Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 22

Geometric Mean and Harmonic Mean

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

http://home.iitk.ac.in/~shalab/sp

Geometric Mean

Geometric mean is useful in calculating the average value of ratio or rate of interest etc.

Not applicable of any of the observation is zero.

Geometric Mean

$x_1, x_2, \ldots, x_n$ observations which are all positive.

The geometric mean for ungrouped or discrete data is

$$\overline{x}_G = \left(x_1 \times x_2 \times \ldots \times x_n\right)^{\frac{1}{n}}$$

Geometric Mean for Ungrouped Data

R Command

`x` : Data vector

Geometric mean for discrete data

```
prod(x)^(1/length(x))
```

(`length(x)` is equal to the number of elements in `x`)

Example

The time (in minutes) taken by a customer to arrive in a shop in a month on different days are recorded as follows:

Day	1	2	3	4	5	6	7	8	9	10
No. of minutes	30	31	30	30	29	29	29	29	29	28
Day	11	12	13	14	15	16	17	18	19	20
No. of minutes	28	28	27	27	27	26	26	26	26	25
Day	21	22	23	24	25	26	27	28	29	30	31
No. of minutes	25	25	25	25	25	24	24	23	22	21	21

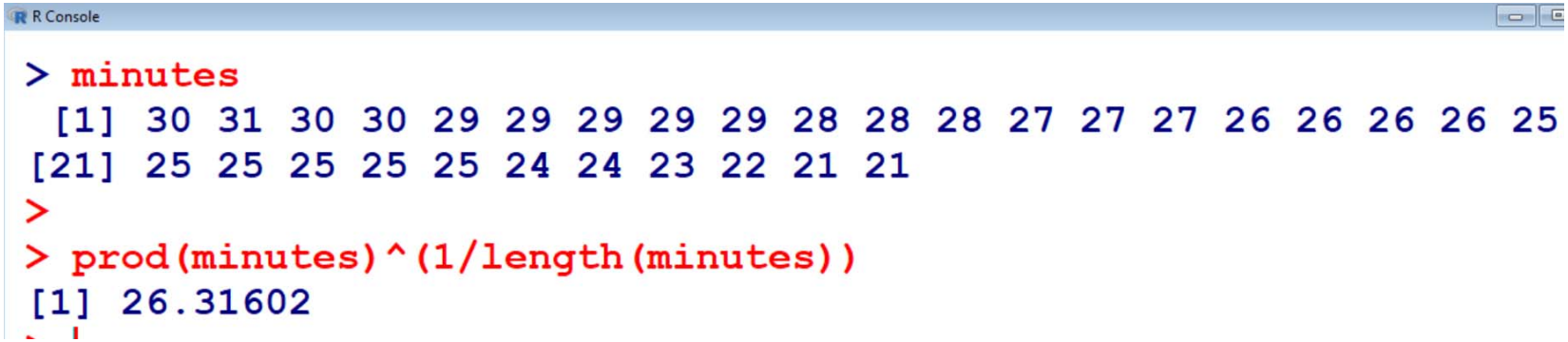
Geometric Mean for Ungrouped Data

Example: Considering it as ungrouped data

```
minutes = c(30,31,30,30,29,29,29,29,29,28,28,
28,27,27,27,26,26,26,26,25,25,25,25,25,25,24,2
4,23,22,21,21)
```

Geometric mean for discrete data

```
> prod(minutes)^(1/length(minutes))
[1] 26.31602
```



Geometric Mean for Grouped Data

Class intervals	Mid point (x_i)	Absolute frequency (f_i)
e_1- e_2	$x_1 = (e_1 + e_2)/2$	f_1
e_2- e_3	$x_2 = (e_2 + e_3)/2$	f_2
…	…	…
e_{K-1}- e_K	$x_K = (e_{K-1} + e_K)/2$	f_K

Grouped or continuous data with frequency distribution is

$$\overline{x}_G = (x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_k^{f_k})^{\frac{1}{N}} \quad \text{where} \quad N = \sum_{i=1}^{k} f_i$$

Note that $x_1, x_2, \ldots, x_k$ are the mid points of class intervals.

