Chapter 4

MEASURES OF CENTRAL TENDENCY

✣ SYNOPSIS ✣

4.1 CENTRAL TENDENCY

It is the property of observations to concentrate in central part of the data. Measure of central tendency is a single number which is considered as a representative of the data. The most commonly used measure of central tendency is average. Here we consider following measures of central tendency.

(i) Arithmetic mean
(ii) Geometric mean
(iii) Harmonic mean
(iv) Mode
(v) Median
(vi) Partition values.

Each of the above measure of central tendency can be obtained for three types of statistical data (i) Discrete Observations (ii) Ungrouped Frequency Distribution (iii) Grouped Frequency Distribution.

4.2 DISCRETE OBSERVATIONS

Suppose $x_1, x_2, \ldots, x_n$ are n given observations then we use the following formulae to calculate different measures.

$$\text{Arithmetic mean} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{Geometric mean} = \text{antilog} \frac{\sum_{i=1}^{n} \log x_i}{n}$$

$$\text{Harmonic mean} = \frac{n}{\sum_{i=1}^{n} \left(\frac{1}{x_i}\right)}$$

Mode = The observation occurring most frequently in the data.

Median = Value of the middlemost observation when the observations arranged in ordered.

Partition Values : There are commonly used three types of partition values *quartiles, deciles and percentiles* which divide the distribution into four, ten, and hundred equal parts respectively.

Example 1 : Monthly sales (in '00 Rs.) of 10 small shops are given below.

100, 190, 210, 160, 150, 160, 190, 200, 170, 152

Calculate arithmetic mean, geometric mean, harmonic mean, mode, median, lower quartile, third decile and 42nd percentile.

Solution :

```
> x=c(100,190, 210,160,150,160,190, 200,170,152)
> n=length(x);am=mean(x);lx=log10(x);gm=10^mean(lx)
> hm=n/sum(1/x)
> tx=table(x);m=which(tx==max(tx));stx=sort(unique(x))
> mo=stx[m]
> me=median(x)
> ql=quantile(x,0.25);d3=quantile(x,0.3); p42=quantile(x,0.42)
> cat("mean = ", am,"\n") # It displays value of  mean
> cat("geometric mean = ",gm, "\n")
> cat("harmonic mean = ",hm, "\n")
> cat("mode = ",mo, "\n")
> cat("median = ",me, "\n")
> cat("quartile1 = ",ql, "\n")
> cat("decile 3 = ",d3,"\n")
> cat("percentile 42 = ",p42, "\n")
```

Each of above cat command displays corresponding value.

4.3 UNGROUPED FREQUENCY DISTRIBUTION

Let (x_i, f_i), $i = 1, 2, \ldots, k$ be a given ungrouped frequency distribution, then we use the following formulae to calculate different measures.

$$\text{Arithmetic mean} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$\text{Geometric mean} = \text{antilog}\left(\frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}\right)$$

$$\text{Harmonic mean} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$$

Mode = The observation with highest frequency in the data.

Median = Value of the middlemost observation, when the observations arranged in order.

Example 2 : For the following frequency distribution

x : 1 2 3 4 5

f : 7 11 9 8 3

Calculate arithmetic mean, geometric mean, harmonic mean, mode, median, upper quartile, seventh decile and 29th percentile.

Solution :

```
> x=1:5 ; f=c(7,11,9,8,3);n=sum(f)
> y=rep(x,f);am=mean(y);ly=log10(y)';
> gm=10^mean(ly);hm=n/sum(f/x)
> m=which(f==max(f));mo=x[m];me=median(y)
> q3=quantile(y,0.75);d7=quantile(y,0.7); p29=quantile(y,0.29)
> cat("Mean= ",am, "\n")
> cat("Geometric mean= ",gm,"\n")
> cat("Harmonic mean= ",hm, "\n")
> cat("Mode  = ", mo, "\n")
> cat("Median = ", me, "\n")
> cat("Quartile 3 = ",q3, "\n")
> cat("Decile 7 = ",d7, "\n")
> cat("Percentile 29 = ",p29,"\n")
```

Each of above cat command displays corresponding value.

