STATISTICAL APPLICATIONS

R includes descriptive statistics & plots for summarizing data. It can compute measures of center, like mean & median, as well as measures of spread, like (sample) standard deviation & range

(i) Let us input marks of a class of 20 students into object 'marks'. We use scan() for this :

```
> marks = scan()
1: 70 87 61 55 92
6: 42 33 88 95 68
11: 93 68 82 71 55
16: 40 50 85 81 72
21:
Read 20 items
```

(ii) Range :- > range (marks)

(iii) Median :- > median (marks)

(iv) mean :- > mean (marks)

(v) standard deviation :- > sd (marks)

(vi) Variance :- > var (marks)

(vii) mad :- > mad (marks) → median absolute value

Sign.

DELTA Pg No. Date / /

(viii) maximum value out of all values :-
> max (marks)
OR
> max (marks, na.rm = TRUE)
{ na.rm is used when there are NA values & one wants to eliminate these values }

(ix) minimum value out of all values :-
> min (marks)
OR
> min (marks, na.rm = TRUE)
{na.rm is used when there are NA values & one wants to eliminate these values }

(x) Gives length of a vector including NA values :- > length (marks)

(xi) Sum of all the elements in 'marks'
> sum (marks)
OR
> sum (marks, na.rm = TRUE)

SUMMARY COMMANDS WITH MULTIPLE RESULTS

One can have commands that produce multiple results :-

(NEXT PAGE)

Sign.

DELTA Pg No. Date / /

```
> data
[1] 3 5 7 5 32 6 8 5 6 9 4 5 7 84
```

*
```
> summary(data)
   Min. 1st Qu. Median  Mean 3rd Qu.  Max.
  2.000   3.750  5.000 5.125   6.250 9.000
```

{ It provides with a summary of all the result. }

*
```
> quantile(data)
  0%  25%  50%  75% 100%
2.00 3.75 5.00 6.25 9.00
```

{ It provides quantiles by default, that is 0%, 25%, 50%, 75% & 100% quantiles.

*
```
> quantile(data, c(0.2, 0.5, 0.8))
 20%  50%  80%
   3    5    7
```

*
```
> cumsum(data)
 [1]  3  8 15 20 23 25 31 39 44 50 59
[12] 63 68 75 78 82
```

{ Determines the cumulative sum of these data }.

DELTA Pg No. Date / /

* cummax (data)

[1] 3 5 7 7 7 7 7 8 8 8 9 9 9 9 9

{Determines the cumulative maximum value of the sample}.

* cummin (data)

[1] 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2

{Determines the cumulative minimum value of the sample 'data'}

* cumprod (data)

[1] 3 15 105 525
[5] 1575 3150 18900 151200
[9] 756000 4536000 40824000 163296000
[13] 816480000 5715360000 17146080000 68584320000

{Determines the cumulative product of the sample 'data'}.

NOTE:
* Cumulative command does not work on character data :-

```
> data_char
[1] "She" "is" "good"
> cummax(data_char)
(1) NA NA NA
```

* If data includes NA items then the following is going to happen :-

DELTA Pg No. Date / /

```
> data.na
[1] 2 5 4 NA 7 3 9 NA 2
> cumprod (data.na)
[1] 2 10 40 NA NA NA NA NA NA
```

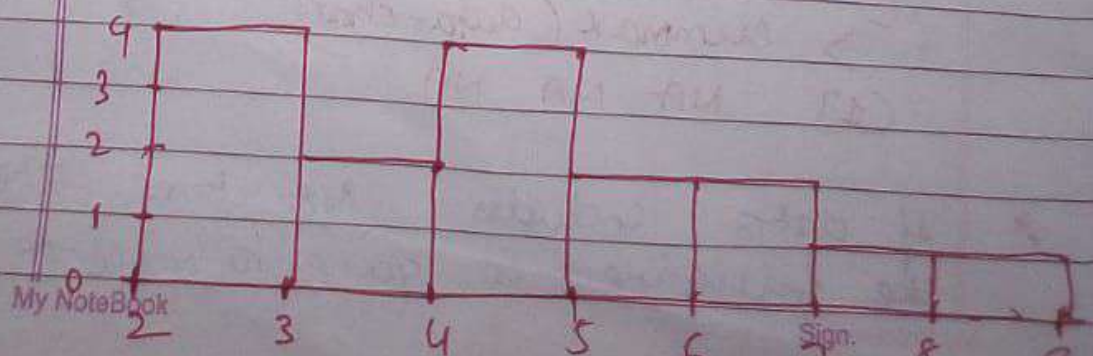
HISTOGRAM

Histogram is a classic way of viewing the distribution of a sample:-
One can create histograms using the graphical command hist() :-

```
> data.hist
[1] 3 5 7 5 3 2 6 8 5 6 9 4 5 7 3 4
> hist (data.hist).
```

* Before using hist(), one can use table() command to see how it is constructed :-

```
> table( data.hist)
data.hist
2 3 4 5 6 7 8 9
1 3 2 4 2 2 1 1
```

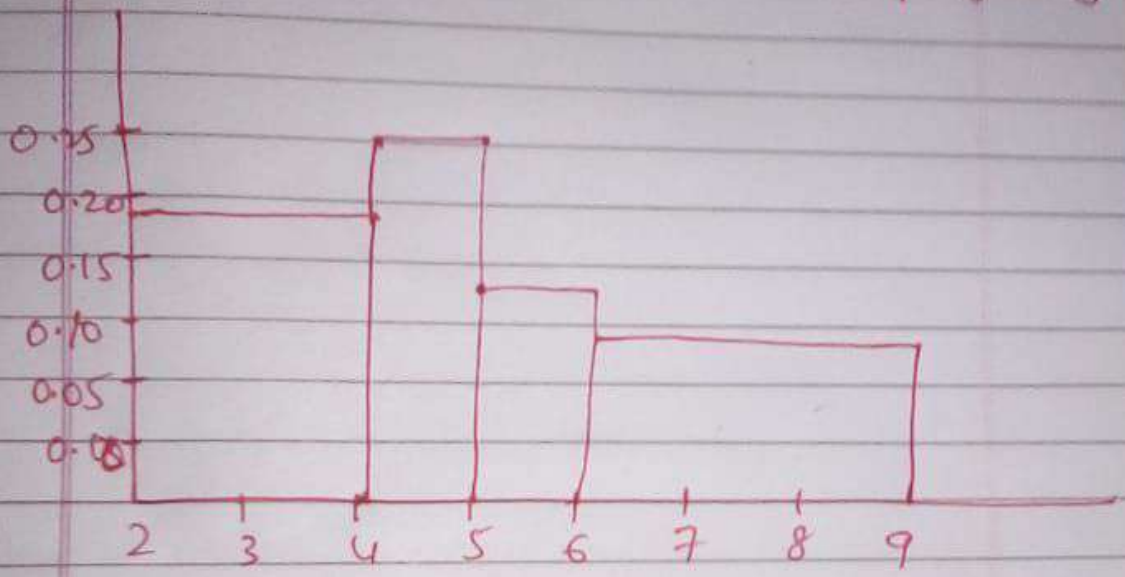


(6)

DELTA Pg No. Date / /

> table (data.hist, breaks = c(2,4,5,6,9))

{ breaks = instruction used to alter the number of columns to be displayed. }



Here y-axis does not show frequency but instead shows the density. The command has attempted to keep the areas of the bars correct & in proportion.