Lab – 4

Objective : To perform different operation on data frames, implement the different statistical measure on dataset and implement different probability Distributions.

Procedure :

Task 1 - To perform different operation on data frames.

Creating data frames

```
> roll <- c(1, 2, 3, 4, 5)
> name <- c("Arsh", "Ankit", "Anish", "Binod", "Bhupinder Jogi")
> marks <- c(69, 89, 98, 3, 100)
> grade <- c('A', 'B', 'C', 'F', 'S')
> student <- data.frame(RollNums=roll, Name=name, Marks=marks, Grades=grade)
> print(student)
  RollNums           Name Marks Grades
1        1           Arsh    69      A
2        2          Ankit    89      B
3        3          Anish    98      C
4        4          Binod     3      F
5        5 Bhupinder Jogi   100      S
> 
```

Fetching column names

```
> # Get column names
> names(student)
[1] "RollNums" "Name"     "Marks"    "Grades"
```

Generating summary of the data

```
> summary(student)
    RollNums     Name               Marks          Grades
 Min.   :1   Length:5           Min.   :  3.0   Length:5
 1st Qu.:2   Class :character   1st Qu.: 69.0   Class :character
 Median :3   Mode  :character   Median : 89.0   Mode  :character
 Mean   :3                      Mean   : 71.8
 3rd Qu.:4                      3rd Qu.: 98.0
 Max.   :5                      Max.   :100.0
> 
```

Task 2 - Implement the different statistical measure on dataset.

Reading data from csv file

```
> marks <- read.csv("D:/B Tech/Sem 6/DA/Marks.csv")
> print(marks)
   Sr.No RollNo    Name English Maths Economics Total Grade
1      1   1001     Sia      34    98        99   231     B
2      2   1002   Priya      65    65        47   177     C
3      3   1003  Nitish      78    85        58   221     B
4      4   1004   Grish      45    78        85   208     C
5      5   1005    Atul      74    25        78   177     C
6      6   1006  Nilesh      12    48        96   156     C
7      7   1007   Raman      47    47        78   172     B
8      8   1008   Satya      98    89        48   235     B
9      9   1009    Adil      47    95        52   194     C
10    10   1010    Ritu      89    65        78   232     B
11    11   1011   Somya      62    78        92   232     B
12    12   1012 Jaiveer      47    45        56   148     E
13    13   1013   Abdul      74    88        87   249     A
>
```

Fetching number of row and columns

```
> # Row fetching
> nrow(marks)
[1] 13
> # Column fetching
> ncol(marks)
[1] 8
>
```

Finding maximum and minimum marks

```
> max(marks$Maths)
[1] 98
> min(marks$Economics)
[1] 47
>
```

Finding student with maximum marks in particular subject by creating subset

```
> subMarks <- subset(marks, marks$Maths == max(marks$Maths))
> print(subMarks)
  Sr.No RollNo Name English Maths Economics Total Grade
1     1   1001  Sia      34    98        99   231     B
>
```

Inserting a new column in the dataset

```
> Physics <- c(98, 48, 78, 58, 48, 98, 45, 88, 77, 65, 14, 65, 78)
> Data <- cbind(marks, Physics)
> print(Data)
   Sr.No RollNo    Name English Maths Economics Total Grade Physics
1      1   1001     Sia      34    98        99   231     B      98
2      2   1002   Priya      65    65        47   177     C      48
3      3   1003  Nitish      78    85        58   221     B      78
4      4   1004   Grish      45    78        85   208     C      58
5      5   1005    Atul      74    25        78   177     C      48
6      6   1006  Nilesh      12    48        96   156     C      98
7      7   1007   Raman      47    47        78   172     B      45
8      8   1008   Satya      98    89        48   235     B      88
9      9   1009    Adil      47    95        52   194     C      77
10    10   1010    Ritu      89    65        78   232     B      65
11    11   1011   Somya      62    78        92   232     B      14
12    12   1012 Jaiveer      47    45        56   148     E      65
13    13   1013   Abdul      74    88        87   249     A      78
> |
```

Finding mean and median of marks

```
> mean(marks$English)
[1] 59.38462
> median(Data$Physics)
[1] 65
> |
```

Finding quartiles for marks

```
> quantile(marks$English)
  0%  25%  50%  75% 100%
  12   47   62   74   98
> quantile(Data$Economics)
  0%  25%  50%  75% 100%
  47   56   78   87   99
> |
```

Finding percentile

```
> quantile(marks$English, c(.4))
40%
 47
> |
```

```
> # Finding 56th, 72th, 89th percentile
> quantile(Data$Maths, c(.56, .72, .89))
  56%   72%   89%
78.00 86.92 93.08
> |
```

Finding interquartile range

```
> # Interquartile range
> IQR(marks$Maths)
[1] 40
>
```

Finding range of marks of a particular subject

```
> # Range
> max(Data$English) - min(Data$English)
[1] 86
>
```

Calculating variance and standard deviation of marks

```
> var(Data$Maths)
[1] 513.2308
>
> sd(Data$Physics)
[1] 23.80422
>
```

Finding covariance of two subjects

```
> # Covariance
> cov(marks$English, marks$Economics)
[1] -202.4103
```

Finding correlation coefficient between two subjects

```
> # Corelation Coefficient
> cor(marks$English, marks$Maths)
[1] 0.1497397
>
```

Task 3 - Implement different probability Distributions.

Binomial Distribution

```
> # Binomial Distribution
> dbinom(3, size = 5, prob = 0.2)
[1] 0.0512
```

```
> pbinom(3, size = 5, prob = 0.2)
[1] 0.99328
```

Cumulative Probability

```
> # Cumulative prob
> dbinom(0, size = 5, prob = 0.2) + dbinom(1, size = 5, prob = 0.2) + dbinom(2, size = 5, prob = 0.2)
[1] 0.94208
```

Uniform Distribution

```
> # Uniform distribution
> runif(10, min = 2, max = 5)
 [1] 3.304867 3.975822 2.105362 4.728240 2.780342 4.229067 3.348587 3.670856 4.458926 2.503305
```

Normal Distribution

```
> # Normal distribution
> pnorm(80, 64, 16.4, lower.tail = FALSE)
[1] 0.1646289
```

Poisson Distribution

```
> # Poisson distribution
> dpois(20, lambda = 5, log = FALSE)
[1] 2.641211e-07
> ppois(10, lambda = 5, lower.tail = TRUE, log = FALSE)
[1] 0.9863047
> rpois(10, 3)
 [1] 4 0 2 4 0 4 4 4 2 5
> y <- c(.10, .5, .1, .2)
> qpois(y, 2)
[1] 0 2 0 1
```

Result : Successfully implemented data frames, operations on dataset and different probability distributions.