4.4 GROUPED FREQUENCY DISTRIBUTION

For a grouped frequency distribution, let x_i denote the class mark of the i^{th} class and let f_i be the frequency of the i^{th} class, $i = 1, 2, \ldots, k$ then we use the following formulae to calculate different measures.

$$\text{Arithmetic mean} = \frac{\sum f_i x_i}{\sum f_i}$$

$$\text{Geometric mean} = \text{antilog}\left(\frac{\sum f_i \log x_i}{\sum f_i}\right)$$

$$\text{Harmonic mean} = \frac{\sum f_i}{\sum\left(\frac{f_i}{x_i}\right)}$$

$$\text{Median} = L + (N/2 - CF)(h/f)$$

$$\text{Mode} = L + ((f_m - f_1)/(2f_m - f_1 - f_2))h$$

$Q_i = L_i + ((i * N/4) - C.F.)(h/f) \quad i = 1, 2, 3$

$D_i = L_i + ((i * N/10) - C.F.)(h/f) \quad i = 1, 2, \ldots 9$

$P_i = L_i + ((i * N/100) - CF)(h/f) \quad i = 1, 2, \ldots 99$

Symbols have their usual meanings.

Example 3 : The frequency distribution of weight (in grams) of mangoes of a certain variety is given below.

Weight	No. of mangoes
410-420	14
420-430	20
430-440	42
440-450	54
450-460	45
460-470	18
470-480	7

Calculate arithmetic mean, geometric mean, harmonic mean, mode, median, upper quartile, sixth decile and 62nd percentile.

Solution :

```
> lb=seq(410,470,10); ub=seq(420,480,10);h=10
> f=c(14,20,42,54,45,18,7);x=(lb+ub)/2;n=sum(f)
> am=sum(f*x)/n; gm=10^(sum(f*log10(x))/n) ;hm=n/sum(f/x)
> lcf=cumsum(f)
> mc=min(which (lcf>=n/2))
> med=lb[mc]+(n/2-lcf[mc-1])*(h/f[mc])
```

```
> qc=min(which (lcf>=3*n/4))
> q3=lb[qc]+(3*n/4-lcf[qc-1])*(h/f[qc])
> dc=min(which(lcf>=6*n/10))
> d6=lb[dc]+(6*n/10-lcf[dc-1])*(h/f[dc])
> pc=min(which(lcf>=62*n/100))
> p62=lb[pc]+(62*n/100-lcf[pc-1])*(h/f[pc])
> moc=which(f==max(f))
> mo=lb[moc]+((f[moc]-f[moc-1])/(2*f[moc]-f[moc-1]f[moc+1]))*h
> cat("mean= ",am,"\n")
> cat("geometric mean= ",gm,"\n")
> cat("harmonic mean= ",hm,"\n")
> cat("mode= " ,mo,"\n")
> cat("median= ",med,"\n")
> cat("quartile 3 = ",q3,"\n")
> cat("decile 6 = ",d6,"\n")
> cat("percentile 62 = ",p62,"\n")
```

Each of above cat command displays corresponding value

4.5 EXERCISE

1. The number of jobs completed in a day for 30 days are given below : 5, 2, 7, 9, 12, 15, 6, 7, 8, 14, 21, 9, 12, 15, 7, 7, 9, 12, 5, 2, 7, 5, 10, 12, 16, 14, 10, 9, 7, 8 Compute arithmetic mean, geometric mean, harmonic mean, mode, median, 8th decile.

2. Find mean, median, modal age of married women at the first child birth.

Age at the first child birth	19	20	21	22	23	24	25	26	27	28	29
No. of married women	37	85	116	225	200	190	320	290	300	190	100

3. Frequency distribution of age (in yrs) of number of employees in a company is given below.

Age	20-24	25-29	30-34	35-39	40-44	45-49	50-54
No. of Persons	50	70	100	180	120	60	25

Determine the values of upper quartile, median, fourth decile and thirty third percentile.

4. Find mean, median, modal wage for the following data :

Wages. (in Rs.)	No. of workers
60-80	4
80-100	6
100-120	10
120-140	16
140-160	12
160-180	7
180-200	3

5. Frequency distribution of number of colds experienced in ten months by the number of persons is given below.

No. of months	0	1	2	3	4	5	6	7	8	9
No. of Persons	25	46	91	162	110	95	82	26	13	2

Calculate (i) lower quartile (ii) seventh decile (iii) fifty eighth percentile.

