

# Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

## Lecture 10

### Data, Frequency and Frequency Distribution

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

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



## **Discrete Variables**

**Variables can take a finite number of values.**

**Informally speaking, variables are "counted".**

**Example:**

- **Number of children in a family – 1, 2, 3, etc.**
- **Number of branches of a school in a city – 4, 6, 7 etc.**

## Continuous Variables

Variables can take an infinite number of values.

Informally speaking, variables are “measured” and not “counted” .

Example:

- Length of a road is 1.5 kms.
- Length of a road is 1.52 kms.
- Length of a road is 1.521 kms.
- ...

## **Grouped Data**

**Sometimes the original values of data are grouped or the data is available in the form of groups.**

**Original values in a group may not be known.**

**Only the category to which the values belong to is known.**

## **Primary Data**

**Data originally collected by an investigator for the first time for any statistical investigation.**

## **Secondary Data**

**Data which has already been collected by some person or agency for any statistical investigation.**

**Some data which is primary for one may be secondary for other.**

# Source of Primary Data

1. Direct personal investigation
2. Indirect oral investigation
3. Questionnaire received through postal mail, email, e-forms  
(google forms), online surveys etc.
4. Questionnaire sent through surveyors.

# Sources of Secondary Data

- 1. Published sources**
- 2. Data collected from survey agencies**
- 3. Places where public reports the data, e.g.; municipalities.**

