Descriptive Statistics Unit-I

Statistics Introduction

Statistics:

Statistics is a tool in the hands of mankind to translate complex facts into simple and understable statements of fact.

The word statistics is derive from the Italian word stato and it means a political state in the singular sense statistics is as defined a science which deals with the scientific methods of collection organization presentation analysis and Interpretation of numerical data. Statistical methods are applied for investigation in every important ~~sides~~ signs

Statistical Methods:

i) collection of data:

The first step of an investigation is the collection of data. careful collection is needed because further analysis is based on this. The success of an enquiry depends on the proper collection of data.

primary data: It is an individual of an officer collects data to study a particular problem. The data are the raw materials of the enquiry. They are the primary data collected by the investigator in self to study any particular problem.

Secondary data: secondary data are those which are already collected by someone for some purpose and are available for the present study. for example, the data collected during senses census operations are primary data to the department of census and the same data if used by a research worker for some study are Secondary data.

Population Vs Sample.

In statistical enquiry all the items which fall within the preview of enquiry are known as univers or population.

1) finite and Infinite population: population can be either finite population or infinite population when the number of observations can be counted and definite it is known as finite population.

Eg: when we are studying the economic background of students of a college, all the students of the college will constitue population and this number will be finite.

when the number of observations cannot be counted and is infinite it is known as Infinite population.

Eg: The number of stars in the sky is infinite population.

Sample method:

In the case of sample enquiry only a part of the whole group of population will be studied. we can study the characterisltic of a population from sampling. A study of the sample wil give correct idea of the universe or population.

Merits:

It reduces the cost of enumeration.

* More relable results can be obtained since there are few chances sampling errors.

* It saves the time.

* when the resultes are urgently required this method is very helpful.

Methods of sampling:

Random sampling Method: A random sample is one where each item in the universe as an equal chance of known opportunity

Non Random sampling: These can be divided in three methods

1) Judgment or purposive Sampling

The choice of the sample item depends on the judgment of the investigation.

2) Quota sampling: To collect data the universe is divided into Quota according to some characteristics

3) convenience sampling or check sampling: The sampling is obtained by selecting convenient population unit

* It is sutable when the population is not clearly defined.

* Sample is not clear.

* complete source list is not available.

Measures of central tendency:

The following are the measures of central tendency which represents the entire set of observations by a single value.

1) Mean or Arithmetic mean

2) Median

3) Mode

Mean: The mean of set of data is defined as the sum of observations divided by the total number of observations. Suppose a set consists $x_1, x_2 --- x_n$ are 'n' observations then the Arithmetic mean is denoted by $A.M = (\bar{x}) = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Arithmetic Mean is denoted by $\bar{x}$

Mean for frequency distribution:

For a frequency distribution the mean of observations defined as $A.M = \bar{x} = \frac{x_1f_1+x_2f_2+\cdots+x_nf_n}{f_1+f_2+\cdots+f_n}$

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

problems

Find the Mean of

1) 20, 22, 25, 28, 30

2) 3, 6, 10, 4, 9, 10

3) 6, 7, 10, 12, 13, 4, 8, 12

4) 40, 50, 60, 70, 80, 90

3) sol let $x_1 = 6$, $x_2 = 7$, $x_3 = 10$, $x_4 = 12$, $x_5 = 13$, $x_6 = 4$

$x_7 = 8$, $x_8 = 12$

$n = 8$

$$A.M = \bar{x} = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8}{8}$$

$$= \frac{6 + 7 + 10 + 12 + 13 + 4 + 8 + 12}{8} = \frac{72}{8} = 9$$

1) let $x_1 = 20$, $x_2 = 22$, $x_3 = 25$, $x_4 = 28$, $x_5 = 30$

$n = \cancel{8}$

$$A.M = \bar{x} = \frac{\sum_{i=1}^{5} x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$= \frac{20 + 22 + 25 + 28 + 30}{5} = \frac{125}{5} = 25$$

2) let $x_1 = 3$, $x_2 = 6$, $x_3 = 10$, $x_4 = 4$, $x_5 = 9$, $x_6 = 10$

$$A.M = \bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

$$= \frac{3 + 6 + 10 + 4 + 9 + 10}{6}$$

$$= \frac{42}{6} = 7$$

4) let $x_1 = 40$, $x_2 = 50$, $x_3 = 60$, $x_4 = 70$, $x_5 = 80$, $x_6 = 90$

$n = 6$

$$A.M = \bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6}$$

$$= \frac{40 + 50 + 60 + 70 + 80 + 90}{6} = \frac{390}{6} = 65$$

Method 2 Find the Mean for the following data..

1)

x_i	10	30	50	70	90
f_i	4	24	28	16	8

2)

x_i	10	11	12	13
f_i	3	12	18	12

3)

x_i	8	10	15	20
f_i	5	8	8	4

1)

x_i	f_i	$f_i x_i$
10	4	40
30	24	720
50	28	1400
70	16	1120
90	8	720
Total	$\Sigma f_i = 80$	$\Sigma f_i x_i = 4000$

$$A.M = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{4000}{80} = 50$$

2)

x_i	f_i	$f_i x_i$
10	3	30
11	12	132
12	18	216
13	12	156
Total	$\Sigma f_i = 45$	$\Sigma f_i x_i = 534$

$$A.M = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{534}{45} = 11.866$$

3)

x_i	f_i	$f_i x_i$
8	5	40
10	8	80
15	8	120
20	4	80
Total	$\Sigma f_i = 25$	$\Sigma f_i x_i = 320$

$$A.M = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \cancel{\frac{25}{320}}\frac{320}{25} = 12.8$$

Method 3 Find the Mean of the following data.

1)

class	0-10	10-20	20-30	30-40	40-50
frequency	7	8	20	10	5

2)

class	0-10	10-20	20-30	30-40	40-50
frequency	5	10	20	5	10

3)

Height in cms	95-105	105-115	115-125	125-135	135-145	145-155
No. of Boys	9	13	26	30	12	10

3)

Class intervals x	Let f_i = no. of Boys	x_i = mid value	$x_i f_i$
95-105	9	$\frac{95+105}{2} = \frac{200}{2} = 100$	900
105-115	13	110	1430
115-125	26	120	3120
125-135	30	130	3900
135-145	12	140	1680
145-155	10	150	1500
Total	$\Sigma f_i = 100$		$\Sigma f_i x_i = 12530$

$$A.M = \bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{12530}{100} = 125.3$$

1)

Class intervals x	frequency f_i	x_i = Mid value	$x_i f_i$
0-10	7	$\frac{0+10}{2} = \frac{10}{2} = 5$	35
10-20	8	15	120
20-30	20	25	500
30-40	10	35	350
40-50	5	45	225
Total	$\Sigma f_i = 50$		$\Sigma f_i x_i = 1230$

$$A.M = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1230}{50} = 24.6$$

2)

Class intervals x	frequency f_i	x_i = Mid value	$x_i f_i$
0-10	5	5	25
10-20	10	15	150
20-30	20	25	500
30-40	5	35	175
40-50	10	45	450
Total	$\Sigma f_i = 50$		$\Sigma f_i x_i = 1300$

$$A.M = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1300}{50} = 26.$$

Median: Median is one of the measures of tendency. Median of distribution is the value that divides the data into two equal parts when the variants are arranged in ascending/descending order of magnitude.

Median of an ungrouped data:

Arrange the variants in ascending/descending order of magnitude. Determine the total number of 'n' variants.

If 'n' is odd then the median is the value of $\left(\frac{n+1}{2}\right)^{th}$ term

If 'n' is even, the median is the Average value of $\left(\frac{n}{2}\right)^{th}$ term and $\left(\frac{n}{2}+1\right)^{th}$ term.

i.e., $\text{median} = \dfrac{\left(\frac{n}{2}\right) + \left(\frac{n}{2}+1\right)}{2}$ (If n is even)

Median of a grouped data:

The Median of grouped data is the following formula.

$$\text{Median} = \frac{l + \frac{N}{2} - m}{f} \times c$$

l = lower limit of the median class

m = cumulative frequency of the class preceding the median class.

N = Total frequency.

f = frequency of the median class.

C = width of the Median class.

Problems:

1) The number of runs Scored by 11 players of a Cricket team of School are 5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27 find. Median.

Given data is 5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27

The observations arranged in ascending (small

0, 5, 11, 19, 21, 27, 30, 36, 42, 50, 52

$n = 11$ is odd

Median $= \left(\frac{n+1}{2}\right)^{th}$ term $= 6^{th}$ term

Median $= 27$

Q, Find the Median 5, 9, 19, 23, 7, 4, 1

Given data is 5, 9, 19, 23, 7, 4, 1

The observations arranged in ascending order

1, 4, 5, 7, 9, 19, 23

$n = 7$ is odd

Median $= \left(\frac{n+1}{2}\right)^{th}$ term $= \left(\frac{7+1}{2}\right)^{th}$ term $= \left(\frac{8}{2}\right)^{th}$ term $= 4$ terms

Median $= 7$

Find the Median of the following

1) 6, 10, 4, 3, 9, 11, 22, 18

2) 57, 58, 61, 42, 38, 65, 72, 66

2) Given data is

57, 58, 61, 42, 38, 65, 72, 66.

These data can be arranged in ascending order -

38, 42, 57, 58, 61, 65, 66, 72

Here $n = 8$ even.

$$\text{Median} = \frac{\left(\frac{n}{2}\right)^{th}\text{term} + \left(\frac{n}{2}+1\right)^{th}\text{term}}{2} = \frac{\left(\frac{8}{2}\right)^{th}\text{term} + \left(\frac{8}{2}+1\right)^{th}\text{term}}{2}$$

$$= \frac{4^{th}\text{ term} + 5^{th}\text{ term}}{2} = \frac{58+61}{2} = 59.5$$

1) Given data is

6, 10, 4, 3, 9, 11, 22, 18

These data can be arranged in ascending order

3, 4, 6, 9, 10, 11, 18, 22

Here $n = 8$ even

$$\text{Median} = \frac{\left(\frac{n}{2}\right)^{th}\text{term} + \left(\frac{n}{2}+1\right)^{th}\text{term}}{2} = \frac{\left(\frac{8}{2}\right)^{th}\text{term} + \left(\frac{8}{2}+1\right)}{2}$$

$$= \frac{4^{th}\text{term} + 5^{th}\text{term}}{2} = \frac{9+10}{2} = \frac{19}{2} = 9.5$$

Calculate the following data
Median for

1)

x	5	7	9	12	14	17	19	21
f	6	5	3	6	5	3	2	4

2)

x	9	20	25	40	50	80
f	4	6	16	8	7	2

1) Sol:

x	f	Cumulative frequency
5	6	6
7	5	11
9	3	14
12	6	20
14	5	25
17	3	28
19	2	30
21	4	34

Total $\Sigma f = N = 34$

Now $\frac{N}{2} = \frac{34}{2} = 17$

Here 17 is nearly 20

The cumulative frequency

Median is 12. (greater than 17 = to 12)

2)

x	f	Cumulative frequency
9	4	4
20	6	10
25	16	26
40	8	34
50	7	41
80	2	43

Total $\Sigma f = N = 43$

Now $\frac{N}{2} = \frac{43}{2} = 21.5$

Here 26 is nearly 21.5

The cumulative frequency

Median is 25

Calculation of Median in continuous series.

In these case $\text{Median} = l + \left(\frac{\frac{N}{2} - m}{f}\right) \times c$ where

l = lower limit of median class

$N = \Sigma f$ sum of frequencies

m = cumulative frequency of the class preceding the median class.

f = frequency of median class

c = size or width of class interval.

Problems.

Calculate median for the following data

1)

Weekly expenditure	0-10	10-20	20-30	30-40	40-50
No. of families	14	23	27	21	15

2)

Height in cms	135-140	140-145	145-150	150-155	155-160	160-165	165-170	170-175
No. of Boys	4	9	18	28	24	10	5	2

3)

Class intervals	10-25	25-40	40-55	55-70	70-85	85-100
frequency	6	20	44	26	3	1

2)
Sol:

Height in cms	No. of Boys	Cumulative frequency (c.f)
135-140	4	4
140-145	9	13
145-150	18	31 (m)
150-155	28	59
155-160	24	83
160-165	10	93
165-170	5	98
170-175	2	100
	Total $\Sigma f = 100$, $N = 100$	

$N = 100$

$\frac{N}{2} = 50$

$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$

$l = 150,\ \frac{N}{2} = 50,\ m = 31,$

$f = 28,\ c = 5$

$\text{Median} = 150 + \frac{50 - 31}{28} \times 5$

$\text{Median} = 153.39$

1)

weekly expenditure	No. of families	Cumulative frequency c.f
0-10	14	14
10-20	23	37 (m)
20-30	27	64
30-40	21	85
40-50	15	100
	Total $\Sigma f = 100$, $N = 100$	

$N = 100$

$\frac{N}{2} = 50$

$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$

$l = 20,\ \frac{N}{2} = 50,\ m = 37$

$f = 27,\ c = 10$

$\text{Median} = 20 + \frac{50 - 37}{27} \times 10$

$\text{Median} = 20 + \frac{13}{27} \times 10$

$= 20 + 4.814$

$= 24.814$

3)
Sol

Class intervals	frequency	Cumulative frequency (c.f)
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100
	Total $\Sigma f = 100$, $N = 100$	

$N = 100,\ \frac{N}{2} = 50$

$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$

$l = 40,\ \frac{N}{2} = 50,\ m = 26$

$f = 44,\ c = 15$

$\text{Median} = 40 + \frac{50 - 26}{44} \times 15$

$= 40 + \frac{24}{44} \times 15$

$= 40 + 8.1818$

Mode: Mode is the value which occurs most frequently in a frequency distribution mode is a value which has maximum frequency.

Problems:

Find the mode of the following data:

1) 850, 750, 600, 825, 850, 725, 600, 850, 640, 530.

Mode = 850 3 times Repeated.

600 2 times Repeated.

Mode = 850.

2) 40, 45, 48, 57, 78

No value is repeated thus there is no mode.

3) 45, 55, 50, 45, 40, 55, 45, 55

45 repeated 3 times

55 repeated 3 times

mode = 45 & 55.

4) for the Series 2, 2, 2, 3, 4, 4, 5, 2, 6, 7, 4, 4 find mode

2 Repeated 4 times

4 Repeated 4 times

Mode = 2 & 4.

Calculation of mode for Continuous Series

$$\text{Mode} = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times C$$

here l = lower limit of the modal class.

f_0 = frequency of modal class

f_1 = frequency of the class preceeding modal class

f_2 = frequency of the class suceeding modal class.

C = size/length/width of the class.

Find the mode of the following distribution.

1)

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	8	7	$\overset{f_1}{12}$	$\overset{f_0}{28}$	$\overset{f_2}{20}$	10	10

2)

Class interval	3-4	4-5	5-6	6-7	7-8	8-9	9-10
frequency	7 83	10 ~~27~~	16 ~~25~~ f_1	32 ~~50~~ f_0	24 75 f_2	22 ~~38~~	8 ~~18~~

3)

Marks	10-25	25-40	40-55	55-70	70-85	85-100
frequency	6	f_1 20	f_0 44	26 f_2	3	1

1) Sol Here the maximum frequency 28

$f_0 = 28$

$f_1 = 12, \ f_2 = 20, \ c = 10$

40 - 50 lower limit $l = 40$ $\quad$ Mode $= l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times c$

Mode $= 40 + \frac{28 - 12}{2(28) - 12 - 20} \times 10$

Mode $= 46.666$

2) Sol Here the maximum frequency 32

$f_0 = 32$

$f_1 = 16, \ f_2 = 24, \ c = 1$

6 - 7 lower limit $l = 6$.

Mode $= l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times c = 6 + \frac{32 - 16}{2(32) - 16 - 24} \times 1$

$= 6 + \frac{16}{64 - 16 - 24} \times 1 = 6 + \frac{16}{24} \times 1 = \frac{20}{3}$

$= 6.666$

3) Here the maximum frequency 44

$f_0 = 44$

$f_1 = 20, \ f_2 = 26, \ c = 15$

40-55 lower limit $l = 40$

Mode $= l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times c = 40 + \frac{44 - 20}{2(44) - 20 - 26} \times 15$

$= 40 + \frac{24}{88 - 20 - 26} \times 15 = 40 + \frac{24}{42} \times 15$

$= 40 + 8.571$

$= 48.571$

Measures of variability or Measures of dispersion

The following are the most commonly used measures of dispersions are

1) Range
2) Mean Deviation
3) Standard deviation

Range :- Range is the one of the Measures of dispersion the Range is defined as the difference between the greatest and the least values of the variate

$$\text{Range} = L - S$$

where L = largest value

S = Smallest value

$$\text{Coefficient of range} = \frac{L-S}{L+S}$$

problems

Find the range and coefficient of range for the following data.