6. Compute arithmetic mean, geometric mean, harmonic mean, mode, median for the following data on number of books read by 20 students in a month.

4, 12, 7, 24, 32, 18, 9, 19, 21, 14, 27, 16, 8, 20, 12, 15, 7, 2, 6, 11

7. Following is a frequency distribution of I.Q. of students.

I. Q.	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
No. of students	2	13	28	32	45	39	15	3

Find mean, median, upper quartile, mode, fourth decile, geometric mean for the above data.

8. Calculate the mean, median and the mode for the following data.

Daily earnings	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
No. of workers	3	8	14	35	28	16	10	5

9. Calculate the median, fourth decile and sixty fifth percentile for the following data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	5	8	7	12	28	24	10	2

10. Verify the relation between arithmetic mean, geometric mean and harmonic mean

for the following data. 24, 45, 78, 69, 56, 52, 31, 43, 84, 92, 74, 62, 51, 48, 86

11. For the data given below, compute mode, geometric mean and harmonic mean.

Output	70-74	75-79	80-84	85-89	90-94	95-99	100-104
No. of workers	3	5	15	12	7	6	2

***

Dept. of Statistics
Ramniranjan Jhunjhunwala College,
Bombay-86.

Chapter 5

MEASURES OF DISPERSION

✣ SYNOPSIS ✣

5.1 INTRODUCTION

Average condenses the information into a single value. However only average is not sufficient to describe the distribution completely. There may be two distributions having same mean but the distributions are not identical. For example, Consider the marks obtained by two students in five tests.

Student 1 : 0, 30, 5, 40, 5
Student 2 : 14, 18, 16, 15, 17

Here the average marks of both students are 16 but the marks of student 1 are more scattered from the average as compared to marks of student 2. Hence for further study and analysis it is essential to measure the extent of variation present in a given data. Observations are scattered or dispersed from the central value. This variation is called as dispersion.

5.2 MEASURE OF DISPERSION

A measure of dispersion gives the degree of extent of such variation (dispersion) present in a given distribution. There are two types of measures of dispersion.

(i) Absolute measures of dispersion
(ii) Relative measures of dispersion.

We shall consider the computations of following measures of dispersion for any of three types of data.

(i) Range and coefficient of range.

Range = M – m
where, M = maximum observation
and m = minimum observation
In case of frequency distribution conventionally
M = upper boundary of last class
m = lower boundary of first class.

Coefficient of range = (M – m) / (M + m)

(ii) Quartile deviation and coefficient of quartile deviation.
Quartile deviation = $(Q_3 - Q_1) / 2$
Coefficient of quartile deviation = $(Q_3 - Q_1) / (Q_3 + Q_1)$

(iii) Mean deviation.
Mean deviation is defined about any arbitrary constant. however usually we define it around central value like mean or median.

Mean deviation about mean = $\sum |x_i - \text{mean}| / n$ (for n observations)
Mean deviation about mean = $\sum f_i |x_i - \text{mean}| / \sum f_i$ (for frequency distribution)

(iv) Variance, standard deviation.
Variance = $\sum ((x_i - \text{mean})^\wedge 2) / n$ (for n observations)
Variance = $\sum (f_i (x_i - \text{mean})^\wedge 2) / \sum f_i$ (for frequency distribution)
Standard deviation (σ) is defined as positive root of variance.

(v) Coefficient of variation.
Coefficient of variation = σ/|mean|

Example 1 : The number of mistakes in a page recorded for 20 pages are as follows.

2, 5, 9, 7, 11, 6, 5, 2, 7, 9, 3, 2, 8, 12, 14, 6, 3, 9, 8, 7.

Find (i) range and coefficient of range (ii) quartile deviation and coefficient of quartile deviation (iii) mean deviation about mean (iv) coefficient of variation.

