Descriptive Statistics With R Software

Frequency Distribution
::
Absolute Frequency, Relative Frequency and Frequency Distribution

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Classification of Data

Process of arranging the data into groups or classes according to resemblance and similarities.

Functions of Classification

(i) Condenses the data

(ii) Facilitates comparisions

(iii) Helps in studying relationships

(iv) Facilitates statistical treatment of the data

Absolute and relative frequencies

Suppose there are 10 persons participated in a test and there results were declared in two categories as Pass (P) and Fail (F).

P, F, P, F, F, P, P, F, P, P.

Use a_1 and a_2 to refer to Pass and Fail categories.

There are 6 persons who passed, denoted as $n_1 = 6$.

There are 4 persons who failed, denoted as $n_2 = 4$.

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

Absolute and relative frequencies

The relative frequency of a_1 is $$f_1 = \frac{n_1}{n_1 + n_2} = \frac{6}{10} = 0.6 = 60\%$$

The relative frequency of a_2 is $$f_2 = \frac{n_2}{n_1 + n_2} = \frac{4}{10} = 0.4 = 40\%$$

This gives us information about the proportions of Pass and Fail persons in the test.

Absolute and relative frequencies

`table(data vector)` creates the absolute frequency of the `data vector` of the given data in the vector.

Enter data as `x`

```
table(x) # absolute frequencies
```

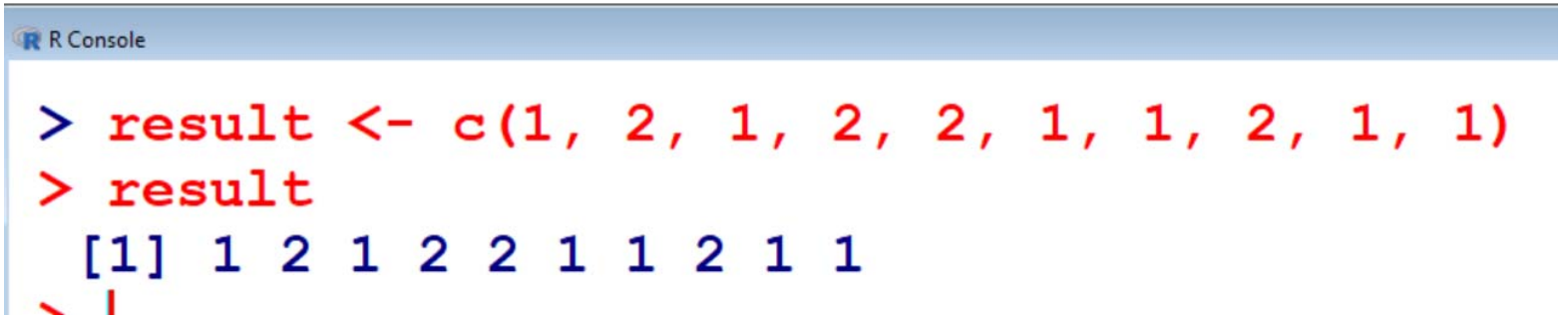
```
table(x)/length(x) # relative frequencies
```

Absolute and relative frequencies

Results of 10 persons declared in two categories as Pass (P) and Fail (F) is categorised as 1 and 2 respectively.

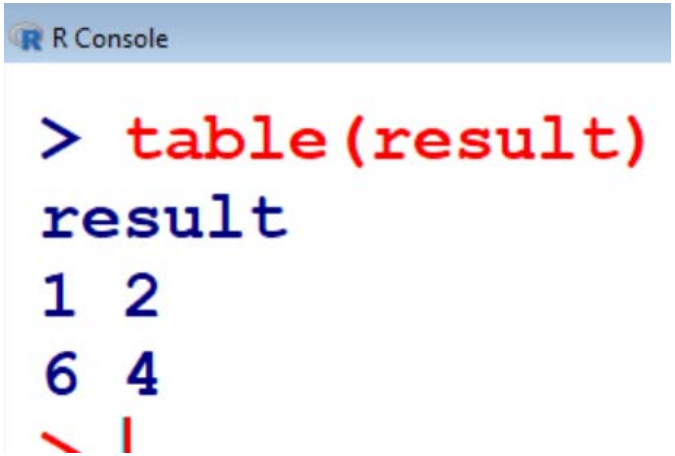
P,	F,	P,	F,	F,	P,	P,	F,	P,	P
1,	2,	1,	2,	2,	1,	1,	2,	1,	1

