

Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 18

Central Tendency of Data : Weighted Arithmetic Mean and Partition Values

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from
<http://home.iitk.ac.in/~shalab/sp>



Arithmetic Mean for Ungrouped Data

The arithmetic mean of observations x_1, x_2, \dots, x_n is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

`mean(x)` provides the value of arithmetic mean of the data in data vector **x**.

Arithmetic Mean for Grouped Data

Example

Following are the time taken (in seconds) by 20 participants in a race: 32, 35, 45, 83, 74, 55, 68, 38, 35, 55, 66, 65, 42, 68, 72, 84, 67, 36, 42, 58.

```
> time
```

```
[1] 32 35 45 83 74 55 68 38 35 55 66 65 42 68  
72 84 67 36 42 58
```

```
> time.na = c(NA, NA, 45, 83, 74, 55, 68, 38,  
35, 55, 66, 65, 42, 68, 72, 84, 67, 36, 42,  
58)
```

Arithmetic Mean for Ungrouped Data

R command of mean

Mean of 20 available values (No missing values)

```
> mean(time)
```

```
[1] 56
```

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i$$

Mean of 18 values (Two values are missing as **NA**)

```
> mean(time.na, na.rm=TRUE)
```

```
[1] 58.5
```

$$\bar{x} = \frac{1}{18} \sum_{i=1}^{18} x_i$$

Arithmetic Mean for Grouped Data

Class intervals	Mid point (m_i)	Absolute frequency (n_i)	Relative frequency (f_i)
$e_1 - e_2$	$m_1 = (e_1 + e_2)/2$	n_1	f_1
$e_2 - e_3$	$m_2 = (e_2 + e_3)/2$	n_2	f_2
...
$e_{K-1} - e_K$	$m_K = (e_{K-1} + e_K)/2$	n_K	f_K

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

Arithmetic Mean for Grouped Data

The arithmetic mean for grouped data, is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^K n_i m_i = \sum_{i=1}^K f_i m_i$$

Another version: **Weighted arithmetic mean**

Weight : w_i

$$\bar{x} = \frac{\sum_{i=1}^K w_i m_i}{\sum_{i=1}^K w_i}$$

Arithmetic Mean for Grouped Data

R command of mean

The arithmetic mean for grouped data is

$$m = c(m_1, m_2, \dots, m_n)$$

$$f = c(f_1, f_2, \dots, f_n)$$

`weighted.mean(m, f)`

Note: Please note that `f` in `weighted.mean(m, f)` is denoting the data vector of absolute frequencies and not the data vector of relative frequencies.

Arithmetic Mean for Grouped Data

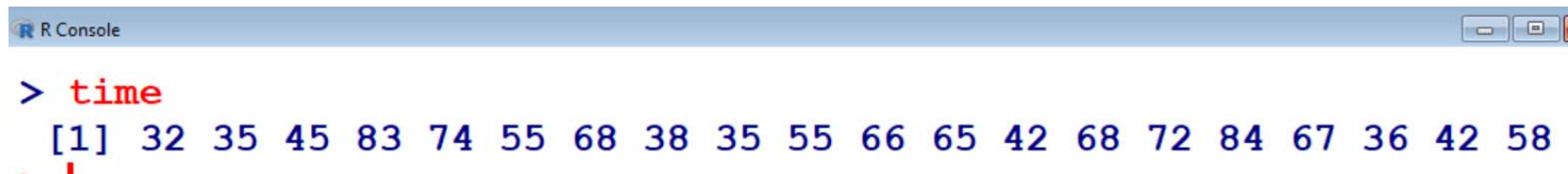
R command of mean

Example

Following are the time taken (in seconds) by 20 participants in a race: 32, 35, 45, 83, 74, 55, 68, 38, 35, 55, 66, 65, 42, 68, 72, 84, 67, 36, 42, 58.

```
> time
```

```
[1] 32 35 45 83 74 55 68 38 35 55 66 65 42 68  
72 84 67 36 42 58
```



```
R Console  
> time  
[1] 32 35 45 83 74 55 68 38 35 55 66 65 42 68 72 84 67 36 42 58
```


Arithmetic Mean for Grouped Data

R command of mean

Example

Class intervals	Mid point	Absolute frequency (or frequency)
31 – 40	35.5	5
41 – 50	45.5	3
51 – 60	55.5	3
61 – 70	65.5	5
71 – 80	75.5	2
81 - 90	85.5	2
	Total	20

Arithmetic Mean for Grouped Data

R command of mean

Example

Frequency distribution

```
> breaks = seq(30, 90, by=10) # sequence at  
                                interval of 10 integers
```

```
> breaks
```

```
[1] 30 40 50 60 70 80 90
```

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

```
> time.cut
```

```
[1] [30,40) [30,40) [40,50) [80,90) [70,80) [50,60) [60,70)  
[8] [30,40) [30,40) [50,60) [60,70) [60,70) [40,50) [60,70)  
[15] [70,80) [80,90) [60,70) [30,40) [40,50) [50,60)  
Levels: [30,40) [40,50) [50,60) [60,70) [70,80) [80,90)
```

Arithmetic Mean for Grouped Data

R command of mean

Example

Frequency distribution

```
> table(time.cut)
```

```
time.cut
```

```
[30,40) [40,50) [50,60) [60,70) [70,80) [80,90)  
      5      3      3      5      2      2
```

Extract frequencies from frequency table using command

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

```
> f = as.numeric(table(time.cut))
```

```
[1] 5 3 3 5 2 2
```

Arithmetic Mean for Grouped Data

R command of mean

Example

Weighted arithmetic mean

```
> m = c(35,45,55,65,75,85)
> f = as.numeric(table(time.cut))
[1] 5 3 3 5 2 2
```

Obtained from `as.numeric(table(time.cut))`

```
> weighted.mean(m,f)
[1] 56
```

Arithmetic Mean for Grouped Data

R command of mean

Example

```
R Console
> table(time.cut)
time.cut
[30,40) [40,50) [50,60) [60,70) [70,80) [80,90)
      5      3      3      5      2      2
> m = c(35,45,55,65,75,85)
> f=as.numeric(table(time.cut))
> f
[1] 5 3 3 5 2 2
> weighted.mean(m, f)
[1] 56
.
```

Partition Values

The frequency distribution is partitioned to have an idea about the concentration of values over the entire frequency distribution.

Several measures: Median, quartiles, deciles, percentiles.