# **Classification of Data**

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



## Categorical Data:

3 4 2 4 4 5 3 2 5 3 3 3 3 3 4 4 5 5 5 4 4 1 1 3 3 1 5 2 3 5 2 3 1 5 4 4 3 5 2 1 3 2 5 5 3 2  
1 3 2 5 5 3 5 3 1 1 3 3 3 4 1 3 3 3 5 5 2 5 2 5 2 1 5 4 5 4 3 2 5 3 1 5 1 4 3 3 4 5 1 3 1 2  
5 4 1 5 5 5 4 4 1 2 4 2 1 2 3 3 5 1 5 2 3 5 5 1 1 3 1 1 3 5 2 4 4 2 5 5 3 5 5 4 3 4 1 2 1 2  
1 2 4 1 1 5 2 4 5 2 2 1 2 3 4 3 5 4 3 5 4 1 3 5 4 1 4 3 3 1 5 3 1 2 4 4 1 2 3 4 2 5 1 5 5 4  
1 1 2 5 5 2 2 4 5 3 2 2 2 1 2 5 1 1 4 1 2 1 4 3 5 4 5 1 2 3 5 5 3 5 2 1 4 1 2 2 4 5 1 2 1 1  
4 2 5 4 3 2 4 5 3 1 2 1 2 1 5 4 5 1 3 4 3 1 3 1 3 5 4 3 2 1 4 3 5 3 3 4 3 4 4 4 4 3 4 1 2 1  
3 2 5 4 2 4 5 2 5 4 2 5 5 4 1 4 3 2 2 1 1 2 2 3 1 2 1 4 5 1 4 1 5 3 5 3 5 2 3 3 3 2 3 4 4 5  
2 3 5 3 5 3 1 5 2 5 4 4 3 2 5 5 1 2 1 1 1 3 5 3 3 4 5 5 1 3 1 2 3 2 2 3 1 3 3 1 4 4 3 5 4 3  
2 4 2 2 5 2 2 5 1 4 3 1 3 1 3 1 5 3 3 4 1 2 4 3 2 1 3 2 2 1 5 1 4 2 5 2 4 3 1 3 3 4 3 5 4 3  
1 4 3 3 4 1 2 5 1 1 5 1 5 4 1 1 1 1 2 5 2 3 2 2 4 4 1 5 5 5 4 5 3 3 5 2 3 3 5 5 3 1 3 3 2 5  
1 3 5 3 3 5 1 5 2 1 5 4 3 5 5 3 3 2 3 2 4 1 1 2 2 4 5 2 2 3 5 2 1 3 3 4 2 2 5 5 3 4 2 4 4 5  
3 2 5 3 3 3 3 3 4 4 5 5 5 4 4 1 1 3 3 1 5 2 3 5 2 3 1 5 4 4 3 5 2 1 3 2 5 5 3 2 1 3 2 5 5 3  
5 3 1 1 3 3 3 4 1 3 3 3 5 5 2 5 2 5 2 1 5 4 5 4 3 2 5 3 1 5 1 4 3 3 4 5 1 3 1 2 5 4 1 5 5 5  
4 4 1 2 4 2 1 2 3 3 5 1 5 2 3 5 5 1 1 3 1 1 3 5 2 4 4 2 5 5 3 5 5 4 3 4 1 2 1 2 1 2 4 1 1 5  
2 4 5 2 2 1 2 3 4 3 5 4 3 5 4 1 3 5 4 1 4 3 3 1 5 3 1 2 4 4 1 2 3 4 2 5 1 5 5 4 1 1 2 5 5 2  
2 4 5 3 2 2 2 1 2 5 1 1 4 1 2 1 4 3 5 4 5 1 2 3 5 5 3 5 2 1 4 1 2 2 4 5 1 2 1 1 4 2 5 4 3 2  
4 5 3 1 2 1 2 1 5 4 5 1 3 4 3 1 3 1 3 5 4 3 2 1 4 3 5 3 3 4 3 4 4 4 3 4 1 2 1 3 2 5 4 2 4  
5 2 5 4 2 5 5 4 1 4 3 2 2 1 1 2 2 3 1 2 1 4 5 1 4 1 5 3 5 3 5 2 3 3 3 2 3 4 4 5 2 3 5 3 5 3  
1 5 2 5 4 4 3 2 5 5 1 2 1 1 1 3 5 3 3 4 5 5 1 3 1 2 3 2 2 3 1 3 3 1 4 4 3 5 4 3 2 4 2 2 5 2  
2 5 1 4 3 1 3 1 3 1 5 3 3 4 1 2 4 3 2 1 3 2 2 1 5 1 4 2 5 2 4 3 1 3 3 4 3 5 4 3 1 4 3 3 4 1  
2 5 1 1 5 1 5 4 1 1 1 1 2 5 2 3 2 2 4 4 1 5 5 5 4 5 3 3 5 2 3 3 5 5 3 1 3 3 2 5 1 3 5 3 3 5  
1 5 2 1 5 4 3 5 5 3 3 2 3 2 4 1 1 2 2 4 5 2 2 3 5 2 1 3 3 4 2 2 5 5 3 4 2 4 4 5 3 2 5 3 3 3  
3 3 4 4 5 5 5 4 4 1 1 3 3 1 5 2 3 5 2 3 1 5 4 4 3 5 2 1 3 2 5 5 3 2 1 3 2 5 5 3 5 3 1 1 3 3  
3 4 1 3 3 3 5 5 2 5 2 5 2 1 5 4 5 4 3 2 5 3 1 5 1 4 3 3 4 5 1 3 1 2 5 4 1 5 5 5 4 4 1 2 4 2  
1 2 3 3 5 1 5 2 3 5 5 1 1 3 1 1 3 5 2 4 4 2 5 5 3 5 5 4 3 4 1 2 1 2 1 2 4 1 1 5 2 4 5 2 2 1  
2 3 4 3 5 4 3 5 4 1 3 5 4 1 4 3 3 1 5 3 1 2 4 4 1 2 3 4 2 5 1 5 5 4 1 1 2 5 5 2 2 4 5 3 2 2  
2 1 2 5 1 1 4 1 2 1 4 3 5 4 5 1 2 3 5 5 3 5 2 1 4 1 2 2 4 5 1 2 1 1 4 2 5 4 3 2 4 5 3 1 2 1  
2 1 5 4 5 1 3 4 3 1 3 1 3 5 4 3 2 1 4 3 5 3 3 4 3 4 4 4 3 4 1 2 1 3 2 5 4 2 4 5 2 5 4 2 5  
5 4 1 4 3 2 2 1 1 2 2 3 1 2 1 4 5 1 4 1 5 3 5 3 5 2 3 3 3 2 3 4 4 5 2 3 5 3 5 3 1 5 2 5 4 4  
3 2 5 5 1 2 1 1 1 3 5 3 3 4 5 5 1 3 1 2 3 2 2 3 1 3 3 1 4 4 3 5 4 3 2 4 2 2 5 2 2 5 1 4 3 1  
3 1 3 1 5 3 3 4 1 2 4 3 2 1 3 2 2 1 5 1 4 2 5 2 4 3 1 3 3 4 3 5 4 3 1 4 3 3 4 1 2 5 1 1 5 1  
5 4 1 1 1 1 2 5 2 3 2 2 4 4 1 5 5 5 4 5 3 3 5 2 3 3 5 5 3 1 3 3 2 5 1 3 5 3 3 5 1 5 2 1 5 4  
3 5 5 3 3 2 3 2 4 1 1 2 2 4 5 2 2 3 5 2 1 3 3 4 2 2 5 5 3 4 2 4 4 5 3 2 5 3 3 3 3 3 4 4 5 5  
5 4 4 1 1 3 3 1 5 2 3 5 2 3 1 5 4 4 3 5 2 1 3 2 5 5 3 2 1 3 2 5 5 3 5 3 1 1 3 3 3 4 1 3 3 3  
5 5 2 5 2 5 2 1 5 4 5 4 3 2 5 3 1 5 1 4 3 3 4 5 1 3 1 2 5 4 1 5 5 5 4 4 1 2 4 2 1 2 3 3 5 1

