DESCRIPTIVE STATISTICS

Experiment-1

JUNE 18, 2021

BIMAL PARAJULI
20BDS0405

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

Descriptive Statistics

AIM:

Using R software, compute all descriptive statistics and interpret the result.

R-Syntax:

R-Code	Description
Mean(X)	To compute the mean of X
Median(X)	To obtain the median of X
Quartile(X)	Find all the quartiles of X.
Range(X)	To find the range of X
Var(X)	To find the variance of X
Table(X)	To create the frequency table of X
X[n]	To obtain the data in nth column of data vector X.
Var(X, Y)	Calculate the covariance of X and Y
IQR(X)	Find the interquartile range of X.
Length(X)	Find the length of vector X.

Tools Used:

R-Studio (IDE)
R (programming language)

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

Problem 1:

Twenty students, graduates and undergraduates, were enrolled in a statistics course. Their ages were 18,19,19,19,19,20,20,20,20,20,21,21,21,21,22,23,24,27,30,36.

a) Find Mean and Median of all students

b) Find median age of all students under 25 years.

c) Find modal age of all student

R- Code:

```
Console | Terminal | Jobs
~/
> x=c(18,19,19,19,19,20,20,20,20,20,21,21,21,21,22,23,24,27,30,36)        #given age data
> x
 [1] 18 19 19 19 19 20 20 20 20 20 21 21 21 21 22 23 24 27 30 36
> mean(x)                                                                  #mean
[1] 22
> md=median(x)                                                             #median
> md
[1] 20.5
> y=x[x<25]                                                               #data of ages under 25
> y
 [1] 18 19 19 19 19 20 20 20 20 20 21 21 21 21 22 23 24
> median(y)
[1] 20
> rx=table(x)                                                             #frequency table of given data
> rx
x
18 19 20 21 22 23 24 27 30 36
 1  4  5  4  1  1  1  1  1  1
> mode=which(xr==max(xr))                                                 #mode
> mode
20
 3
```

Twenty students, graduates and undergraduates, were enrolled in a statistics course. Their ages were :-

18, 19, 19, 19, 19, 20, 20, 20, 20, 20, 21, 21, 21, 21, 22, 23, 24, 27, 30, 36.

a). Find the mean and median of all students.

b) Find median age of all students under 25 years.

c). Find the modal age of all students.

R code :-

```
> x = c(
> mean(x)   # mean
 [1] 22
> median(x)   # median
 [1] 20.5
> y = x[x<25]   # median of under 25.
> md = median(y)
> md
 [1] 20

> xr = table(x)
> mode = which(xr == max(xr))   # mode.
> mode
 20
  3
```

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

Problem 2:

A survey of 25 faculty members is taken in a college to study their vocational mobility. They were asked the question "In addition to your present position, at how many educational institutes have served on the faculty? Following is the frequency distribution of their responses.

X	0	1	2	3
f	8	11	5	1

Find mean and median of the distribution

R- Code:

```
Console  Terminal  Jobs
~/
> x=c(0, 1, 2, 3)                                  #given X data
> x
[1] 0 1 2 3
> f=c(8,11,5,1)                                    # Corresponding frequency data
> f
[1]  8 11  5  1
> y=rep(x,f)                                       # Individual data
> y
 [1] 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3
> mean=(sum(y)/sum(f))                             #mean
> mean
[1] 0.96
> median(y)                                        #median
[1] 1
>
```

Measures of Central Tendency for frequency table:-

Problem 2: A survey of 25 faculty members is taken in a college to study their vocational mobility. They were asked the question "In addition to your present position, at how many educational institutes have served on the faculty? Following is the frequency distribution of their responses.

X	0	1	2	3
f	8	11	5	1

Find the mean and median of the distribution:

Rcode:-

```
>x= c(0,1,2,3)
> f = c( 8,11,5,1)
> y = rep (x,f)
> Mean = (sum(y) / length(y))   #mean
>Mean
 [1] 0.96
```

sum (y)

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

Problem 3:

Compute mean, median and mode of for the following frequency Distribution:

Height in Cm	145-150	150-155	155-160	160-165	165-170	170-175	175-180	180-185
No. of Adult men	4	6	28	58	64	30	5	5

R- Code:

```
Console  Terminal  Jobs
~/
> x=seq(147.5,182.5,5)
> x
[1] 147.5 152.5 157.5 162.5 167.5 172.5 177.5 182.5
> f=c(4,6,28,58,64,30,5,5)
> f
[1]  4  6 28 58 64 30  5  5
> mean=sum(x*f)/sum(f)            #mean
> mean
[1] 165.175
>
>
```

```
Console  Terminal  Jobs
~/
>
> #For median:
> cumulative= cumsum(f)              #cumulative frequency og given distribution
> cumulative
[1]   4  10  38  96 160 190 195 200
>
> N=sum(f)                          #total number of observations
> N
[1] 200
> position_Q2=min(which(cumulative>n/2))  #serial number of median class
Error in which(cumulative > n/2) : object 'n' not found
> position_Q2=min(which(cumulative>N/2))  #serial number of median class
> position_Q2
[1] 5
>
> h=5
> h
[1] 5
>
> f_q2=f[position_Q2]              #frequency corresponding to median
> f_q2
[1] 64
>
> cumulative_q2=cumulative[position_Q2-1]
> cumulative_q2
[1] 96
>
> lower=x[position_Q2]-h/2
> lower
[1] 165
>
> median=lower+(((n/2)-cumulative_q2)/f_q2)*h
Error: object 'n' not found
> median=lower+(((N/2)-cumulative_q2)/f_q2)*h
> median
[1] 165.3125
> #Hence the median is calculated.
>
```

```
>
>
> #For Mode:
> position_mode=which(f==max(f))
> position_mode
[1] 5
>
> f0=f[position_mode-1]
> f0
[1] 58
>
> f1=f[position_mode]
> f1
[1] 64
>
> f2=f[position_mode+1]
> f2
[1] 30
>
> lower_mo=x[position_mode]-h/2
> lower_mo
[1] 165
>
> mode=lower_mo+ ((f1-f0)/(2*f1-f0-f2))*h
> mode
[1] 165.75
>
> #Hence the mode of data is calculated
>
```

