Introduction to R Software

Introduction to Statistical Functions

:::

Introduction, Frequencies and Partition Values

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

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
```

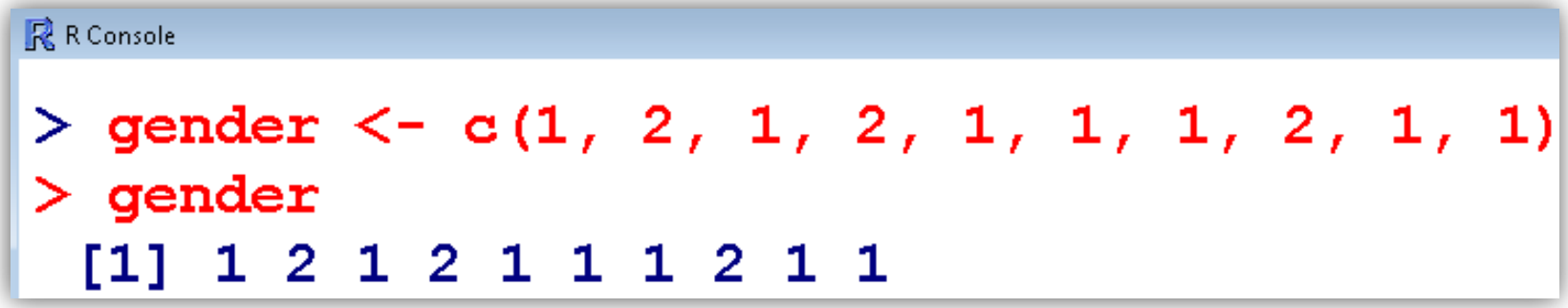
```
table(x)/length(x) # relative frequencies
```

Absolute and relative frequencies:

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

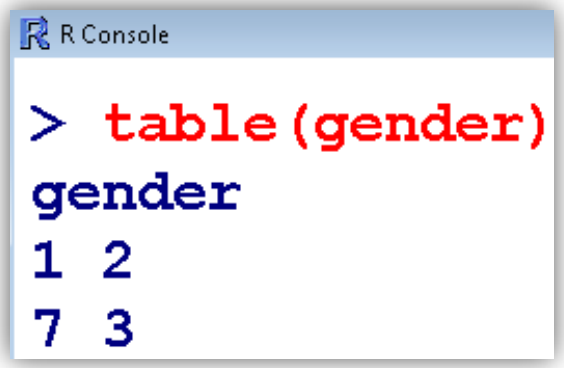
M,	F,	M,	F,	M,	M,	M,	F,	M,	M
1,	2,	1,	2,	1,	1,	1,	2,	1,	1

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

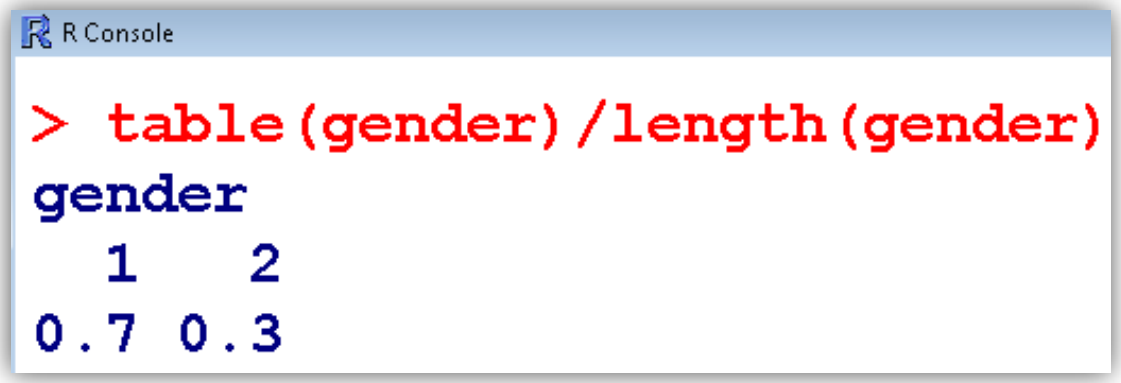


Absolute and relative frequencies:

```
> table(gender) # Absolute frequencies
gender
1 2
7 3
```




```
> table(gender)/length(gender) #Relative freq.
gender
  1   2
0.7 0.3
```



Example:

`'pizza_delivery.csv'` **contains the simulated data on pizza home delivery.**

- **There are three branches (East, West, Central) of the restaurant.**
- **The pizza delivery is centrally managed over phone and delivered by one of the five drivers.**
- **The data set captures the number of pizzas ordered and the final bill**

```
> setwd("C:/Rcourse")
> pizza <- read.csv('pizza_delivery.csv')
```

Example:

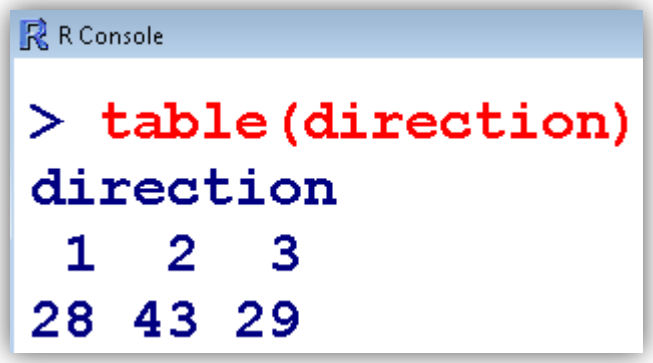
Consider data from Pizza. Take first 100 values from Direction and code Directions as

- **East: 1**
- **West: 2**
- **Centre: 3**

```
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)
```

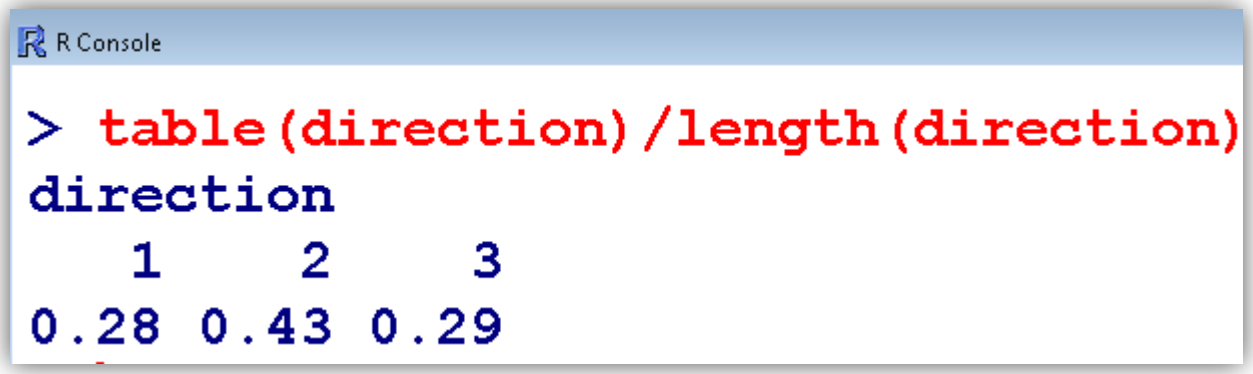
Example:

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



Example:

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



Partition values:

Such values divides the total frequency given data into required number of partitions.

Quartile: Divides the data into 4 equal parts.

Decile: Divides the data into 10 equal parts.

Percentile: Divides the data into 100 equal parts.

Partition values:

`quantile` function computes quantiles corresponding to the given probabilities.

The smallest observation corresponds to a probability of 0 and the largest to a probability of 1.

