BIOSTATS MANUAL

"C BATCH"

LT. BRKMGMC JAGDALPUR

NOTE: SOME OF THE GRAPHS etc ARE MISSING.

		$\sqrt{\frac{p_x q}{n_1} + \frac{p_x q}{n_2}}$
		Where p = $\frac{n_1 \times p_1}{n_1} + \frac{n_2 \times p_2}{n_2}$
± 't'	Difference divided by its	't' = $\frac{M - \bar{X}}{SE}$ (paired 't' test)
	Standard Deviation (Error)	$\frac{\bar{X}_1 - \bar{X}_2}{SE}$ (Unpaired 't' test)
Df	Degree of Freedom	= (c - 1) (r - 1)
p	Probability (Usually stated as a decimal but also as percent)	
χ^2	Chi-square for 2x2 table (chi is to be pronounced as kye rhyming with eye.)	If *df* = 1
	with yate's correction	$\sum \frac{(IO - EI - ½)^2}{E}$
		OR
		$\frac{(ab - bc\,I - ½K)^2 K}{efgh}$
O	Observed Value	
E	Expected Value	If *df* > 1
		$\sum \frac{(O - E)^2}{E}$

1. INTRODUCTION: LEVELS OF MEASUREMENT

Statistics as a plural noun means masses of figures and as a singular noun it means the science statistics.

Statistics refers to scientific methods of collecting, processing, analyzing and interpretation off numerical data.

Data are information in numerical form generated by observations or experimentation.

The data are of two types:

(1) Quantitative data (measurement data)

(2) Qualitative data (categorical or enumeration data)

It is essential to know type of data because they require separate statistical treatment. Quantitative data can be either continuous (value can be fractional e.g. height 1.83 meter) or discrete (only full integer values e.g. No. of children, white blood cell count). Qualitative (categorical) variables are measured either on a nominal or an ordinal scale and quantitative variables are measured on an interval or ratio scale.

Nominal : Observations are placed into broad categories which may be denoted by symbols or labels or names. e.g. Diagnostics groups like cancer, heart disease etc.

Ordinal : Categories are ranked or ordered. Each category is in unique position in relation to other categories but distances between the categories are not known. e.g. Severity of illness.

Interval : In addition to ordinal level of measurement, distance between any two numbers (value of the variable) is fixed and equal. The origin is arbitrary. e.g. Temperature in °C or °F.

Ratio : In addition to interval level of measurement it has true Zero point as its origin. e.g. weight in kg. or pounds.

Exercise:

1) Identify level (scale) of measurement of following variables.

i) ABO blood group system. Qualitative, Nominal

ii) Height (in cms.). Quantitative, Interval

iii) IQ. Qualitative

iv) Rank of a student in class. Ordinal, Quantitative

v) Time taken to complete given task. Qualitative

vi) Level of education-primary, secondary etc. Ordinal, Qualitative.

vii) Number of cases of malaria in a given P.H.C. area. Qualitative, Nominal

viii) Birth place of an individual. Nominal, Qualitative

ix) Degree of malnutrition of a child. Ordinal, Qualitative

x) Area of a town. Quantitative, Ordinal

xi) Score on a depression scale. Qualitative, ordinal

xii) Social status of a person measured as middle class, lower class etc. Qualitative, ordinal

xiii) Classification of a place: urban, rural etc. Qualitative, Nominal

xiv) Classification of employee into clerical, supervisory etc. categories. Qualitative Nominal

xv) Gender or patients. Qualitative, Nominal.

2. COLLECTION CLASSIFICATION AND ORGANIZATION OF STATISTICS (DATA): -

The word statistics is used with two distinct meanings. Statistics as a plural noun, as it is use above, means masses of figures. As a singular noun, it means the science of statistics. The science of statistics is concerned with the collection, presentation, analysis and interpretation of numerical data affected by multiple factors.

Data are of two types as the observations are made in two ways. These two types require separate statistical treatment.

1. **Measurement data:** - The data are quantitative, Measurements may be fractional e.g. blood sugar, body weight etc.
2. **Enumeration data:** - Figures obtained on counting the data are qualitative and represent particular attribute of characteristic, e.g. 'cured' and 'not cured' birth etc.

Care and thoroughness are essential requirements in clinical, observational and experiment, techniques and in recording. The units of measurement must be clearly mentioned. Records should be correct, complete, sufficiently concise and so arranged that they are easy to comprehend.

Qualitative / Discrete / Ungrouped / Nominal	Quantitative / Grouped data / Continuous data / Measurable
1. Bar diagram 2. Pie diagram / Sector Diagram 3. Pictogram 4. Epidemic / Map diagram	1. Histogram 2. Frequency polygon 3. Ogive 4. Line diagram 5. Scattered diagram

A. Collection of data: -

N.B. in each of the following example, mention the type of data.

1. How much is the heamoglobin level in your blood?
2. When was it measured and by which method?
3. Count your resting pulse rate. Repeat after exercise.
4. What is the number of students wearing spectacles in your batch?
5. Tick your family history of: -

   Diabetes / Hypertension / Asthma / Allergy / Epilepsy / Obesity / Tuberculosis / Any other / Nil.

What is your blood group? (ABO and Rh).

B. Discussion on the Data collection by the class: -

Principles of presentation: -

i) To arrange the data in such a way that it will arouse interest in a reader.
ii) To make the day a sufficiently concise without losing important details.
iii) To present the data in simple form to enable the reader to form quick impressions and to draw some conclusions, directly or indirectly.
iv) To facilitate further statistical analysis.

Tabulation: -

First step in presentation and analysis of data.

Frequency Distribution Table: -

Frequency distribution table is a table showing the frequency with which the values (observations) are distributed into different mutually exclusive groups with some defined characteristic or characteristics.

Rules for Making the Table: -

(1) Divisions (groups) should not be too broad or too narrow.
(2) Number of groups should be ordinarily between 10 to 20.
(3) Class interval must be the same throughout.
(4) Heading must be clear, sufficient and fully defined.
(5) If data include rates, mention the denominator.
(6) Rates or proportions should not be given alone without information as to the number of observations form which these were derived.
(7) Full details of deliberate exclusion of observations in a collected series must be given.

Exercise

Example 1: In a study of susceptibility to diphtheria 982 children were studied in Dantewada. Of the 494 boys 63 were Schick positive i.e. susceptible. Eight hundred and seven children were Schick negative, and out of these 376 were girls. Prepare a table.

Susceptible	Boys	Girls	Total
Schick +ve	63	112	175
Schick –ve	431	376	807
Total	494	488	982

Example 2: A group of 3579 persons was protected against post-traumatic tetanus. of the 1546 who were given Benzathine Penicillin intramuscularly, 578 has punctured wounds, 271 had lacerated or contused wounds and 306 had incised wounds. Of the 962 with other types of injuries 571 received A.T.S. which was also administered to 767 with punctured wounds and 381 with incised wounds. Present this information in a tabular form.

