Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 19

Partition Values – Median and Quantiles

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

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

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.

Median

Median is the value which divides the observations into two equal parts such that

- **at least 50% of the values are greater than or equal to the median and**
- **atleast 50% of the values are less than or equal to the median.**

Median is a better average than arithmetic mean in case of extreme observations.

Median for Ungrouped Data

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

Order the Observations : $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$

where $x_{(1)} = \min(x_1, x_2, \ldots, x_n)$

$x_{(n)} = \max(x_1, x_2, \ldots, x_n)$

$$\overline{x}_{med} = \begin{cases} x_{((n+1)/2)} & \text{if } n \text{ is odd integer} \\ \dfrac{x_{(n/2)} + x_{(n/2+1)}}{2} & \text{if } n \text{ is even integer.} \end{cases}$$

Median for Grouped Data

For grouped data, median is calculated assuming the values within each class are equally distributed $A_1, A_2, \ldots, A_K$: K classes

n_i = number of observations in i^{th} class A_i

Determine the median class A_m , i.e., the class which includes the median.

Median class A_m is the class for which

$$\sum_{i=1}^{m-1} f_i < 0.5 \quad \text{and} \quad \sum_{i=1}^{m} f_i \geq 0.5$$

Median for Grouped Data

Then median is

$$\overline{x}_{med} = e_m + \frac{0.5 - \sum_{i=1}^{m-1} f_i}{f_m} d_m$$

where e_m **: lower limit of** A_m

d_m **: width of** A_m

f_m **: relative frequency of** A_m

Median

Example: Median for ungrouped odd and even data

The number of 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

Median

Example: Median for ungrouped odd and even data

Consider this as ungrouped data

$$n = 31, \quad \frac{n+1}{2} = 16$$

$$\overline{x}_{med} = \overline{x}_{((n+1)/2)} = \overline{x}_{(16)} = 26.$$

Considering only 30 observations

$$n = 30, \quad \frac{n}{2} = 15, \quad \frac{n}{2} + 1 = 16$$

$$\overline{x}_{med} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2} = \frac{x_{(15)} + x_{(16)}}{2} = \frac{1}{2}(27 + 26) = 26.5$$

Median

Example: Median for grouped data

Counsedering the data as grouped data, we can present the data as

Class intervals	Absolute frequency (n_i)	Relative frequency (f_i)
$e_1 \equiv x_i < 20$	$n_1 = 0$	$f_1 = 0$
$e_2 \equiv 20 < x_i \leq 25$	$n_2 = 12$	$f_2 = 12/31$
$e_3 \equiv 25 < x_i \leq 30$	$n_3 = 18$	$f_3 = 18/31$
$e_4 \equiv 30 < x_i \leq 35$	$n_4 = 1$	$f_4 = 1/31$
$e_5 \equiv x_i > 35$	$n_5 = 0$	$f_5 = 0$
Total	$n = 31$	1

Median class : $m = 3 : e_3$ $\sum_{i=1}^{3-1} f_i = \frac{12}{31} < 0.5$ and $\sum_{i=1}^{3} f_i = \frac{30}{31} \geq 0.5$

Median

Example: Median for grouped data

$e_m = e_3 = 25$

$f_m = f_3 = 18/31$

$d_m = d_3 = 30 - 25 = 5$

$$\overline{x}_{med} = e_m + \frac{0.5 - \sum_{i=1}^{m-1} f_i}{f_m} d_m$$

$$= 25 + \frac{0.5 - 12/31}{18/31} \times 5$$

$$\approx 25.97$$

Median

R command

The R command for median is

```
median(x)
```

```
median(x, na.rm = TRUE, ...)
```

