Descriptive Statistics With R Software

Central Tendency of Data

::

Arithmetic Mean

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Measure of Central Tendency

Data set contains many variables. Every variable has many observations.

Difficult to handle each observation and dig out the information from every observation.

Our interest is in summary of information hidden inside the data.

Measure of Central Tendency

Example:

Suppose the last year‘s temperature (in degree centegrades) of following two cities in the month of May for 5 days are recorded as follows:

Lucknow: 35, 37, 36, 40, 38

Srinagar: 20, 18, 17, 22, 23

What type of clothings are needed to visit these two cities in the month of May?

Measure of Central Tendency

Natural human dendency is to compile the information in term of average.

For example, the average marks in a subject in a class are 60%.

A medicine tablet controls the fever for 6 hours.

Statistical concept referes to the “average“ or the central tendency of the data.

Measure of Central Tendency

- **Arithmetic mean**
- **Geometric mean**
- **Harmonic mean**
- **Median**
- **Quantiles**
- **Mode etc.**

Arithmetic Mean for Ungrouped Data

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

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

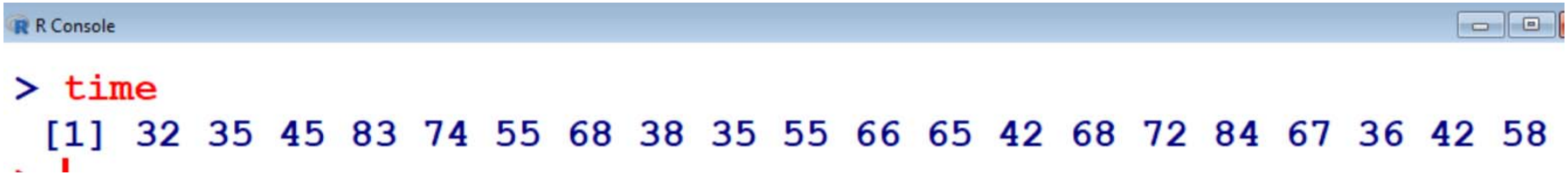
`mean(x)` provides the value of arithmetic mean of the data in data vector `x` .

Arithmetic Mean Arithmetic Mean for Ungrouped 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 = c(32, 35, 45, 83, 74, 55, 68, 38, 35,
55, 66, 65, 42, 68, 72, 84, 67, 36, 42, 58)
```

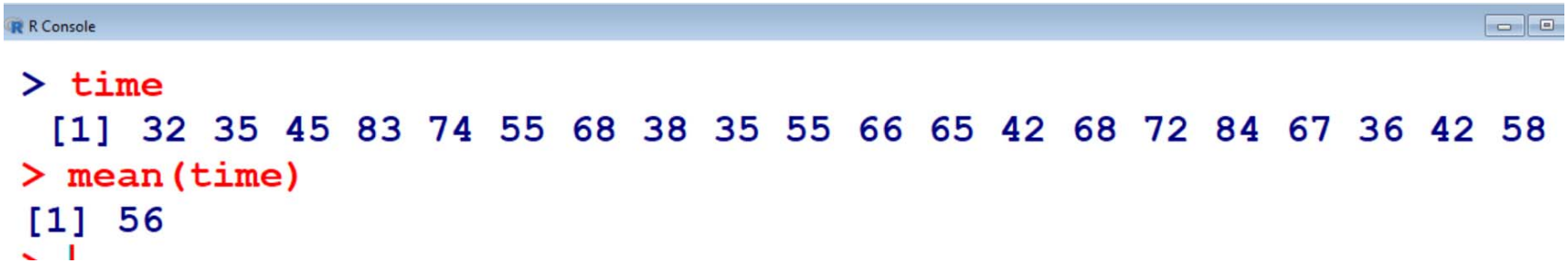


Arithmetic Mean Arithmetic Mean for Ungrouped Data

Example:

```
> mean(time)