Solution :

```
> x=c(2,5,9,7,11,6,5,2,7,9,3,2,8,12,14,6,3,9,8,7)
> mx=mean(x);ql=quantile(x,0.25);q3=quantile(x,0.75)
> mi=min(x);ma=max(x);r=ma-mi;cr=r/(ma+mi);qd=(q3-ql)/2
> cqd=(q3-ql)/(q3+ql);n=length(x)
> vl=var(x) # var function uses denominator (n-1)
> v=((n-1)/n)*vl
> sd=v ^0.5
> cv=sd*100/abs(mx) # coeff.of variation is in percentage
> md=sum(abs(x-mx))/n
> cat("mean = ",mx, "\n")
> cat("range = ",r, "\n")
> cat("coefficient of range = ",cr,"\n")
> cat("quartile deviation = ",qd, "\n" )
> cat("coefficient of variation = ",cv,"\n")
> cat(" mean deviation about mean = ",md, "\n")
```

Each of above cat command displays corresponding value.

Example 2 : Find (i) quartile deviation and coefficient of quartile deviation (ii) mean deviation about median (iii) coefficient of variation for the following data.

No. of goals scored in a match	No. of matches
0	27
1	9
2	8
3	5
4	4

Solution :

```
> x=0:4;  f=c(27,9,8,5,4)  ; n=sum(f);y =rep(x,f)
> mx=sum(f*x)/n;q1=quantile(y,0.25); q3=quantile(y,0.75);
> me=quantile(y,0.5)
> qd=(q3-q1)/2
> cqd=(q3-q1)/(q3+q1);v=sum(f*(x-mx)^2)/n
> sd=v^0.5;cv=sd* 100/abs(mx)
> md=sum(f*abs(x-me))/n
> cat("mean= ",mX,"\n")
> cat("median= ",me,"\n")
> cat("quartile deviation = ",qd,"\n")
> cat("coefficient of quartile deviation = ",cqd, "\n")
> cat( "coefficient of variation = ",CV,"\n")
> cat("mean deviation about median= ",md,"\n")
```

Each of above cat command displays corresponding value.

Example 3 : Two batsmen A and B scored the following runs in ten innings.

A : 101, 27, 0, 36, 82, 45, 7, 13, 65, 14

B : 97, 12, 40, 96, 13, 8, 85, 8, 56, 15

Who is the more consistent batsman ?

Solution :

```
> a=c(101,27,0,36,82,45,7,13,65,14)
> b=c(97,12,40,96,13,8,85,8,56,15)
> ma=mean(a);mb=mean(b);n=length(a)
> va=((n-1)/n)*var(a);sda=va^0.5
> vb=((n-1)/n)*var(b);sdb=vb^0.5
> cva=(sda/abs(ma))*100; cvb=(sdb/abs(mb))*100
> cat("mean of A = ",ma,"\n")
> cat("mean of B = ",mb,"\n")
> cat("coefficient of variation of A = ",cva, "\n" )
> cat("coefficient of variation of B = ",cvb,"\n")
> if (cva<cvb){cat("A is more consistent batsman\n");}
else {cat("B is more consistent batsman\n");}
```


Example 4 : For the following frequency distribution of size of holdings, calculate (i) quartile deviation (ii) standard deviation (iii) coefficient of variation.

Size of holdings (hectares)	No. of farms
2.5-3.5	1000
3.5-4.5	2300
4.5-5.5	3600
5.5-6.5	2400
6.5-7.5	1700
7.5-8.5	1200
8.5-9.5	520

Solution :

```
> lb=2.5:8.5; ub=3.5:9.5;h=1
> f=c(1000,2300,3600,2400,1700,1200,520);x=(lb+ub)/2;n=sum(f)
> am=sum(f*x)/n;v=sum(f*(x-am)^2)/n;sd=v^0.5;cv=sd*100/abs(am)
> lcf=cumsum(f)
> q1c=min(which(lcf>=n/4))
> q3c=min(which(lcf>=3*n/4))
> q3=lb[q3c]+(3*n/4-lcf[q3c-1])*(h/f[q3c])
> q1=lb[q1c]+(3*n/4-lcf[q1c-1])*(h/f[q1c])
> qd=(q3-q1)/2
> cat("quartile deviation = ",qd,"\n")
> cat("standard deviation = ",sd,"\n")
> cat("coefficient of variation = ",cv, "\n")
```

Each of above cat command displays corresponding value.