Geometric Mean for Grouped Data

R Command

`x` : Data vector of mid values c($x_1, x_2, \ldots, x_k$)

`f` : Frequency vector c($f_1, f_2, \ldots, f_k$)

where $x_1, x_2, \ldots, x_k$ are the mid points with frequencies f_1, $f_2, \ldots, f_k$ respectively.

Geometric mean for continuous data

```
prod(x^f)^(1/sum(f))
```

(`sum(f)` is equal to the sum of elements in `f`)

Geometric Mean for Grouped Data

Example: Considering it as grouped data

```
minutes = c(30,31,30,30,29,29,29,29,29,28,28,
28,27,27,27,26,26,26,26,25,25,25,25,25,25,24,2
4,23,22,21,21)
```

Geometric Mean for Grouped Data

Example: Considering the data as grouped data, we can present the data as

Class intervals	Mid point (x_i)	Absolute frequency (f_i)
15 – 20	17.5	$f_1 = 0$
20 – 25	22.5	$f_2 = 12$
25 – 30	27.5	$f_3 = 18$
30 – 35	32.5	$f_4 = 1$
35 – 40	37.5	$f_5 = 0$

Geometric Mean for Grouped Data

R command

Example

Frequency distribution

```
> breaks = seq(15, 40, by=5) # sequence at
                        interval of 5 integers
> breaks
[1] 15 20 25 30 35 40

> minutes.cut = cut(minutes,breaks,right=FALSE)
```

Geometric Mean for Grouped Data

R command

Example

Frequency distribution

```
> minutes.cut = cut(minutes,breaks,right=FALSE)

> minutes.cut
 [1] [30,35) [30,35) [30,35) [30,35) [25,30) [25,30) [25,30) [25,30)
 [9] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[17] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[25] [25,30) [20,25) [20,25) [20,25) [20,25) [20,25) [20,25)
Levels: [15,20) [20,25) [25,30) [30,35) [35,40)
```

Recall

```
minutes  =  c(30,31,30,30,29,29,29,29,29,28,28,
28,27,27,27,26,26,26,26,25,25,25,25,25,25,24,24
,23,22,21,21)
```

Geometric Mean for Grouped Data

R command Example

Frequency distribution

```
> table(minutes.cut)

minutes.cut

[15,20) [20,25) [25,30) [30,35) [35,40)

      0       6      21       4       0
```

Extract frequencies from frequency table using command

```
as.numeric(frequency table data)
```

```
> f = as.numeric(table(minutes.cut))
> f
[1]  0  6 21  4  0
```

Geometric Mean for Grouped Data

R command

Example

```
> x = c(17.5,22.5,27.5,32.5,37.5) # Mid values
> f = as.numeric(table(minutes.cut))
> f
[1]  0  6 21  4  0
```

Geometric mean for continuous data

```
> prod(x^f)^(1/sum(f))
[1] 27.02877
```

Mode for Grouped Data

R command of mode

Example

R Console

```
> breaks = seq(15, 40, by=5)
> breaks
[1] 15 20 25 30 35 40
> minutes = c(30,31,30,30,29,29,29,29,29,28, 28,28,27,27,27,26,26,26,2$
> minutes
 [1] 30 31 30 30 29 29 29 29 29 28 28 28 27 27 27 26 26 26 26 25 25 25
[23] 25 25 25 24 24 23 22 21 21
> minutes.cut = cut(minutes,breaks,right=FALSE)
> minutes.cut
 [1] [30,35) [30,35) [30,35) [30,35) [25,30) [25,30) [25,30) [25,30)
 [9] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[17] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[25] [25,30) [20,25) [20,25) [20,25) [20,25) [20,25) [20,25)
Levels: [15,20) [20,25) [25,30) [30,35) [35,40)
> table(minutes.cut)
minutes.cut
[15,20) [20,25) [25,30) [30,35) [35,40)
      0       6      21       4       0
```

Mode for Grouped Data

R command of mode

Example

R Console

```
> f = as.numeric(table(minutes.cut))
> f
[1]  0  6 21  4  0
> x = c(17.5,22.5,27.5,32.5,37.5)
> x
[1] 17.5 22.5 27.5 32.5 37.5
>
> prod(x^f)^(1/sum(f))
[1] 27.02877
>
```