Problem 3: Compute mean, median and mode for the following frequency distribution

Height in cm	145-150	150-155	155-160	160-165	165-170	170-175	175-180	180-185
No. of Adult men	4	6	28	58	64	30	5	5

Rcode:-

```
>mid = seq(147.5, 182.5, 5)
>mid
 [1] 147.5 152.5 157.5 162.5 167.5 172.5 177.5 182.5
>f=c(4, 6, 28, 58, 64, 30, 5, 5) :
>f
 [1] 4 6 28 58 64 30 5 5
>mean=sum(x*f)/sum(f)          # mean
>mean
 [1] 165.175
>#For Median:
```

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

```
>cl=cumsum(frequency)
>cl
 [1] 4 10 38 96 160 190 195 200
>n=sum(frequency)
>n
 [1] 200
>ml=min(which(cl>=n/2))      # The serial number of the median class.
>ml
 [1] 5
>h=5
>h
 [1] 5
>f=frequency[ml]             # frequency of the median class.
>f
 [1] 64
>C=cl[ml-1]                  # Cumulative frequency of median class.
>C
 [1] 96
>l=mid[ml]-h/2
>l
 [1] 165
>median=l+h*((n/2)-C)/f      # median.
>median
 [1] 165.3125
```

Mode:-

```
>m=which(frequency==max(frequency))   # serial number of modal class.
>m
 [1] 5
>fm=frequency[m]             # frequency of modal class.
>fm
 [1] 64
>f1=frequency[m-1]           # frequency of pre modal class.
>f2=frequency[m+1]           # frequency of post modal class.
>f1
 [1] 58
>f2
 [1] 30
```

```
>l=mid[m]-h/2
>l
 [1] 165
>mode=l+(fm-f1)/(2*fm-f1-f2))*h
>mode
 [1] 165.75
```

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

Problem 4:

An entomologist studying morphological variation in species of mosquito recorded the following data on body length: 1.2, 1.4, 1.3, 1.6, 1.0, 1.5, 1.7, 1.1, 1.2, and 1.3.

Compute all the measures of dispersion.

```
Console ~/R/
> x=c(1.2,1.4,1.3,1.6,1.0,1.5,1.7,1.1,1.2,1.3)
> x
 [1] 1.2 1.4 1.3 1.6 1.0 1.5 1.7 1.1 1.2 1.3
>
> res=range(x)                                #range of given data
> res
[1] 1.0 1.7
>
> diff(res)
[1] 0.7
>
> var(x)                                      #Variance of X
[1] 0.049
>
> sd(x)                                       #standard Deviation of X
[1] 0.2213594
>
> quartile(x)                                 #quartiles of x
Error in quartile(x) : could not find function "quartile"
> quantile(x)                                 #quartiles of x
   0%   25%   50%   75%  100%
1.000 1.200 1.300 1.475 1.700
>
> IQR(x)                                      #inter-quartile_range of X
[1] 0.275
>
>
> #For mean deviation about mean, median, mode:
> y=abs(x-mean(x))                            #absolute deviations from the mean.
> y
 [1] 0.13 0.07 0.03 0.27 0.33 0.17 0.37 0.23 0.13 0.03
>
> md_mean=sum(y)/length(y)                    # Mean Deviation from mean
> md_mean
[1] 0.176
>
> md_median=abs(x-median(x))                  #absolute deviations from median.
> md_median
 [1] 0.1 0.1 0.0 0.3 0.3 0.2 0.4 0.2 0.1 0.0
>
>
> #Since, this problem is a bimodal problem, mode calculation is not possible as of now.
>
```

Name:Bimal parajuli Date: 18/06/2021 Class:Statistics(LAB)

An entomologist studying morphological variation in species of mosquito recorded the following data on body length:

1.2, 1.4, 1.3, 1.6, 10, 1.5, 1.7, 1.1, 1.2, 1.3

Compute all the measures of dispersion.

Rcode:-

```
>x=c(1.2,1.4,1.3,1.6,1.0,1.5,1.7,1.1,1.2,1.3).
>x
 [1] 1.2 1.4 1.3 1.6 1.0 1.5 1.7 1.1 1.2 1.3
>res=range(x)
>res
[1] 1.0 1.7
>diff(res)
 [1] 0.7
>var(x)
 [1]. 0.049
>sd(x)
  [1]. 0.02213594
>quartile(x)
 0%  25%  50%  75%  100%.
1.000 1.200 1.300 1.475 1.700.
>IQR(x)                          # Interquartile range of X.
 [1] 0.275
>y=abs(x-mean(x))                # absolute deviations from mean.
>md_mean=sum(y)/length(y)        # Mean deviation from mean.
>md_median=sum(abs(x-median(x)))/length(x)  # Mean deviation from Q2.
># since, this is bimodal, mode is not possible to calculate.
```