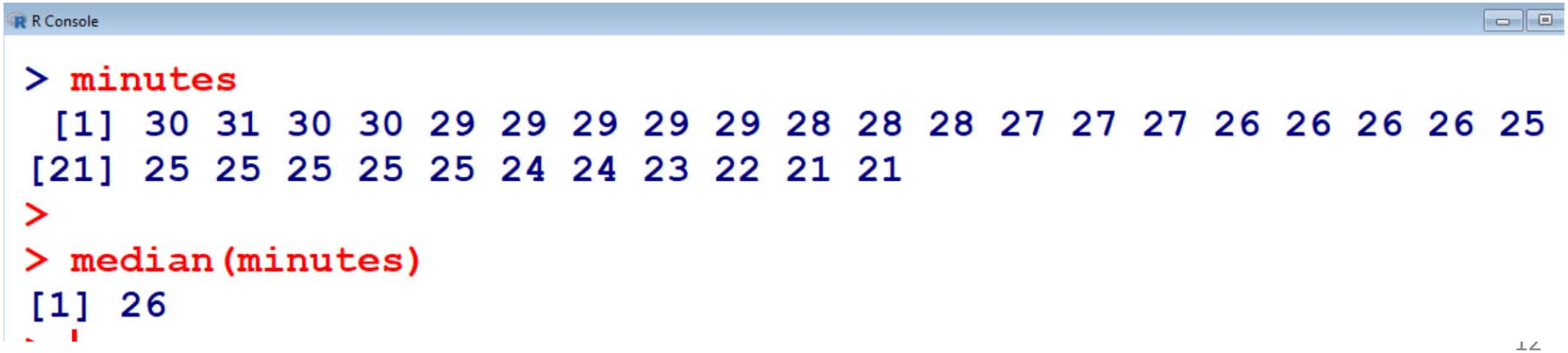
if obervations are missing as `NA`

Median

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)

> median(minutes)
[1] 26
```

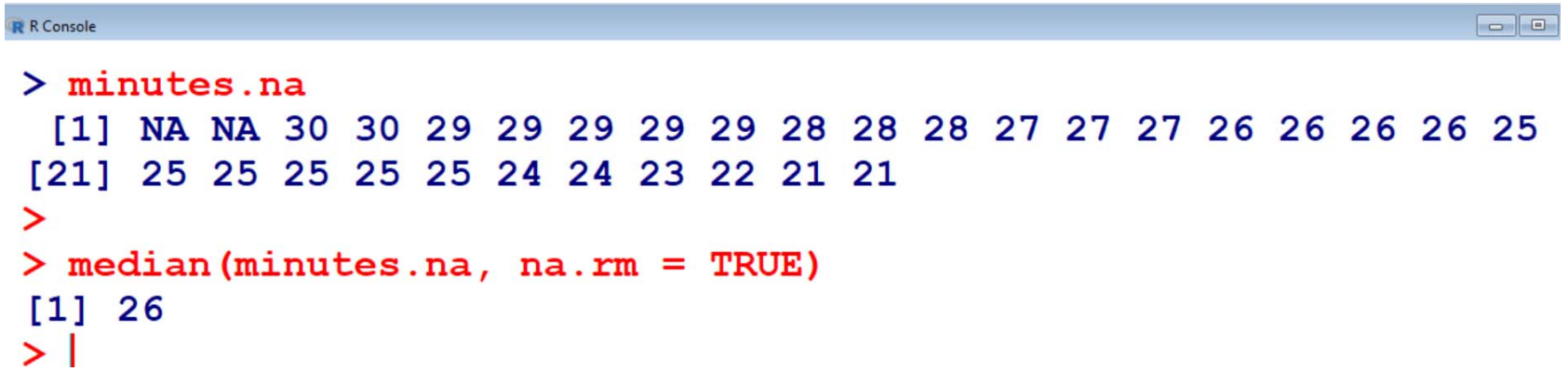


Median

Example with missing data

```
> minutes.na = c(NA,NA,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)


> median(minutes.na, na.rm = TRUE)
[1] 26
```



Quantiles

Median: Value which splits the data into two equal parts.

Quantile: Partitions the data into other proportions.