

Introduction to R Software

Swayam Prabha

Lecture 33

Introduction and Frequencies

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

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



Descriptive statistics:

First hand tools which gives first hand information.

- **Central tendency of data**
- **Variation in data**
- **Structure and shape of data tendency**
- **Relationship study**

Graphical as well as analytical tools are used.

Graphical tools:

Graphical tools- various type of plots

- 2D & 3D plots,**
- scatter diagram**
- Pie diagram**
- Histogram**
- Bar plot**
- Stem and leaf plot**
- Box plot ...**

Absolute and relative frequencies:

Suppose there are 10 persons coded into two categories as male (M) and female (F).

M, F, M, F, M, M, M, F, M, M.

Use a_1 and a_2 to refer to male and female categories.

There are 7 male and 3 female persons,

denoted as $n_1 = 7$ and $n_2 = 3$

The number of observations in a particular category is called the absolute frequency.

Absolute and relative frequencies:

The relative frequencies of a_1 and a_2 are

$$f_1 = \frac{n_1}{n_1 + n_2} = \frac{7}{10} = 0.7 = 70\%$$

$$f_2 = \frac{n_2}{n_1 + n_2} = \frac{3}{10} = 0.3 = 30\%$$

This gives us information about the proportions of male and female persons.

Absolute and relative frequencies:

`table(variable)` creates the absolute frequency of the `variable` of the data file.

Enter data as `x`

`table(x)` # absolute frequencies

`table(x)/length(x)` # relative frequencies

Absolute and relative frequencies:

Example: Code the 10 persons by using, say 2 for male (M) and 1 for female (F).

M, F, M, F, M, M, M, F, M, M
2, 1, 2, 1, 2, 2, 2, 1, 2, 2

```
> gender <- c(2,1,2,1,2,2,2,1,2,2)
```

```
> gender
```

```
[1] 2 1 2 1 2 2 2 1 2 2
```

R Console

```
> gender <- c(2,1,2,1,2,2,2,1,2,2)
```

```
> gender
```

```
[1] 2 1 2 1 2 2 2 1 2 2
```

```
>
```

Absolute and relative frequencies:

```
> table(gender) # Absolute frequencies
```

```
gender
```

```
1 2
```

```
3 7
```

```
R Console
> table(gender)
gender
1 2
3 7
>
```

```
> table(gender)/length(gender) #Relative freq.
```

```
gender
```

```
1 2
```

```
0.3 0.7
```

```
R Console
> table(gender)/length(gender)
gender
1 2
0.3 0.7
>
```

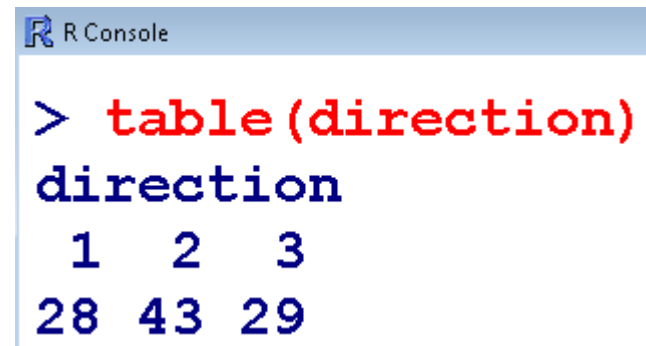

Example

Consider a data set on home delivery of grocery items. The home delivery is centrally managed over phone and delivered by one of the three branches (East- denoted as 1, West- denoted as 2, Central- denoted as 3) of the shop and the 100 data values are recorded on the directions where the grocery items are delivered.

```
direction <-  
c(1,1,2,1,2,3,2,2,3,3,3,1,2,3,2,2,3,1,  
1,3,3,1,2,1,3,3,3,2,2,2,2,1,2,2,1,1,1,3,2,2,1,2  
,3,2,2,1,2,3,3,2,1,2,2,3,1,1,2,1,2,3,2,3,2,2,3,  
1,2,3,3,3,2,1,1,1,2,1,1,2,1,2,3,3,1,2,3,3,2,1,2  
,3,2,1,3,2,2,2,2,3,2,2)
```

Example:

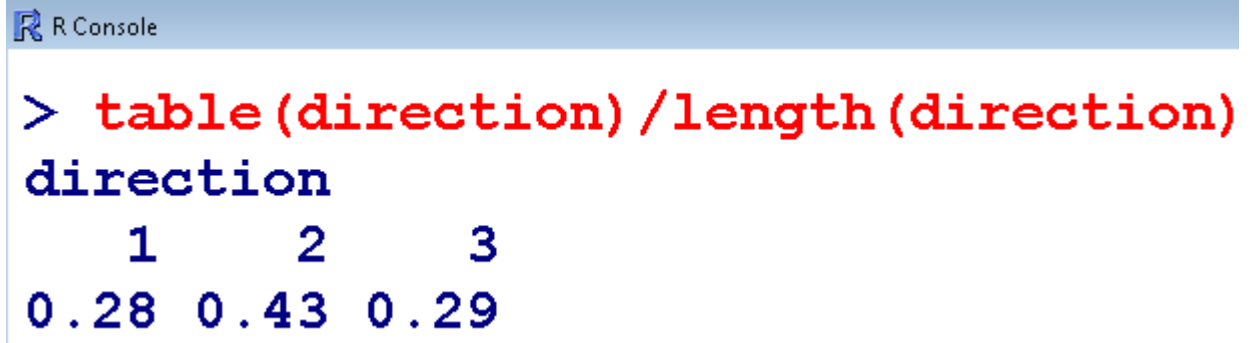
```
> table(direction)
direction
 1  2  3
28 43 29
```

A screenshot of an R console window. The title bar reads "R Console". The console shows the command "> table(direction)" in red text, followed by the output "direction" in blue text, and a table of counts: "1 2 3" on one line and "28 43 29" on the next line, both in blue text.

```
R Console
> table(direction)
direction
 1  2  3
28 43 29
```

Example:

```
> table(direction)/length(direction)
direction
      1      2      3
0.28 0.43 0.29
```



```
R Console
> table(direction)/length(direction)
direction
      1      2      3
0.28 0.43 0.29
```