```
quantile(x, ...)

quantile(x, probs = seq(0, 1, 0.25),...)
```

Arguments

`x` **numeric vector whose sample quantiles are wanted,**

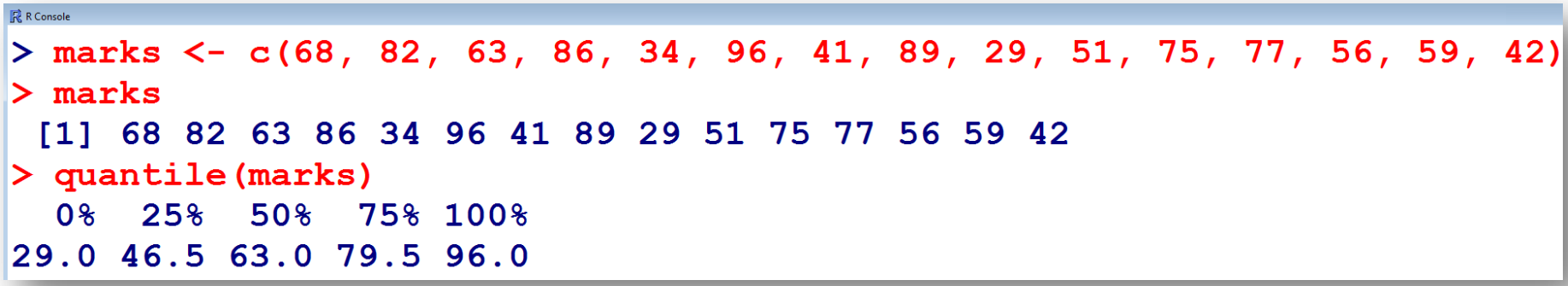
`probs` **numeric vector of probabilities with values in [0, 1].**

Partition values:

Example: Marks of 15 students are

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89,
  29, 51, 75, 77, 56, 59, 42)


>  quantile(marks)
  0%  25%  50%  75% 100%
29.0 46.5 63.0 79.5 96.0
```



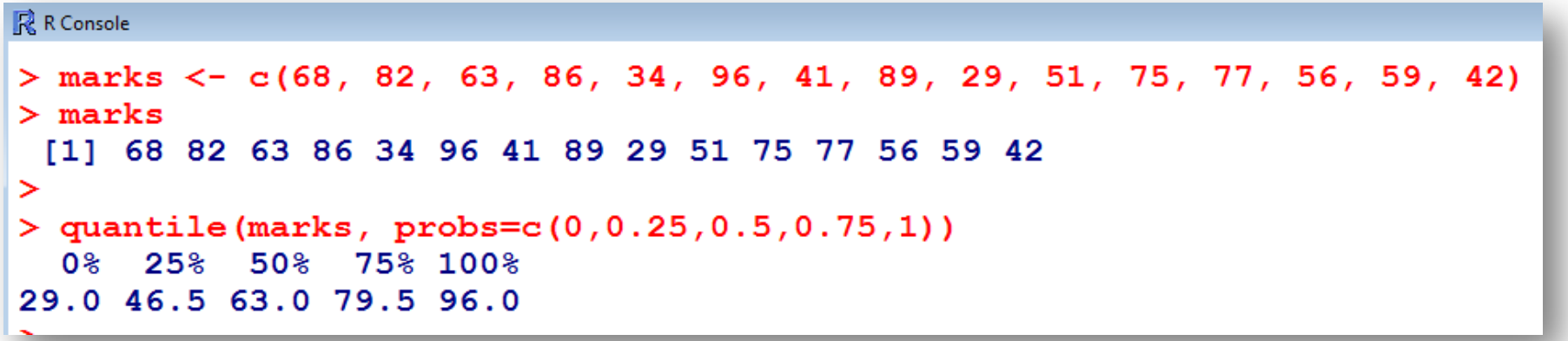
Partition values:

Example: Marks of 15 students are

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89,
  29, 51, 75, 77, 56, 59, 42)

> quantile(marks, probs=c(0,0.25,0.5,0.75,1))
  0%  25%  50%  75% 100%
29.0 46.5 63.0 79.5 96.0
```

Default values



Partition values:

Example: Marks of 15 students are

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89, 29,
  51, 75, 77, 56, 59, 42)
```

Defining probabilities

```
> quantile(marks, probs=c(0,0.20,0.4,0.6,0.8,1))
  0%  20%  40%  60%  80% 100%
29.0 41.8 57.8 70.8 82.8 96.0
```