Types of wounds	Protected By Benzathine Penicillin	Protected By A.T.S.	
Punctured wound	578	767	1345
Lacerated wound	271	314	585
Incised wound	306	381	687
Others	391	571	962
Total	1546	2033	3579

Example 3: Heamoglobin values in grams percent of 100 first M.B.B.S. first term medical students of a Medical College Jagdalpur are given below. The data were collected in 2005 June. Prepare a frequency distribution table and comment on what you notice from it.

11.4	12.4	12.8	13.0	12.0	11.0	12.9	12.4	11.5	11.2
11.3	01.8	12.4	12.7	11.5	13.0	12.8	11.0	10.5	12.4
13.2	11.9	11.4	12.7	12.8	12.5	12.1	13.0	11.8	12.3
11.8	12.2	12.3	12.4	12.1	12.4	12.9	13.0	11.0	11.5
12.9	13.2	13.1	12.0	12.5	12.8	13.4	12.5	12.2	10.9
11.3	11.5	12.1	10.8	13.9	13.5	13.2	11.8	12.8	12.5
12.0	11.5	11.9	11.8	11.7	11.6	12.2	11.9	11.3	10.6
10.5	12.0	13.1	10.9	13.8	13.6	11.5	11.9	10.8	10.1
12.3	12.6	13.9	10.1	10.9	11.1	10.9	11.8	12.5	12.7
11.8	11.9	11.4	11.9	12.0	13.4	12.2	13.6	12.3	13.4

Maximum value – 13.9 gm%

Minimum value – 10.1 gm%

Range = Maximum - Minimum = 13.9 – 10.1 = 3.8 gm%

$\frac{3.8}{10}$ = 0.38 , Round figure = 0.4

S.no.	Class Interval	Tally marks	Frequency
1.	10.1 – 10.5	II	2
2.	10.5 – 10.9	~~IIII~~ I	6
3.	10.9 – 11.3	~~IIII~~ IIII	9
4.	11.3 – 11.7	~~IIII~~ ~~IIII~~ III	13
5.	11.7 – 12.1	~~IIII~~ ~~IIII~~ ~~IIII~~ III	18
6.	12.1 – 12.5	~~IIII~~ ~~IIII~~ ~~IIII~~ II	17
7.	12.5 – 12.9	~~IIII~~ ~~IIII~~ IIII	14
8.	12.9 – 13.3	~~IIII~~ ~~IIII~~ II	12
9.	13.3 – 13.7	~~IIII~~ I	6
10.	13.7 – 14.1	III	3
	Total		100

Inference :-

The maximum no. of students have Haemoglobin concentration in the range of 11.7 – 12.1 gm%

And there are least no. of students in the range of 10.1 – 10.5 gm%

4. MEASURES AND LOCATION OF CENTRAL TENDENCY

The measures, also called 'centering constants', provide a summary or a single representing value the observation.

Mean : The arithmetic Mean.

Median : The Central value when the observations are arranged in the order of magnitude.

Mode : The most common or frequent value.

Mean = $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Mean of grouped data, $\bar{x} = \dfrac{\sum_{i=1}^{n} f_i x_i}{N}$, where $N = \sum_{i=1}^{n} f_i$ = total frequency.

Median Md = $\dfrac{L + \frac{N}{2} - C.f. \times i}{f_m}$

where, L = Lower limit of median class interval
N = total frequency
C.f = cummulative frequency of class proceeding the median-class interval
fm = frequency of median class interval
i = width of median class interval

Mode : Mode of a series of values of a variable is that value which which occurs with the maximum frequency i.e. it is the most frequent value.

Median
for eg.
In series of values – 2, 4, 7, 5, 2, 7, 5, 8, 9, 5
we see that 5 occur most frequently i.e. 5 is mode, which represents true characteristics of series.

Mean
Symbolically $\bar{x} = \sum_{i=1}^{n} x_i / n$

Q. Find the mean incubation period of 9 polio cases given below :-
17, 20, 18, 24, 16, 19, 21, 22, 23

Ans :- $\bar{x} = \sum_{i=1}^{9} x_i / 9$

$\bar{x} = \dfrac{17+20+18+24+16+19+21+22+23}{9}$

$\bar{x} = \dfrac{180}{9} = 20$

∴ Mean incubation period = 20

$\bar{x} = \dfrac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \dfrac{\sum_{i=1}^{n} f_i x_i}{N}$

Q. Find the mean weight of 100 people from the following frequency distribution.

weight in kg = 45, 50, 55, 60, 65, 70, 75
No of persons = 5 12 18 20 33 10 2

Sol:-

wt. in kg x	No. of persons (f)	fx
45	5	225
50	12	600
55	18	990
60	20	1200
65	33	2145
70	10	700
75	2	150
Total.	100	6010

$\sum f = 100$

$\sum fx = 6010$

Mean, $\bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{6010}{100} = 60.1$

Hence, mean weight of 100 people is 60.1 kg.

Median

Q. In a community survey of 100 families, the following distribution of no. of children was obtained :-

No. of children	1 - 3	3 - 5	5 - 7	7 - 9
No. of families	20	42	30	8

find the median of above distribution.

Class Interval No. of children	No. of families frequency	Cummulative frequency
1 - 3	20	20
3 - 5	42	62
5 - 7	30	92
7 - 9	8	100

$N = 100$
$c.f = 20$
$f.m = 42$
$i = 2.$

$$\text{Median} = \frac{L + \frac{N}{2} - c.f.}{fm} \times i$$

$$= \frac{3 + \left(\frac{100}{2} - 20\right)}{42} \times 2$$

$$= 3 + \left(\frac{50 - 20}{42} \times 2\right)$$

$$= 3 + 1.5$$

$$= 4.5$$

$\therefore$ Median = 4.5.

19

5. MEASURES OF VARIABILITY

Variability is a biological characteristic. There is no number which is the normal for a group but there is a range of normal values.

Variability may be because of: -

1. Biological Variability
   a. By category of individual
   b. Which categories from individual and from time to time.
2. Measurement Error.
   a. Observer error
   b. Measuring device or instrument error
3. Sampling Variability or Error

Measures of Variability

Range: It is the distance between the lowest and the highest values in the distribution.

Standard Deviation: This is a most commonly used measure.

Coefficient of Variation: This is used to compare to scatter (variability) of two distributions with different units of measurement.

Standard Deviation

$$SD / \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x-\bar{x})^2}{N}}$$

where N = Total no. of values. when sample size > 30, when < 30 use n-1

$\bar{x}$ = mean

Or, it can be expressed as

$$\sigma = \sqrt{\frac{\sum fd^2}{fi} - \left(\frac{fd}{fi}\right)^2}$$

Coefficient of Variation $\frac{\sigma}{x} \times 100$

Q. Find the quartile deviation of weight (in kg) of 7 persons given below –
50, 62, 55, 58, 65, 43, 52

Sol:- Arranging in ascending order —
43, 50, 52, 55, 58, 62, 65

N = 7

$Q_1 = \frac{N+1}{4} = \frac{7+1}{4} = \frac{8}{4} = 2^{nd}$ value = 50

$Q_3 = \frac{3(N+1)}{4} = 3(2) = 6^{th}$ value = 62

Hence, quartile deviation = $\frac{Q_3 - Q_1}{2} = \frac{62 - 50}{2} = 6$ i.e. 6kg of weight.