1) 41, 20, 15, 65, 73, 84, 53, 35, 71, 55.

2) 60, 72, 96, 28, 35, 10, 40, 9, 85, 25.

3)

x	20	30	40	50	60	70	80
f	3	61	132	153	140	51	3

2) Sol: Here l = largest value = 96

S = smallest value = 9

$\text{Range} = l - s = 96 - 9 = 87$

$\text{Coefficient of range} = \frac{l-s}{l+s} = \frac{96-9}{96+9} = \frac{87}{105} = 0.8285$

3) $x_{max} = L = 80$

$x_{min} = S = 20$

$\text{Range} = L - S = x_{max} - x_{min} = 80 - 20 = 60$

$\text{Range} = 60$

$\text{Coefficient of range} = \frac{L-S}{L+S} = \frac{80-20}{80+20} = \frac{60}{100} = \frac{3}{5} = 0.6$

1) Sol: Here L = largest value = 84

S = smallest value = 15

$\text{Range} = l - s = 84 - 15 = 69$

$\text{Coefficient of range} = \frac{l-s}{l+s} = \frac{84-15}{84+15} = \frac{69}{99} = 0.6970$

The following table ~~gives~~ gives is the daily Sales (rupees) of two firms A and B for 5 days.

Firm A	Firm B
5050	4900
5025	3100
4950	2200
4835	1800
5140	1300
$\bar{x}_A = 5000$	$\bar{x}_B = 5000$

find Range from A and firm B.

Sol firm A $L = 5140$

$S = 4835$

Range $= L - S = 5140 - 4835 = 305$

Firm B $L = 13000$

$S = 1800$

Range $= L - S = 13000 - 1800 = 11200$

Mean deviation : Mean deviation is defined as Arithmetic Average of absolute values of the deviations of the variants measured from an average [Median, Mode, Mean].

The absolute value of the deviation denoted by |deviation| is the numerical value of the deviation with positive sign.

Mean deviations for ungrouped data:

Let $x_1, x_2 \ldots x_n$ are n observations.

1) Mean $\bar{x}$ for ungrouped data is $\bar{x} = \frac{\Sigma x_i}{n}$ then mean deviation with respect to mean is given by

~~Medin~~ Mean deviation $= \frac{\Sigma |x_i - \bar{x}|}{n}$

2) Mean deviation is the Median of ungrouped data then mean deviation with respect to Median is given by Mean deviation $= \frac{\Sigma |x_i - M_d|}{n}$

Calculate the mean deviation ~~for~~ from the mean for the following data

1) 40, 62, 54, 68, 76

2) 38, 70, 48, 40, 42, 55, 63, 46, 54, 44.

3) 6, 7, 10, 12, 13, 4, 12, 16.

2)

Sol: $A.M = \bar{x} = \frac{\Sigma x_i}{n} = \frac{38+70+48+40+42+55+63+46+54+44}{10}$

$= \frac{500}{10} = 50$

$\bar{x} = 50$.

Mean deviation $= \frac{\Sigma|x_i - \bar{x}|}{n}$

$= \frac{|38-50|+|70-50|+|48-50|+|40-50|+|42-50|+|55-50|+|63-50|+|46-50|+|54-50|+|44-50|}{10}$

$= \frac{12+20+2+10+8+5+13+4+4+6}{10} = \frac{84}{10} = 8.4$

Mean deviation $= 8.4$.

1)

Sol: $A.M = \bar{x} = \frac{\Sigma x_i}{n} = \frac{40+62+54+68+76}{5} = \frac{300}{5} = 60$

$\bar{x} = 60$

Mean deviation $= \frac{\Sigma|x_i - \bar{x}|}{n}$

$= \frac{|40-60|+|62-60|+|54-60|+|68-60|+|76-60|}{5}$

$= \frac{20+2+6+8+16}{5} = \frac{52}{5} = 10.4$

Mean deviation $= 10.4$

3)

Sol: $A.M = \bar{x} = \frac{\Sigma x_i}{n} = \frac{6+7+10+12+13+4+12+16}{8} = \frac{80}{8} = 10$

$\bar{x} = 10$

Mean deviation $= \frac{\Sigma|x_i - \bar{x}|}{n}$

$= \frac{|6-10|+|7-10|+|10-10|+|12-10|+|13-10|+|4-10|+|12-10|+|16-10|}{8}$

$= \frac{4+3+0+2+3+6+2+6}{8} = \frac{26}{8} = 3.25$

find the Mean deviation about 1) Mean 2) Median for the following data.

i)

x_i	6	7	8	9	10	11	12
f_i	3	6	9	13	8	5	4

H.W

2)

x_i	5	10	15	20	25
f_i	7	4	6	3	5

n=5

3)

x	2	5	7	8	10	35
f	6	8	10	6	8	2

n=6

1) Sol

1) Mean

x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}_i\|$	$f_i\|x_i - \bar{x}_i\|$
6	3	18	$\|6-9\| = 3$	9
7	6	42	$\|7-9\| = 2$	12
8	9	72	$\|8-9\| = 1$	9
9	13	117	0	0
10	8	80	1	8
11	5	55	2	10
12	4	48	3	12
	$\Sigma f_i = 48$	$\Sigma f_i x_i = 432$		$\Sigma f_i\|x_i - \bar{x}_i\| = 60$

$$\bar{x}_1 = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{432}{48} = 9$$

$$\text{Mean deviation} = \frac{\Sigma f_i |x_i - \bar{x}|}{\Sigma f_i} = \frac{60}{48} = 1.25$$

2) Median

x_i	f_i	Cumulative frequency	$\|x_i - \text{Median}\|$	$f_i\|x_i - \text{median}\|$
6	3	3	$\|6-9\| = 3$	9
7	6	9	$\|7-9\| = 2$	12
8	9	18	$\|8-9\| = 1$	9
$\boxed{9}$	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
	$\Sigma f_i = 48$			$\Sigma f_i\|x_i - \text{median}\| = 60$

Here $N = 48$

$\frac{N}{2} = \frac{48}{2} = 24$.

Cumulative frequency just greater than $\frac{N}{2}$ is 31 then the corresponding value of $x = 9$ is the Median

$$\text{Median deviation} = \frac{\Sigma f_i |x_i - \text{median}|}{\Sigma f_i} = \frac{60}{48} = 1.25$$

2) 1) Mean

Sol

x_i	f_i	$f_i x_i$	$\lvert x_i - \overline{x}_i \rvert$ $\lvert x_i - 14 \rvert$	$f_i \lvert x_i - \overline{x}_i \rvert$
5	7	35	$\lvert 5-14 \rvert = 9$	63
10	4	40	4	16
15	6	90	1	6
20	3	60	6	18
25	5	125	11	55
	$\Sigma f_i = 25$	$\Sigma f_i x_i = 350$		$\Sigma f_i \lvert x_i - \overline{x}_i \rvert = 158$

$\overline{x}_i = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{350}{25} = 14$

Mean deviation $= \frac{\Sigma f_i \lvert x_i - \overline{x} \rvert}{\Sigma f_i} = \frac{158}{25} = 6.32$

2) Median

x_i	f_i	c.f	$\lvert x_i - \text{Median} \rvert$	$f_i \lvert x_i - \text{Median} \rvert$
5	7	7	$\lvert 5-15 \rvert = 10$	70
10	4	11	$\lvert 10-15 \rvert = 5$	20
15	6	17	0	0
20	3	20	5	15
25	5	25	10	50
	$\Sigma f_i = 25$			$\Sigma f_i \lvert x_i - \text{Median} \rvert = 155$

Here $N = 25$, $\frac{N}{2} = \frac{25}{2} = 12.5$

Median deviation $= \frac{\Sigma f_i \lvert x_i - \text{Median} \rvert}{\Sigma f_i} = \frac{155}{25} = 6.2$

3)

Sol

x_i	f_i	$f_i x_i$	$\lvert x_i - \overline{x}_i \rvert$	$f_i \lvert x_i - \overline{x}_i \rvert$
2	6	12	$\lvert 2-8 \rvert = 6$	36
5	8	40	$\lvert 5-8 \rvert = 3$	24
7	10	70	$\lvert 7-8 \rvert = 1$	10
8	6	48	0	0
10	8	80	2	16
35	2	70	27	54
	$\Sigma f_i = 40$	$\Sigma f_i x_i = 320$		$\Sigma f_i \lvert x_i - \overline{x}_i \rvert = 140$

$\overline{x}_i = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{320}{40} = 8$

Mean deviation $= \frac{\Sigma f_i \lvert x_i - \overline{x}_i \rvert}{\Sigma f_i} = \frac{140}{40} = 3.5$

2) Median

x_i	f_i	c.f	$\lvert x_i - 7\rvert$ $\lvert x_i - \text{median}\rvert$	$f_i\lvert x_i - \text{median}\rvert$
2	6	6	$\lvert 2-7\rvert = 5$	30
5	8	14	$\lvert 5-7\rvert = 2$	16
7	10	24	$\lvert 7-7\rvert = 0$	~~24~~0
8	6	30	$\lvert 8-7\rvert = 1$	6
10	8	38	$\lvert 10-7\rvert = 3$	24
35	2	40	$\lvert 35-7\rvert = 28$	56
	$\Sigma f_i = 40$			$\Sigma f_i\lvert x_i - \text{median}\rvert = 1$

$\Sigma f_i = N = 40$

$\frac{N}{2} = \frac{40}{2} = 20.$

Mean deviation $= \frac{\Sigma f_i \lvert x_i - \text{median}\rvert}{\Sigma f_i} = \frac{132}{40} = 3.3$

calculation of mean deviation for continuous frequency distribution.

Mean deviation $= \frac{\Sigma f_i \lvert x_i - \overline{x}_i\rvert}{\Sigma f_i}$

Problems.

Calculate mean deviation about mean for the following data.

1)

Class intervals	0-10	10-20	20-30	30-40	40-50
f_i	5	8	15	16	6

2)

~~x_i~~ C.I	0-4	4-8	8-12	12-16	16-20	20-24
f_i	8	12	35	25	13	7

2) Sol)

C.I	f_i	x_i = midpoint	$f_i x_i$	$\lvert x_i - \overline{x}_i\rvert$	$f_i\lvert x_i - \overline{x}_i\rvert$
0-4	8	$\frac{0+4}{2} = 2$	16	$\lvert 2 - 11.76\rvert = 9.76$	78.08
4-8	12	$\frac{4+8}{2} = 6$	72	$\lvert 6 - 11.76\rvert = 5.76$	69.12
8-12	35	10	350	1.76	61.6
12-16	25	14	350	2.24	56
16-20	13	18	234	6.24	81.12
20-24	7	22	154	10.24	71.68
	$\Sigma f_i = 100$		$\Sigma f_i x_i = 1176$		$\Sigma f_i\lvert x_i - \overline{x}_i\rvert = 417.60$

$\overline{x}_i = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1176}{100} = 11.76$

Mean deviation $= \frac{\Sigma f_i \lvert x_i - \overline{x}_i\rvert}{\Sigma f_i} = \frac{417.60}{100} = 4.176$

1)

Sol

C.I	f_i	x_i = midpoint	$f_i x_i$	$\lvert x_i - \bar{x}_i \rvert$	$f_i \lvert x_i - \bar{x}_i \rvert$
0-10	5	$\frac{0+10}{2} = 5$	25	22	110
10-20	8	15	120	12	96
20-30	15	25	375	2	30
30-40	16	35	560	8	128
40-50	6	45	270	18	108
	$\Sigma f_i = 50$		$\Sigma f_i x_i = 1350$		$\Sigma f_i \lvert x_i - \bar{x}_i \rvert = 472$

$\bar{x}_i = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1350}{50} = 27$

Mean deviation

$= \frac{\Sigma f_i \lvert x_i - \bar{x}_i \rvert}{\Sigma f_i} = \frac{472}{50}$

$= 9.44$

Calculate the mean deviation from the median of the following data.

wages week (Rs)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of workers	4	6	10	20	10	6	4

Sol:

wages	No. of workers f_i	Midpoint x_i	c.f	$\lvert x_i - md \rvert$	$f_i \lvert x_i - md \rvert$
10-20	4	$\frac{10+20}{2} = 15$	4	$\lvert 15-45 \rvert = 30$	120
20-~~20~~	6	$\frac{20+30}{2} = 25$	10	$\lvert 25-45 \rvert = 20$	120
30-40	10	35	20	10	100
40-50	20	45	40~~30~~	0	0
50-60	10	55	50~~40~~	10	100
60-70	6	65	56	20	120
70-80	4	75	60	30	120
	$\Sigma f_i = 60$				$\Sigma f_i \lvert x_i - md \rvert = 680$

$\Sigma f_i = 60 = N, \quad \frac{N}{2} = \frac{60}{2} = 30$

$\Sigma f_i \lvert x_i - md \rvert = 680$

Median $= l + \frac{\frac{N}{2} - m}{f} \times c$

$l = 40, \frac{N}{2} = 30, m = 20, f = 20, c = 100$

$= 40 + \frac{30-20}{20} \times 10 = 40 + \frac{\cancel{10}}{\cancel{20}_2} \times \cancel{10}^5 = 45$

Median $= 45$

Mean deviation $= \frac{\Sigma f_i \lvert x_i - md \rvert}{\Sigma f_i}$

$= \frac{680}{60}$

$= 11.3$

3) Standard Deviation

Let $x_1, x_2 ---- x_n$ are 'n' observations and $\bar{x}$ is their mean, the Variants and standard deviation for the given series is Variance $= \sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$.

Standard deviation $= \sqrt{\text{Variance}} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

Method-1

Find variance and standard deviation for the following data.

1) 5, 12, 3, 18, 6, 8, 2, 10.

2) 4, 5, 2, 8, 7

3) 6, 7, 10, 12, 13, 4, 8, 12

3) Sol

x_i	$x_i - \bar{x}$ ($\bar{x} = 9$)	$(x_i - \bar{x}_i)^2$
6	$6 - 9 = -3$	9
7	$7 - 9 = -2$	4
10	$10 - 9 = 1$	1
12	$12 - 9 = 3$	9
13	$13 - 9 = 4$	16
4	$4 - 9 = -5$	25
8	$8 - 9 = -1$	1
12	$12 - 9 = 3$	9
		$\Sigma(x_i - \bar{x}_i)^2 = 74$

$n = 8$

$\bar{x} = \frac{\Sigma x}{n} = \frac{6+7+10+12+13+4+8+12}{8}$

$= \frac{72}{8} = 9$

Variance $= \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{74}{8} = 9.25$

$S.D = \sqrt{\text{Variance}} = \sqrt{9.25}$

$= 3.04$

1)

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5	-3	9
12	4	16
3	-5	25
18	10	100
6	-2	4
8	0	0
2	-6	36
10	2	4
		$\Sigma(x_i - \bar{x})^2 = 194$

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{64}{8} = 8$

Variance $= \sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{194}{8} = 24.25$

2)

x_i	$x_i - \bar{x}$	$(x_i-\bar{x})^2$
4	-1.2	1.44
5	-0.2	0.04
2	-3.2	10.24
8	2.8	7.84
7	1.8	3.24
		$\Sigma(x_i-\bar{x})^2 = 22.8$

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{26}{5} = 5.2$

variance $= \sigma^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$

$= \frac{22.8}{5} = 4.56$

$S.D = \sqrt{4.56} = 2.1354$

Method-2

calculation of variance and standard deviation for group data :

<u>Discrete Series :</u>

Mean $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$

Variance $= \sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}$ $\quad S.D = \sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{\Sigma f_i}}$

Calculate the standard deviation for the following distribution :

1)

x_i	4	8	11	17	20	24	32
f_i	3	5	9	5	4	3	1

2)

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

2) Sol

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
6	2	12	6-19 = -13	169	2×169 = 338
10	4	40	10-19 = -9	81	324
14	7	98	14-19 = -5	25	175
18	12	216	18-19 = -1	1	12
24	8	192	24-19 = 5	25	200
28	4	112	28-19 = 9	81	324
30	3	90	30-19 = 11	121	363
	$\Sigma f_i = 40$	$\Sigma x_i f_i = 760$			$\Sigma f_i(x_i-\bar{x})^2 = 1736$

$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{760}{40} = 19$

Variance $= \sigma^2 = \frac{\Sigma f_i(x_i-\bar{x})^2}{\Sigma f_i} = \frac{1736}{40} = 43.4$

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
4	3	12	-10	100	300
8	5	40	-6	36	180
11	9	99	-3	9	81
17	5	85	3	9	45
20	4	80	6	36	144
24	3	72	10	100	300
32	1	32	18	324	324
	$\Sigma f_i = 30$	$\Sigma f_i x_i = 420$			$\Sigma f_i(x_i - \bar{x})^2 = 1374$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{420}{30} = 14$

variance $= \sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} = \frac{1374}{30} = 45.8$

$S.D = \sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}} = \sqrt{45.8} = 6.7675$

Method : 3

Calculate the continuous series or continuous distribution

In these case, the mean $\bar{x}$ is given by $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$

$$\text{Variance} = \sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}$$

$$S.D = \sigma = \sqrt{\text{Variance}} = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}}$$

Calculate variance and standard deviation of the continuous frequency distribution.

