Ex Following table gives the height of 1000 adult male. Calculate mean, median and mode. Compare with empirical formula.

Sol.

Height (in)	frequency (f_i)	Cumulative frequency	mid-value x_i
57.875 – 59.875	2	2	58.875
59.875 – 61.875	28	30	60.875
61.875 – 63.875	125	155	62.875
63.875 – 65.875 ← l	270 ← f_1	425 ← m	64.875
65.875 – 67.875	303	728 ← 500	66.875
67.875 – 69.875	197 ← f_2	925	68.875
69.875 – 71.875	65	990	70.875
71.875 – 73.875	10	1000	72.875
	$\sum f_i = 1000$		

$$\text{mean} = \bar{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{66365}{1000} = 66.365$$

$\frac{N}{2} = 500$

$$\text{median} = M = l + \left(\frac{N}{2} - m\right)\frac{c}{f} = 65.875 + (500 - 425)\frac{2}{303}$$

$$M = 66.37$$

$$\text{mode} = M_o = l + \frac{c f_2}{f_1 + f_2} = 65.875 + \frac{2 \times 197}{270 + 197} = 66.719$$

using empirical formula:

$$M_o = 3M - 2\bar{X}$$

$$= 3 * 66.37 - 2 * 66.365 = 66.38$$

[redacted] Coding Method

This method can be used to find the arthmetic mean, by change of origin and change of scale.

Change of origin is based on arbitrary fixed number A.

$$x_i = X_i + A$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i (X_i + A)}{\sum f_i} = \frac{\sum f_i X_i}{\sum f_i} + \frac{A \sum f_i}{\sum f_i}$$

$$\bar{x} = \bar{X} + A$$

and $X_i = x_i - A$

Chang of scale is based on the highest common facter c of the values of X_i.

$$X_i = c\, y_i$$

$$\therefore \quad \bar{x} = A + \bar{X} = A + c\bar{y} = A + c\,\frac{\sum f y}{\sum f}$$

Ex Calculate the value of mean for the following distribution :-

x (variate)	0	1	2	3	4	5	6
f (frequency)	20	150	61	36	25	7	1

Sol.

x	f	$X = x - 3$	$y = \frac{X}{c}$	fy
0	20	-3	-3	-60
1	150	-2	-2	-300
2	61	-1	-1	-61
3	36	0	0	0
4	25	1	1	25
5	7	2	2	14
6	1	3	3	3

$X = x - A$ choose $A = 3 \Rightarrow \therefore X = x - 3$

c length of $X \Rightarrow \therefore c = 1$

$\sum fy = -379 \qquad \sum f = 300$

$$\therefore \quad \bar{x} = A + \bar{X} = A + c\bar{y} = A + c\,\frac{\sum fy}{\sum f}$$

$$\bar{x} = 3 + (1)\left(-\frac{379}{300}\right) = 1.7366$$

3. Measures of Dispersion

There are four measures of dispersion :

1. Range.
2. Quartile deviation.
3. Mean deviation.
4. standard deviation.

1. Range : Is the simplest measure of dispersion. It is the difference between the maximum and minimum of the given variate values.

2. Quartile deviation

The first three quartiles of a grouped frequency distribution are given by

$$Q_i = l_i + \left(\frac{iN}{4} - m\right)\frac{c}{f}$$

l_i : lower limit of class in which Q_i lies
N : total frequency
m : cumulative frequency
$i = 1, 2, 3$
c : class length

$\therefore$ quartile deviation $Q = \dfrac{Q_3 - Q_1}{2}$.

ملاحظة في حالة $i=1$ اذن نجد قيمة $\frac{N}{4}$

Ex Calculate the range and the quartile deviation, and coefficient of dispersion.