Q. Find out standard Deviation of daily temperature in a city happens to be-

38, 40, 42, 41, 39, 42, 40, 38, 41, 39.

Sol:- Mean $= \frac{38+40+41+42+39+42+40+38+41+39}{10}$

$= \frac{400}{10} = 40$

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
38	40	38 − 40 = −2	4
40	40	40 − 40 = 0	0
42	40	42 − 40 = 2	4
41	40	41 − 40 = 1	1
39	40	39 − 40 = −1	1
42	40	42 − 40 = 2	4
40	40	40 − 40 = 0	0
38	40	38 − 40 = −2	4
41	40	41 − 40 = 1	1
39	40	39 − 40 = 1	1
			20

$n = 10$, $(x-\bar{x})^2 = 20$.

$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{20}{10-1}} = \sqrt{20/9} = 1.49$$

Q. Find S.D. of respiration rate per minute found to be 16, 18, 19, 17, 21, 24, 22, 23 in 8 individuals.

Sol:- Mean $= \frac{16+18+19+17+21+24+22+23}{8}$

$= \frac{160}{8} = 20$

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
16	20	16 − 20 = −4	16
18	20	18 − 20 = −2	4
19	20	19 − 20 = −1	1
17	20	17 − 20 = −3	9
21	20	21 − 20 = 1	1
24	20	24 − 20 = 4	16
22	20	22 − 20 = 2	4
23	20	23 − 20 = 3	9

21

$n = 8$, $(x-x)^2 = 60$.

$$S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{60}{8-1}} = \sqrt{60/7} = 2.9$$

Q. Find standard deviation for the following distribution :-

Age in years -	25	30	35	40	45	50	55	60
No. of persons -	15	28	42	50	30	20	10	5

Arithmetic Mean = 40.

Age in years	Deviation $d = x - \bar{x}$	frequency (f)	d^2	$f \times d$	$(f \times d)^2$
25	−15	15	225	−225	3375
30	−10	28	100	−280	2800
35	−5	42	25	−210	1050
40	0	50	0	0	0
45	5	30	25	150	750
50	10	20	100	200	2000
55	15	10	225	150	2250
60	20	5	400	100	2000
		200		+115	14225

$$\sigma = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{14225}{200} - \left(\frac{-115}{200}\right)^2}$$

$$= \sqrt{71.1 - 0.3} = \sqrt{70.8} = 8.41 \text{ approx.}$$

Q. In a series of 100 individuals, mean blood glucose level in mg/dl was found to be 155 with S.D = 52. In some individuals the mean cholesterol level in mg/dl was found to be 210 with SD = 36. Find which character shows greater variation.

Sol:- Coefficient of variation for Blood glucose

$$= \frac{\sigma}{\bar{x}} \times 100 = \frac{52}{\cancel{155}_{31}} \times \cancel{100}^{20} = \frac{1040}{31} = 33.5\%$$

Coefficient of variation for cholesterol

$$= \frac{\cancel{36}^{12}}{{}_{7}\cancel{210}} \times 10\cancel{0} = \frac{120}{7} = 17.1\ \%$$

Q. ~~find the S.D. of following data.~~
Thus, Blood glucose is found to be more variable character as compared to serum cholesterol.

Q. find the S.D. of following data :—

Height —	95-105	105-115	115-125	125-135	135-145
No. of children —	20	55	95	70	60

Height	Mid value	Frequency	Deviation (d)	d^2	fd	fd^2
95-105	100	20	-20	400	-400	8000
105-115	110	55	-10	100	-550	5500
115-125	120	95	0	0	0	0
125-135	130	70	10	100	700	7000
135-145	140.	60	20	400	1200	24000
		300			950	44500

Arithmetic Mean = 120. , n = 300

$$S.D. = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \Sigma\left(\frac{fd}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{445\cancel{00}}{3\cancel{00}} - \left(\frac{95\cancel{0}}{30\cancel{0}}\right)^2}$$

$$= \sqrt{148.33 - 10.24}$$

$$= \sqrt{138.09}$$

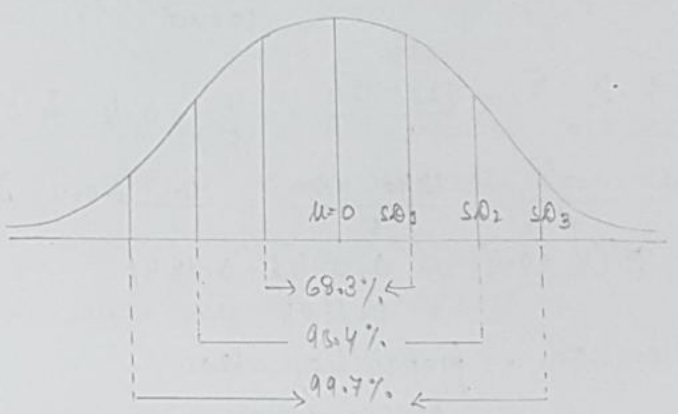
$$= 11.75$$

6. NORMAL DISTRIBUTION

The 'Normal' curve is one particular form of a symmetrical distribution. This mathematical curve has a great importance in statistical theory. It is fundamental to the tests of significance. These tests are based on the assumption that the distribution is more or less 'Normal'

Normal = Standard or usual and not 'not abnormal'

Apart from being symmetrical an important property of the curve is, 'the proportion of the area lying under the curve between the mean and a point located at any multiple of S.D. from it, is constant for every normal curve.



Q. In a health survey of 1200 persons in Raipur area, mean systolic BP of a healthy male is 120 mm Hg. with S.D. of 5 mm Hg. Calculate.

a) What % of person who has BP > 133.5 mm Hg?

(b) No. of persons having BP in between 122 - 124.5 mm Hg.

(c) find no. of persons who has BP < 119 mm Hg.

Sol:- (a) $z = \frac{x - \bar{x}}{S.D} = \frac{133.5 - 120}{5} = \frac{13.5}{5} = 2.7$

here $x = 133.5$, $\bar{x} = 120$, $S.D = 5$

Actual value of the % required
$= 0.5 - 0.4965$
$= 0.0035$

Area of normal curve = 0.4965

In percentage, value = $0.0035 \times 100 = 0.35\%$

% of person having BP > 133.5 mm Hg in 1200 persons
$= 0.0035 \times 1200 = \frac{35}{10000} \times 1200 = 4.20$

(b) To find $z = \frac{x - \bar{x}}{\sigma} = \frac{122 - 120}{5} = \frac{2}{5} = 0.4$ is 0.3446

$z = \frac{x - \bar{x}}{\sigma} = \frac{124.5 - 120}{5} = \frac{4.5}{5} = 0.9$ is 0.1841

$P(z < 0.4) - P(z < 0.9) = 0.3446 - 0.1841$
$= 0.1605$

No. of persons covered = probability × 1200
$= 0.1605 \times 1200$
$= 192.6 \simeq 193$ people.