1)

Class interval	30-40	40-50	50-60	60-70	70-80	80-90	90-100
frequency	3	7	12	15	8	3	2

2)

Wages (Rs)	125-175	175-225	225-275	275-325	325-375	375-425	425-475	475-525	525-575
No. of workers	2	22	19	14	3	4	6	1	1

1)

Class Interval	frequency	Mid values x_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
30-40	3	$\frac{30+40}{2} = 35$	105	35-62 = -27	729	3×729 = 2187
40-50	7	45	315	45-62 = -17	289	2023
50-60	12	55	660	55-62 = -7	49	588
60-70	15	65	975	65-62 = 3	9	135
70-80	8	75	600	75-62 = 13	169	1352
80-90	3	85	255	85-62 = 23	529	1587
90-100	2	95	190	95-62 = 33	1089	2178
	$\Sigma f_i = 50$		$\Sigma x_i f_i = 3100$			$\Sigma f_i(x_i - \bar{x})^2 = 10050$

$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{3100}{50} = 62$, Variance $= \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} = \frac{10050}{50} = 201$

$S.D = \sigma = \sqrt{\text{variance}} = \sqrt{201} = 14.17$

** Measures of skewness

The measures of skewness gives the direction and extent of skewness. Mean of data most away from the mode these skewness is lack of symmetry.

The measures of central tendency and dispersion are inadequate to characterize a distribution is completely. They may be supported by two or more measures skewness and kurtosis.

Test of skewness:

If the distribution is symmetric the following conditions are observed.

1) The values of mean, mode, median coincide (the values are equal).

2) Q_3 - Median = Median - Q_1

3) The sum of positive deviations = the sum of negative deviations.

4) The frequencies on either of the mode are equal.

Similarly a skewed distribution will have the following characterisitics

(i) Mean ≠ median ≠ mode

(ii) Q_3 - median ≠ median - Q_3

(iii) the sum of positive deviations is not equal to the sum of negative deviations.

Measures of skewness:

*Absolute skewness = Mean - Mode = (+ positive skewness)
= (-negative skewness)

*If the values of mean is greater than mode then the skewness is positive.

*If the values of mode is greater than mean then the skewness is negative.

*There are three important measures of relative skewness.

1) Karl pearson's coefficient of skewness.

2) Bowley's coefficient of skewness

3) Kelly's coefficient of skewness.

Karl pearson's coefficient of skewness

the Karl pearson's coefficient of skewness Skp

$$Skp = \frac{\bar{x} - mode}{\sigma} = \frac{Mean - mode}{S.D}$$

or

$$\frac{3(Mean - median)}{\sigma}$$

Calculate Karl pearson's Coefficient of skewness for the following data

1) 25, 15, 23, 40, 27, 25, 23, 25, 20

Calculate table of mean and S.D

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
25	25-24.77=0.23	0.0529
15	-9.77	95.4529
23	-1.77	3.1329
40	15.23	231.9529
27	2.23	4.9729
25	0.23	0.0529
23	-1.77	3.1329
25	0.23	0.0529
20	-4.77	22.7529

$\bar{x} = \frac{\Sigma x_i}{n}$, $n = 9$

$= \frac{223}{9} = 24.77$

$$Skp = \frac{Mean(\bar{x}) - Mode}{\sigma}$$

Mode = 25 (3 Times Repeated)

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} = \sqrt{\frac{361.5561}{9}} = 6.3382$$

$$Skp = \frac{24.77 - 25}{6.3382} = -0.036$$

Here Skewness is negative

Method-2

Calculate Karl pearson Coefficient of Skewness or Measures of Skewness for the following distribution

x	4	8	11	17	20	24	32
f	3	5	9	8	4	3	1

x_i	f_i	$f_i x_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
4	3	12	4-14=-10	100	300
8	5	40	8-14=-6	36	180
mode 11	9	99	-3	9	81
17	5	85	3	9	45
20	4	80	6	36	144
24	3	72	10	100	300
32	1	32	18	24	324
					$\Sigma f_i(x_i-\bar{x})^2 = 1374$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{420}{30} = 14$

$S.D = \sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{\Sigma f_i}} = \sqrt{\frac{1374}{30}} = 6.7$

$Skp = \frac{\text{Mean} - \text{Mode}}{S.D}$

In the above table, Maximum frequency occurs at $x = 11$

$Skp = \frac{14-11}{6.7675} = 0.4433$

Here skewness is positive

Method-3.

Calculate Karl pearson's coefficient of skewness (or) Measures of skewness for the following data

Marks	10-25	25-40	40-55	55-70	70-85	85-100
f_i	6	20	44	26	3	1

Marks	f_i	Midpoint x_i	$x_i f_i$	C.F	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
10-25	6	17.5	105	6	-30.45	927.2025	5563.215
25-40	20	32.5	650	26	-15.45	238.7025	4774.05
40-55	44	47.5	2090	70	-0.45	0.2025	8.91
55-70	26	62.5	1025	96	14.55	211.7025	5504.265
70-85	3	77.5	232.5	99	29.55	873.2025	2619.6075
85-100	1	92.5	92.5	100	44.55	1964.7025	1964.7025
	$\Sigma f_i = 100$		$\Sigma f_i x_i = 4795$				$\Sigma f_i(x_i-\bar{x})^2 = 20454.75$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{4745}{100} = 47.45$

$N = \Sigma f_i = 100, \quad \frac{N}{2} = \frac{100}{2} = 50 (\geq) C.F$

$Mode = L + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times C \qquad L = 40, f_0 = 44, f_1 = 26, f_2 = 96$

$C = 15$

$mode = 40 + \frac{44-26}{2(44)-26-96} \times 15 = 32.058$

$\sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{\Sigma f_i}} = \sqrt{\frac{20454.75}{100}} = 14.302$

$\sigma = 14.302$

$Skp = \frac{mean(\bar{x}) - mode}{\sigma} = \frac{47.95 - 32.058}{14.302}$

$= 1.1111$

Here the skewness is positive

2) Calculate Skp for the following data

Variable	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
frequency	2	5	7	13	21	16		

Variable	f_i	C.F	midvalues x_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
0 - 5	2	2	2.5	5	-19.4	376.36	752.72
5 - 10	5	7	7.5	37.5	-14.4	207.36	1036.8
10 - 15	7	14	12.5	87.5	-9.4	88.36	618.52
15 - 20	13	27	17.5	227.5	-4.4	19.36	251.68
20 - 25	21	48	22.5	472.5	0.6	0.36	7.56
25 - 30	16	64	27.5	440	5.6	31.36	501.76
30 - 35	8	72	32.5	260	10.6	112.36	398.88
35 - 40	3	75	37.5	112.5	15.6	243.36	730.08
	$\Sigma f_i = 75$			$\Sigma f_i x_i = 1642.5$			$\Sigma f_i(x_i-\bar{x})^2 = 4798$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1642.5}{75} = 21.9$

$\sigma = \sqrt{\frac{\Sigma f_i (x_i-\bar{x})^2}{\Sigma f_i}} = \sqrt{\frac{4798}{75}} = \sqrt{63.9733} = 7.9983$

$N = \Sigma f = 75 \quad \frac{N}{2} = 37.5 (\geq)$

$f_0 = 21, \ f_1 = 27, \ f_2 = 64, \ C = 5, \ l = 20$

$\text{mode} = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times C = 20 + \frac{21-27}{2(21)-27-64} \times 5$

$= 20 + \frac{-6}{42-91} \times 5 = 20 + \frac{+6}{+49} \times 5 = 20 + 0.1224 \times 5$

$= 20.612$

$S_{KP} = \frac{\text{Mean}(\bar{x}) - \text{Mode}}{\sigma} = \frac{21.9 - 20.612}{7.9983} = \frac{1.288}{7.9983} = 0.1610$

2)

wages	No. of works	midvalues x_i	$x_i f_i$	$x_i - \bar{x}$ ($x_i - 278.47$)	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
125-175	2	150	300	-128.47	16504.54	33009.08
175-225	22	200	4400	-78.47	6157.54	135465.88
225-275	19	250	4750	-28.47	810.54	15400.26
275-325	14	300	4200	21.53	463.54	6489.56
325-375	3	350	1050	71.53	5116.54	15349.62
375-425	4	400	1600	121.53	14769.54	59078.16
425-475	6	450	2700	171.53	29422.54	176535.24
475-525	1	500	500	221.53	49075.54	49075.54
525-575	1	550	550	271.53	73728.54	73728.54
	$\Sigma f_i = 72$		$\Sigma f_i x_i = 20050$			$\Sigma f_i(x_i-\bar{x})^2 = 564131.88$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{20050}{72} = 278.4722$

$\text{Variance} = \sigma^2 = \frac{\Sigma f_i (x_i-\bar{x})^2}{\Sigma f_i} = \frac{564131.88}{72} = 7835.165$

$S.D = \sqrt{7835.165} = 88.5164$

Kurtosis :- The degree of kurtosis of a distribution is measured relative to the peakedness of a normal curve. The measure of kurtosis denotes the shape of the top of a frequency curve

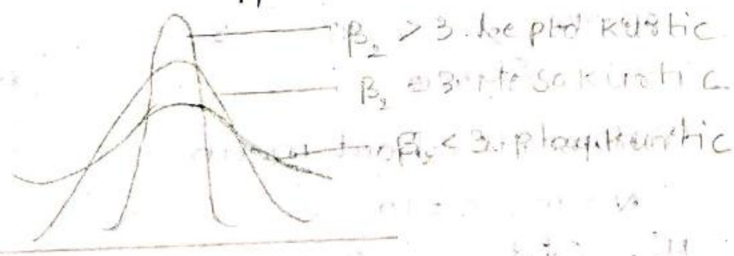
Measures of Kurtosis :

Kurtosis is measured by coefficient of β_2 where β_2 is equal to the 4th moment about the mean divided by the square of the 2nd moment about the mean. Symbolical it can be written as

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

where $\mu_4 = \frac{\Sigma(x_i - \bar{x})^4}{n}$, $\mu_2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$ for ungrouped data

for grouped data $\mu_4 = \frac{\Sigma f_i(x_i - \bar{x})^4}{n}$ & $\mu_2 = \frac{\Sigma f_i(x_i - \bar{x})^2}{n}$



Calculate kurtosis for the following data.

1)

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$	$(x_i-\bar{x})^4$	$f_i(x_i-\bar{x})^4$
6	2	12	-13	169	338	28561	57122
10	4	40	-9	81	324	6561	26244
14	7	98	-5	25	175	625	4375
18	12	216	1	1	12	1	12
24	8	192	5	25	200	625	5000
28	4	112	9	81	324	6561	26244
30	3	90	11	121	363	14641	43923
Total	$\Sigma f_i = 40$	$\Sigma f_i x_i = 760$			$\Sigma f_i(x_i-\bar{x})^2 = 1736$		$\Sigma f_i(x_i-\bar{x})^4 = 162920$

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{760}{40} = 19$

$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$, $\mu_4 = \frac{\Sigma f_i(x_i - \bar{x})^4}{\Sigma f_i} = \frac{162920}{40} = 4073$

$\mu_2 = \frac{\Sigma f_i(x_i - \bar{x})^2}{\Sigma f_i} = \frac{1736}{40} = 43.4$, $\beta_2 = \frac{4073}{(43.4)^2} = 2.1623$

Calculate

a) first 4 moments about the mean.

b) Skewness based on moments.

c) Kurtosis.

Income	0-10	10-20	20-30	30-40
frequency	1	3	4	2

calculation of moments, skewness and kurtosis

Income	frequency f	Midvalue x	$d=\frac{x-A}{h}=\frac{x-15}{10}$	fd	fd^2	fd^3	fd^4
0-10	1	$\frac{0+10}{2}=5$	$\frac{5-15}{10}=\frac{-10}{10}=-1$	-1	1	-1	1
10-20	3	15	0	0	0	0	0
20-30	4	25	1	4	4	4	4
30-40	2	35	2	4	8	16	32
	$\Sigma f=N=10$			$\Sigma fd=7$	$\Sigma fd^2=13$	$\Sigma fd^3=19$	$\Sigma fd^4=37$

a) Moments about mean

$N=10,\ C=10$

$\mu_1'=\frac{\Sigma fd}{N}\times C=\frac{7}{10}\times 10=7$

$\mu_2'=\frac{\Sigma fd^2}{N}\times C^2=\frac{13}{10}\times 100=130$

$\mu_3'=\frac{\Sigma fd^3}{N}\times C^3=\frac{19}{10}\times 1000=1900$

$\mu_4'=\frac{\Sigma fd^4}{N}\times C^4=\frac{37}{10}\times 10000=37000$

first moments about mean $\mu_1=\mu_1'-\mu_1'=7-7=0$

Second moments about mean $\mu_2=\mu_2'-(\mu_1')^2=130-7^2=81$

Third moments about mean $\mu_3=\mu_3'-3\mu_2'\mu_1'+2\mu_1'^3$

$\mu_3=1900-3(130)(7)+2(7^3)=-144$

fourth moments about mean $\mu_4=\mu_4'-4\mu_3'\mu_1'+6\mu_2'\mu_1'^2-3(\mu_1')^4$

$=37000-4(1900)(7)+6(130)(7)^2-3(7$

$=37000-53200+38220-7203$

$\mu_4=14817$

b) $\beta_1=\frac{\mu_3^2}{\mu_2^3}=\frac{(-144)^2}{(81)^3}=0.039$

c) $\beta_2=\frac{\mu_4}{\mu_2^2}=\frac{14817}{(81)^2}=2.2583$

since $\beta_2<3$ this is a play kurtic

Correlation: correlation is a statistical analysis which measures and analysis the degree or extent to the two variables with reference to each other.

The correlation expresses the relationship between two sets of variables upon to each other. In this one variable may be called subject (Independent) and the other relative (Dependent)

Types of correlation:

correlation is classified into following types

1) positive correlation: The correlation between two variables in which with increase in the value of one variable the values of the other variables also increases as vice versa both the variable tend to move in the same direction and is said to be positive correlation.

2) Negative correlation: If two variables tend to move together in opposite directions so that an increase or decrease in the values of one variable is accompanied by a decrease or increase in the value of other variable. Then the correlation is called Negative or Inverse correlation.

3) simple and multiple correlation: when we study only two variables then the relationship is described as simple correlation

Ex: Quantity of money, price level, Demand & price etc.

Multiple correlation we study more than two variables simultaneously

Ex: The relationship of price, Demand & Supply of a commodity.

4) Partial and Total correlation: The study of two variables excluding some other variables is called partial correlation.

Ex: We study price & Demand, eliminating the supply sign. In total correlation all the facts are taken into account.

5, Linear and non-linear correlation: If the variation in the values of two variables have the constant ratio the correlation between them is a linear.

In a ~~le~~ non linear correlation the amount of change in one variable does not be a constant ratio of the amount of change in the other variable.

Ex: The series of height and income of individual.

Methods of studying correlation.

Karl pearson's coefficient of correlation:

The Karl pearson coefficient of correlation is defined as $r = \frac{\text{cov}(x,y)}{\sigma x \, \sigma y} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}}$

$\sigma x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$, $\sigma y = \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}$

The values of r is always lies between -1 and $+1$

If $r = +1$ is said to be perfect positive correlation

If $r = -1$ is said to be perfect Negative correlation

If $r = 0$ then there is no correlation between x & y

If r is positive value between 0 & 1 then it is positive correlation

If r is negative value lying between -1 and 0 then the correlation is negative.

Probleme.

Calculate Karl Pearson's coefficient of correlation from the following data.

x	50	50	55	60	65	65	65	60	60	60
y	11	13	14	16	16	15	15	14	13	13

2)

x	12	9	8	10	11	13	7
y	14	8	6	9	11	12	3

3)

x	10	12	18	24	23	27
y	13	18	12	25	30	10

4).

x	9	8	7	6	5	4	3	2	1
y	15	16	14	13	11	12	10	8	9

1)

x	y	$x_i-\bar{x}$ x_i-59	$(x_i-\bar{x})^2$	$y_i-\bar{y}$ y_i-14	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
50	11	-9	81	-3	9	27
50	13	-9	81	-1	1	9
55	14	-4	16	0	0	0
60	16	1	1	2	4	2
65	16	6	36	2	4	12
65	15	6	36	1	1	6
65	15	6	36	1	1	6
60	14	1	1	0	0	0
60	13	1	1	-1	1	-1
60	13	1	1	-1	1	-1

$n = 10$

$\bar{x} = \frac{\Sigma x}{n} = \frac{590}{10} = 59$

$\bar{y} = \frac{\Sigma y}{n} = \frac{140}{10} = 14$

$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$ $\cdot$ $\text{cov } x,y = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{n}$

$\sigma_x = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$ & $\sigma_y = \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}$

$r = \frac{60}{10\sqrt{\frac{290}{10}} \times \sqrt{\frac{22}{10}}} = \frac{60}{\cancel{10}\frac{\sqrt{290\times 22}}{\cancel{10}}} = \frac{60}{\sqrt{290 \times 22}} = 0.7511$

$r = 0.7511$ is lies b/w 0 & 1
Hence the relation b/w x & y is said to positive correlation.