5.3 EXERCISE

1. The weekly wages (in Rs.) of 20 unskilled workers are as given below.
   350, 320, 410, 360, 520, 290, 300, 305, 260, 310
   290, 320, 400, 370, 450, 360, 620, 590, 520, 410
   Find (i) range and coefficient of range (ii) quartile deviation, (iii) mean deviation about mean (iv) mean deviation about median (v) variance (vi) coefficient of variation.
2. The following table gives the number of finished articles turned out per day by number of workers in a factory.

No. of articles (x)	18	19	20	21	22	23	24	25	26	27
No. of workers (f)	3	7	11	14	18	17	13	8	5	4

   Find (i) quartile deviation and coefficient of quartile deviation (ii) mean deviation about median (iii) standard deviation (iv) coefficient of variation.
3. Frequency distribution of number of families according to their monthly medical expenses (in Rs.) is given below.

Expenses	300-399	400-499	500-599	600-699	700-799	800-899
No. of Families	30	46	58	76	50	20

   Find (i) quartile deviation (ii) mean deviation about mean (iii) mean deviation about median (iv) coefficient of variation.

4. Frequency distribution of age (in yrs.) of number of employees in a company is given below.

Age	20-24	25-29	30-34	35-39	40-44	45-49	50-54
No. of Persons	50	70	100	180	120	60	25

Determine the values of upper quartile, median, fourth decile and thirty third percentile.

5. Frequency distribution of life (in years) of two models of radio is given below.

Life (in years)	No. of set Examined	
	Model A	Model B
0-2	5	2
2-4	16	7
4-6	13	12
6-8	7	19
8-10	5	9
10-12	4	1

Which model is more uniform ?

6. Calculate mean deviation from mean and median for the following data : 100, 200, 160, 240, 360, 490, 672, 423, 567

7. Calculate mean deviation about mean and median for the following data :

Class	5-10	10-15	15-20	20-25	25-30	30-35
Frequency	8	17	30	26	12	7

8. The following table gives the profit (in '000 Rs.) of two companies X and Y.

Year	2000	2001	2002	2003	2004	2005	2006	2007
X	700	625	725	650	700	658	720	731
Y	550	600	610	590	630	640	630	645

Which of the two companies has greater consistency in profits ?

9. Daily sales (in '000 Rs) of a certain firm are as given below.

Daily sales	100-105	105-110	110-115	115-120	120-125	125-130
No. of days	2	8	24	34	20	9

Calculate coefficient of variation for the above data.

10. Following are the data on length of life (in hours) for two types of electric lamps A and B.

Length of life	No. of lamps (A)	No. of lamps (B)
500-700	5	4
700-900	11	30
900-1100	26	12
1100-1300	10	8
1300-1500	8	6

Compare the variability of life of two types of lamps.

* * *

Chapter 6

MOMENTS, SKEWNESS AND KURTOSIS

✦ SYNOPSIS ✦

6.1 MOMENTS

Moments : Moments for all types of data can be obtained by using the following formulae.

Raw moments :

1. Discrete observations : $\mu_r' = \frac{\Sigma x^r}{n}$, $r = 1, 2, \ldots$

2. Ungrouped frequency distribution: $\mu_r' = \frac{\Sigma f x^r}{\Sigma f}$, $r = 1, 2, \ldots$

3. Grouped frequency distribution : $\mu_r' = \frac{\Sigma f x^r}{\Sigma f}$, $r = 1, 2, \ldots$

Central moments :

1 : Discrete observations : $\frac{\mu_r = \Sigma\ (x - \text{mean})^r}{n}$, $r = 1, 2, \ldots$

2 : Ungrouped frequency distribution : $\frac{\mu_r = \Sigma\ f\,(x - \text{mean})^r}{\Sigma f}$, $r = 1, 2, \ldots$

3 : Grouped frequency distribution : $\frac{\mu_r = \Sigma\ f\,(x - \text{mean})^r}{\Sigma f}$, $r = 1, 2, \ldots$

Corrected central moments (For grouped frequency distribution only) :

$\mu_2 = \mu_2 - \frac{h^2}{12}$, $\mu_4 = \mu_4 - \frac{h^2}{2}\mu_2 + \frac{7}{240}h^4$

Measure of skewness based on moments :

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\gamma_1 = \sqrt{\beta_1}$ (Attach sign of μ_3)

Symbols have their usual meanings.