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



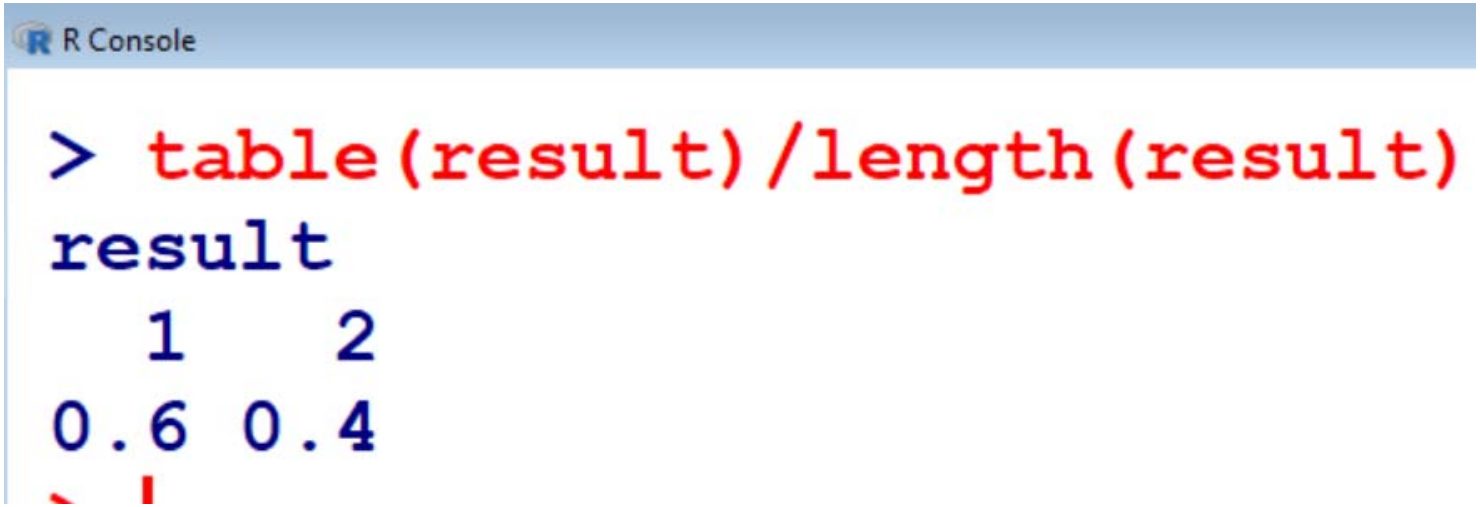
Absolute and relative frequencies:

```
> table(result) # Absolute frequencies
result
1 2
6 4
```



Absolute and relative frequencies:

```
> table(result)/length(result) #Relative freq.
result
  1   2
0.6 0.4
```



Frequency Distribution

- **Arrangement of ungrouped data in the form of group is called frequency distribution of data.**

- **Classify the data into different classes by dividing the entire range of the values of variables into suitable number of groups called class.**

Frequency Distribution

- **Lower and upper boundary figures of a class are called the lower limit and upper limit respectively.**

- **Difference between the limits is called the width of the class or class interval.**

- **The value of variate lies in the middle of lower and upper limits.**

Frequency Distribution

- **The number of observations in a particular class is called absolute frequency or frequency.**

- **The number of observations in a particular class divided by total frequency is called relative frequency.**

Frequency Distribution

- **The cumulative frequency corresponding to any variate value is the number of observations less than or equal to that value.**

- **The cumulative frequency corresponding to a class is the total number of observations less than or equal to the upper limit of the class.**

Frequency Distribution

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.

The data is summarized in class intervals

31-40, 41-50, 51-60, 61-70, 71-80 and 81-90

Frequency Distribution

Example (contd.):

Class intervals	Mid point	Absolute frequency (or frequency)	Relative Frequency	Cumulative Frequency
31 – 40	35.5	5	5/20 = 0.25	5
41 – 50	45.5	3	3/20 = 0.15	5+3 = 8
51 – 60	55.5	3	3/20 = 0.15	5+3+3 = 11
61 – 70	65.5	5	5/20 = 0.25	5+3+3+5 = 16
71 – 80	75.5	2	2/20 = 0.01	5+3+3+5+2 = 18
81 - 90	85.5	2	2/20 = 0.01	5+3+3+5+2+2 = 20
	Total	20	1	

Frequency Distribution

General, if there are k class intervals, n observations are divided into k class intervals $a_1, a_2,\ldots,a_k$ containing $n_1, n_2,\ldots,n_k$ observations respectively.

Relative frequency of j^{th} class : $f_j = n_j/n$

Frequency distribution:

Class interval (a_j)	a_1	a_2	…	a_k
Absolute frequency (n_j)	n_1	n_2	…	n_k
Relative frequency (f_j)	f_1	f_2	…	f_k