## Contineous Data:

84 73 133 38 115 183 157 96 108 146 58 144 56 15 65 65 154 103 181 10 25 130 160  
147 151 12 131 97 76 94 185 93 41 171 66 55 175 104 80 131 38 86 81 182 179 47  
178 28 154 147 135 70 145 200 87 57 137 173 143 148 126 98 114 98 195 112 27 116 125  
185 39 133 87 153 179 134 10 110 62 122 88 102 59 184 21 15 178 93 107 155 123 22  
119 82 164 58 170 46 88 13 42 38 126 153 187 65 182 155 172 198 46 115 145 152 138  
200 18 161 57 182 173 194 153 190 97 132 70 53 170 16 54 148 20 155 58 125 83 30  
146 163 166 133 129 184 168 173 139 119 69 105 173 109 175 124 59 196 94 73 188 155 41  
158 94 131 22 137 14 36 159 166 67 181 64 63 167 118 87 95 95 101 65 110 116 155  
117 60 74 151 107 77 66 126 42 30 196 51 30 32 17 172 100 161 193 54 184 65 32  
128 41 26 131 111 101 13 137 56 29 164 41 42 32 60 54 153 121 22 92 75 121 147  
58 104 188 124 82 37 39 200 34 109 142 50 14 176 137 50 83 168 117 12 85 158 12  
188 114 88 13 140 109 144 26 21 149 165 32 195 42 164 164 60 136 86 41 189 145 182  
30 87 110 41 132 156 156 172 29 199 103 185 79 86 140 59 70 183 114 53 169 172 13  
197 143 152 87 72 125 11 197 63 67 64 38 53 160 118 177 106 151 84 36 137 77 166  
116 27 162 51 23 109 63 145 37 45 179 112 177 124 153 52 42 185 45 36 76 101 151  
82 181 126 118 184 200 130 25 71 131 38 98 83 171 53 97 62 90 48 152 130 128 71  
181 98 84 174 160 110 44 70 48 40 94 134 25 38 104 21 196 104 198 151 65 111 161  
181 127 77 175 150 113 172 79 66 144 127 158 82 42 66 56 119 133 100 105 177 33 192  
157 150 108 82 166 28 76 19 15 161 173 158 51 90 84 168 120 44 120 163 161 49 162  
164 141 142 170 43 71 71 119 41 13 90 85 78 79 106 176 178 192 64 108 187 162 19  
33 175 118 174 14 24 53 34 180 169 187 85 175 19 44 99 139 190 13 151 157 144 143  
123 28 145 185 38 94 57 13 164 34 26 169 52 138 195 128 105 73 10 158 37 188 99  
117 137 13 139 115 91 16 151 21 193 153 50 91 35 124 54 152 197 92 73 136 71 138  
74 22 197 198 151 69 199 200 142 123 123 35 183 191 194 174 173 190 102 29 59 85 165  
29 159 147 71 150 115 26 58 63 131 126 140 45 45 124 192 39 200 27 126 192 160 84  
114 171 156 72 112 10 71 112 110 188 91 111 115 183 125 187 136 129 158 134 65 33 146  
70 141 196 177 107 200 146 76 89 176 117 192 141 182 194 39 197 94 138 133 85 111 165  
27 36 87 134 107 38 47 118 130 129 154 85 149 116 150 27 143 134 130 34 162 161 92  
159 190 188 169 23 45 95 181 109 156 14 162 32 40 189 68 99 151 22 169 187 150 124  
147 182 37 11 127 23 173 31 81 140 174 57 77 126 118 86 25 109 15 189 179 191 143  
69 177 74 197 30 70 183 112 160 79 31 113 139 66 122 87 74 48 173 200 58 92 57  
168 197 59 45 79 79 16 31 47 13 198 182 67 152 148 185 166 48 182 31 74 105 58  
125 119 59 66 69 197 106 146 181 149 42 168 150 58 34 184 127 145 40 65 23 197 163  
53 155 103 147 82 75 190 29 92 164 41 131 152 38 195 110 86 164 198 33 60 87 141  
121 98 19 150 133 184 158 34 94 129 40 101 101 54 109 159 67 172 197 60 157 132 122