[1] 56
```



Arithmetic Mean for Ungrouped Data

R command of mean

`mean(x, na.rm = TRUE)` provides the value of arithmetic mean when the data in data vector `x` is not available (`NA`).

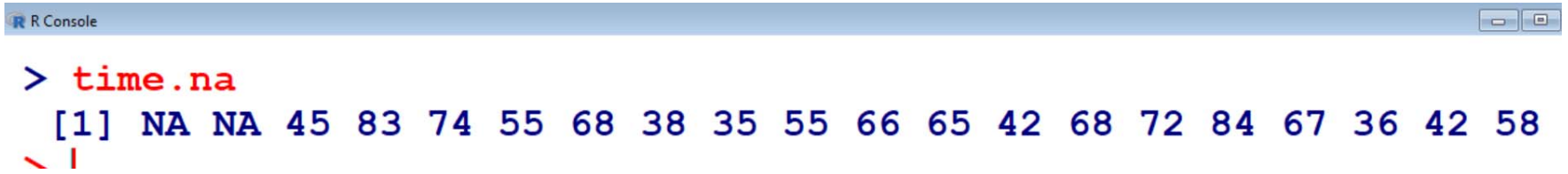
Arithmetic Mean for Ungrouped Data

R command of mean : Example

Suppose two data points are missing in the earlier example where the time taken (in seconds) by 20 participants in a race. They are recorded as NA

NA, NA, 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 : Example

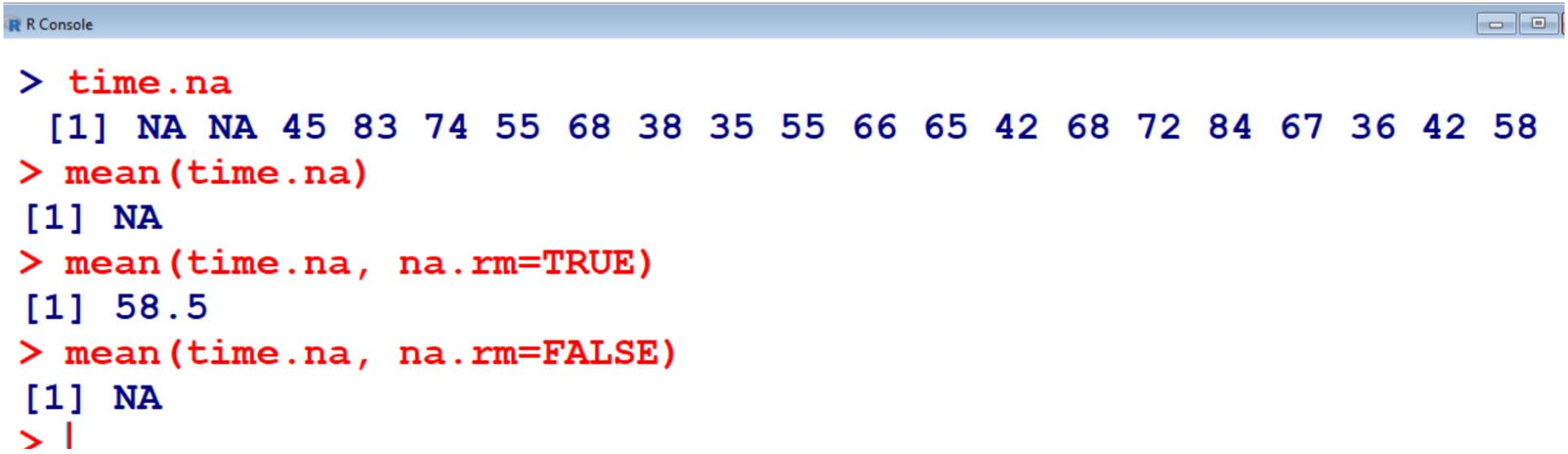
```
> mean(time.na)
[1] NA

> mean(time.na, na.rm=TRUE)
[1] 58.5

> mean(time.na, na.rm=FALSE) # default mean
[1] NA
```

Arithmetic Mean for Ungrouped Data

R command of mean : Example



Arithmetic Mean for Ungrouped Data

R command of mean : Example

Difference between `mean(time)` and `mean(time.na, na.rm=TRUE)`

Mean of 20 values

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

```
> mean(time)
[1] 56
```

Mean of 18 values

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

```
> mean(time.na, na.rm=TRUE)
[1] 58.5
```

Arithmetic Mean for Grouped Data

Class intervals	Mid point (m_i)	Absolute frequency (n_j)	Relative frequency (f_j)
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

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

Another version: Weighted arithmetic mean

Weight : w_i

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

Arithmetic Mean for Grouped Data

R command of mean

The arithmetic mean for grouped data is

$\mathtt{m = c(m_1, m_2, \ldots, m_n)}$

$\mathtt{f = c(f_1, f_2, \ldots, f_n)}$

```
weighted.mean(m,f)
```

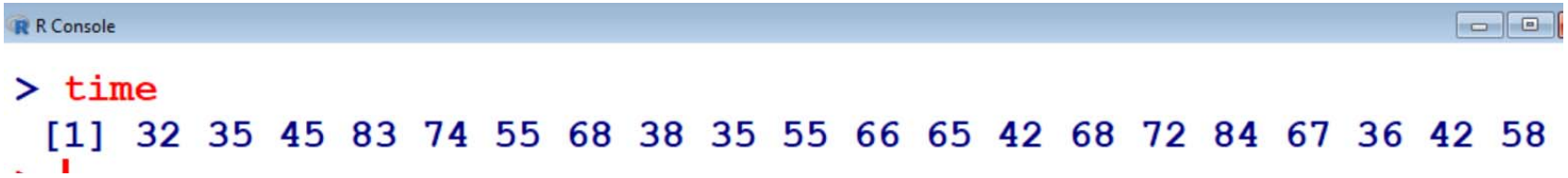
Arithmetic Mean for Grouped Data

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
```



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
```