4)

x	y	$x_i-\bar{x}$ x_i-5	$(x_i-\bar{x})^2$	$y_i-\bar{y}$ y_i-12	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
9	15	4	16	3	9	12
8	16	3	9	4	16	~~64~~ 12
7	14	2	4	2	4	~~28~~ 4
6	13	1	1	1	1	~~13~~ 1
5	11	0	0	-1	1	0
4	12	-1	1	0	0	0
3	10	-2	4	-2	4	4
2	8	-3	9	-4	16	12
1	9	-4	16	-3	9	12
$\Sigma x=45$	$\Sigma y=108$		$\Sigma(x_i-\bar{x})=60$		$\Sigma(y_i-\bar{y})^2=60$	$\Sigma(x_i-\bar{x})(y_i-\bar{y})=57$

$n = 9$

$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$

$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$

$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$ $\text{cov } x,y = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{n}$

$\sqrt{\Sigma(y_i-\bar{y})^2}$

$$r = \frac{57}{9\sqrt{\frac{60}{9}}\sqrt{\frac{60}{9}}}$$

$$= \frac{57}{\frac{\cancel{9}\sqrt{60\times60}}{\cancel{9}}} = \frac{57}{\sqrt{60\times 60}} = 0.95$$

$r = 0.95$ lies b/w 0 & 1

Hence the relation b/w x & y is said to be positive correlation.

2) sol.

x	y	$x_i-\bar{x}$ (x_i-10)	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$ (y_i-9)	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
12	14	2	4	5	25	10
9	8	-1	1	-1	1	1
8	6	-2	4	-3	9	6
10	9	0	0	0	0	0
11	11	1	1	2	4	2
13	12	3	9	3	9	9
7	3	-3	9	-6	36	18
$\Sigma x = 70$	$\Sigma y = 63$		$\Sigma(x_i-\bar{x})^2 = 28$		$\Sigma(y_i-\bar{y})^2 = 84$	$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = 46$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{70}{7} = 10$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{63}{7} = 9$$

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y} \qquad cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{n}$$

$$\sigma_x = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}} \qquad \sigma_y = \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}$$

$$r = \frac{46}{\cancel{7}\sqrt{\frac{28}{7}}\times\sqrt{\frac{84}{7}}} = \frac{46}{\frac{\cancel{7}\sqrt{28\times84}}{\cancel{7}}} = \frac{46}{\sqrt{28\times84}} = 0.948$$

$r = 0.948$ lies b/w 0 & 1

Hence the relation b/w x & y is said to be positive correlation.

3)

x	y	$x_i-\bar{x}$ (x_i-19)	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
10	13	-9	81	~~-8~~ -5	~~64~~ 25	~~72~~ 45
12	18	-7	49	~~-6~~ 0	~~36~~ 0	~~42~~ 0
18	12	-1	1	~~0~~ -6	~~0~~ 36	~~0~~ 6
24	25	5	25	~~6~~ 7	~~36~~ 49	~~30~~ 35
23	30	4	16	~~5~~ 12	~~25~~ 144	~~20~~ 48
27	10	8	64	~~9~~ -8	~~81~~ 64	~~72~~ -64
$\Sigma\bar{x} = 19$	$\Sigma\bar{y} = 18$		$\Sigma(x_i-\bar{x})^2 = 236$		$\Sigma(y_i-\bar{y})^2 =$ ~~[illegible]~~ 318	$\Sigma(x_i-\bar{x})(y_i-\bar{y}) =$ ~~236~~ 70

$\bar{x} = \frac{\Sigma x}{n} = \frac{114}{6} = 19$

$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{6} = 18$

$r = \frac{cov(x,y)}{\sigma_x \sigma_y}$ $cov(x,y) = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n}$

$\sigma_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$, $\sigma_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n}}$

$r = \frac{\cancel{236}\ 70}{6\sqrt{\frac{236}{6}}\sqrt{\frac{\cancel{348}}{6}}} = \frac{\cancel{236}\ 70}{\frac{\cancel{6}\sqrt{236 \times \cancel{348}}}{\cancel{6}}} = \cancel{0.9875}\ 0.2555.$

$r = \cancel{0.9875}\ 0.2555$ lies b/w 0 & 1

Hence the relation b/w x & y is said to be positive correlation.

HW

5)

x	38	45	46	38	35	38	46	32	36	38
y	28	34	38	34	36	26	28	29	25	36

x	y	$x_i - \bar{x}$ ($x_i - 39.2$)	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$ ($y_i - 31.4$)	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
38	28	$38 - 39.2 = \cancel{6.6}\ -1.2$	1.44	-3.4	11.56	4.08
45	34	$\cancel{13.6}$ 5.8	33.64	2.6	6.76	15.08
46	38	6.8	46.24	6.6	43.56	44.88
38	34	-1.2	1.44	2.6	6.76	-3.12
35	36	-4.2	17.64	4.6	21.16	-19.32
38	26	-1.2	1.44	-5.4	29.16	6.48
46	28	6.8	46.24	-3.4	11.56	-23.12
32	29	-7.2	51.84	-2.4	5.76	17.28
36	25	-3.2	10.24	-6.4	40.96	20.48
38	36	-1.2	1.44	4.6	21.16	-5.52
$\Sigma \bar{x} = 392$	$\Sigma \bar{y} = 314$		$\Sigma (x_i - \bar{x})^2 = 211.6$		$\Sigma (y_i - \bar{y})^2 = 198.4$	$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 57.2$

$\bar{x} = \frac{\Sigma x}{n} = \frac{392}{10} = 39.2$, $\bar{y} = \frac{\Sigma y}{n} = \frac{314}{10} = 31.4$

$r = \frac{cov(x,y)}{\sigma_x \sigma_y}$ $cov\ x,y = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n}$

$\sigma_x = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$ $\sigma_y = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n}}$

$r = \frac{57.2}{10\sqrt{\frac{211.6}{10}}\sqrt{\frac{198.4}{10}}} = \frac{57.2}{\frac{\cancel{10}\sqrt{211.6}\sqrt{198.4}}{\cancel{10}}} = 0.2791$

$r = 0.2791$ lies b/w 0 & 1

Hence the relation b/w x & y is said to be positive correlation

2) Rank correlation

Let us suppose that a group of n individuals is arranged in the order of merit or professionce in possesion of two characteristic a & b.

Let (x_i, y_i) $i=1,2,3 \cdots n$ be the ranks of i^{th} individual in two characteristic a and b respectively. Pearson coefficient of correlation b/w the ranks x_i's and y_i's is called the rank correlation co-efficient between a and b for that group of individuals -

Spearman Rank correlation:

Spearman rank correlation is the technique of determining the degree of correlation between two variables in case of ordinal data. where ranks are given to the different values of the variables. The coefficient is determinent as here D_i = difference between the ranks.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad \text{(If ranks are not repeated)}$$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2 + \frac{m(m^2-1)}{12}}{n(n^2-1)} \quad \text{(If ranks are repeated)}.$$

Problem:

Q) A random sample of 5 college students is selected and there grades in mathematics and statistics are found to be.

	1	2	3	4	5
Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

Calculate Spearman rank correlation coefficient

Marks in mathematics X	Marks in Statistic Y	Rank in X_i	Rank in y_i	$d_i = x_i - y_i$	d_i^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	-1	1
40	50	5	5	0	0
90	80	1	2	-1	1
					$\sum d_i^2 = 4$

Spearman Rank correlation $\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$

Here $n=5$

$$\rho = 1 - \frac{6(4)}{5(5^2-1)} = 1 - \frac{24}{5(24)} = 1 - \frac{1}{5} = 0.8$$

$\rho = 0.8$

Q, Following of the marks obtained by 10 Students in 2subjects. Statistics & Mathematics to what extent the knowledge of the student in 2 subject is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

Sol

x	y	$d_i = x-y$	d_i^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4

$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$ Here $n=10$

$$\rho = 1 - \frac{6(40)}{10(99)} = 0.75$$

Q, The ranks of 16 Students in maths and statistics are as follows:

(1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (9,8) (10,11)
(11,15) (12,9) (13,14) (14,12) (15,16) (16,13)

Sol

x	y	$d_i = x-y$	d_i^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9
			$\Sigma d_i^2 = 136$

Here $n = 16$

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6(136)}{16(16^2-1)}$$

$$= 1 - \frac{6(136)}{16(256-1)}$$

$$= 1 - \frac{6(136)}{16(255)}$$

$$\rho = 0.8$$

Q, 10 compatitors in a musical test were ranked by the 3 judges a, b and c in the following order.

Ranks by A	1	6	5	10	3	2	4	9	7	8
Ranks by B	3	5	8	4	7	10	2	1	6	9
Ranks by C	6	4	9	8	1	2	3	10	5	7

using rank correlation method discuss which pair of judges has the nearest approach to common likings in music -

Here $N = 10$

Rank by A (x)	Ranks by B(y)	Ranks by C(z)	$D_1 = x-y$	$D_2 = y-z$	$D_3 = z-x$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	1	~~8~~ 16
10	4	8	6	-4	-2	36	16	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	-1	1	4	1
						$\Sigma d_1^2 = 200$	$\Sigma d_2^2 = 214$	$\Sigma d_3^2 = 60$

$n = 10$

$$\rho_1 = 1 - \frac{6\Sigma d_1^2}{n(n^2-1)} = 1 - \frac{6(200)}{10(99)} = -0.2121$$

$$\rho_2 = 1 - \frac{6\Sigma d_2^2}{n(n^2-1)} = 1 - \frac{6(214)}{10(99)} = -0.2969$$

$$\rho_3 = 1 - \frac{6\Sigma d_3^2}{n(n^2-1)} = 1 - \frac{6(60)}{10(99)} = 0.6363$$

Since $\rho_3(x, z)$ is maximum we can conclude that the pair of Judges A and C has the nearest approach to common linkinges in music.

Repeated Rank problems :

Q) Obtain the rank correlation coefficient for the following data

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Sol

x	y	Ranks x	Ranks y	D = x - y	D^2
68	62	4	5	-1	1
64	~~64~~ 58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16

$$\rho = 1 - \frac{6\Sigma d^2 + \frac{m(m^2-1)}{12}}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\left(72 + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}\right)}{10(10^2-1)}$$

$\rho = 0.561$

100 pending

From the following data calculate the Karl Parleson coefficient after making adjustment of tid vons

x 48 33 40 9 16 16 65 24 16 57

y 13 13 24 6 15 4 20 9 6 19

A)

x	Rank in x	y	Ranks in y	$d=x-y$	d^2
48	3	13	5.5	-2.5	6.25
33	5	13	5.5	-0.5	0.25
40	4	24	1	3	9
9	10	6	8.5	1.5	2.25
16	8	15	4	4	16
16	8	4	10	-2	4
65	1	20	2	-1	1
24	6	9	7	-1	1
16	8	6	8.5	+0.5	0.25
57	2	19	3	-1	1
					$\Sigma d^2 = 41$

16 Repeated 3 times in x then $m=3$

13 Repeated 2 times in y then $m=2$

6 Repeated 2 times in y then $m=2$

$$\rho = 1 - \frac{6\left[\overset{(41)}{\Sigma d^2} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12}\right]}{10(10^2-1)} \Rightarrow \rho = 0.7479$$

$$\boxed{\rho = 1 - \frac{6\left[\Sigma d^2 \frac{m(m^2-1)}{12}\right]}{\ }}$$

Calculate coefficient of rank correlation

fathers(x)	65	63	67	64	68	62	70	66	68	67	69	71
Sons (y)	68	66	68	65	69	66	68	65	71	67	68	70

x	Rank in x	y	Rank in y	d = x - y	d^2
65	9	68	5	4	16
63	11	66(9)	9.5	1.5	2.25
67	6.5	68(7)	5	1.5	2.25
64	10	65(11)	11.5	-1.5	2.25
68	4.5	69	3	1.5	2.25
62	12	66(10)	9.5	2.5	6.25
70	2	68	5	-3	9
66	8	65(12)	11.5	-3.5	12.25
68	4.5	71	1	3.5	12.25
67	6.5	67	8	-1.5	2.25
69	3	68	5	-2	4
71	1	70	2	-1	1
					82

$$\rho = 1 - \frac{6 \Sigma d^2 \frac{m(m^2-1)}{12}}{n(n^2-1)}$$

$$\rho = 1 - \frac{6(82) + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{4(4^2-1)}{12} + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12}}{12(12^2-1)}$$

Regression line :

Regression : In Regression we can estimate the value of one variable with the value of the other variable which is known. The statistical method which helps the to estimate the unknown value of one variable from the known value the related variable is called regression. The line described in the average relation between two variables is known as line of regression.

Fitting a straight line:

$y = a + bx$ is called straight line.

The normal equations are $\Sigma y = na + b\Sigma x$,

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Determine the equation of straight line which best [fits] the data.

1)

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

2)

x	0	1	2	3	4
y	1	1.8	3.3	4.5	6.3

2) sol

x	y	x^2	xy
0	1	0	0
1	1.8	1	1.8
2	3.3	4	6.6
3	4.5	9	13.5
4	6.3	16	25.2
$\Sigma x=10$	$\Sigma y=16.9$	$\Sigma x^2=30$	$\Sigma xy=47.1$

The normal equations are

$\Sigma y = na + b\Sigma x$, $\Sigma xy = a\Sigma x + b\Sigma x^2$, Here $n=5$

$16.9 = 5a + 10b$ — ①

$47.1 = 10a + 30b$ — ②

Solving eq's ① & ② but multiplying ① × 2, ② × 1

$33.8 = 10a + 20b$

$\underline{\underset{(-)}{47.1} = \underset{(-)}{10a} \underset{(-)}{+30b}}$

$\cancel{-}13.3 = \cancel{-}10b$

$10b = 13.3 = \frac{13.3}{10} \Rightarrow b = 1.3$

Sub $b = 1.3$ in ① we get

$16.9 = 5a + 10(1.3)$

$16.9 = 5a + 13$

$16.9 - 13 = 5a$

$5a = 3.9 \Rightarrow a = \frac{3.9}{5} = 0.78$

WKT $y = a + bx$ is a straight line

$y = 0.78 + (1.3)x$

This is the required straight line.

This is also Regression equation of y on x.

2m **Regression lines**

1) The Regression line of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$ wh

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

2) Regression line of x on y is $x - \bar{x} = b_{xy}(y - \bar{y})$ where bx

2m **Regression coefficient**

Regression coefficient is denoted by 'r' and is defined as $r = \sqrt{b_{yx} \times b_{xy}}$ where $b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$,

$b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$

Q) For the following data find eq's of the 2 regression lin[illegible]

1)

x	1	2	3	4	5
y	15	25	35	45	55

2) Criticise the following Regression coefficient of y on x i[illegible] 0.7 and x on y is 3.2

Regression coefficient $r = \sqrt{byx + bxy}$

Given y on x ie $byx = 0.7$

Given x on y i.e $bxy = 3.2$

$\therefore$ Regression coefficient $r = \sqrt{0.7 \times 3.2} = 1.4966$

But correlation coefficient cannot exceed '1'

Hence there is some inconsistency in the informati[on] given.

1)

sol:

x	y	$X = x - \bar{x}$ ($X = x - 3$)	X^2	$Y = y - \bar{y}$ ($Y = y - 35$)	Y^2	XY
1	15	-2	4	-20	400	40
2	25	-1	1	-10	100	10
3	35	0	0	0	0	0
4	45	1	1	10	100	10
5	55	2	4	20	400	40
$\Sigma x = 15$	$\Sigma y = 175$		$\Sigma X^2 = 10$		$\Sigma Y^2 = 1000$	$\Sigma XY = 100$

$\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$

$\bar{x} = \frac{15}{5} = 3$, $\bar{y} = \frac{175}{5} = 35$

$byx = \frac{\Sigma XY}{\Sigma X^2} = \frac{100}{10} = 10.$

$bxy = \frac{\Sigma XY}{Y^2} = \frac{100}{1000} = 0.1$

Regression line of y on x i[s]

$y - \bar{y} = byx(x - \bar{x})$

$y - 35 = 10(x - 3)$

$y - 35 = 10x - 30$

$10x - y - 30 + 35 = 0$

$10 - y + 5 = 0$

Regression line of x on y is

$x - \bar{x} = bxy(y - \bar{y})$

$x - 3 = 0.1(y - 35)$

$x - 3 - 0.1y + 3.5 = 0$

$x - 0.1y + 0.5 = 0.$

Q) Price indices of cotton and wool are given below for the 12 Months of a year obtain the equations of lines of regression between the indices.