Harmonic mean

Observations: $x_1, x_2, \ldots, x_n$

For discrete data

$$\overline{x}_H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

Harmonic mean for discrete data

R Command

`x` : Data vector

Harmonic mean for discrete data

```
1/mean(1/x)
```

Example

The time (in minutes) taken by a customer to arrive in a shop in a month on different days are recorded as follows:

Day	1	2	3	4	5	6	7	8	9	10
No. of minutes	30	31	30	30	29	29	29	29	29	28
Day	11	12	13	14	15	16	17	18	19	20
No. of minutes	28	28	27	27	27	26	26	26	26	25
Day	21	22	23	24	25	26	27	28	29	30	31
No. of minutes	25	25	25	25	25	24	24	23	22	21	21

Harmonic mean for Ungrouped Data

Example

```
minutes = c(30,31,30,30,29,29,29,29,29,28,28,
28,27,27,27,26,26,26,26,25,25,25,25,25,25,24,2
4,23,22,21,21)
```

```
>  1/mean(1/minutes)
[1] 26.17633
```

Harmonic Mean for Grouped Data

Class intervals	Mid point (x_i)	Absolute frequency (f_i)
e_1- e_2	$x_1 = (e_1 + e_2)/2$	f_1
e_2- e_3	$x_2 = (e_2 + e_3)/2$	f_2
…	…	…
e_{K-1}- e_K	$x_K = (e_{K-1} + e_K)/2$	f_K

For grouped or continuous data with frequency distribution

$$\bar{x}_H = \frac{1}{\frac{1}{N}\sum_{i=1}^{K}\left(\frac{f_i}{x_i}\right)} \qquad \text{where} \qquad N = \sum_{i=1}^{K} f_i$$

Note that $x_1, x_2, \ldots, x_k$ are the mid points of class intervals.

Harmonic mean for contineous data

R Command

x : Data vector c($x_1, x_2, \ldots, x_k$)

f : Frequency vector c($f_1, f_2, \ldots, f_k$)

where $x_1, x_2, \ldots, x_k$ occur with frequencies $f_1, f_2, \ldots, f_k$ respectively.

Harmonic mean for continuous data

```
sum(f)/sum(f/x)
```

Harmonic Mean for Grouped Data

Example: Considering the data as grouped data, we can present the data as

Class intervals	Mid point (x_i)	Absolute frequency (f_i)
15 – 20	17.5	$f_1 = 0$
20 – 25	22.5	$f_2 = 12$
25 – 30	27.5	$f_3 = 18$
30 – 35	32.5	$f_4 = 1$
35 – 40	37.5	$f_5 = 0$

Harmonic Mean for Grouped Data

R command

Example

Frequency distribution

```
> breaks = seq(15, 40, by=5) # sequence at
                        interval of 5 integers
> breaks
[1] 15 20 25 30 35 40

> minutes.cut = cut(minutes,breaks,right=FALSE)
```

Harmonic Mean for Grouped Data

R command

Example

Frequency distribution

```
> minutes.cut = cut(minutes,breaks,right=FALSE)

> minutes.cut
 [1] [30,35) [30,35) [30,35) [30,35) [25,30) [25,30) [25,30) [25,30)
 [9] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[17] [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30) [25,30)
[25] [25,30) [20,25) [20,25) [20,25) [20,25) [20,25) [20,25)
Levels: [15,20) [20,25) [25,30) [30,35) [35,40)
```

Recall

```
minutes  =  c(30,31,30,30,29,29,29,29,29,28,28,
28,27,27,27,26,26,26,26,25,25,25,25,25,25,24,24
,23,22,21,21)
```

Harmonic Mean for Grouped Data

R command Example

Frequency distribution

```
> table(minutes.cut)

minutes.cut

[15,20) [20,25) [25,30) [30,35) [35,40)

      0       6      21       4       0
```

Extract frequencies from frequency table using command

```
as.numeric(frequency table data)
```

```
> f = as.numeric(table(minutes.cut))
> f
[1]  0  6 21  4  0
```

Harmonic mean

Example

```
> x = c(17.5,22.5,27.5,32.5,37.5) # Mid values
> f = as.numeric(table(minutes.cut))
> f
[1]  0  6 21  4  0
```

Harmonic mean for continuous data

```
> sum(f)/sum(f/x)
[1] 26.87753
```

Harmonic mean

Example