## **Functions of Classification**

**(i) Condenses the data**

**(ii) Facilitates comparisons**

**(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

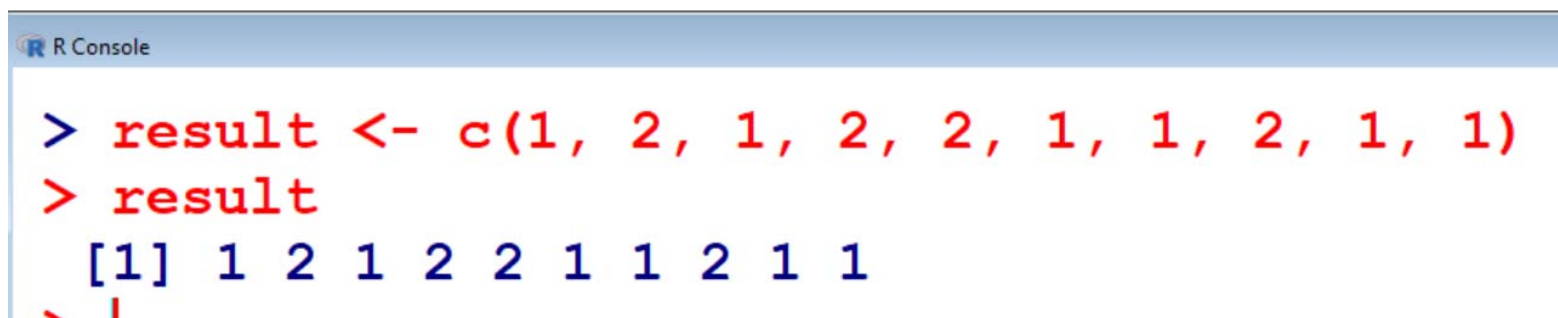
Results of 10 persons declared in two categories as Pass (P) and Fail (F) is categorised as 1 and 2 respectively.

