Total price of all the types of Mailing}}{\text{Total gm delivered (in millions)}}$

$$= \frac{14433}{99450} = 0.15$$

Quartiles, Deciles, Percentiles

Quartiles

- If the items in a series are arranged in ascending order of their magnitudes then those values of the variable that divide the total frequency in to **four equal parts** are called **quartiles**.
- There are three quartiles denoted by $\boldsymbol{Q_1, Q_2, \text{and } Q_3}$. The second quartile ($Q_2$) coincides with the **median**. The lower quartile (Q_1) is the point such that one-fourth of the total frequency is less than Q_1 and three-fourth is greater than Q_3.
- Note: Median divides the dataset into 2 equal halves, thus coinciding with $Q_{2.}$

Determination of Quartile

- For raw data,
  - If $i = 1,2,3$
  - $n = number\ of\ values$

  Then, $i^{th} quartile,\ Q_i = \frac{1}{2}\ [\ \left(\frac{in}{4}\right)^{th}\ value + (\frac{in}{4} + 1)^{th}\ value)],\ \boldsymbol{when\ \frac{in}{4}\ is\ an\ integer}$

  Or, $i^{th} quartile,\ Q_i = next\ integer^{th} value\ of\ \frac{in}{4},\ \boldsymbol{when\ \frac{in}{4}\ is\ not\ an\ integer}$

Deciles and Percentiles

- **Deciles** (Denoted by $D_1, D_2, \ldots, \ldots, \ldots, D_9$) are the quantities of the variable that divides the total frequencies into **10 equal parts**.
- **Percentiles** (Denoted by $P_1, P_2, \ldots, \ldots, \ldots, P_{99}$) are the quantities of the variable that divides the total frequencies into **100 equal parts**.
- Note: **Median** = $\boldsymbol{Q_2 = D_5 = P_{50}}$ >> All these quantities divide the total frequencies in to two equal halves.

Determination of Decile and Percentile

- For raw data,
  - If $i = 1,2,3,\ldots\ldots 9$ (For Decile) or $1,2,3,\ldots\ldots\ldots\ldots,99$
  - $n = number\ of\ values$

Then, $i^{th} decile,\ D_i = \frac{1}{2}\ [\ \left(\frac{in}{10}\right)^{th}\ value + (\frac{in}{10} + 1)^{th}\ value)],\ \boldsymbol{when\ \frac{in}{10}\ is\ an\ integer}$

Or, $i^{th}\ decile,\ D_i = next\ integer^{th} value\ of\ \frac{in}{10},\ \boldsymbol{when\ \frac{in}{10}\ is\ not\ an\ integer}$

And, $i^{th}\ percentile,\ P_i = \frac{1}{2}\ [\ \left(\frac{in}{100}\right)^{th}\ value + (\frac{in}{100} + 1)^{th}\ value)],\ \boldsymbol{when\ \frac{in}{100}\ is\ an\ integer}$

Or, $i^{th}\ percentile,\ P_i = next\ integer^{th} value\ of\ \frac{in}{100},\ \boldsymbol{when\ \frac{in}{100}\ is\ not\ an\ integer}$

Example

- Following is the marks in STA-201 obtained by 20 students in Summer 2020.

99	75	84	33	45	66	97	69	55	61
72	91	74	93	54	76	62	91	77	68

- Find out:
  - 1^{st} and 3^{rd} Quartiles
  - 3^{rd}, 6^{th}, 8^{th} Deciles
  - 20^{th}, 37^{th}, 60^{th}, 86^{th} Percentiles

Solution

- First, arrange the data.

33	45	54	55	61	62	66	68	69	72
74	75	76	77	84	91	91	93	97	99

Here, n =20

1st Quartile:

For *i=1*, $\frac{in}{4} = \frac{1*20}{4}$ = 5, an integer

So, $1^{st}\ quartile,\ Q_1 = \frac{1}{2}\ [\ 5^{th}\ value + (5+1)^{th}\ value)]$ = ½ (61+62) = 61.5

Solution

6^{th} Decile:

For $i=6$, $\frac{in}{10} = \frac{6*20}{10} = 12$, an integer

So, $6^{th}\ decile,\ D_6 = \frac{1}{2}\ [\,12^{th}\ value + 13^{th}\ value]$ = ½ (75+76) = 75.5

37^{th} Percentile:

For $i=37$, $\frac{in}{100} = \frac{37*20}{100} = 7.4$, not an integer

Next integer = 8

So, $37^{th}\ percentile,\ P_{37} = 8^{th}\ value = 68$