Descriptive Statistics With R Software

Central Tendency of Data

::

Mode, Geometric Mean and Harmonic Mean

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Mode

Examples:

- **A fruit juice shop owner wants to know that which of the fruit juice is more preferred.**
- **A clothing shop owner wants know that the size of the shirt and or trouser is highest in demand.**

Mode of n observations $x_1, x_2, \ldots, x_n$ is the value which occurs the most, compared with all other values.

Mode

Mode is the value which occurs most frequently in a set of observations.

Mode is not at all affected by extreme observations.

Distributions having only one mode are called <u>unimodal</u> and the one with two modes is called <u>bimodal</u>.

Mode for Ungrouped Data

For discrete variables, the mode of a variable is the value of the variable having the highest frequency in a unimodal distribution.

Mode for Ungrouped Data

R command

Step 1: Create a table of given data vector or matrix `data`, say `modetab` as follows:

```
modetab = table(as.vector(data))
```

The first row of `modetab` is a sorted list of all unique values in `data`.

Step 2: Following returns the names of the values having the highest count in second row of `modetab` which is the mode.

```
names(modetab)[modetab == max(modetab)]
```

Mode for Ungrouped Data

R command

```
> data = c(10,10,10,10,2,2,3,4,5,6)
```

Create a table of given data vector `data`

```
> modetab = table(as.vector(data))
> modetab

 2  3  4  5  6 10
 2  1  1  1  1  4
```

sorted list of all unique values in `data`

names of values having highest count

```
> names(modetab)[modetab == max(modetab)]
[1] "10"
```

Mode for Ungrouped Data

R command

R Console

```
> data = c(10,10,10,10,2,2,3,4,5,6)
>
> modetab = table(as.vector(data))
> modetab

 2  3  4  5  6 10
 2  1  1  1  1  4
>
> names(modetab)[modetab == max(modetab)]
[1] "10"
```

Mode for Ungrouped Data

R command

```
> data = matrix(nrow= 3, ncol=3, data=c(1,2,2,
3,3,4,5,6,6))
> data
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    3    6
[3,]    2    4    6
```

Mode for Ungrouped Data

R command

Create a table of given data matrix `data`

```
> modetab = table(as.vector(data))

> modetab

1 2 3 4 5 6

1 2 2 1 1 2

> names(modetab)[modetab == max(modetab)]

[1] "2" "3" "6"
```

Mode for Ungrouped Data

R command

R Console

```
> data = matrix(nrow=3, ncol=3, data=c(1,2,2,3,3,4,5,6,6))
> data
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    3    6
[3,]    2    4    6
> modetab = table(as.vector(data))
> modetab

1 2 3 4 5 6
1 2 2 1 1 2
> names(modetab)[modetab == max(modetab)]
[1] "2" "3" "6"
>
```

Mode for Grouped Data

For <u>continuous variable</u>, the mode is the value of the variable with the highest frequency density corresponding to the ideal distribution which would be obtained if the total frequency were increased indefinitely and if, at the same time, the width of the class intervals were decreased indefinitely.

Mode for Grouped Data

Class intervals	Mid point (m_i)	Absolute frequency (f_i)
$e_1 - e_2$	$m_1 = (e_1 + e_2)/2$	f_1
$e_2 - e_3$	$m_2 = (e_2 + e_3)/2$	f_2
…	…	…
$e_{K-1} - e_K$	$m_K = (e_{K-1} + e_K)/2$	f_K

$$\sum_{i=1}^{K} f_i = n$$

Modal class: Class corresponding to the maximum frequency.

Mode for Grouped Data

$$\overline{x}_{mo} = e_l + \frac{f_0 + f_{-1}}{(f_0 - f_1) + (f_0 - f_{-1})} d_l$$

e_l : **lower limit of modal class**

d_l : **class width**

f_0 : **frequency of modal class**

f_{-1} : **frequency of the class just before the modal class**

f_1 : **frequency of the class just after the modal class**

Mode for Grouped Data

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

Mode 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	$f_5 = 0$

Modal class: Class corresponding to the maximum frequency.

$l = 3 : 25 - 30$

Mode for Grouped Data

Example:

e_l = 25 : lower limit of modal class

d_l = 5 : class width

f_0 = 18 : frequency of modal class

f_{-1} = 12 frequency of the class just before the modal class

f_1 = 1: frequency of the class just after the modal class

$$\overline{x}_{mo} = e_l + \frac{f_0 + f_{-1}}{(f_0 - f_1) + (f_0 - f_{-1})} d_l$$

$$= 25 + \frac{18 + 12}{(18 - 1) + (18 - 12)} \times 5 \approx 31.52$$

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**

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

- **Grouped or continuous data with frequency distribution is**

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

where $x_1, x_2, \ldots, x_n$ occur with frequencies $f_1, f_2, \ldots, f_n$ respectively.

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`)

Geometric Mean for Grouped Data

R Command

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

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

where $x_1, x_2 ,\ldots, x_n$ occur with frequencies $f_1, f_2 ,\ldots, f_n$ 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 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
```

R Console

```
> minutes
 [1] 30 31 30 30 29 29 29 29 29 28 28 28 27 27 27 26 26 26 26 25
[21] 25 25 25 25 24 24 23 22 21 21
>
> prod(minutes)^(1/length(minutes))
[1] 26.31602
```

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

```
> 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)
```

R Console

```
> minutes
 [1] 30 31 30 30 29 29 29 29 29 28 28 28 27 27 27 26 26 26 26 25
[21] 25 25 25 25 25 24 24 23 22 21 21
>
> median(minutes)
[1] 26
```

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)

> 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)
```

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

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

For continuous data having frequency distribution

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

where $N = \sum_{i=1}^{n} f_i$

Harmonic mean for discrete data

R Command

`x` : **Data vector**

Harmonic mean for discrete data

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

Harmonic mean for contineous data

R Command

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

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

where $x_1, x_2, \ldots, x_n$ occur with frequencies $f_1, f_2, \ldots, f_n$

respectively.

Harmonic mean for continuous data

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

Harmonic mean

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)
```

Harmonic mean for discrete data

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

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

```
> 1/mean(f/x)
[1] 4.335085
```

Harmonic mean

Example

R Console

```
> 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
> 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
>
> 1/mean(f/x)
[1] 4.335085
> 
```