6.2 SKEWNESS

The measures of central tendency give us the representative value of a series. Measure of dispersion give us idea about the spread or scatter of observations about mean or central value but it fails to reveal whether the scatter of values on either side of mean is symmetrical or not. These measures also do not describe the frequency distribution completely because there may be two distributions with same mean, variance but still different from each other regarding shape or pattern. Skewness and kurtosis are the additional characteristics of a frequency distribution for studying shape and pattern. Skewness is the property which describes the lack of symmetry in a distribution. A distribution may be skewed or symmetric. There are two types of skewness namely positive and negative. We study the following measures of skewness.

(i) Karl Pearson's coefficient of skewness.

$$= \frac{(\text{Mean} - \text{Mode})}{\text{Standard deviation}}$$

(ii) Bowley's coefficient of skewness.

$$= \frac{(Q_3 + Q_1 - 2\,Q_2)}{(Q_3 - Q_1)}$$

(iii) Coefficient of skewness based on moments (γ_1) .

Example 1 : For the following observations.

10, 12, 14, 16, 12, 8, 23, 10, 12, 20, 30, 25, 27, 12.

find (i) Karl Pearson's coefficient of skewness (ii) Bowley's coefficient of skewness (iii) coefficient of skewness based on moments.

Solution :

```
> x=c(10,12,14,16,12,8,23,10,12,20,30,25,27,12);n=length (x)
> tx=table(x);m=which(tx==max(tx));stx=sort(unique(x));mo=stx[m]
> mx = mean(x); v =((n - 1)/n)*var(x);sd = v ^ 0.5
> skp =(mx-mo)/sd
> q1=quantile(x,0.25); q3=quantile(x,0.75); q2=quantile(x,0.5)
> skb=(q3+q1-2*q2)/(q3-q1)
> cm2=sum((x-mx)^2)/n; cm3=sum((x-mx)^3)/n
> m1=cm3^2/cm2^3;ms=sqrt(m1)
> if(cm3<0){ms=-ms}
> cat("Mean = ", mx,"\n")
> cat("Mode = ",mo,"\n")
> cat("Standard deviation =",sd,"\n")
> cat("Karl Pearson's coefficient of skewness = ",skp, "\n" )
> cat("Bowley's coefficient of skewness = ",skb, "\n" )
> cat("Coefficient of skewness based on moments = ",ms,"\n")
```

Each of above cat command displays corresponding value.

Example 2 : Frequency distribution of number of absent days in a week of college students is given below.

No-of absent days	0	1	2	3	4	5	6
No. of students	3	9	17	23	10	7	1

Calculate (i) Karl Pearson's coefficient of skewness, (ii) Bowley's coefficient of skewness, (iii) Coefficient of skewness based on moments for the above data.

Solution :

```
> x = 0:6; f = c (3, 9, 17, 23, 10, 7, 1)
> m = which(f==max(f)); mo=x[m]: y=rep(x,f)
> mx=mean(y); v=((n-1)/n)*var(y);sd=v^0.5
> skp =(mx-mo)/sd
> q1=quantile(y,0.25);q3=quantile(y,0.75);q2=quantile(y,0.5)
> skp=(q3+q1-2*q2)/(q3-q1);n=sum(f)
> cm2=sum(f(x-mx)^2)/n;cm3=sum(f*(x-mx)^3)/n
> m1=cm3^2/cm2^3;ms=sqrt(m1)
> if(cm3<0){ms=-ms}
> cat("Mean = " ,mx, "\n")
> cat("Mode = ",mo,"\n")
> cat("Standard deviation =" ,sd, "\n")
> cat("Karl Pearson's coefficient of skewness = ",skp,"\n")
> cat("Bowley'S coefficient of skewness = ",skb,"\n)
> cat("Coefficient of skewness baed on moments =",ms,"\n")
```

Each of above cat command displays corresponding value.

Example 3 : Frequency distribution of number of persons according to their monthly expenditure on transport is given below.

Monthly expenditure	300-350	350-400	400-450	450-500	500-550	550-600
No. of persons	12	23	38	15	10	2

Calculate (i) Karl Pearson's coefficient of skewness. (ii) Bowley's coefficient of skewness, (iii) Coefficient of skewness based on moments for the above data.

