

R tutorial 3 Statistical Analyzation

Index No: - 18001149

- 1) A company has 50 employees with employee numbers 101-150

```
1 #create column vector for register employee Numbers|
2 M1 <- matrix(data = 101:150, nrow = 50, ncol = 1)
3 M1
4
5
6
```

	V1
1	101
2	102
3	103
4	104
5	105
6	106
7	107
8	108
9	109
10	110
11	111
12	112
13	113
14	114
15	115
16	116
17	117

- 2) The gender of an employee can be male or female with equal chance.

```
#column vector for Creating Gender Distribution among the columns|
M2 <- matrix(rbinom(50*1,1,0.5),50,1)
```

	V1
1	Female
2	Female
3	Female
4	Male
5	Male
6	Male
7	Female
8	Male
9	Male
10	Female
11	Male
12	Female
13	Female
14	Female
15	Male
16	Female
17	Male

- 3) Employees are randomly assigned to Sales department, HR department and Accounts department. The company policy is such that the chance of assigning an employee to the Sales department is twice as the chance of assigning to Accounts department and the chance of assigning an employee to the HR department is same as the chance of assigning to Accounts department.

```
#Assign departments randomly according to given probabilities|
M3 <- matrix(sample(c("Sales", "Accounts", "HR"), size=50, replace=TRUE, prob=c(.50,.25,.25))
M3
```

- 4) The salaries of the employees in the Sales department are known to be normally distributed with mean Rs. 15000/= and standard deviation of Rs.1250/=. The salaries of the employees in the Accounts department are equally likely to distribute between Rs. 15000/= and Rs.20000/=. The salaries of the employees in the HR department are normally distributed with mean Rs. 25000/= and standard deviation of Rs.2500/=.

```
3 #copy M3 vector to M4
4 M4 <- M3
5
6 #get random salary values for Sales division with given mean & standard deviation
7 Msales <- round(rnorm(n=sum(M3 == "Sales"), mean=15000, sd=1250))
8 Msales
9
0 #replace Sales strings with random salary values in Msales vector
1 M4[M4 == "Sales"] <- sample(Msales)
2
3 #get random salary values for Accounts division within given min and max values
4 Maccounts <- round(seq(from = 15000, to = 20000, length.out = sum(M3 == "Accounts")))
5 Maccounts
6
7 #replace Accounts strings with random salary values in Maccounts vector
8 M4[M4 == "Accounts"] <- sample(Maccounts)
9
0 #get random salary values for HR division with given mean & standard deviation
1 Mhr <- round(rnorm(n=sum(M3 == "HR"), mean=25000, sd=2500))
2 Mhr
3
4 #replace HR strings with random salary values in Mhr vector
5 M4[M4 == "HR"] <- sample(Mhr)
6
7 #convert M4 from string to numeric vector
8 M4 <- matrix(apply(M4, 1, as.numeric), 50, 1)
9 M4

> Msales <- round(rnorm(n=sum(M3 == "Sales"), mean=15000, sd=1250))
> Msales
[1] 13802 13992 15547 15640 16700 14296 15140 14416 14738 15946 13318 16214 16765 12731 16104
[16] 14804 13947 14938 14173 11686 12938 15035
> #replace Sales strings with random salary values in Msales vector
> M4[M4 == "Sales"] <- sample(Msales)
> #get random salary values for Accounts division within given min and max values
> Maccounts <- round(seq(from = 15000, to = 20000, length.out = sum(M3 == "Accounts")))
> Maccounts
[1] 15000 15500 16000 16500 17000 17500 18000 18500 19000 19500 20000
> #replace Accounts strings with random salary values in Maccounts vector
> M4[M4 == "Accounts"] <- sample(Maccounts)
> #get random salary values for HR division with given mean & standard deviation
> Mhr <- round(rnorm(n=sum(M3 == "HR"), mean=25000, sd=2500))
> Mhr
[1] 27574 21494 23951 28054 30604 28427 20368 26554 25820 25302 26684 23748 28205 22688 25029
[16] 23604 24640
> #replace HR strings with random salary values in Mhr vector
```

	Filter
1	Sales
2	Sales
3	Accounts
4	HR
5	HR
6	HR
7	HR
8	Accounts
9	HR
10	Sales
11	Accounts
12	Sales
13	Accounts
14	Sales
15	HR
16	Sales
17	HR

Showing 1 to 17 of 50 entries.

```

2
3
4 #combine 4 column vectors into a data frame
5 df <- data.frame(col1 = M1, col2 = M2, col3 = M3, col4 = M4)
6
7 #rename data frame columns
8 colnames(df) <- c("EmpNo", "Gender", "Department", "Salary")
9 df
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72

```

```

> #rename data frame columns
> colnames(df) <- c("EmpNo", "Gender", "Department", "Salary")
> df

```

	EmpNo	Gender	Department	Salary
1	101	Male	Sales	13992
2	102	Male	Sales	14938
3	103	Female	Accounts	19500
4	104	Female	HR	22688
5	105	Male	HR	20368
6	106	Male	HR	26554
7	107	Male	HR	30604
8	108	Female	Accounts	16000
9	109	Male	HR	25820
10	110	Male	Sales	14738
11	111	Male	Accounts	18000
12	112	Female	Sales	11686
13	113	Female	Accounts	15500
14	114	Male	Sales	14416

Give a display of first 6 observations and the last 6 observations of the data set you created using the appropriate commands.

```

63
64 #get first 6 observations
65 head(df)
66
67 #get last 6 observations
68 tail(df)
69
70 #summary of the data frame
71 summary(df)
72

```

```

> #get first 6 observations
> head(df)

```

	EmpNo	Gender	Department	Salary
1	101	Male	Sales	13992
2	102	Male	Sales	14938
3	103	Female	Accounts	19500
4	104	Female	HR	22688
5	105	Male	HR	20368
6	106	Male	HR	26554

```

> #get last 6 observations
> tail(df)

```

	EmpNo	Gender	Department	Salary
45	145	Female	HR	23748
46	146	Male	Sales	14296
47	147	Female	Sales	16700
48	148	Male	HR	23604
49	149	Male	Sales	15547
50	150	Male	Accounts	19000

```

> #summary of the data frame

```

Compare salaries of employee's gender wise using suitable summary statistics and graphs.

III. Compare salaries of employee's department wise using suitable summary statistics and graphs.

```
> #summary of the data frame
> summary(df)
      EmpNo      Gender      Department      salary
Min.   :101.0  Length:50  Length:50  Min.   :11686
1st Qu.:113.2  Class :character  Class :character  1st Qu.:14954
Median :125.5  Mode  :character  Mode  :character  Median :16733
Mean   :125.5                                     Mean   :18962
3rd Qu.:137.8                                     3rd Qu.:23712
Max.   :150.0                                     Max.   :30604
> |
```

```
df
plot(df$EmpNo,df$Salary,xlab="Employee",ylab = "salary",main = "Employess salary Distribution")
|
```