Price index of cotton x	78	77	85	88	87	82	81	77	76	83	97	93
Price index of wool y	84	82	82	85	89	90	88	92	83	89	98	99

x	y	$x = X - \bar{X}$ ($x-84$)	x^2	$y = Y - \bar{Y}$ ($y-88$)	y^2	xy
78	84	-6	36	-4	16	24
77	82	-7	49	-6	36	42
85	82	1	1	-6	36	-6
88	85	4	16	-3	9	-12
87	89	3	9	1	1	3
82	90	-2	4	2	4	-4
81	88	-3	9	0	0	0
77	92	-7	49	4	16	-28
76	83	-8	64	-5	25	40
83	89	-1	1	1	1	-1
97	98	13	169	10	100	130
93	99	9	81	11	121	99
$\Sigma X = 1004$	$\Sigma Y = 1061$		$\Sigma x^2 = 488$		$\Sigma y^2 = 365$	$\Sigma xy = 287$

$\Sigma x = 1004$, $n = 12$, $\Sigma y = 1061$

$\bar{x} = \frac{\Sigma x}{n} = \frac{1004}{12} = 83.66 = 84$

$\bar{y} = \frac{\Sigma y}{n} = \frac{1061}{12} = 88.41 = 88$.

$\Sigma x^2 = 488$, $\Sigma y^2 = 365$, $\Sigma xy = 287$

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{287}{488} = 0.5881$

$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{287}{365} = 0.7863$

The regression line of y on x is

$y - \bar{y} = b_{yx}(x - \bar{x})$.

$y - 88 = 0.5881(x-4) \Rightarrow y = 0.5881x - 84(0.5881) + 88$

$y = 0.5881x + 38.5996$.

The regression line of x on y is

$x - \bar{x} = b_{xy}(y - \bar{y})$

$x - 84 = 0.7863(y-88) \Rightarrow x = 0.7863y - 0.7863 \times 88 + 84$

$x = 0.7863y + 14.8056$.

H.W) The height of the mothers & daughters are given in the following table. From the two tables of regression estimate the expected Average height of daughter when the height of the Mother is 64.5 inches

Height of the Mother (x)	62	63	64	64	65	66	68	70
Height of the	64	65	61	69	67	68	71	65

2) Calculate the regression equations of y on x from the following data. Taking deviations from actual means of x & y.

Price (₹) x Rs	10	12	13	12	16	15
Amount demanded (y)	40	38	43	45	37	43

Estimate the likely demand when the price is 20 Rupees

sol

xy	X	y	$X = x - \bar{x}$ ($x-13$)	x^2	$y = y - \bar{y}$ ($y-41$)	y^2
~~400~~ 3	10	40	-3	9	-1	1
~~456~~ 3	12	38	-1	1	-3	9
~~559~~ 0	13	43	0	0	+2	4
~~540~~ -4	12	45	-1	1	4	16
~~592~~ -12	16	37	3	9	-4	16
~~645~~ 4	15	43	2	4	2	4
$\Sigma xy = -6$	$\Sigma x = 78$	$\Sigma y = 246$		$\Sigma x^2 = 24$		$\Sigma y^2 = 50$

$\bar{x} = \frac{\Sigma x}{n} = \frac{78}{6} = 13$

$\bar{y} = \frac{\Sigma y}{n} = \frac{246}{6} = 41$

$b_{yx} = \frac{\Sigma xy}{x^2} = \frac{-6}{24} = -0.25$

$b_{xy} = \frac{\Sigma xy}{y^2} = \frac{-6}{50} = -0.12$

Regression line of y on x

$y - \bar{y} = b_{yx}(x - \bar{x})$

$y - 41 = -0.25(x - 13)$

$y - 41 = -0.25x + 3.25$

$y = -0.25x + 3.25 + 41$

$y = -0.25x + 44.25$

given $x = 20$

$y = -0.25(20) + 44.25 \Rightarrow y = -5 + 44.25 \Rightarrow y = 39.25$

1)
sol

X	y	$X = x - \bar{x}$ ($x-65$)	x^2	$y = y - \bar{y}$ ($y-66$)	y^2	xy
62	64	-3	9	-2	4	6
63	65	-2	4	1	1	2
64	61	-1	1	-5	25	5
64	69	-1	1	3	9	-3
65	67	0	0	1	1	0
66	68	1	1	2	4	2
68	71	3	9	5	25	15
70	65	5	25	-1	1	-5
$\Sigma x = 522$	$\Sigma y = 530$		$\Sigma x^2 = 50$		$\Sigma y^2 = 70$	$\Sigma xy = 22$

$\bar{x} = \frac{522}{8} = 65.25 = 65$

$\bar{y} = \frac{530}{8} = 66.25 = 66$

$\bar{x} = \frac{522}{8} = 65.25 = 65$

$\bar{y} = \frac{530}{8} = 66.25 = 66$,

Regression of y on x

$y - \bar{y} = byx(x - \bar{x})$

$byx = \frac{\Sigma xy}{x^2} = \frac{22}{50} = 0.44$

$y - 66 = 0.44(x - 65)$

$y - 66 = 0.44x - 0.44(65)$

$y = 0.44x - 28.6 + 66$

$y = 0.44x + 37.4$

given $x = 64.5$

$y = 0.44(64.5) + 37.4$

$y = 28.38 + 37.4$

$y = 65.78$.

Q) from the following data calculate
i) Correlation coefficient
ii) standard deviation of y (σy)
$bxy = 0.85$, $byx = 0.89$ $\sigma x = 3$

Sol i) Correlation coefficient $r = \sqrt{bxy \times byx} = \sqrt{0.85 \times 0.89} = 0.8697$
$= 0.87$

ii) Standard deviation of y is $r \times \frac{\sigma x}{\sigma y} = bxy$

$$0.87 \times \frac{3}{\sigma y} = 0.85$$

$$\sigma y = \frac{0.87 \times 3}{0.85} = 3.07$$

The following data based on 450 students are given for marks in statistic & economics at a certain examination Mean marks in statistic = 40, Mean marks in economics = 8, Standard Deviation of marks in Statistics = 12, variance of marks(economics) = $\sigma y^2 = 256$. Sum of products of deviation of marks from this respective Mean 42075 give the equations of the two lines of regression and estimate the Average Marks in economics of candidates who obtained 50 Marks in statistics.

Given, $\bar{x}=40$, $\bar{y}=8$,

$\sigma_y^2 = 256$

$\sigma_y = \sqrt{256} = 16$.

r = coefficient of correlation $= \dfrac{\Sigma xy}{N\sigma_x \sigma_y}$

$\Sigma xy = 42075$ given

~~N = 0.4869 = 0~~ $N = 450$ students

$r = \dfrac{42075}{450 \times 12 \times 16} = 0.4869 = 0.49$

Regression line of y on x is

$y - \bar{y} = r \dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$

$y - 8 = \dfrac{0.49 \times 16}{12}(x - 40)$

$y = 0.65x - 26 + 8$

$y = 0.65x - 18$

Given $x = 50$

$y = 0.65 \times 50 - 18$

$\boxed{y = 14.5}$

Regression line of x on y

$x - \bar{x} = r \dfrac{\sigma_x}{\sigma_y}(y - \bar{y})$

$x - 40 = \dfrac{0.49(12)}{16}(y - 8)$

$x = 0.3675y - 2.94 + 40 \Rightarrow x = 0.3675y + 37.06$

The following calculations have been made for prices of 12 stocks (x) in stock exchange on a certain day along with the volume of the sales in thousands of shares (y). From these calculations find the regression equation of prices of stock, on volume of the sales of shares. $\Sigma x = 580$, $\Sigma y = 370$, $\Sigma xy = 11499$, $\Sigma x^2 = 41658$, $\Sigma y^2 = 17.206$

Given $\Sigma x = 580$, $\Sigma y = 370$, $\Sigma x^2 = 41658$, $\Sigma y^2 = 17.206$, $\Sigma xy = 11499$

Given $N = 12$, $\bar{x} = \dfrac{\Sigma x}{N} = \dfrac{580}{12} = 48.33$, $\bar{y} = \dfrac{\Sigma y}{N} = \dfrac{370}{12} = 30.83$

$b_{xy} = \dfrac{\Sigma xy - N\bar{x}\bar{y}}{\Sigma y^2 - N(\bar{y})^2} = \dfrac{11499 - (12)(48.33)(30.83)}{17.206 - 12(30.83)^2} = 0.5603$

Regression line of x on y is

$x - \bar{x} = b_{xy}(y - \bar{y})$

$x - 48.33 = 0.5603(y - 30.83)$

$x = 0.5603y - 17.274 + 48.33$

Q) Find the Mean Values of the variables X & y, & Correlation Coefficient from the following regression lines:

$2y - x - 50 = 0$,

$3y - 2x - 10 = 0$,

Given equations can be written as $2y - x = 50$ — (1)

$3y - 2x = 10$ — (2)

Solving (1) & (2), (1) $\times 2$, (2) $\times 1$

$$4y - \cancel{2x} = 100$$

$$3y - \cancel{2x} = 10$$

$$\underline{(-)\quad (+)\quad\quad (-)}$$

$$\boxed{y = 90}$$

sub $y = 90$ in eq (1)

$$2(90) - x = 50$$

$$180 - 50 = x$$

$$\boxed{x = 130}$$

we get means $\bar{x} = 130$, $\bar{y} = 90$.

$$2y - x = 50$$

$$2y = x + 50$$

$$y = \frac{1}{2}x + \frac{50}{2}$$

$$y = \frac{1}{2}x + 25 \quad \text{— (3)}$$

b_{yx}

$r = \pm\sqrt{b_{yx} \times b_{xy}}$

Hly $3y - 2x = 10$

$$3y - 10 = 2x$$

$$\frac{3}{2}y - \frac{10}{2} = x$$

$$x = \frac{3}{2}y - 5 \quad \text{— (4)}$$

Let $r = \frac{\sigma_y}{\sigma_x} = \frac{1}{2}$ from (3) — (5)

from (4) $r = \frac{\sigma_x}{\sigma_y} = \frac{3}{2}$ — (6)

Multiply (5) $\times$ (6). $r^2 = \frac{1}{2} \times \frac{3}{2} = \frac{3}{4}$

$$r = \frac{\sqrt{3}}{2} = 0.8660.$$

1Q) Test whether the equations $2x + 3y = 4$ & $x - y = 5$ represent valid regression lines.

2Q) Given the bivariant data

x	1	5	3	2	1	1	7	3
y	6	1	0	0	1	2	1	5

a) find the regression line of y on x & hence predict y if $x = 10$

if $y = 2.5$

c) calculate Karl pearson's correlation coefficient (r) $r = \sqrt{byx + bxy}$

Sol:

x	y	$X = x - \bar{x}$ ($x-3$)	X^2	$Y = y - \bar{y}$ ($y-2$)	y^2	xy
1	6	-2	4	4	16	-8
5	1	2	4	-1	1	-2
3	0	0	0	-2	4	0
2	0	-1	1	-2	4	2
1	1	-2	4	-1	1	2
1	2	-2	4	0	0	0
7	1	4	16	-1	1	-4
3	5	0	0	3	9	0
$\Sigma x = 23$	$\Sigma y = 16$		$\Sigma x^2 = 33$		$\Sigma y^2 = 36$	$\Sigma xy = -10$

$\bar{x} = \frac{\Sigma x}{n} = \frac{23}{8} = 2.8 = 3$

$\bar{y} = \frac{\Sigma y}{n} = \frac{16}{8} = 2$

$byx = \frac{\Sigma xy}{\Sigma x^2} = \frac{-10}{33} = -0.3030$

$bxy = \frac{\Sigma xy}{y^2} = \frac{-10}{36} = -0.2777$

i) Regression line of y on x is

$y - \bar{y} = byx(x - \bar{x})$

$y - 2 = -0.303(x - 3)$

$y = -0.303x + 0.909 + 2$

$y = -0.303x + 2.909$

given $x = 10$

$y = -3.03 + 2.909$

$y = -0.121$

ii) Regression line of x on y is

$x - \bar{x} = bxy(y - \bar{y})$

$x - 3 = -0.2777(y - 2)$

$x - 3 = -0.2777y + 0.5554$

$x = -0.2777y + 0.5554 + 3$

$x = -0.2777y + 3.5554$

given $y = 2.5$

$x = -0.2777(2.5) + 3.5554$

$x = -0.69425 +$

$x_1 = 2.8612$

c) $r = \sqrt{byx + bxy}$

$r = \sqrt{(-0.3030) + (-0.2778)} = 0.2901$

1). Given equations can be $2x+3y=4$ — (1)

Sol. $x-y=5$ — (2)

Multiply (2) $\times 2$, (1) $\times 1$

$$\cancel{2x}+3y=4$$
$$\cancel{2x}-2y=10$$
$$\underline{\quad - \quad + \quad - \quad}$$
$$5y=-6 \Rightarrow \boxed{y=\frac{-6}{5}}$$

sub $y=\frac{-6}{5}$ in eq (2)

$$x-\left(-\frac{6}{5}\right)=5 \Rightarrow x+\frac{6}{5}=5 \Rightarrow x=5-\frac{6}{5} \Rightarrow x=\frac{25-6}{5}=\frac{19}{5}$$

$$\boxed{x=\frac{19}{5}}$$

$\bar{x}=\frac{19}{5}$, $\bar{y}=\frac{-6}{5}$

$2x+3y=4$

$3y=4-2x$

$y=\frac{4}{3}-\frac{2}{3}x$ — (3)

Illy $x-y=5$

$x=5+y$

$x=y+5$ — (4)

Let $r=\frac{\sigma_y}{\sigma_x}=-\frac{2}{3}$ from (5)

$r=\frac{\sigma_x}{\sigma_y}=1$ from (6)

Multiply (5) & (6) $r^2=\frac{-2}{3}\times 1$

$r^2=-\frac{2}{3}$

$\therefore r^2$ is negative

Hence the given eq's does not represent valid regression lines.

Unit-2

Probability and Random variable

Random Experiment

If an experiment is conducted any number of times under essentially identical conditions there is a set of all possible outcomes associated with it. If the result is not certain & is any one of the several possible outcomes. The experiment is called a random trial or random experiment. The outcomes are known as elementary events & a set of outcomes is an even.

Simple event: An event in a trial that cannot be further split is called a simple event or an elementary event

Sample space: The set of all possible simple events in a trial is called a sample space for the trial. sample point is generally denoted by 'E'. Thus, a simple event is a sample point. sample space is denoted by 'S'.

Eg: Two coins are tossed.

Then the possible simple events of the trial are HH, HT, TH, TT.

The sample space of the trial S = {HH, HT, TH, TT}

Mutually exclusive Events:

Two events E_1, E_2 of a sample space 'S' are said to be mutually exclusive if they have no sample points in common i.e., $E_1 \cap E_2 = \phi$. Mutually exclusive events are sometimes called disjoint events.

probability: The ratio of number of elementary events in E and total no. of elementary events in the random experiment then let E be an event of the experiment then the probability of E is defined as $P(E) = \frac{m}{n}$ =

$$\frac{\text{number of elementary in } E}{\text{Total number of elementary events in the random experiment}}$$

The probability is always between 0 and 1

Axioms of probability:

1) Axioms of positivity: $P(E) \geq 0$ for every subset $E(S)$

2) Axioms of certainity: $P(S) = 1$

3) Axioms of ... : If E_1, E_2 are disjoint subsets of

S' then $p(E_1 \cup E_2) = p(E_1) + p(E_2)$

problems:

What is the probability for a leap year to have 52 Mondays & 53 Sundays.

A leap year has 366 days i.e., 52 weeks & 2 days. These 2 days can be any one of the following (seven) 7 ways.

1) Mon & Tue 2) Tue & wed 3) wed & Thurs 4) Thurs & Fri

5) Fri & Sat 6) Sat & Sun 7) Sun & Mon

Let 'E' be an event having 52 Mondays & 53 Sundays in the year.

Total no. of possible cases $n = 7$

Number of favourable cases to E is $m = 1$ (Sat & Sun is the only favourable case).

$$P(E) = \frac{m}{n} = \frac{1}{7}$$

In a class there are 10 Boys & 5 girls. A committee of 4 students is to be selected from the class. Find the probability for the committee to contain atleast 3 girls.