(c). To find $P(x < 119)$

$z = \frac{x - \bar{x}}{\sigma} = \frac{119 - 120}{5} = \frac{-1}{5} = P(z < 0.2) = 0.4207$

No. of persons with BP < 119 mm Hg $= 0.4207 \times 1200$
$= 504.84$
$= 505$ persons.

Q. The mean height of 500 students is 165 cm & S.D = 5 cm. Assuming that heights are normally distributed, find how many students will have height between 153 - 180 cm. 23

Sol:- $\bar{x} = 165$ cm

$\sigma = SD = 5$ cm

$z = \frac{x - \bar{x}}{S.D} = \frac{153 - 165}{5} = \frac{-12}{5} = -2.4$

& $z = \frac{x - \bar{x}}{SD} = \frac{180 - 165}{5} = \frac{15}{5} = 3$

∴ Proportion of students whose height are in range between 153 & 180 (cm)
= Area under standard distribution curve between z -2.4 to 3
= (Area between $z = -2.4$ & $z = 0$) + (Area between $z = 0$ & $z = 3$)
$= 0.4918 + 0.4987$
$= 0.9905$

Hence required no. of students whose height are in range b/w 153 & 180 cm $= 0.9905 \times 500 = 495$

Q. A hospital records the weight of every newborn child at hospital. The distribution of weight is normally shaped, was mean $(\mu) = 2.9$ kg, SD = 0.45.

Find -

(a) % of newborn whose weight is 2.1 kg.

$\mu = 2.9$ kg, $\sigma = 0.45$ kg.

$z = \frac{x - \mu}{\sigma} = \frac{2.1 - 2.9}{0.45} = \frac{-0.8}{0.45} = -1.78$.

Area beyond z score of 1.78 = 0.5 − 0.4625 = 0.0375

Hence, % of newborn with wt. < 2.1 kg = $0.0375 \times 100 = 3.75\%$

(b) % of newborn who weigh 1.8 kg and 4 kg.

z score of 1.8 kg $= \frac{1.8 - 2.9}{0.45} = -2.4$

z score of 4 kg $= \frac{4 - 2.9}{0.45} = 2.4$

To locate case corresponding to $z = -2.4$ to $z = 2.4$
= Area between -2.4 to 0 & from 0 to 2.4.

$= 0.4918 + 0.4918$
$= 0.9836$.

Hence % of new born who weighed between 1.8 & 4kg $= 98.4\%$

Q. Assume that the age of onset of disease X is distributed normally with a mean of 50 years & SD of 12 years. What is the probability that an individual affected with 'X' has developed it before 35 years.

Sol:- $\mu = 50$ years
$\sigma = SD = 12$ years
z corresponding to 35 years
$z = \frac{35-50}{12} = \frac{-15}{12} = -1.25$

Area under curve of normal distribution $z = -1.25$ is 0.3944
∴ Area under standard normal curve below $z = -1.25$
$= 0.5 - ($ Area Between $z = -1.25$ & $z = 0)$
$= 0.5 - 0.3944$
$= 0.1056$
Required probability $P(x < 35) = z(-1.25)$
$=$ Area $= 0.5 - 0.3944 = 0.1056$.

Q.

	Healthy women	Anaemic women
Mean Hb% →	13.8	10.1
SD →	0.6	1.4

Assuming that each group is normally distributed →

(a) What proportion of healthy women has Hb b/w 12.6 to 15.0?
For 68% healthy women, means 1 SD
Mean ± 1 S.D $= 13.8 \pm 1 \times 0.6 = 14.4$ to 13.2.
At 95% ⇒ $13.8 \pm 2 \times 0.6 = 13 \pm 1.2$ ⇒ 15 and 12.6.
Hence - 12.6 to 15 represents mean ± 2 SD which include 95% value of given data.

(b) What Hb % will include 68% anaemic women & 95% anaemic women?
For 68% Anaemic women $=$ Mean ± 1 SD
$= 10.1 \pm 1 \times 1.4 = 10.1 + 1.4, 10.1 - 1.4$
$= 8.7, 11.5$.

For 95% Anaemic women $= 10.1 \pm 2$ SD
$= 10.1 \pm 2 \times 1.4 = 10.1 + 2.8, 10.1 - 2.8$

⇒ 12.9, 7.3.

24

(c) What Hb % will include 68% healthy women & 95% healthy women?

For 68% healthy women $= 13.8 \pm 1$ SD
$= 13.8 + 0.6, 13.8 - 0.6$
$= 14.4, 13.2$

For 95% healthy women $= 13.8 \pm 2$ SD
$= 13.8 + 1.2, 13.8 - 1.2$
$= 15$ and 12.6.

Q. Mean Height of 48 Students is 147 cm, SD $= 6.5$ cm.
(a) What % of students will have height > 158 cm?
$\bar{x} = 147$ cm
$\sigma = 6.5$ cm
$z = \frac{x - \bar{x}}{\sigma} = \frac{158 - 147}{6.5} = 1.69$

Area corresponding to 1.69 is 0.4545
% of children $= 0.5 - 0.4545$
$= 0.0455$ of area.
% of children > 158 cm height $= 0.0455 \times 100 = 4.55 \sim 5$ students.

(b) How many students will have height between 158 - 165 cm?
$P(158 > x > 165)$
$z = \frac{158-147}{6.5} = 1.69$ at area $= 0.4545$
$z = \frac{165-147}{6.5} = \frac{18}{6.5} = 2.769 = 2.77$ area is 0.4972

Height between is $0.4972 - 0.4545 = 0.0427$.
In 48 students $= 48 \times 0.0427$
$= 2.0496$
~ 2 students.

27

Thus it can be interpreted will 95% confidence that the population of school children proportion of hearing defects is more than 11.16% but less than 20.84%.

Example:- Systolic BP of 100 males was taken, Mean BP was found to be 128 mm Hg., and S.D = 13 mm Hg.
find - (i) 95%.
(ii) 99%.

Sol:- $\bar{x} = 128$ mm Hg, $n = 100$

$S.E$ of $\bar{x} = \frac{SD}{\sqrt{n}} = \frac{13}{\sqrt{100}} = \frac{13}{10} = 1.3$

(i). 95% confidence limits-

$= \bar{x} \pm 1.96\,(SE\,\bar{x})$

$= 128 + 1.96(1.3)$ & $128 - 1.96 \times 1.3$

$= 128 + 2.5$ & $128 - 2.5$

$= 125.5$ & 130.5

(ii) 99% of confidence limit

$= \bar{x} \pm 2.58\,(SE\,\bar{x})$

$= 128 + 2.58(1.3)$ & $128 - 2.58(1.3)$

$= 128 + 3.4$ & $128 - 3.4$

$= 124.6$ & 131.4

28

Example:- Average Hb of pregnant women in 1 area is 10.5 gm%. with SD = 1 gm%. If Hb level are normally distributed then,
(a) Hb of 8.4 gm% as abnormal
(b) Hb of 11.5 gm% as normal.