Class	frequency	Cumulative frequency
9.5 – 19.5	1	1
19.5 – 29.5	0	1
29.5 – 39.5	4	5
39.5 – 49.5 ← l_1	6	11 ← m_1
49.5 – 59.5	7 ← f_1	18
59.5 – 69.5 ← l_2	12	30 ← m_2
69.5 – 79.5	16 ← f_2	46
79.5 – 89.5	10	56
89.5 – 99.5	4	60

49.5 – 59.5
$\frac{iN}{4} = \frac{N}{4} = 15$

$\frac{iN}{4} = \frac{3N}{4} = 45$
وهي [illegible]

Sol. Range $= 99.5 - 9.5 = 90$

$N = 60$, $\frac{N}{4} = 15$, $\frac{3}{4}N = 45$ In cumulative frequenc column, 15 corresponding to class 49.5 – 59.5

$$Q_1 = 49.5 + \left(\frac{N}{4} - m\right)\frac{c}{f}$$

$$= 49.5 + (15 - 11)\frac{10}{7} = 55.214$$

$$Q_3 = 69.5 + \left(\frac{3}{4}N - m\right)\frac{c}{f} = 69.5 + (45 - 30)\frac{10}{16}$$

$$= 78.875$$

$\therefore$ quartile deviation $Q = \frac{Q_3 - Q_1}{2} = \frac{78.875 - 55.214}{2} = 11.83$

تشتت

Coefficient of dispersion (q) $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{78.875 - 55.214}{78.875 + 55.214}$

$\therefore$ $q = 0.18$

3- Mean deviation

The mean deviation is :—

$$\text{mean deviation} = \frac{\sum f|x-\bar{x}|}{\sum f}$$

where $\bar{x}$ is arthmatic mean

Ex. Compute the mean deviation for :

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	8	12	17	14	9	7	4

Sol-

center x	f	fx	$\lvert x-30.8\rvert$	$f\lvert x-30.8\rvert$
5	8	40	25.8	206.4
15	12	180	15.8	189.6
25	17	425	5.8	98.6
35	14	490	4.2	58.6
45	9	405	14.2	127.8
55	7	385	24.2	169.4
65	4	260	34.2	136.8
Total	71	2185		987.4

mean $\bar{x} = \frac{2185}{71} = 30.8$

$$\text{mean deviation} = \frac{\sum f|x-\bar{x}|}{\sum f} = \frac{987.4}{71} = 13.9$$

ملاحظة يمكن مجموعتين من البيانات ان يكون لهما نفس المعدل mean
لكن يختلفون بالانحراف المعياري .

4. The Standard Deviation

Is the root mean square deviation from the mean. It is usually denoted by σ. σ^2 is known as the variance.

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{\sum f} \quad \text{and} \quad \sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}}$$

Ex Find standard deviation (S.D.) and mean deviation (M.D) for the following data.

x	2	4	6	8	10
f	1	4	6	4	1

Sol. Arthmatic mean $\bar{x} = \frac{\sum fx}{\sum f} = 6$

x	f	$\lvert x-6 \rvert$	$f\lvert x-6 \rvert$	$f(x-6)^2$
2	1	4	4	16
4	4	2	8	16
6	6	0	0	0
8	4	2	8	16
10	1	4	4	16
Total	16		24	64

standard deviation $\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} = \sqrt{\frac{64}{16}} = 2$

The mean deviation $= \frac{\sum f\lvert x-\bar{x} \rvert}{\sum f} = \frac{24}{16} = 1.5$

Moment : The r^{th} moment of a distribution about (a) is the aveange of the r^{th} powers of deviation from (a).

$$\mu'_r = \frac{\sum f (x-a)^r}{\sum f}$$

The r^{th} moment about the mean is :