A committee of 4 students out of 15 (10 Boys & 5 girls) can be formed in ${}^{15}C_4$ ways i.e. $n = {}^{15}C_4$

Let 'E' be the event of forming a committee with atleast 3 girls.

Now the committee can have 1 boy 3 girls, 4 girls no boys

So the no. of ways forming a committee is equal to the no. of favourable ways 'E'

$= {}^{10}C_1 \times {}^{5}C_3 + {}^{10}C_0 \times {}^{5}C_4 = 105 = m$

$$p(E) = \frac{m}{n} = \frac{105}{{}^{15}C_4} = 0.0769$$

A class consist of 6 girls & 10 boys. If a committee of 3 is choosen at random from the class find the probability that

i) 3 boys are selected

ii) Exactly 2 girls are selected.

Total no. of students = 16

$n(S)$ = no. of ways of choosing 3 from 16 = ${}^{16}C_3$ ways

i) Suppose 3 boys are selected this can be done

$n(E) = {}^{10}C_3$

$P(E) = \frac{n(E)}{n(S)} = \frac{{}^{10}C_3}{{}^{16}C_3} = \frac{120}{560} = 0.2143$.

ii, Suppose exactly two girls are selected then

$n(E) = {}^6C_2 \times {}^{10}C_1$

$P(E) = \frac{n(E)}{n(S)} = \frac{{}^6C_2 \times {}^{10}C_1}{{}^{16}C_3} = 0.2678$

Q A and B throw alternatively with a pair of ordinary dice. A wins if he throws 6 before B throws 7. and B wins if he throws 7 before A throws 6. If A begins show that is chance of winning is $\frac{30}{61}$.

when two dice are thrown we have $n(S) = 36$

The probability of ~~that~~ A throwing 6 i.e. (1,5) (2,4) (3,3), (4,2) (5,1) $= \frac{5}{36}$

The probability of A not throwing 6 is and is given by $P(\bar{A}) = 1 - P(A) = 1 - \frac{5}{36} = \frac{36-5}{36} = \frac{31}{36}$

The probability of B throwing 7 = (1,6) (2,5) (3,4) (4,3) (5,2) (6,1).

$P(B) = \frac{6}{36}$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{6}{36} = \frac{30}{36}$

Chances of winning of A

$= P(A) + P(\bar{A})P(\bar{B})P(A) + P(\bar{A})P(\bar{B})P(\bar{A})P(\bar{B})P(A) + \cdots$

$= \frac{5}{36} + \left(\frac{31}{36}\right)\left(\frac{30}{36}\right)\left(\frac{5}{36}\right) + \left(\frac{31}{36}\right)\left(\frac{30}{36}\right)\left(\frac{31}{36}\right)\left(\frac{30}{36}\right)\left(\frac{5}{36}\right) + \cdots$

$= \frac{5}{36}\left[1 + \left(\frac{31}{36}\right)\left(\frac{30}{36}\right) + \left(\frac{31}{36} \times \frac{30}{36}\right)^2 + \cdots\right]$

$1 + x + x^2 + \cdots = (1-x)^{-1} = \frac{1}{1-x}$ Here $x = \frac{31}{36} \times \frac{30}{36}$

$= \frac{5}{36}\left[\frac{1}{1 - \frac{31}{36} \times \frac{30}{36}}\right] = \frac{5}{36}\left[\frac{1}{\frac{36^2 - 30 \times 31}{36^2}}\right] = \frac{5}{36} \times \frac{36^2}{366} = \frac{36 \times 5}{366} = \frac{30}{61}$

Determine the probability for each of the following events a non defective bolt will be found if out of 600 bolt already examined 12 were defective bolt

the probability of defective bolt, $P(B) = \frac{12}{600}$

The probability of finding non defective bolt, $P(\bar{B}) = 1 - P(B)$

$= 1 - \frac{12}{600}$

$= \frac{49}{50}$

What is the probability of picking and ace and a king from 52 cards deck

Number of ways of picking two cards (1 ace & 1 king) from 52 cards $= {}^{52}C_2$ ways $= n(S)$

The no. of ways 1 ace and a king $= {}^4C_1 \times {}^4C_1 = {}^{16}C_1$ ways $= n(E)$

Probability of picking 1 ace and a king $= P(E) = \frac{n(E)}{n(S)}$

$$= \frac{16}{{}^{52}C_2} = \frac{16}{\frac{52 \times 51}{2 \times 1}} = \frac{\overset{4}{\cancel{16}} \times 2}{\underset{13}{\cancel{52}} \times 51} = \frac{8}{663}$$

Q) Out of 15 items 4 are not in good condition. 4 are selected at random. Find the probability that

i) All are not good

ii) 2 are not good.

Total no. of items $= 15$

The number of ways of picking 4 items $= {}^{15}C_4 = n(S)$

i) Suppose 4 items are choosen which are not good

Number of ways of selecting 4 items which are not good $= n(E) = {}^4C_4$

$$P(E) = \frac{n(E)}{n(S)} = \frac{{}^4C_4}{{}^{15}C_4} = \frac{1}{1365}$$

ii) Suppose two items are not good

The no. of ways of selecting 2 bad items $= {}^4C_2$

Number of ways of selecting 2 good items $= {}^{11}C_2$

Required probability of getting 2 items which are not good $= P(2 \text{ good } \& \text{ } 2 \text{ bad items}) = P(E) = \frac{n(E)}{n(S)} = \frac{{}^4C_2 \times {}^{11}C_2}{{}^{15}C_4}$

$$= \frac{330}{1365} = \frac{22}{91}$$

Q) A card is drawn from a well shuffled pack of cards. What is the probability that it is either get a spade or an ace.

Let 'S' is the sample space of all the simple events

$n(S) = 52$

'A' denote the event of getting a spade & 'B' denote the event of getting an ace

$P(A) = \frac{13}{52}$, $P(B) = \frac{4}{52}$

$P(A \cap B) = \frac{1}{52}$

$P(A\cup B) = P(A) + P(B) - P(A\cap B)$

$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$

P, 3 students A, B, C are in running race. A & B have the same probability of winning & each is twice as likely to win as C. find the probability that B or C wins.

$A\cup B\cup C = S$

$P(A) = P(B)$

$P(A) = 2P(C)$

we have, $P(A) + P(B) + P(C) = 1$

$2P(C) + P(A) + P(C) = 1$

$2P(C) + 2P(C) + P(C) = 1$

$5P(C) = 1 \Rightarrow P(C) = \frac{1}{5}$

$P(A) = 2P(C) = 2\left(\frac{1}{5}\right) = \frac{2}{5}$

$P(A) = P(B) = \frac{2}{5}$

The probability that B or C wins $= P(B\cup C)$

$= P(B) + P(C) - P(B\cap C)$

$= \frac{2}{5} + \frac{1}{5} - 0$

$\boxed{P(B\cup C) = \frac{3}{5}}$

2M

Conditional event:

If E_1, E_2 are events of a sample space 'S' ~~of a Sample Space~~ and if E_2 occurs after the occurence of E_1 then the event of occurrence of E_2 after the event E_1 is called conditional event of E_2 given E_1. It is denoted by $\frac{E_2}{E_1}$ (E_2 given E_1) lly $\frac{E_1}{E_2}$

Eg: Two coins are tossed. The event of getting two tails is a conditional event given that there is atleast 1 tail is a conditional event.

Conditional probability: If E_1 and E_2 are two events in a sample space 'S' and $P(E_1) \neq 0$ then the probability of E_2 after the event E_1 has occured is called the conditional probability of the event E_2 given E_1 and is denoted by $P\left(\frac{E_2}{E_1}\right)$ and we define $P\left(\frac{E_2}{E_1}\right) = \frac{P(E_2\cap E_1)}{P(E_1)}$ lly $P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1\cap E_2)}{P(E_2)}$

Multiplication theorem of probability:

Statement: In a random experiment if E_1, E_2 are two events such that $P(E_1) \neq 0$ & $P(E_2) \neq 0$ then $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2/E_1)$, $P(E_2 \cap E_1) = P(E_2) P(E_1/E_2)$

Q) Determine 1) $P(B/A)$ & 2) $P(A/B^c)$ if A & B are events if $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$, $P(A \cup B) = \frac{1}{2}$

Given $P(A) = \frac{1}{3}$

$P(B) = \frac{1}{4}$

$P(A \cup B) = \frac{1}{2}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$= \frac{1}{3} + \frac{1}{4} - \frac{1}{2} = \frac{4+3-6}{12} = \frac{1}{12}$

$P(A \cap B) = \frac{1}{2}$

1) $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{12}}{\frac{1}{3}} = \frac{3}{12} = \frac{1}{4}$, $\boxed{P(B^c) = 1 - P(B) = 1 - \frac{1}{4} = \frac{3}{4}}$

2) $\boxed{P(A \cap B^c)} = P(A) - P(A \cap B)$

$= \frac{1}{3} - \frac{1}{12} = \frac{4-1}{12} = \frac{3}{12} = \frac{1}{4}$

$P(A/B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$

Q) Box A contains 5 red and 3 white marbles & box B contains two red & 6 white marbles. If a marble is drawn from each box what is the probability that they are both of the same colour.

Suppose E_1 = the event that the marble is drawn from box A and is red.

$P(E_1) = \frac{1}{2} \cdot \frac{5}{8} = \frac{5}{16}$

E_2 = the event that the marble is from box B & is red.

$P(E_2) = \frac{1}{\cancel{2}} \cdot \frac{\cancel{2}}{8} = \frac{1}{8}$.

The probability that both the marbles are red is

$P(E_1 \cap E_2) = P(E_1) P(E_2) = \frac{5}{16} \cdot \frac{1}{8} = \frac{5}{128}$

Let E_3 = the event that the marble drawn from box A and is white

$P(E_3) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16}$

Let E_4 = The event that the marbles from box B' & is white

$P(E_4) = \frac{1}{2} \cdot \frac{\cancel{6}3}{\cancel{8}4} = \frac{3}{8}$

$P(E_3 \cap E_4) = P(E_3) P(E_4) = \frac{3}{16} \cdot \frac{3}{8} = \frac{9}{128}$

The probability that the marble are of same Colour

$= P(E_1 \cap E_2) + P(E_3 \cap E_4)$

$= \frac{5}{128} + \frac{9}{128} = \frac{14}{128} = \frac{7}{64}$

H.W If $P(A \cap B) = \frac{1}{6}$, $P(A) = \frac{1}{2}$ then find $P(B/A)$

Given $P(A \cap B) = \frac{1}{6}$, $P(A) = \frac{1}{2}$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{\cancel{6}3} \times \cancel{2} = \frac{1}{3}$$

If $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{4}$ & $P(A \cap B) = \frac{1}{8}$ then find $P(A \cup B)$

Given $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{4}$, $P(A \cap B) = \frac{1}{8}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{4+2-1}{8}$$

$$P(A \cup B) = \frac{5}{8}$$

If $P(A) = 0.25$, $P(B) = 0.50$ & $P(A \cup B) = 0.59$ then find $P(A \cap B)$

$P(A) = 0.25$, $P(B) = 0.50$, $P(A \cup B) = 0.59$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.25 + 0.50 - 0.59 = 0.16$

Two Aeroplanes bomb a target in Successen the probability of each correctly Scoring a hit is 0.3 & 0.2 respectively. The second will bomb only if the first misses the target. Find the probability that (i) target is hit (ii) both fails to score hits.

Let A be the event of first plane hitting the target, and B be the event of second plane hitting the target

$P(A) = 0.3$, $P(B) = 0.2$

$P(\bar{A}) = 1 - P(A) = 1 - 0.3 = 0.7$

$P(\bar{B}) = 1 - P(B) = 1 - 0.2 = 0.8$

i, P(Target is hit) = P(A hits) OR (A fails and B hits)

$= P(A \cup (\bar{A} \cap B))$

$[P(A \cap (\bar{A} \cap B)) = 0$

$= P(A) + P(\bar{A} \cap B)$

$= P(A) + P(\bar{A})P(B) = 0.3 + (0.7)(0.2) = 0.44$

iii) $P(\text{both fails}) = P(\bar{A} \cap \bar{B}) = P(\bar{A})P(\bar{B}) = (0.7)(0.8) = 0.56$

Two marbles are drawn in successen from a box containing 10Red, 30 white, 20 blue and 15 orange marbles. with replacement being made after each draw. Find the probability that ① Both are white ② first is red and second is white.

Total number of marbles in the box = 75

1) Let E_1 = event of the first drawn marble is white.

$P(E_1) = \frac{30}{75}$

Let E_2 = event of second drawn marble is also white

$P(E_2) = \frac{30}{75}$.

$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) = \frac{\cancel{30}^{2}}{\cancel{75}_{5}} \times \frac{\cancel{30}^{2}}{\cancel{75}_{5}} = \frac{4}{25}$.

2) Let E_1 = the event that the first drawn marble is red

$P(E_1) = \frac{10}{75}$

Let E_2 = the event that the draw marble is white

$P(E_2) = \frac{30}{75}$

$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) = \frac{\cancel{10}^{2}}{\cancel{75}_{15}} \times \frac{\cancel{30}^{2}}{\cancel{75}_{5}} = \frac{4}{75}$

The probabilities that the students A, B, C, D solve a Problem are $\frac{1}{3}, \frac{2}{5}, \frac{1}{5}, \frac{1}{4}$ respectively. If all of them try to solve the problem, what is the probability that the problem is solved.

Let $P(A) = \frac{1}{3}$, $P(B) = \frac{2}{5}$, $P(C) = \frac{1}{5}$, $P(D) = \frac{1}{4}$

$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}$, $P(\bar{B}) = 1 - \frac{2}{5} = \frac{3}{5}$, $P(\bar{C}) = 1 - P(C)$
$= 1 - \frac{1}{5} = \frac{4}{5}$

$P(\bar{D}) = 1 - P(D) = 1 - \frac{1}{4} = \frac{3}{4}$

The probability that the problem is not solved when A, B, C, D try together independently.

$P(\bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D}) = P(\bar{A}) P(\bar{B}) P(\bar{C}) P(\bar{D}) = \frac{2}{3} \times \frac{3}{5} \times \frac{4}{5} \times \frac{3}{4}$

$= \frac{6}{25}$

The probability that the problem is solved $= 1 - \frac{6}{25} = \frac{19}{25}$

H.W
A can hit a target 3 times in 5 shots, B hits target 2 times in 5 shots, C hits target 3 times in 4 shots. Find the probability of the being hits when all of them try.

$P(A) = \frac{3}{5}$, $P(B) = \frac{2}{5}$, $P(C) = \frac{3}{4}$

$P(\bar{A}) = 1 - P(\bar{A}) = 1 - \frac{3}{5} = \frac{2}{5}$, $P(\bar{B}) = 1 - \frac{2}{5} = \frac{3}{5}$

$P(\bar{C}) = 1 - \frac{3}{4} = \frac{1}{4}$

$P(\bar{A}\cap\bar{B}\cap\bar{C}) = P(\bar{A})P(\bar{B})P(\bar{C}) = \frac{2}{5}\cdot\frac{3}{5}\cdot\frac{1}{4} = \frac{\cancel{6}^{3}}{\cancel{100}_{50}} = \frac{3}{50}$

The probability of being hits when all of them try

$= 1 - P(\bar{A}\cap\bar{B}\cap\bar{C}) = 1 - \frac{3}{50} = \frac{47}{50}$

Imp
A problem in statistics is given to the 3 students whose chances of solving hit are $\frac{1}{2}$, $\frac{3}{4}$ & $\frac{1}{4}$ respecti[vely]. What is the probability that the problem is solved?

$P(A) = \frac{1}{2}$, $P(B) = \frac{3}{4}$, $P(C) = \frac{1}{4}$

$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$, $P(\bar{B}) = 1 - \frac{3}{4} = \frac{1}{4}$, $P(\bar{C}) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(\bar{A}\cap\bar{B}\cap\bar{C}) = P(\bar{A})P(\bar{B})P(\bar{C}) = \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{3}{4} = \frac{3}{32}$

The probability that the problem is solved $= 1 - P(\bar{A}\cap\bar{B}\cap\bar{C})$

$= 1 - \frac{3}{32} = \frac{29}{32}$

H.W
The probabilities of 3 students to solve a problem in Mathematics are $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$. Find the probability that the problem to be solved.