Ans: (a) $z = \frac{x - \mu}{SD} = \frac{8.4 - 10.5}{1} = -2$

Confidence limit = Mean $\pm$ 2 SD

$= 10.5 \pm 2 \times 1$

$= 10.5 \pm 2$

$= 8.5$ & 12.5.

So. 8.4 gm% Hb is abnormal

(b) $z = \frac{x - \mu}{SD} = \frac{11.5 - 10.5}{1} = 1$

Confidence limit = Mean $\pm$ 2 SD

$= 10.5 \pm 2 \times 1$

$= 8.5$ & 12.5.

So, Hb of 11.5 gm% is normal.

Q. A newer drug has duration of action in 50 days with a mean of 10 hours & S.D of 2 hours. How frequently can accept a duration of action :-

(a) 4 hours or less

(b) 9 hours or more

Ans :- (a) $z = \frac{x - \mu}{SD} = \frac{4 - 10}{2} = \frac{-6}{2} = -3$

Confidence limit = Mean $\pm$ 2SD

$= 10 \pm 2(2)$

$= 10 \pm 4$

$= 14$ and 6.

$z = -3$, Area $= 0.49865$

Frequency accept a duration of 4 hour or less

$= 0.5 - 0.49865 = 0.00135$

$\therefore$ 0.135 % of frequency accept a duration of 4 hour.

(b) $z = \frac{9-10}{2} = \frac{-1}{2} = -0.5$.

Area for $z = -0.5$ is 0.1915

30

Standard Error of Proportion is a measure off variability of proportions in taking repeated samples from same population.

Note: -

The formula gives the correct result when p = q and n is a large number. Whenever p and q are very divergent or n is a small number it is best to refer to a table or graph for the **Binomial Confidence Limit.**

(d) STANDARD ERROR OF DIFFERENCE BETWEEN TWO PROPORTIONS

With the help of this measure we can calculate the size of difference between two sample proportions that might be expected to occur by chance. In actual practice we do not know the value off population proportion. So we have to substitute it by the value noticed in the sample or samples. Such a substitution is valid only if the sample is large.

Note: - Where as a difference between two proportions may fall within the value of p ± 2 S.E. it does not conclusively prove that the observation (result) is only due to play of chance and not due to play of the particular factor under study. It could have been due to that factor but it is more likely to be due to play off chance than due any particular causal factor. Similarly a value outside the range of p ± 2 S.E. could be due to chance also but it is less likely to be so. It is more likely to occur due to the factor under study which differentiates the two groups (being similar in all other respects). Remember that in the test of significance we are weighing probabilities and not offering mathematical proofs.

* Standard Error of Mean

SE of Sample mean $= \bar{x} = \frac{\sigma}{\sqrt{n}}$

* Standard Error of Sample Proportion

$p = \sqrt{\frac{pq}{n}}$ where $q = 1 - p$.

* Standard Error of difference between 2 means $\bar{x}_1$ and $\bar{x}_2$.

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$SE \text{ of } P_1 - P_2 = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

34

Q. A random sample of size 64 is drawn from a finite population consisting of 122 units. If population standard deviation is 16.8, find S.E.

Sol:- $S.D = \sigma = 16.8$

$n = 64$

So, Standard Error of sample $= \dfrac{\sigma}{\sqrt{n}}$

$= \dfrac{16.8}{\sqrt{64}}$

$= \dfrac{16.8}{8}$

$= 2.1$

Q. A random sample of 800 articles is drawn from a large consignment, containing 128 defective articles. Find SE of proportion of defective articles in a sample.

Sol:- $n = 800$,

p = proportion of defective articles

$= \dfrac{128}{800} \quad \dfrac{32}{200} \quad \dfrac{16}{100}$

$= 0.16$

& $q = 1 - p = 1 - 0.16 = 0.84$

Hence, standard Error of proportion

$$SE(P) = \sqrt{\frac{Pq}{n}}$$

$$= \sqrt{\frac{0.16 \times 0.84}{800}}$$

$$= \sqrt{\frac{16 \times 84}{8000000}}$$

$$= \frac{\cancel{4}\,2}{1000 \times \cancel{2}} \sqrt{\frac{\cancel{84}\,42}{2}}$$

$$= \frac{2 \times 6.48}{1000}$$

$$= \frac{12.96}{1000} = \frac{1296}{100000} = 0.0129$$

$= 0.013$ approx

Example - Index of Brightness of 50 boys and 50 girls have following values —

	Mean	SD
Boys	91.2	5.23
Girls	90.8	4.41

Q. find 90 % confidence limit of mean index of brightness in population of both.

Sol. $\overline{x_1} = 91.2$ $\quad \sigma_1/S_1 = 5.23$

$\overline{x_2} = 90.8$ $\quad \sigma_2/S_2 = 4.41$

Hence, $SE(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$= \sqrt{\frac{(5.23)^2}{50} + \frac{(4.41)^2}{50}}$

$= \sqrt{\frac{27.35}{50} + \frac{19.45}{50}}$

$= \sqrt{0.547 + 0.389}$

$= \sqrt{0.936}$

$= 0.97$

Now at 90% C.I. $= \overline{x} \pm 1.64 \; SE(x_1 - x_2)$

$= (91.2 - 90.8) \pm 1.64 \times 0.97$

$= 0.4 \pm 1.6$

$= 1.2$ to $2.$

35

10. CHI – SQUARE TEST (χ^2 – TEST)

This test is applied for analysing the enumeration data, resulting from counting the number of individuals, etc., that possesses a certain attribute.

Note: -

1. Working of χ^2 – test is done with the actual numbers, observed and expected. It cannot be done by using percentages.
2. The degrees of freedom are related not to the total number of observations (n) but to the number of categories. In many tables of χ^2, the column for the degrees of freedom is denoted as n (and not *d.f.*).
3. Yate's correction (reduction of the absolute difference between O and E by half)' is applied only to the fourfold (2x2) tables, or whenever *d.f.* = 1. The correction is not used where *d.f.* > 1.
4. In case of very small samples it may be misleading. If the expected value in any cell is less than 5, it is best not to use χ^2 – test, unless the χ^2 – calculated is very large.

* Chi-Square test is denoted by χ^2

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

where O = observed frequency
E = expected frequency

Event	A_1	A_2	A_n
Observed ν	O_1	O_2	O_n
Expected ν	E_1	E_2	E_n

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$$

* Degree of Freedom :-

The no. of degree of freedom is defined as total no. of observation – no. of independent constraints imposed on observations

$$d.f = (c-1)(r-1)$$

where c = no. of columns
r = no. of rows.

Q. Out of 250 diarrhoea cases with dehydration, 150 were treated with home made fluids, remaining 100 were treated with ORS. 30 have not recovered from each groups. Find out is there any difference in treatment outcome of ORS and home made fluids.

Sol:-

	ORS	Home made fluids	
Recovered	70	120	190
Not Recovered	30	30	60
Total	100	150	250

Null Hypothesis H_0
There is no significant difference in degree between both.

Alternate Hypothesis H_1
There is significant difference between both ORS and homemade fluids.