$$\mu_r = \frac{\sum f (x-\bar{x})^r}{\sum f}$$

(i) The first moment about the orgin is the mean

$$\mu'_1 = \frac{\sum f x}{\sum f}$$

(ii) The second moment about the mean is the variance

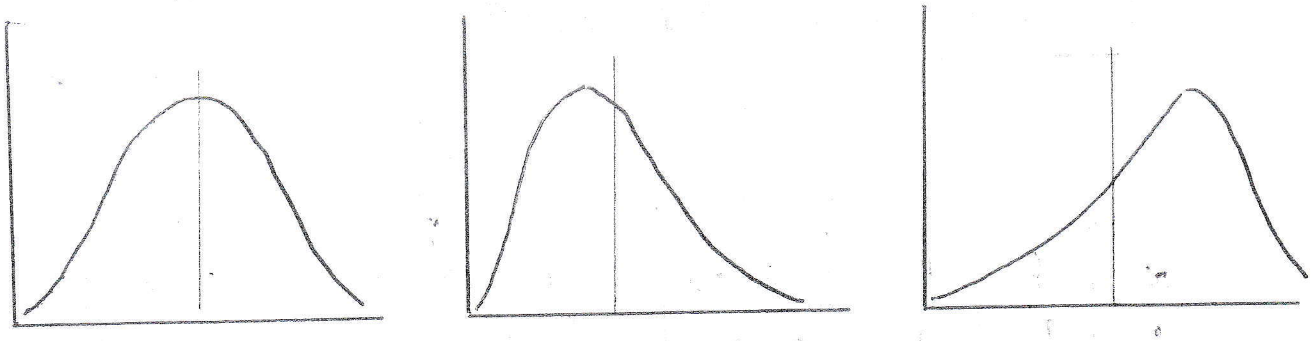
$$\mu'_2 = \frac{\sum f (x-\bar{x})^2}{\sum f} = \sigma^2$$

iii) The third moment about the mean is

$$\mu_3 = \frac{\sum f (x-\bar{x})^3}{\sum f}$$

التواء

Skewness : mean lack of symmetry or departure from symmetry



(a) skewness is zero $\mu_3 = 0$

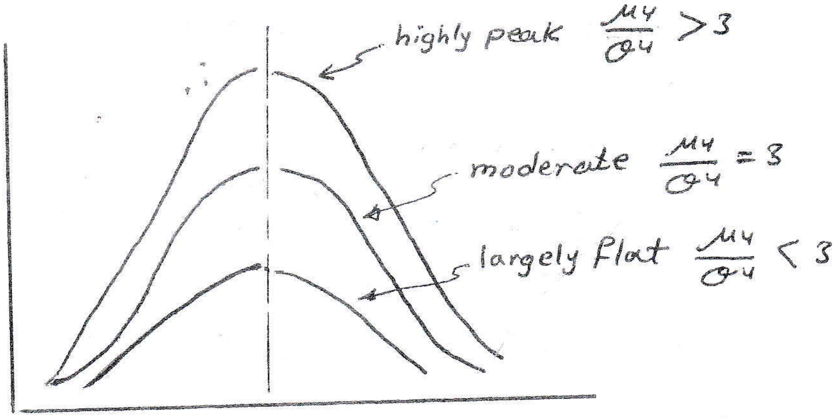
(b) skewness +ve $\mu_3 > 0$

(c) skewnes −ve $\mu_3 < 0$

$$\text{coefficient of skewness} = \frac{\mu_3}{\sigma^3}$$

تدبب

Kurtosis : indicate the flatness or peakedness of a frequency distribution



$$\text{coefficient of Kurtosis} = \frac{\mu_4}{\sigma^4}$$

Ex Discuss the skewness and kurtosis of the following:

x	1	2	3	4	5	6	7	8	9
f	1	6	13	25	30	22	9	5	2

Sol

x	f	fx	$(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
1	1	1	-3.9	15.21	-59.32	231.34
2	6	12	-2.9	50.46	-146.33	242.373
3	13	39	-1.9	46.93	-89.16	169.41
4	25	100	-0.9	20.25	-18.22	16.4
5	30	150	0.1	0.3	0.03	0.003
6	22	132	1.1	26.62	29.28	32.21
7	9	63	2.1	39.69	83.34	175.03
8	5	40	3.1	48.05	148.96	461.76
9	2	18	4.1	33.62	137.84	565.15
Total	113	555		281.13	86.42	1893.6

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{555}{113} = 4.9$$

$$\mu_2 = \frac{\sum f(x-\bar{x})^2}{\sum f} = 2.49 = \sigma^2 \Rightarrow \sigma = 1.577$$

$$\therefore \sigma^3 = 3.922 \qquad \sigma^4 = 6.2$$

$$\therefore \mu_3 = \frac{\sum f(x-\bar{x})^3}{\sum f} = \frac{86.42}{113} = 0.764$$

and $$\mu_4 = \frac{\sum f(x-\bar{x})^4}{\sum f} = \frac{1893.6}{113} = 16.8$$

coefficient of skewness $= \dfrac{\mu_3}{\sigma_3} = \dfrac{0.764}{3.922} = 0.194 > 0$

coefficient of kurtosis $= \dfrac{\mu_4}{\sigma^4} = \dfrac{16.86}{6.2} = 2.77 < 3$

4. Statistical Tests

هل هناك فرق في معدلات داخل البيانات ؟

① The t-test : this test compares the observed difference between aveargе within the data, and find whether the difference is significant.