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, $P(C) = \frac{1}{4}$

$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$, $P(\bar{B}) = 1 - \frac{1}{3} = \frac{2}{3}$, $P(\bar{C}) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(\bar{A}\cap\bar{B}\cap\bar{C}) = P(\bar{A})P(\bar{B})P(\bar{C}) = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4} = \frac{6}{24}$

The probability that the problem is solved $= 1 - P(\bar{A}\cap\bar{B}\cap\bar{C})$

$= 1 - \frac{6}{24} = \frac{18}{24} = \frac{\cancel{9}^{3}}{\cancel{12}_{4}} = \frac{3}{4}$

Imp
A & B are two events such that $P(A) = \frac{1}{3}$ & $P(B) = \frac{3}{4}$ & $P(A\cup B) = \frac{11}{12}$ find $P(A/B)$ & $P(B/A)$. -2016

Given $P(A) = \frac{1}{3}$, $P(B) = \frac{3}{4}$ and $P(A\cup B) = \frac{11}{12}$

$P(A\cup B) = P(A) + P(B) - P(A\cap B)$

$P(A\cap B) = P(A) + P(B) - P(A\cup B) = \frac{1}{3} + \frac{3}{4} - \frac{11}{12} = \frac{1}{6}$

i) $p(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{4}} = \frac{1}{\cancel{6}_3} \times \frac{\cancel{4}^2}{3} = \frac{2}{9}$

ii) $p(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{\cancel{6}_2} \times \cancel{3} = \frac{1}{2}$

Q) 3 Machines I, II, III produce 40%, 30%, 30% of the total no. of items of the factory. The percentage of defective items of this machines are

A) Let A, B & C be the events of the Machines I, II & III respectively and let D be the event which denotes the defective item.

Given $p(A) = \frac{40}{100}$, $p(B) = \frac{30}{100}$, $p(C) = \frac{30}{100}$

from the given data we have $p(D/A) = \frac{4}{100}$, $p(D/B) = \frac{2}{100}$, $p(D/C) = \frac{3}{100}$.

l^{ly} The probability that the selected item at random is defective is.

$P(D) = p(A)p(D/A) + p(B)p(D/B) + p(C)p(D/C)$

$= \frac{4\cancel{0}}{10\cancel{0}} \times \frac{4}{100} + \frac{3\cancel{0}}{10\cancel{0}} \times \frac{2}{100} + \frac{3\cancel{0}}{10\cancel{0}} \times \frac{3}{100}$

$= \frac{16+6+9}{1000} = \frac{31}{1000}$

A can hit a target once in 5 shots, B can hit two targets in 3shots. C can hit

Bayes Theorem

Defined Bayes theorem

$E_1, E_2, \dots E_n$ are n mutually exclusion and exhaustive ~~exections~~ events such that $p(E_i) > 0$ $(i = 1, 2, 3 \dots n)$ is in a sample space S and A is any other event in S intersetting with $F \subset D \subset E_i$ such that $p(E) > 0$ if (E_i is any of events of $E_1, E_2, \dots n$) where p(Events, E_2, E_n) is any $p\left(\frac{A}{E_1}\right) p\left(\frac{A}{E_2}\right) \dots p(A/E_n)$ are known that

$$p(E_K/A) = \frac{P(E_K)\, p\left(\frac{A}{E_K}\right)}{P(G)\, p\left(\frac{A}{G}\right) + p(E_2)\, p\left(\frac{A}{E_2}\right) + \dots + p(E_n)\, p(A/E_n)}$$

Problems

Q) Suppose 5 men out of 100 & 26 women out of 10,000 are colour blind a colour blind person is choose at random. what is the probability of the person being a male (assume male & female to be in equal number)

The probability of the choose person is male

$P(m) = \frac{1}{2}$

probability of person is female $p(w) = \frac{1}{2}$

$P[B/m) = \frac{5}{100} = 0.05$

$p(B/w) = \frac{25}{10,000} = 0.0025$

the probability is choose person is male given by

$$P(m/B) = \frac{P(B/m)\, P(m)}{P(m)\, P(B/m) + P(w)\, P(B/w)} = \frac{(0.05)(0.5)}{(0.5)(0.05) + (0.5)(0.0025)}$$

$$= 0.95$$

Q) In a Bold-factory A, B, C manufacturing 20%, 30%, 50% of the total of there out & 6%, 3%, 2% are defected a bold is drawn at random & bold to be detected. find the probability the is manufacturing from machine A, machine B and machine C.

P(A) P(B) P(C) be the probability of the events the bold are manufacture by the machine P(A) P(B) P(C) are respectively

$P(A) = \frac{20}{100}$ $\quad p(B) = \frac{30}{100}$ $\quad P(C) = \frac{50}{100}$

$P(D/A) = \frac{6}{100}$ $\quad p(D/B) = \frac{3}{100}$ $\quad P(D/C) = \frac{2}{100}$

1) $p(A/D) = \frac{P(D/A)\, p(A)}{\dots}$

$$= \frac{\frac{20}{100} \times \frac{6}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}} = \frac{\frac{3}{250}}{\frac{3}{350} + \frac{9}{1000} + \frac{1}{100}} = \frac{\frac{3}{250}}{\frac{31}{1000}}$$

$$= \frac{3}{250} \times \frac{1000}{31} = \frac{12}{31}$$

2) $$P(B/D) = \frac{P(B)P(D/B)}{P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)}$$

$$= \frac{\frac{30}{100} \times \frac{3}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}} = \frac{\frac{9}{\cancel{1000}}}{\frac{31}{\cancel{1000}}} = \frac{9}{31}$$

3) $$P(C/D) = \frac{P(C)\,P(C/B)}{P(A)P(D/B) + P(B)P(D/B) + P(C)P(D/C)} = \frac{\frac{50}{100} \times \frac{2}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{3}{100} \times \frac{30}{100} + \frac{50}{100} \times \frac{20}{100}}$$

$$= \frac{1}{\cancel{100}} \times \frac{\cancel{1000}}{31} = \frac{10}{31},$$

A bag A contains 2 white and 3 red balls and bag B contains 4 white and 5 Red balls. one ball is drawn for one of the bags and its found to be red. find the probability the red ball drawn is from bag B.

$P(A) = 1/2$, $P(B) = 1/2$

$P(R/A) = 3/5$ similarly, $P(R/B) = 5/9$ it is founded to be red then the probability the selected bag is B

$$P(B/R) = \frac{P(B) \cdot P(B/R)}{P(A)\,P(R/A) + P(B) \cdot P(R/B)} = \frac{3/5 \times \frac{5}{9}}{1/2 \times 3/5 + 1/2 \times \frac{5}{9}}$$

$$= \frac{\frac{1}{2} \times \frac{5}{9}}{\frac{1}{2} \times \frac{3}{5} + 1/2 \times \frac{5}{9}} = \frac{\frac{5}{18}}{\frac{3}{10} + \frac{5}{18}} = \frac{\frac{5}{18}}{\frac{26}{45}} = \frac{5}{18} \times \frac{45}{26}$$

$$= \frac{25}{52}.$$

A business man goes to hotels x, y, z 20%, 50%, 30% of the time respectively it is known that 5%, 4%, 8% of the rooms in x, y, z hotels have faulty plumbings what is the probability that business man room having faulty plumbing is assign to hotel z.

let the probabilities of businessman going to hotels x, y, z respectively.

$P(x), P(y), P(z)$ then $P(x) = \frac{20}{100} = \frac{1}{5}$

$$P(y) = \frac{50}{100} = \frac{1}{2}$$

$$P(z) = \frac{30}{100} = \frac{3}{10}$$

$P(E/x) = \frac{5}{100} = \frac{1}{20}$, $P(E/y) = \frac{4}{100} = \frac{1}{25}$

$P(E/z) = \frac{8}{100} = \frac{2}{25}$

The probability that the business man room having faulty plumbings is assign to hotel z.

$$P(z/E) = \frac{P(z) \cdot P(E/z)}{P(x)P(E/x) + P(y) \cdot P(E/y) + P(z)P(E/z)}$$

$$= \frac{\frac{3}{10} \times \frac{2}{25}}{\frac{1}{5} \times \frac{1}{20} + \frac{1}{2} \times \frac{1}{25} + \frac{3}{10} \times \frac{2}{25}} = \frac{\frac{6}{250}}{\frac{27}{500}}$$

$$= \frac{6}{250} \times \frac{500}{27} = \frac{4}{9}.$$

The 3 men the chances that a politican, business man (or) an academician will be apointed as a vice chanceller (vc) of a university are 0.5, 0.3, 0.2 respectively. probability that research promoted by this person if they are appointed as vc are 0.3, 0.7, 0.8 respectively.

a) Determine the probability that research is promoted

b) If research is promoted what is the probability that vc is academician.

let A, B, C be the events that a politician, business man (or) an academician will be apointed as v-c of the three men then $P(A) = 0.5$, $P(B) = 0.3$, $P(C) = 0.2$

$P(R/A) = 0.3$.

$P(R/B) = 0.7$

$P(R/C) = 0.8$

a) The probability that the research is promoted when the V-C is an academician

$$P(C/R) = \frac{P(C) \cdot P(R/C)}{P(A) \cdot P(R/A) + P(B) \cdot P(R/B) + P(C) \cdot P(R/C)}$$

$$= \frac{(0.8)(0.8)}{(0.5)\cdot(0.3) + (0.3)(0.7) + (0.2)(0.8)} = \frac{0.16}{0.52}$$

$$= 0.30769.$$

Random variables:

Definition: A Real variables 'x' whose value is determined by the outcome of a Random experiment is called a random variable.

Types of random variables:

Random variable is of two types

1) Discrete Random variable
2) continuous Random variable.

1) Discrete Random variable: A random variable 'x' which can take a only a finite number of discrete values in an interval of Domain is called a discrete random variable.

~~2 continuous R~~

Ex: The Random variable denoting a no. of students in a class is $X(x) = \{x : x$ is a positive integer$\}$.

2) continuous Random variable: A Random variable 'x' which can take values continuously i.e. which takes all possible values in a given interval. is called a continuous Random variable.

Ex: The height, age and weight of individuals are examples of continuous Random variable.

Formulas:

Expectation (or) mean:

$$E(x) = \sum_{i=1}^{n} P_i x_i$$

$$\text{variance} = E(x^2) - [E(x)]^2$$

$$S.D = \sqrt{\text{variance}}$$

problems:

let x denotes the no. of heads in a single task of 4 fails coins. Determine

a) $p(x<2)$ b) $p(1<x\leq 3)$

the required probability distribution.

x	0	1	2	3	4
$p(x)$	$1/16$	$4/16$	$6/16$	$\frac{4}{16}$	$1/16$

a) $p(x<2) = p(x=0) + p(x=1) = 1/16 + 4/16 = \frac{5}{16}$

b) $p(1<x\leq 3) = p(x=2) + p(x=3) = \frac{6}{16} + 4/16 = \frac{10}{16} = \frac{5}{8}$

For the descrete probability distribution.

Find a) k

b) mean

c) variance.

a) Since the total probability is unity.

$$\sum_{k=0}^{6} p(x) = 1$$

$$0+2k+2k+3k+k^2+2k^2+7k^2+k=1$$

$$10k^2+8k-1=0$$

$$k = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

$a=10$

$b=8$

$c=-1$

$$k = \frac{-8 \pm \sqrt{64-4(10)(-1)}}{2(10)}$$

$$= \frac{-8 \pm \sqrt{64+40}}{20} = \frac{-8 \pm \sqrt{104}}{20} = \frac{-8 + \sqrt{104}}{20}$$

$= 0.1099.$

b) Mean $= \sum_{k=0}^{6} P_i x_i$

$= 0\times 0 + 1\times 2k + 2\times 2k + 3\times 3k + 4\times k^2 + 5\times 2k^2 + 6(7k^2+k)$

$= 56k^2 + 21k$

$= 56(0.1099)^2 + 21(0.1099)$

mean $= \mu = E(x) = 2.9643.$

c) variance $= \Sigma p_i x_i^2 - \mu^2 = E(x^2) - (E(x))^2$

$= 0 + 2k(1)^2 + (2k)^2 2^2 + (3k)3^2 + (k^2)4^2 + (2k^2)5^2 + (7k^2 + k)6^2 - (2.9843)^2$

$= 0 + 2k + 8k + 27k + 16k^2 + 50k^2 + 25k^2 + 36k - (2.9843)^2$

$= 318k^2 + 73k - 8.9060$

$= 318(0.1099)^2 + 73(0.1099) - 8.9060 = 2.9575$

A random variable x has the following probability

Q) A random variable x has the following probability

x	0	1	2	3	4	5	6	7
$p(x)$	0	1	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

find 1) k 2) $p(x<6)$, $p(x\geq 6)$, $p(0<x<5)$, $p(0\leq x\leq 4)$

3) Determine the distribution function of x 4) mean

5) variance.

1) $\sum_{k=0}^{7} p(x) = 1$

$0+k+2k+2k+3k+k^2+2k^2+7k^2+k=1$

$10k^2+9k-1=0$

$10k^2+10k-k-1=0$

$10k(k+1)-1(k+1)=0$

$(10k-1)(k+1)=0$

$10k-1=0 \Rightarrow 10k=1 \Rightarrow k=\frac{1}{10}$

$k=-1,\ k=\frac{1}{10}$

∴ Only take positive values $p(x)\geq 0$ then $k=\frac{1}{10}$ or 0.1

2) $p(x<6)=p(x=0)+p(x=1)+p(x=2)+p(x=3)+p(x=4)+p(x=5)$

$= 0+k+2k+2k+3k+k^2 = k^2+8k = (0.1)^2+8(0.1)$

$p(x<6)=0.81$

$p(x\geq 6)=p(x=6)+p(x=7)$ or $1-p(x<6)$

$=2k^2+7k^2+k = 9k^2+k = 9(0.1)^2+0.1 = 0.19$

$p(0<x<5) = p(x=1)+p(x=2)+p(x=3)+p(x=4)$

$= k+2k+2k+3k = 8k = 8(0.1) = 0.8$

$p(0\leq x\leq 4) = p(x=0)+p(x=1)+p(x=2)+p(x=3)+p(x=4)$

$= 0+k+2k+2k+3k = 8k = 8(0.1) = 0.8$

3) The distribution function of x is given by the following table

x	$F(x)=p(X\leq x)$
0	0
1	$k+0=k=0.1/\frac{1}{10}$
2	$2k+k+0=3k=\frac{3}{10}$
3	$5k=\frac{5}{10}$
4	$8k=\frac{8}{10}$
5	$8k+k^2=\frac{8}{10}+\frac{1}{100}=\frac{81}{100}$
6	$8k+3k^2=\frac{8}{10}+\frac{3}{100}=\frac{83}{100}$
7	$9k+10k^2=\frac{9}{10}+\frac{10}{100}=\frac{100}{100}=1$

4) Mean $= (0)(0)+(1)(K)+(2)(2K)+3(2K)+4(3K)+5K^2+12K^2+49K^2+7K$

$\mu = E(x) = 66K^2+30K = 66(0.1)^2+30(0.1) = 3.66$

5) Variance $= \Sigma p_i x_i^2 - \mu^2$

$= 0+(K)1^2+(2K)2^2+2K(3^2)+3K(4^2)+K^2(5^2)+2K^2(6)^2+(7K^2+K)(7)^2$

$= 440K^2+124K-(3.66)^2$

Variance $= \sigma^2 = 440(0.1)^2+124(0.1)-(3.66)^2 = 3.4044$

Q) A variate x has been following probability distribution -2016

x	-3	6	9
$p(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

find a) $E(x)$ b) $E(x^2)$ c) $E(2x+1)^2$ d) Mean & Variance

a) $E(x) =$ Mean $= \Sigma x p(x) = (-3)\left(\frac{1}{6}\right)+6\left(\frac{1}{2}\right)+9\left(\frac{1}{3}\right) = \frac{-1}{2}+3+3$

$= \frac{-1}{2}+\frac{6}{1} = \frac{11}{2}$

b) $E(x^2) = \Sigma x^2 p(x) = 9\left(\frac{1}{6}\right)+36\left(\frac{1}{2}\right)+81\left(\frac{1}{3}\right) = \frac{93}{2}$

c) $E(2x+1)^2 = E(4x^2+1+4x) = 4E(x^2)+E(1)+4E(x)$

$= 4\left(\frac{93}{2}\right)+1+4\left(\frac{11}{2}\right) = 209$

d) Mean $= E(x) = \frac{11}{2}$

Variance $= \Sigma x^2 p(x) - \mu^2$ (or) $E(x^2)-(E(x))^2$

$= \frac{93}{2} - \left(\frac{11}{2}\right)^2 = \frac{93}{2} - \frac{121}{4} = \frac{65}{4}$

find the mean & variance of uniform probability distribution given by $f(x) = \frac{1}{n}$ for $x = 1, 2, 3 \cdots n$.