Expected values = $\frac{\text{row wise total} \times \text{Column wise total}}{\text{Grand total}}$

E for cell 1 = $\frac{190 \times 100}{250} = \frac{380}{5} = 76$

cell 2 = $\frac{190 \times 150}{250} = \frac{570}{5} = 114$

cell 3 = $\frac{60 \times 100}{250} = 24$

cell 4 = $\frac{60 \times 150}{250} = 36$

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(70-76)^2}{76} + \frac{(120-114)^2}{114} + \frac{(30-24)^2}{24} + \frac{(30-36)^2}{36}$$

$$= \frac{36}{76} + \frac{36}{114} + \frac{36}{24} + \frac{36}{36}$$

$$= 1.5 + 1 + 0.5 + 0.3 = 3.3$$

37

degree of freedom $= (c-1)(r-1)$
$= (2-1)(2-1)$
$= 1$

At 5% significant value = 3.84

Calculated value is < tabulated value at 1 freedom.

Conclusion — we can say "ACCEPT THE TEST" = null hypothesis is accepted at 5% level of significance & at 1 freedom.

Hence there is no significante difference between the ORS & home made fluids in the treatment of diarrhoea.

Q. Among 200 alcoholic, 50 have developed cirrhosis. Among 300 non-alcoholic, 50 have developed cirrhosis in a cohort study. Find out is there any association between alcoholism & cirrhosis.

Sol:-

	Alcoholic	Non Alcoholic	Total
cirrhosis developed	50	50	100
cirrhosis didn't developed	150	250	400
	200	300	500

Null hypothesis (H_0) - Assume that there is no association between alcohol and cirrhosis.

Alternate hypothesis (H_1) - assuming that there is association between alcoholism & cirrhosis.

Expected value = $\frac{\text{total rows} \times \text{total column}}{\text{Grand total}}$

E for cell 1 = $\frac{100 \times 200}{500} = \frac{200}{5} = 40$

E for cell 2 = $\frac{100 \times 300}{500} = \frac{300}{5} = 60$

E for cell 3 = $\frac{400 \times 200}{500} = 160$

E for cell 4 = $\frac{400 \times 300}{500} = 240$.

$\chi^2 = \sum \frac{(O-E)^2}{E}$

$= \cancel{40-4} \frac{(50-40)^2}{40} + \frac{(50-60)^2}{60} + \frac{(150-160)^2}{160} + \frac{(250-240)^2}{240}$

$= \frac{100}{40} + \frac{100}{60} + \frac{100}{160} + \frac{100}{240}$

$= 2.5 + 1.67 + 0.625 + 0.42$

$= 5.215.$

degree of freedom $= (c-1)(r-1)$
$= (2-1)(2-1)$
$= 1.$

At 5% level of significance, and d.f at 1; $\chi^2 = 3.84$ (tabulated)

χ^2 calculated $> \chi^2$ tabulated.

Conclusion - we say 'Reject the test' - null hypothesis rejected at 5% level of significance and 1° freedom.

Hence, there is significant difference between alcoholism & cirrhosis.

Q.

Group	Attacked	Non-Attacked	Total
Inoculated	10	90	100
Non-inoculated	26	74	100
Total	36	164	200

Null hypothesis - be that inoculation is ineffective

Alternate hypothesis - is that inoculation is effective.

The effective frequencies are as..

Expected $= \frac{\text{Row total} \times \text{Column total}}{\text{grand total}}$

E of cell 1 $= \frac{100 \times 36}{200} = 18$

E of cell 2 $= \frac{100 \times 164}{200} = 82$

E of cell 3 $= \frac{100 \times 36}{200} = 18$

E of cell 4 $= \frac{100 \times 164}{200} = 82.$

38

$\chi^2 = \sum \frac{(O-E)^2}{E}$

$= \frac{(10-18)^2}{18} + \frac{(90-82)^2}{(82)} + \frac{(26-18)^2}{18} + \frac{(74-82)^2}{82}$

$= \frac{64}{18} + \frac{64}{82} + \frac{64}{18} + \frac{64}{82}$

$= 3.6 + 3.6 + 0.78 + 0.78$

$= 8.76.$

Degree of freedom $= (C-1)(R-1)$
$= (2-1)(2-1)$
$= 1.$

Tabulated value of χ^2 at 5% level of significance & 1° freedom is 3.84

χ^2 calculated $> \chi^2$ tabulated

Thus, the null hypothesis is rejected & concluded that the inoculation is effective.

Paired t-test

40

Q. The height of 10 students selected at random from a school had a mean 116 cm & variance 96 cm. Test at 5% level of significance, the hypothesis is that the students are on average 120 cm in all degree of freedom $t = 0.05$ is 1.83.

Sol:- $n = 10$, $S.D = 9.79$, $\bar{x} = 116$ cm

$\mu = 120$ cm

$\bar{x} = 116$ cm

$$t = \frac{\bar{x} - \mu}{SE}$$

Null Hypothesis $\mu = 120$ cm

Alternate hypothesis $\mu \neq 120$ cm.

$$SE\bar{x} = \frac{S}{\sqrt{n-1}} = \frac{9.79}{\sqrt{10-1}} = \frac{9.79}{\sqrt{9}} = \frac{9.79}{3} = 3.26$$

$$t = \frac{\bar{x} - \mu}{SE\bar{x}} = \frac{116 - 120}{3.26} = \frac{-4}{3.26} = -1.23$$

$t = -1.23$.

As t calculated < t tabulated.

Null hypothesis is accepted & rejected the alternate hypothesis at 5% level of significance.

It means average height of students is not less than 120 cm.

Q: A group of 7 patients treated with medicine weighs 35, 39, 40, 42, 51, 48, 60 kg. Another group of 9 patients from same ward of hospital ttt. with medicine weighs 53, 56, 60, 62, 67, 43, 64, 45, 54 kg. Do you agree with the claim that medicine B increases the weight significantly? $t = 0.05 = 2.15$ 42

Sol:- Let the null hypothesis be $H_0 \Rightarrow \mu_1 = \mu_2$

& alternate hypothesis be $H_1 \Rightarrow \mu_1 \neq \mu_2$. or $\mu_1 < \mu_2$

Group I => 35, 39, 40, 42, 51, 48, 60

$\boxed{\text{Mean } \bar{X}_1} = 35+39+40+42+51+48+60 / 7 = 315/7 = 45.$

Group II => 53, 56, 60, 62, 67, 43, 64, 45, 54

$\boxed{\text{Mean } \bar{X}_2} = 53+56+60+62+67+43+64+45+54 / 9 = 504/9 = 56$

$$\boxed{t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{SD^2}{n_1} + \frac{SD^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{SD\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \text{where} = SD^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

We calculate SD:-

x_1	$(x-\bar{x}_1)$	$(x-\bar{x}_1)^2$	x_2	$(x-\bar{x}_2)$	$(x-\bar{x}_2)^2$
35	-10	100	53	-3	9
39	-6	36	56	0	0
40	-5	25	60	4	16
42	-3	9	62	6	36
51	6	36	67	11	121
48	3	9	43	-13	169
60	15	225	64	8	64
		440.	45	-11	121
			54	-2	4
					540