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
```

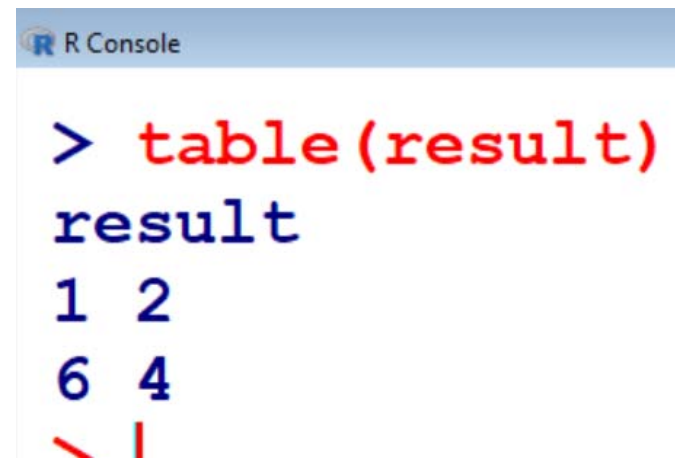
```
[1] 1 2 1 2 2 1 1 2 1 1
```



```
R Console
> 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
```

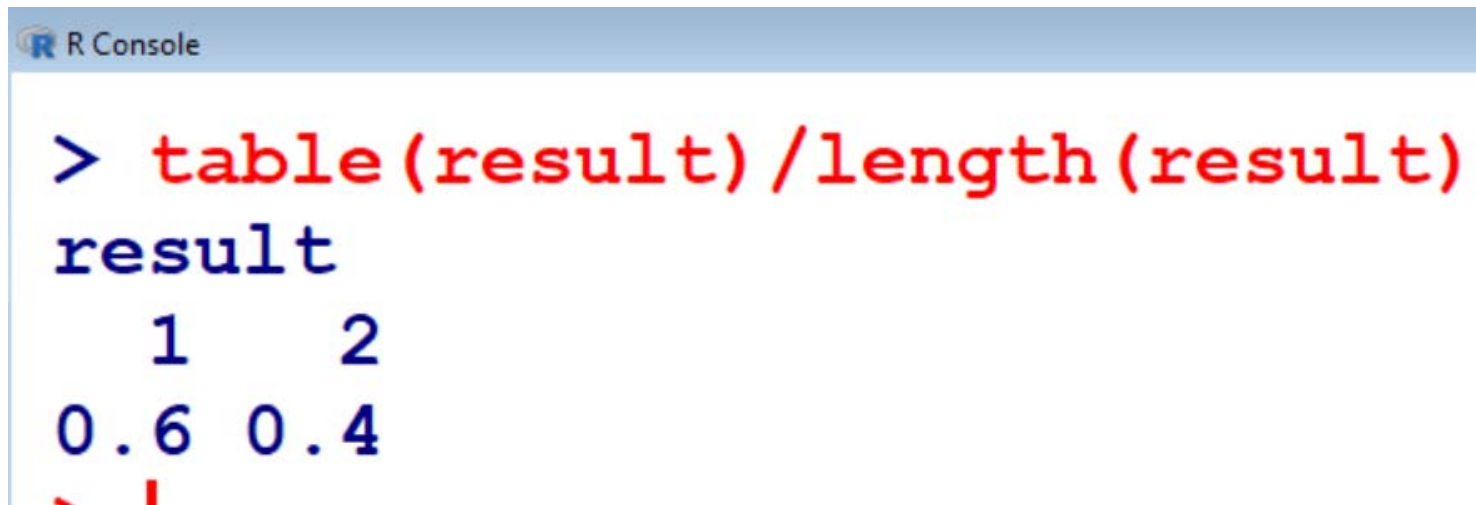


```
R Console
> table(result)
result
1 2
6 4
# |
```



## Absolute and relative frequencies:

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



```
R Console  
  
> table(result)/length(result)  
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.