Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 17

Central Tendency of Data : Arithmetic Mean

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

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

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

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

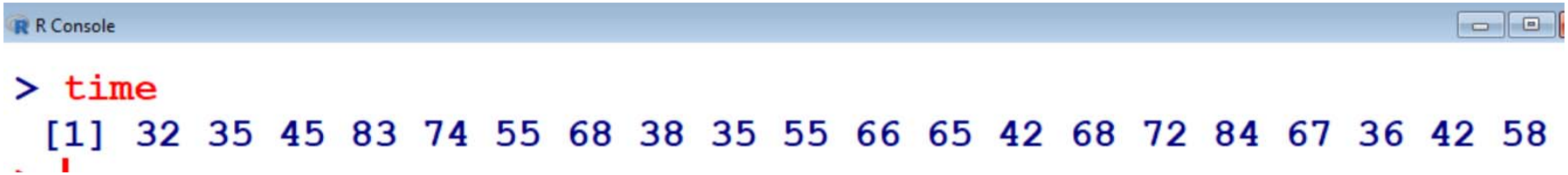
`mean(x)` provides the value of arithmetic mean of the data in data vector `x` .

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)
```

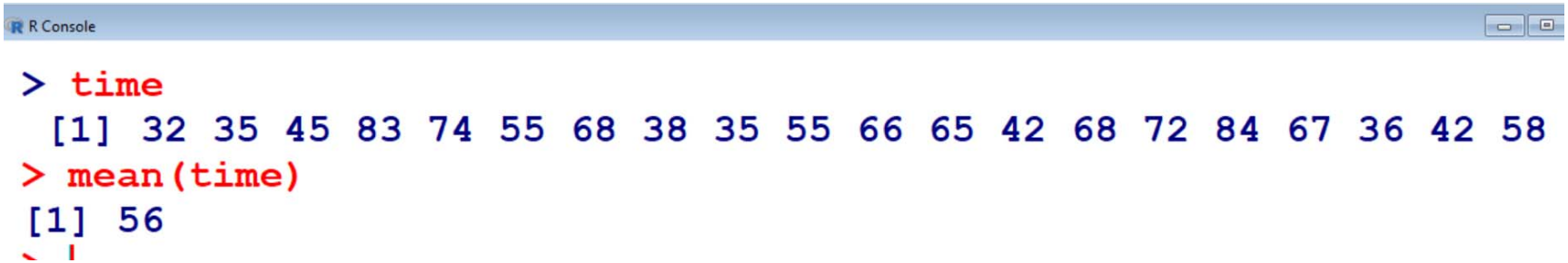


Arithmetic Mean Arithmetic Mean for Ungrouped Data

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`).

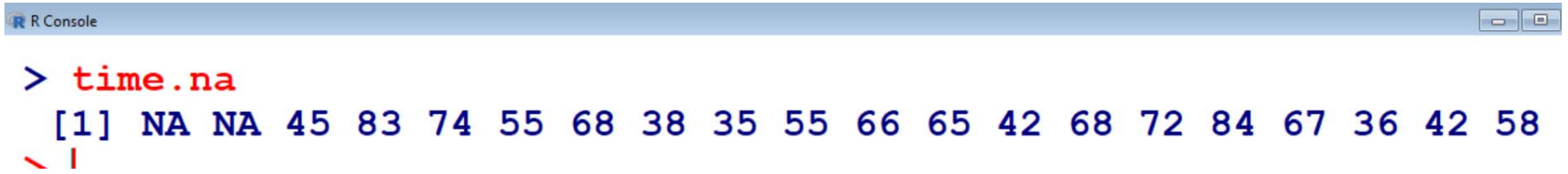
Arithmetic Mean for Ungrouped Data

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

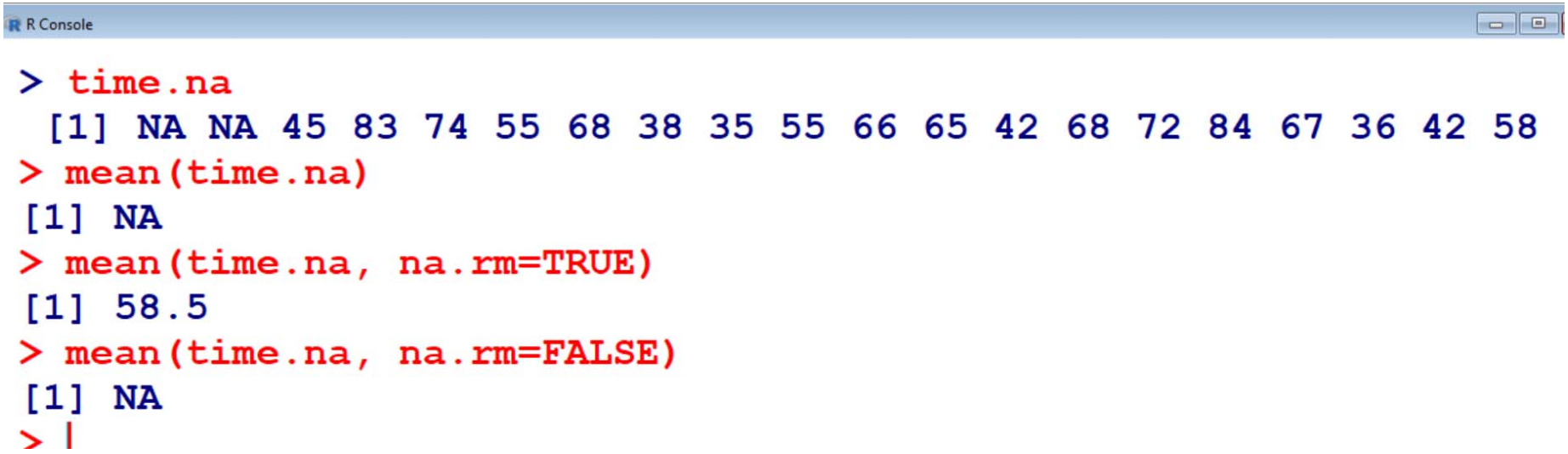
```
> mean(time.na)
[1] NA

> 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
```