Solution :

```
> lb = seq(300, 550, 50); ub = seq (350, 600, 50)
> f = c (12,23,38,15,10,2); x =(lb+ub)/2;n = sum(f)
> mx=sum(f*x)/n;v=sum(f*(x-mx)^2)/n;sd=v^0.5;h=50
> moc=which(f==max(f))
> mo=lb[moc]+((f[moc]-f[moc-1])/(2*f[moc]-f[moc-1]-f[moc+1]))*h
> skp=(mx-mo)/sd
> lcf=cumsum(f)
> q1c=min(which(lcf>=n/4)); q2c=min(which (lcf>=n/2))
> q3c=min(which(lcf>=3*n/4))
> q3=lb[q3c]+(3*n/4-`lcf[q3c-1])*(h/f[q3c])
```

```
> q2=lb[q2c]+(n/2-lcf[q2c-1])*(h/f[q2c])
> q1=lb[q1c]+(n/4-lcf[q1c-1])*(h/f[q1c])
> skb=(q3+q1-2*q2)/(q3-q1)
> cm2=sum(f*(x-mx) ^ 2)/n; cm3=sum(f*(x-mx)^3)/n
> ccm2 = cm2 - h^2/12
> m1=cm3^2/ ccm2^3;ms=sqrt(m1)
> if(cm3<0){ms=-ms}
> cat("Mean = ",mx,"\n")
> cat("Mode = ",mo,"\n")
> cat("standard deviation =" ,sd, "\n")
> cat("Karl Pearson's coefficient of skewness =",skp, "\n")
> cat("Bowley's coefficient of skewness = ",Skb,"\n")
> cat("Coefficient of skewness based on moments =",ms, "\n")
```

Each of above cat command displays corresponding value.

6.3 KURTOSIS

Kurtosis enables us to have an idea about the shape and nature of peakedness of the curve of frequency distribution. Measure of kurtosis gives the extent to which the distribution is peaked as compared to normal curve. Normal curve is symmetric and is known as mesokurtic. If curve of a frequency distribution is having a high peak than a normal curve then it is known as leptokurtic. If curve of a frequency distribution is flat than a normal curve then it is known as platykurtic.

Measure of kurtosis based on moments is defined as follows :

$\beta_2 = \mu_4/\mu_2^2$,

$\gamma_2 = \beta_2 - 3$

If $\gamma_2 > 0$ distribution is leptokurtic

If $\gamma_2 < 0$ distribution is platykurtic

If $\gamma_2 = 0$ distribution is mesokurtic

Symbols have their usual meanings

Example 1 : For the following observations.

3, 6, 3, 8, 12, 9, 4, 2, 10, 11, 9, 7

Compute (i) first four central moments (ii) measure of kurtosis based on moments.

Solution :

```
> x=c(3,6,3,8,12,9,4,2,10,11,9,7);n=length(x)
> mx=mean(x)
> cm2=sum(x-mx)^2/n;cm3=sum((x-mx)^3);cm4=sum((x-mx)^4)/n
> m2=cm4/cm2^2;mk=m2-3
> cat("Second central moment = ",cm2, "\n")
> cat("Third central moment=",cm3,"\n");
> cat("Fourth central moment= ",cm4, "\n")
> cat("Measure of kurtosis based on moments = ",mk,"\n")
```

Each of above cat command displays corresponding value.

Example 2 : Frequency distribution of number of attempts required to pass the final CA examination is given below :

No. of attempts	1	2	3	4	5	6	7
No. of students	12	23	45	32	12	5	2

Compute (i) first four central moments (ii) measure of kurtosis based on moments.

Solution :

```
> x=1:7;f=c(12,23,45,32,13,5,2);n=sum(f)
> mx=sum(f*x)/n
> cm2=sum(f*(x-mx)^2)/n;cm3=sum(f*(x-mx)^3)/n
> cm4=sum(f*(x-mx)^4)/n
> m2=cm4/cm2^2;mk=m2-3
> cat("Second central moment = ",cm2, "\n")
> cat("Third central moment = ",cm3, "\n")
> cat("Fourth central + moment = ",cm4, "\n")
> cat("Measure of kurtosis based on moments = " ,mk, "\n")
```

Each of above cat command displays corresponding value.