We know that $SD = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}$

$\Rightarrow SD^2 = \frac{\Sigma (x-\bar{x})^2}{n}$

$\Rightarrow SD^2 = \frac{n_1 SD_1^2 + n_2 SD_2^2}{n_1+n_2-2} = \frac{440+540}{7+9-2} = \frac{980}{14} = 70$

$\Rightarrow SD = \sqrt{70} = 8.4$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{SD\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{45-56}{8.4\sqrt{\frac{1}{7}+\frac{1}{9}}} = \frac{(45-56)10}{84\sqrt{\frac{9+7}{63}}}$$

$$= \frac{-110}{84\sqrt{\frac{16}{63}}} = \frac{-110}{84\sqrt{0.25}} = \frac{-110}{84 \times 0.5} = \frac{-1100}{5 \times 84}$$

-220 -110 -55 42 21

$$= \frac{-55}{21} = -2.61$$

Conclusion → $t_{\text{calculated}}$ (2.61) > $t_{\text{tabulated}}$ (2.15)

→ Hence, our null hypothesis is rejected which means medicine B increases the weight significantly.

12. STATISTICAL FALLACIES

Like any other discipline, statistics can be misused. International manipulations as well as blunders due to ignorance both are not uncommon in statistics. I remember a study in which 23, 28, and 22 patients were studied. They were given different medicines to see effect on blood pressure and pulse rate. To my surprise there was only one "t" value obtained for three groups! It is not uncommon to see a medical representative telling us that tablet x of company A contains 0.3 micrograms of oestrogen but our tablet x has 0.3 mg of oestrogen and hence it is causing lesser problems! Of course intelligent manipulations and ignorant mistakes both have damaging effect as far as basic purpose of research is concerned.

Comment on each statement stating with reason whether you agree or disagree or partially agree.

Exercise: -

1. Incidence of cardiovascular diseases in the persons who were in executive cadre was found to be 30% and the same was found to be 10% in the persons who were working as clerks. Hence a clerk runs only 1/3 the risk of getting cardiovascular diseases.

   Comment:-

2. Only10% of road accidents in a city occurred due to female drivers and that is because they drive more carefully.

   Comment:-

3. Five years following up study off breast cancer cases who were given operative treatment or chemotherapy showed that of the patients who were given operative treatment 60% were still alive as against only 30% who received chemotherapy. This clearly shows the superiority of the operative treatment over chemotherapy.

   Comment:-

4. Period Prevalence rate = $\dfrac{\text{No. of all existing cases (old \& new) of a specified disease during a given period of time interval}}{\text{Population at risk during that period}} \times 1000$

5. Average duration per spells of sickness = $\dfrac{\text{The total of the entire duration of all spells of sickness ending during a given period}}{\text{No. of spell ending during that period}} \times 1000$

6. Average duration per sick person = $\dfrac{\text{The total of the entire duration of all spells of sickness ending during a given period}}{\text{No. of persons who experienced at least one spells ending during that period}} \times 100$

7. Fatality rate = $\dfrac{\text{No. of deaths from a disease recorded during a period}}{\text{No. of cases of that disease recorded during that same period}} \times 100$

Q. The mid year population of a city was 6,50,000. There were 20500 live births, 400 still births, 850 deaths within 1 week of birth & 1600 death in 1st month of life in a particular year, 2400 in 1st year of life. Calculate -

(a) Still Birth Rate = $\dfrac{\text{Foetal death after 28 weeks of gestation}}{\text{Total live birth + Still birth}} \times 1000$

$= \dfrac{400}{20500+400} \times 1000 = \dfrac{400}{209\cancel{00}} \times 10\cancel{00} = \dfrac{4000}{209}$

$= 19.14$ per thousand.

(b) Perinatal Mortality Rate (PMR) = $\dfrac{\text{Still Birth + death < 1 week of age}}{\text{Total live Birth + Still Birth}} \times 1000$

$= \dfrac{400+850}{20500+400} \times 1000 = \dfrac{1250}{209\cancel{00}} \times 10\cancel{00} = 59.8$ per thousand live birth

(c) Neonatal Mortality Rate (NMR) = $\dfrac{\text{No. of deaths under 28 days of age}}{\text{Total live births}} \times 1000$

$= \dfrac{1600 \times 10\cancel{00}}{205\cancel{00}} = 78.05$ per 1000 live births.

(d) Postnatal M.R. = $\dfrac{\text{No. of deaths of children between 28 days and 1 year of age}}{\text{Total live birth}} \times 1000$

$= \dfrac{2400-1600}{20500} \times 1000 = \dfrac{8000\cancel{00}}{205\cancel{00}} = 39/1000$ live birth.

Q. The mid year population of a city was 4,00,000. In the same year, the no. of live birth was 12000, No. of deaths 6400, no. of infant death 1500.

Calculate :—

(a) Crude Birth Rate (CBR) = $\dfrac{\text{No. of live births}}{\text{Mid year population}} \times 1000$

$= \dfrac{120\cancel{00}}{4\cancel{00000}} \times 1\cancel{000} = 30$ per thousand Mid year population.

(b) Crude Death Rate (CDR) = $\dfrac{\text{No. of deaths during a year}}{\text{Mid year population of year}} \times 1000$

$= \dfrac{64\cancel{00}}{4{,}00{,}\cancel{000}} \times 1\cancel{000} = 16$ per 1000 mid year population.

(c) Infant Mortality Rate (IMR) = $\dfrac{\text{No. of deaths < 1 year of age}}{\text{No. of live births during the year}} \times 1000$

$= \dfrac{1500}{12\cancel{000}} \times 1\cancel{000} = 125$ per 1000 live births.

General Fertility Rate (GFR) = No. of live birth per 1000 women in reproductive age group (15 - 49 years) in a given year.

GFR = $\dfrac{\text{No. of live births in a year in an area}}{\text{Mid year female population of age 15 - 44 years}} \times 1000$

Total Fertility Rate (TFR) = $\dfrac{i \times \Sigma \text{ASFR}}{1000}$. i = class Interval.

Q. Calculate general fertility Rate, total & Gross Reproductive rate from following data → assuming that for every 100 girls - 104 boys :-

Age group	No. of women	Age specific (ASFR) fertility Rate.	Live Births.
15 - 19	21000	92	1932
20 - 24	19000	165	3135
25 - 29	16500	152	2500
30 - 34	14600	140	2030
35 - 39	12300	90	1107
40 - 44	10200	45	457
45 - 49.	4500	16	72
	98000	700	11243

$$ASFR = \frac{\text{No. of live birth}}{\text{No. of women of same age group}} \times 1000$$

$$\Rightarrow \text{No. of live Birth} = \frac{ASFR \times \text{No. of women}}{1000}$$

$$GFR = \frac{\text{No. of live births in an area during a year}}{\text{mid year female population of age 15-49 yrs}} \times 1000$$

$$\text{Total Fertility Rate (TFR)} = \frac{i \times \Sigma ASFR}{1000}$$

$$= \frac{\cancel{5} \times 7\cancel{00}}{2\,1\cancel{000}} = \frac{7}{2} = 3.5 \text{ per 1000 mid year population.}$$

$$\text{Gross Reproductive Rate} = \frac{\text{No } i \times \Sigma ASFR}{1000} \text{ female live birth}$$

$$= \frac{5 \times 700}{1000} = 3.5 \text{ per 1000 mid year population.}$$

Q. The no. of live births occuring in a city in the year 2000 is given below. classify according to age of mother together with the no. women in each age group of reproductive period.