(d.f) is degree of freedom $= n-1$

Hypothesis A (one sample) $\mu = \mu_0$

$$t_{cal} = \frac{\bar{X} - X_0}{S(\bar{X})} \quad , \quad S(\bar{X}) = \frac{S(X)}{\sqrt{n}}$$

$$S(X) = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

Hypothesis B (two samples test - matched pair) $\mu_1 - \mu_2$

$d.f = n-1$

$$t = \frac{\bar{X} - 0}{S(\bar{X})}$$

Hypothesis C (two sample) $\mu_1 = \mu_2$

$d.f = n_1 + n_2 - 2$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\bar{S}(\bar{X})} \quad , \quad S(\bar{X}) = \frac{S(X)}{\sqrt{n}}$$

$$\bar{S}(\bar{X}) = \bar{S}(X)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad , \quad \bar{S}(X) = \sqrt{\frac{\sum(\bar{X}_1 - X_1)^2 + \sum(\bar{X}_2 - X_2)^2}{n_1 + n_2 - 2}}$$

For all hypothesis if $t_{cal} > t_{tab}$, the hypothesis is rejected.

Ex Five analysis, 92.3, 92.4, 92.8, 93 and 92.5 %
Is there is a significant difference from assumed mean of 92.5%. Take significant level of 95%.-

sol. hypothesis A (one sample), $\mu = \mu_0 = 92.5$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{463}{5} = 92.6 \qquad n=5$$

$$S(X) = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} = 0.292 \quad , \quad S(\bar{X}) = \frac{S(X)}{\sqrt{n}} = 0.13$$

$$t_{cal.} = \frac{\bar{X} - \mu_0}{S(\bar{X})} = \frac{92.6 - 92.5}{0.13} = 0.77$$

from t-test table at 95% significant and d.f = 4

$t_{tabel} = 2.132$

$\therefore$ $t_{cal.} < t_{tab.}$ there is no significant difference

Ex seven samples. Is there any difference between the two analysis?

sample	1	2	3	4	5	6	7
Analysis 1	15.1	14.7	15.2	15.3	14.9	14.7	15.1
Analysis 2	14.6	14.2	15	15.3	14	14.6	14.5
difference	-0.5	-0.5	-0.2	0	-0.9	-0.1	-0.6

sol. Hypothesis B $\mu_1 - \mu_2 = 0$ d.f = 6 $n = 7$

$$\bar{X}_{diff.} = \frac{\sum X_i}{n} = -0.4 \qquad S(X) = \sqrt{\frac{\sum (X_{diff} - \bar{X}_{diff.})}{n-1}} = 0.316$$

$$S(\bar{X}) = S(X)/\sqrt{n} = 0.316/\sqrt{7} = 0.119$$

$$t_{cal.} = \frac{0 - \bar{\mu}}{S(\bar{X})} = \frac{0.4}{0.119} = 3.36$$

. from table $t_{tab} = 1.95$ at 95%

) The CHI - squared test :

The chi - squared test provides a test for determine ether the observed count differs significantly from , expected count.

هل هنالك فرق بين الملاحظ والمتوقع

CHI - squared test $X^2 = \frac{\sum (f - f')^2}{f'}$

where :

f : is the observed frequency

f' : is the expected frequency

Tank is handled by four pump, these pumps suffer from 40 failures during the past year. The failures distribution as follows:

$pump_1 = 16$, $pump_2 = 9$, $pump_3 = 6$, $pump_4 = 9$

The expected failure for each pump is (10).

use $X^2_{0.95}$ and $X^2_{0.75}$

$$X^2 = \frac{\sum (f - f')^2}{f'} = \frac{(16-10)^2 + (9-10)^2 + (6-10)^2 + (9-10)^2}{10} = 5.4$$

$$d.f = n - 1 = 4 - 1 = 3$$

From table $X^2_{0.95} = 7.815$

$\therefore$ $X^2_{cal} < X^2_{tab.}$ difference is not significant

gain, from table $X^2_{0.75} = 4.11$

$\therefore$ $X^2_{cal} > X^2_{tab.}$ significant

Ex Hydrogen gas from two source, and have the following Hydrogen content. Is there a significant difference between the two stream.