Example 3 : Frequency distribution of duration of advertisements on television is as follows :

Duration (in seconds)	30-35	35-40	40-45	45-50	50-55	55-60
Number of Advertisements	6	18	25	12	5	3

Compute (i) first four central moments (ii) measure of kurtosis based on moments.

Solution :

```
> lb=seq(30,55,5);ub=seq(35,60,5)
> f=c(6,18,25,12,5,3);x=(lb+ub)/2;n=sum(f);h=5
> mx=sum(f*x)/n
> cm2=sum(f*(x-mx)^2)/n; cm3=sum(f*(x-mx)^3)/n
> cm4=sum(f*(x-mx)^4)/n
> ccm2=cm2-h^2/12;ccm4=cm4-(h^2/2)*cm2+(7/240)*h^4
> m2=ccm4/ccm2^2;mk=m2-3
> cat("Second central moment = ",cm2,"\n")
> cat("Corrected second central moment = ",ccm2, "\n")
> cat("Third central moment = ",cm3,"\n");
> cat("Fourth central moment = ",cm4,"\n")
> cat("Corrected fourth central moment = ",ccm4,"\n")
> cat ("Measure of kurtosis based on moments = ",mk,"\n")
```

Each of above cat command displays corresponding value.

6.4 EXERCISE

1. For the following observations
23, 8, 12, 22, 71, 54, 68, 32, 8, 57
Compute (i) Karl Pearson's coefficient of skewness (ii) Bowley's coefficient of skewness (iii) Coefficient of skewness based on moments.
2. In a group of 110 students each one was given TRUE/FALSE type questions. The following table gives the frequency distribution of number of correct answers given by them.

No. of correct answers	0	1	2	3	4	5
No. of Students	2	13	19	44	28	4

Compute (i) Karl Pearson's coefficient of skewness (ii) Bowley's coefficient of skewness (iii) Coefficient of skewness based on moments.
4. The following data represent the distribution of number of students according to their monthly pocket money (in Rs.)

Pocket money	100-150	150-200	200-250	250-300	300-350	350-400
No. of students	13	31	472	27	18	5

Compute (i) Karl Pearson's coefficient of skewness (ii) Bowley's coefficient of skewness (iii) Coefficient of skewness based on moments.
4. The scores in aptitude test of 10 students are given below.
15, 9, 18, 20, 21, 26, 14, 13, 17, 12
Compute (i) first four central moments (ii) measure of kurtosis based on moments. (iii) measure of skewness based on moments.
5. Using following data

x	3	6	9	12	15	18
f	2	13	17	29	10	4

Compute (i) first four central moments (ii) measure of kurtosis based on moments. (iii) measure of skewness based on moments.
6 The frequency distribution of balance (in Rs) at the end of month in savings account of 80 accounts in a certain bank is given below.

Balance	1000-1200	1200-1400	1400-1600	1600-1800	1800-2000
No. of Accounts	45	18	10	5	2

Compute (i) first four central moments (ii) measure of kurtosis based on moments. (iii) measure of skewness based on moments.

***

Chapter 7

PROBABILITY AND PROBABILITY DISTRIBUTIONS

✦ SYNOPSIS ✦

7.1 PROBABILITY

In real life, experiments are classified into two categories.
(i) Deterministic experiment
(ii) Non-deterministic or random experiment.

In probability theory we are concerned with random experiments. The set of all possible outcomes of a random experiment is called as a *sample space*.

If the chance of occurrence of any outcome is same for all outcomes then the corresponding sample space is called equiprobable sample space and sample points are said to be equally likely. For example, In an experiment of tossing a fair die, probability of occurring any number on uppermost face is same and is equal to 1/6.

An event is a subset of sample space. In an experiment having equiprobable sample space the probability of an event A is defined as

$$P(A) = \frac{\text{No. of sample points in event A}}{\text{No. of sample points in sample space S}} = \frac{n(A)}{n(S)}$$

In computing probabilities of different events using R software we use function choose (n, r) which gives the value of number of combination of n objects taken r at a time. For example, ${}_5C_2$ = choose (5, 2).