Age group	No. of women	No. of live birth	ASFR.
15 - 19	22000	780	34.10
20 - 24	28000	2650	101.29
25 - 29	26000	2200	83.02
30 - 34	20500	1550	76.01
35 - 39	15000	500	33.33
40 - 44	1500	120	16.00
45 - 49	5500	80	5.45
	125000	8000	349.30

54

Total population of city during the year 2000 ⇒ 500,000.
Calculate CBR, GFR, ASFR, TFR.

Sol:- $$CBR = \frac{\text{No. of live Birth}}{\text{Total no. of population}} \times 1000 = \frac{\overset{16}{8\cancel{000}}}{5\cancel{00000}} \times 1\cancel{000} = 16 \text{ per 1000 population}$$

$$GFR = \frac{\text{No. of live Birth}}{\text{Total No. of women in reproductive age group}} \times 1000$$

$$= \frac{8000}{125\cancel{000}} \times 1\cancel{000} = 64 \text{ per 1000 women of reproductive age group in year 2000.}$$

$$ASFR = \frac{\text{No. of live birth in a particular age group}}{\text{No. of female of all age group}} \times 1000$$

$$= i \times \Sigma 349.30.$$

$$TFR = \frac{i \times \Sigma ASFR}{1000}$$

$$= \frac{5 \times 349.30}{1000}$$

$= 1.75$ per 1000 no. of females of all age groups.

Morbidity Statistics

Q. The population & no. of death of people in localities according to age group are given below :-

Age Group.	Locality A Population	Locality A No. of deaths	Locality B Population	Locality B No. of deaths
1. <5	2000	215	35000	844
2. 5-15	1500	280	125000	806
3. 15-35	12500	295	180000	610
4. 35-40	4500	210	80000	1200
5. 50	6500	320	64400	1430.

Compare the crude death rates.

Sol:- CDR of locality A

⇒ 1. 30.7 2. 16.7 3. 23.6 4. 38.2 5. 49.2

$$\text{Total CDR} = \frac{\text{No. of death}}{\text{Total no. of population in locality}} \times 1000.$$

$$= \frac{1290}{465000} \times 1000 = 27.7 \text{ per thousand population.}$$

CDR of locality B ⇒

1. 16.5 2. 12.0 3. 15.8 4. 20.6 5. 22.2

$$\text{Total CDR} = \frac{8840}{534400} \times 1000 = 16.5 \text{ per thousand population.}$$

Q. The mid year population in 2011 is 1020000. Following events occured during year 2011. Calculate.

(a) CBR, (b) CDR (c) IMR (d) MMR (e) NMR

(f) PNMR (g) SBR (h) PMR.

Total live birth = 30,000, Total death = 12000

Maternal death = 120, Infant death = 1600

Death within 28 days = 850, Still Birth = 280

Death within 1 week = 500.

55

(a) $$\text{CBR} = \frac{\text{No. of live births}}{\text{Mid year population}} \times 1000$$

$$= \frac{30000}{1020000} \times 1000 = 29.4 \text{ per 1000 Mid year population.}$$

(b) $$\text{CDR} = \frac{\text{No. of deaths}}{\text{Mid year population}} \times 1000$$

$$= \frac{12000}{1020000} \times 1000 = 11.76 \text{ per thousand mid year population.}$$

(c) $$\text{IMR} = \frac{\text{Total no. of infant death}}{\text{Total live birth}} \times 1000$$

$$= \frac{1600}{30000} \times 1000 = 53.33 \text{ per thousand live births.}$$

(d) $$\text{MMR} = \frac{\text{Total no. of maternal death}}{\text{Total live birth during year.}} \times 1000$$

$$= \frac{120}{30000} \times 1000 = 4 \text{ per 1000 live birth.}$$

(e) $$\text{NMR} = \frac{\text{No. of deaths under 28 days of age}}{\text{Total live birth}} \times 1000$$

$$= \frac{850}{30000} \times 1000 = 28.33 \text{ per 1000 live birth.}$$

(f) $$\text{PNMR} = \frac{\text{Still Birth} + \text{death within 1 week of age}}{\text{Total live birth} + \text{Still Birth}} \times 1000$$

$$= \frac{280+500}{30000+280} \times 1000 = \frac{780}{30280} \times 1000 = \cdot 26 \text{ per 1000 live + Still Birth (Total Birth)}$$

or

(f) $$\text{PNMR} = \frac{\text{Still Birth} + \text{death within 1 week of age}}{\text{Total live birth.}} \times 1000$$

$$= \frac{780}{30000} \times 1000 = \frac{78}{3} = 26 \text{ per 1000 live birth.}$$

(g) $$\text{SBR} = \frac{\text{Foetal death after 28 weeks of gestation}}{\text{Total live birth} + \text{Still Birth}} \times 1000$$

$$= \frac{280}{30000+280} \times 1000 = \frac{280000}{30280} = 9.25 \text{ per 1000 total birth}$$

(ii) PNeonatal M.R. = $\dfrac{\text{No. of death of children between 28 days and 1 year of age in a given year}}{\text{total live birth in same year}} \times 1000$

$= \dfrac{1600 - 850}{30000} \times 1000 = \dfrac{750}{30000} \times 1000 = 25$ per thousand live birth.

15. MEASUREMENT OF RISK

Epidemiological hypothesis generally specifies a cause – effect relationship between two variables. For determination of such relationship it is necessary to prove statistical association and also causal nature of association.

Following types of studies could be conducted to study the association.

1) **Prospective (cohort) studies:** Study is made between group or a cohort exposed to a particular factor and a group not exposed to that factor. The groups are followed over a period of time to observe the outcome or effect of the factor. This study is made from cause to effect.
2) **Retrospective (case - control) studies:** Study is made between two groups which are affected (cases) and non affected (controls). Presence or absence of the factor is as curtained retrospectively in both the groups. This study is made from effect to cause.

From these studies, strength of association and the risk associated with particular factor can be measured.

Prospective Study:

Risk Factor	Disease		Total
	Developed	Not Developed	
Present	A	B	A + B
Absent	C	D	C + D
Total	A + C	B + D	A + B + C +D

The risk of disease developing among exposed is A / (A+B) and that among not exposed is C / (C+D).

$$\text{Relative Risk (R.R.)} = \frac{A/(A+B)}{C/(C+D)}$$

The relative risk indicates the strength of the causal relationship.

Attributable risk (A.R.) is A / (A+B) - C / (C+D), and Attributable risk percentage is:

$$\frac{A/(A+B) - C/(C+D)}{A/(A+B)} \times 100$$

The attributable risk is the rate of disease in exposed individuals that can be attributed to the exposure.