Source 1	65	64.5	74.5	64	75	74	67
Source 2	64	69	61.5	69	67.5		

Sol. For Source 1; (two sample) $n = 7$

$$\bar{X}_1 = \frac{\sum X_i}{n} = 69.1$$

For Source 2, $n = 5$

$$\bar{X}_2 = 66.2$$

Hypothesis C $d.f = n_1 + n_2 - 2 = 7 + 5 - 2 = 10$

$$\bar{S}(X) = \sqrt{\frac{\sum(\bar{X}_1 - X_1)^2 + \sum(\bar{X}_2 - X_2)^2}{n_1 + n_2 - 2}} = 4.47$$

and $\bar{S}(\bar{X}) = \bar{S}(X)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 4.47\sqrt{\frac{1}{7} + \frac{1}{5}} = 2.62$

$$\therefore\ t_{cal.} = \frac{\bar{X}_1 - \bar{X}_2}{\bar{S}(\bar{X})} = \frac{69.1 - 66.2}{2.62} = 1.11$$

From table at 95% significant

$t_{tab.} = 1.812$

$\therefore\ t_{cal.} < t_{tab.}$ no significant

) The CHI - squared test :

The chi - squared test provides a test for determine wether the observed count (مبلغ او كمية) differs significantly from an expected count.

هل هناك فرق بين الملاحظ والمتوقع

CHI - squared test $X^2 = \dfrac{\sum (f-f')^2}{f'}$

where :

f : is the observed frequency

f' : is the expected frequency

Tank is handled by four pump, these pumps suffer from 40 failures during the past year. The failures distribution as follows:

$pump_1 = 16$, $pump_2 = 9$, $pump_3 = 6$, $pump_4 = 9$

The expected failure for each pump is (10). معدل

use $X^2_{0.95}$ and $X^2_{0.75}$

$$X^2 = \frac{\sum (f-f')^2}{f'} = \frac{(16-10)^2+(9-10)^2+(6-10)^2+(9-10)^2}{10} = 5.4$$

$d.f = n-1 = 4-1 = 3$

From table $X^2_{0.95} = 7.815$

$\therefore X^2_{cal} < X^2_{tab.}$ difference is not significant

ain, from table $X^2_{0.75} = 4.11$

$\therefore X^2_{cal} > X^2_{tab.}$ significant

(3) F - test :

The analysis of variance as performed by the F-test makes it possible to separate the total variances of a process into its component parts.

$$F = \frac{S_1^2}{S_2^2} \qquad \text{where} \quad S_1^2 > S_2^2$$

$(d.f)_1 = V_1 = n_1 - 1$, $(d.f) = V_2 = n_2 - 1$

n_1, n_2 = sample size.

EX The data of two test are as follows

method I : 79.7 79.5 79.6 79.5 79.7

method II : 79.2 79.7 79.5 79.4 80 79.6 79.8

Sol. for method I :

$$\bar{X}_I = \frac{\sum X_i}{n} = \frac{398}{5} = 79.6 \quad , \text{where } n = 5$$

$$d.f = n-1 = 5-1 = 4 \quad , \text{and} \quad S_I^2(X) = \frac{\sum (X_1 - \bar{X}_1)^2}{n-1} = 0.01$$

for method II :

$$\bar{X}_{II} = \frac{\sum X_i}{n} = 79.6 \quad , \text{where} \quad n = 7$$

$$d.f = n-1 = 7-1 = 6 \qquad S_{II}^2(X) = \frac{\sum (X_2 - \bar{X}_2)^2}{n-1} = 0.07$$

From Table at $(d.f)_{II} = 6$ and $(d.f)_I = 4$ and 95% , $F_{tab.} = 6.16$

$$F_{cal.} = \frac{S_{II}^2(X)}{S_I^2(X)} = \frac{0.07}{0.01} = 7$$

$F_{cal} > F_{tab.}$ The difference variance is significant