The probability distribution is

x	1	2	3	----	n
$f(x)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	----	$\frac{1}{n}$

Mean $= E(x) = \Sigma x_i f(x_i)$

$= (1)\left(\frac{1}{n}\right)+2\left(\frac{1}{n}\right)+3\left(\frac{1}{n}\right)+\cdots+n\left(\frac{1}{n}\right)$

$= \frac{1}{n}(1+2+3+\cdots+n)$

Mean $= \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$

Variance $= \Sigma x^2 f(x) - E(x)^2$

$= 1^2\left(\frac{1}{n}\right)+2^2\left(\frac{1}{n}\right)+3^2\left(\frac{1}{n}\right)+\cdots+n^2\left(\frac{1}{n}\right)-\left(\frac{n+1}{2}\right)^2$

$= \frac{1}{n}\left[1^2+2^2+3^2+\cdots+n^2\right]-\frac{(n+1)^2}{4}$

$= \frac{1}{\cancel{n}}\frac{\cancel{n}(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$

$= (n+1)\left[\frac{2n+1}{6}-\frac{n+1}{4}\right]$

$= (n+1)\left[\frac{4n+2-3n-3}{12}\right]$

$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$

1)

x	1	2	3	4	5
$P(x)$	k	$2k$	$3k$	$4k$	$5k$

- 2019

find a) k b) mean c) variance

2)

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

find a) k b) Mean c) Variance.

3)

x	-2	-1	0	1	2	3
$P(x)$	0.1	k	0.2	$2k$	0.3	k

find 1) k 2) mean 3) variance 4) $P(x \geq 2)$ 5) $P(x < 2)$
6) $P(-1 < x < 3)$

continuous probability Distribution

when a random variable takes every value in an interval it gives rise to continuous distribution of x. The distribution defined by the variants like temperature heights & weights are continuous Distributions.

Probability Density function.

for continuous variable the probability distribution is called probability Density function.

Properties of Probability Density function $f(x)$:-

1) $f(x) \geq 0 \ \forall x \in R$

2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

3) $P(a < x < b) = \int_a^b f(x)\,dx$

4) $P(a < x \leq b) = P(a \leq x < b) = P(a \leq x \leq b) = F(b) - F(a)$

5) Mean of distribution is given by $\mu = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$

6) variance of distribution is given by $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

Problems:

Q) If x is a continuous random variable & $y = ax+b$ prove that $E(y) = a\,E(x) + b$ and $V(y) = a^2 V(x)$ where V stands for variance & a, b are constants.

Given $y = ax+b$

$$E(y) = E(ax+b) \quad \because E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_{-\infty}^{\infty} (ax+b) f(x)\,dx$$

$= a\int_{-\infty} x f(x)\,dx + b\int_{-\infty} f(x)\,dx \quad \because \int_{-\infty} f(x)\,dx = 1$

$= a\,E(x) + b(1)$

$\boxed{E(y) = a\,E(x) + b}$

Given $E(y) = aE(x) + b$ — ①

$y = ax + b$ — ②

eq ② − eq ①

$y - E(y) = (ax+b) - (aE(x)+b) = ax + \cancel{b} - aE(x) - \cancel{b}$

$= a(x - E(x))$

squaring on b.s & taking exception we get.

$E(y - E(y))^2 = a^2 E((x - E(x))^2$

$V(y) = a^2 V(x)$

where, $V(y) = E(y - E(y))^2$

$V(x) = E(x - E(x))^2$

Q) The probability density $f(x)$ of a continuous random variable is given by $f(x) = ce^{-|x|}$, $-\infty < x < \infty$ show that $c = \frac{1}{2}$ & find the mean & variance of the distribution also find the probability that the variate lies between 0 and 4.

Given $f(x) = ce^{-|x|}$, $-\infty < x < \infty$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$\int_{-\infty}^{\infty} ce^{-|x|}\,dx = 1$

$\left[\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx \right.$ $f(x)$ is even. $\left.\right]$

$2c\int_{0}^{\infty} e^{-|x|}\,dx = 1$

$2c\int_{0}^{\infty} e^{-x}\,dx = 1 \Rightarrow 2c\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = 1$

$-2c\left[e^{-x}\right]_0^{\infty} = 1 \Rightarrow -2c\left[e^{-\infty} - e^{0}\right] = 1 \quad [\because e^{-\infty} = 0,\ e^{\infty} = \infty,\ e^{0} = 1]$

$\Rightarrow -2c[0 - 1] = 1$

$\Rightarrow 2c = 1$

$\boxed{c = \frac{1}{2}}$

$f(x) = ce^{-|x|} = \frac{1}{2}e^{-|x|}$

Mean $= \mu = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} x\,\frac{1}{2}\,e^{-|x|}\,dx = \frac{1}{2}\int_{-\infty}^{\infty} x e^{-|x|}\,dx.$

since the integral is odd function then

$f(x) = x$, $f(-x) = -x$

$\mu = E(x) = \frac{1}{2}(0) = 0.$

Variance $= \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx$ (or) $\int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$= \int_{-\infty}^{\infty} x^2 f(x)dx = 2\int_0^{\infty} x^2 f(x)dx$ [$\because$ The function is even]

$= \cancel{2}\int_0^{\infty} x^2 \frac{1}{\cancel{2}} e^{-|x|}dx$

$= \int_0^{\infty} x^2 e^{-|x|}dx = \int_0^{\infty} x^2 e^{-x}dx$ [Bernoullis Rule]

$= \left[x^2\left(\frac{e^{-x}}{-1}\right) - 2x\left(\frac{e^{-x}}{(-1)^2}\right) + 2\frac{e^{-x}}{(-1)^3}\right]_0^{\infty}$

$= \infty^2\left(\frac{e^{-\infty}}{-1}\right) - 2\infty\frac{e^{-\infty}}{(-1)^2} + \frac{2e^{-\infty}}{(-1)^3}\Big] - \left[0 - 0 + \frac{2e^{-0}}{(-1)^3}\right]$

$e^{-\infty} = 0, e^0 = 1$

$\sigma^2 = 0 + \frac{2(1)}{(-1)} = 2$ $\boxed{\sigma^2 = \text{Variance} = 2}$

4) $P(0 \le x \le 4) = \int_0^4 f(x)dx = \int_0^4 \frac{1}{2}e^{-|x|}dx = \frac{1}{2}\int_0^4 e^{-x}dx$

$= \frac{1}{2}\left[\frac{e^{-x}}{-1}\right]_0^4 = \frac{1}{2}\left[\frac{e^{-4}}{-1} - \frac{e^{-0}}{-1}\right]$

$= \frac{-1}{2}\left[e^{-4} - e^{-0}\right] = -\frac{1}{2}\left[e^{-4} - 1\right] = 0.4908$

Q) probability Density function $f(x) = kx^2e^{-x}$, $x \ge 0$ find a) k b) Mean c) Variance - 2013-16

Given $f(x) = kx^2e^{-x}$, $x \ge 0$

W.K.T $\int_{-\infty}^{\infty} f(x)dx = 1$

$\int_0^{\infty} f(x)dx = 1$ ($x \ge 0$ is given)

$\int_0^{\infty} kx^2e^{-x}dx = 1 \Rightarrow k\int_0^{\infty} x^2e^{-x}dx = 1$

$k\left[x^2\left(\frac{e^{-x}}{(-1)}\right) - 2x\left(\frac{e^{-x}}{(-1)^2}\right) + 2\left(\frac{e^{-x}}{(-1)^3}\right)\right]_0^{\infty} = 1$

$e^{-\infty} = 0, e^0 = 1$

$k\left[0 - \left(0 - 0 + \frac{2e^{-0}}{(-1)^3}\right)\right] = 1 \Rightarrow k\left(0 + \frac{2(1)}{(+1)}\right) = 1$

$2k = 1 \Rightarrow \boxed{k = \frac{1}{2}}$

b) Mean $= E(x) = \int_{-\infty}^{\infty} x f(x)dx$ $\quad f(x) = kx^2e^{-x}$ $(x \ge 0)$

$= \int_0^{\infty} x k x^2 e^{-x}dx$

$= \int_0^{\infty} x\frac{1}{2}x^2e^{-x}dx = \frac{1}{2}\int_0^{\infty} x^3e^{-x}dx$

$= \frac{1}{2}\left[x^3\left(\frac{e^{-x}}{(-1)}\right) - 3x^2\left(\frac{e^{-x}}{(-1)^2}\right) + 6x\frac{e^{-x}}{(-1)^3} - \frac{6e^{-x}}{(-1)^4}\right]_0^{\infty}$

$e^{-\infty} = 0, e^0 = 1$

$= \frac{1}{2}\left[0 - \left(0 - 0 + 0 - 6\frac{e^0}{(-1)^4}\right)\right] = \frac{1}{2}\left[6(1)\right] = 3$

$$\text{Variance} = \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 \quad (x \geq 0$$

$$= \int_0^{\infty} x^2 \frac{1}{2} x^2 e^{-x} dx - 3^2 = \frac{1}{2}\int x^4 e^{-x} dx - 9$$

$$= \frac{1}{2}\left[+x^4 \frac{e^{-x}}{(-1)} - 4x^3 \frac{e^{-x}}{(-1)^2} + 12x^2 \frac{e^{-x}}{(-1)^3} - 24x \frac{e^{-x}}{(-1)^4} + 24 \frac{e^{-x}}{(-1)^5} \right] - 9$$

$$= \frac{1}{2}\left[0 + (0 - 0 + 0 - 0 + \frac{24e^{0}}{(+1)}\right] - 9$$

$$= \frac{1}{2}(24) - 9 = 3.$$

$\text{Variance} = \sigma^2 = 3$

$S.D = \sqrt{\text{Variance}} = \sqrt{3}.$

Q) A continuous random variable x has a probability density function $f(x) = 3x^2$ $0 \leq x \leq 1$. find a & b such that 1) $P(x \leq a) = P(x > a)$ & 2) $p(x \geq b) = 0.05$

The probability density function of a continuous random variable is given by $f(x) = 3x^2$ $0 \leq x \leq 1$

W.K.T $P(x \leq a) = P(x \geq a) = \frac{1}{2}$

$P(x \leq a) = \frac{1}{2}$ $\quad \left[P(x \leq a) = \int_0^a f(x)dx\right]$

$\int_0^a f(x)dx = \frac{1}{2} = \int_0^a 3x^2 dx = \frac{1}{2} = 3\int_0^a x^2 dx = \frac{1}{2}$

$\Rightarrow 3\left(\frac{x^3}{3}\right)_0^a = \frac{1}{2} \Rightarrow a^3 = \frac{1}{2} \Rightarrow a = \left(\frac{1}{2}\right)^{\frac{1}{3}} = \frac{1}{2^{\frac{1}{3}}}.$

2) $P(x \geq b) = 0.05$

$\int_b^1 f(x)dx = 0.05 \Rightarrow \int_b^1 3x^2 dx = 0.05 \Rightarrow 3\left(\frac{x^3}{3}\right)_b^1 = 0.05$

$1 - b^3 = 0.05 \Rightarrow 1 - 0.05 = b^3 \Rightarrow b^3 = 0.95$

$b = (0.95)^{\frac{1}{3}}$ (or) $\left(\frac{19}{20}\right)^{\frac{1}{3}}$

Q) The length of the time (in meters) that a certain lady speaks on the telephone is found to be random phenomenal with a probability function specified with the function $f(x) = Ae^{-x/5}$ $x \geq 0$ 1) find the value of A 2) what is the probability that the no. of minutes that she will take over the phone is more than 10 minutes.

Given $f(x) = A e^{-x/5}, x \geq 0$
$= 0$ otherwise

W.K.T $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^{\infty} A e^{-x/5} dx = 1 \;(x \geq 0)$

$A\int_0^{\infty} e^{-x/5} dx = 1 \Rightarrow A\left[\frac{e^{-x/5}}{\left(\frac{-1}{5}\right)}\right]_0^{\infty} = 1$

$$A(-5)\left[e^{-x/5}\right]_0^{\infty}=1$$

$$-5A\left[e^{-\infty}-e^{0}\right]=1 \Rightarrow -5A[0-1]=1$$

$$5A=1 \Rightarrow \boxed{A=\frac{1}{5}}$$

2) $p(x\geq 10)=\int_{10}^{\infty} f(x)\,dx=\int_{10}^{\infty} A e^{-x/5}\,dx = \int_{10}^{\infty}\frac{1}{5}e^{-x/5}\,dx$

$$=\frac{1}{5}\int_{10}^{\infty} e^{-x/5}\,dx=\frac{1}{5}\left[\frac{e^{-x/5}}{\left(\frac{-1}{5}\right)}\right]_{10}^{\infty}=-\left[e^{-x/5}\right]_{10}^{\infty}$$

$$=-\left[e^{-\infty}-e^{-\frac{10}{5}}\right]=e^{-2}=\frac{1}{e^2}$$

Q) If the Random Variable has the p.d.f $f(x)=k(x^2-1)\ -1\leq x\leq 3$
0 otherwise find the value of k & $P(\frac{1}{2}\leq x\leq 5/2)$

Given $f(x)=k(x^2-1)\quad -1\leq x\leq 3$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx=1$

$$\int_{-1}^{3} f(x)\,dx=1 \Rightarrow \int_{-1}^{3} k(x^2-1)=1$$

$$k\int_{-1}^{3} x^2-1=1 \Rightarrow k\left[\frac{x^3}{3}-x\right]_{-1}^{3}=1 \Rightarrow k\left[(9-3)-\left(\frac{-1}{3}+1\right)\right]=1$$

$$\Rightarrow k\left[6+\frac{1}{3}-1\right]=1 \Rightarrow k\left[\frac{18+1-3}{3}\right]=1 \Rightarrow k\left[\frac{16}{3}\right]=1$$

$$\Rightarrow k=\frac{3}{16}$$

$$P\left(\frac{1}{2}\leq x\leq \frac{5}{2}\right)$$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx \Rightarrow \int_{\frac{1}{2}}^{5/2} k(x^2-1) \Rightarrow k\left[\frac{x^3}{3}-x\right]_{\frac{1}{2}}^{5/2}=$

$$\Rightarrow k\left[\left(\frac{125}{8\times 3}-\frac{5}{2}\right)-\left(\frac{1}{3(8)}-\frac{1}{2}\right)\right]$$

Let x be a continuous random variable with distribution
$f(x)=\ \frac{1}{8}$ if $0\leq x\leq 8$
0 else where.

find 1) $P(2\leq x\leq 5)$ 2) $P(3\leq x\leq 7)$ 3) $P(x\leq 6)$

$f(x)=\frac{1}{8}$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx=1$

1) $\int_{2}^{5}\frac{1}{8}\,dx=1$

$\frac{1}{8}\int_{2}^{5} dx=1 \Rightarrow \frac{1}{8}(x)_2^5=1$

$\Rightarrow \frac{1}{8}(5-2)=1 \Rightarrow \frac{3}{8}$

2) $P(3\leq x\leq 7)$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx$

$\int_{3}^{7}\frac{1}{8}\,dx$

$\frac{1}{8}\int_{3}^{7} dx$

$\frac{1}{8}(x)_3^7$

$\frac{1}{8}(7-3)$

$=\frac{4}{8}$

3) $P(x\leq 6)$

W.K.T $\int_{-\infty}^{\infty} f(x)\,dx$

$\int_{0}^{6}\frac{1}{8}\,dx$

$=\frac{1}{8}\int_{0}^{6} dx$

$=\frac{1}{8}(x)_0^6$

$=\frac{1}{8}(6-0)=\frac{6}{8}$

Unit - 3

Probability Distributions

Binomial Distribution : A random variable 'x' has a binomial distribution if it assumes only non negative values and its probability Density function is given by

$$p(x=r)=p(r)=\begin{cases} {}^nC_r\, p^r q^{n-r} & r=0,1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

Mean of binomial distribution $= np$.

Variance of binomial distribution $= npq$

Standard deviation of binomial distribution $= \sqrt{npq}$

Q) If the probability of a defective bolt is $\frac{1}{8}$ find 1) mean 2) variance of the distribution of defective bolts of 640

Given $n = 640$

p = the probability of a defective bolt $= \frac{1}{8}$

1) mean $= np = 640 \times \frac{1}{8} = 80$

2) variance $= npq$ $\quad [q = 1-p]\ [q = 1 - \frac{1}{8} = \frac{7}{8}]\ [p+q=1]$

$$= 640 \times \frac{1}{8} \times \frac{7}{8} = 70$$

Q) In 8 throws of a die 5 or 6 is consider a success find the mean ~~no~~ number of success & the standard Deviation

Given $n = 8$

$$p = p(5) + p(6)$$

$$p = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

Mean $= np = 8 \times \frac{1}{3} = \frac{8}{3}$

Variance $= npq = 8 \times \frac{1}{3} \times \frac{2}{3} = \frac{16}{9}$

S.D $= \sqrt{\text{variance}} = \sqrt{npq} = \sqrt{\frac{16}{9}} = \frac{4}{3}$.

Q) 10 coins are thrown simultaneously. Find the probability of getting atleast 1) 7 heads 2) 6 heads.

Given $n = 10$

$$p = \frac{1}{2}$$

$$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$$

w.k.T $p(x=r) = p(r) = \begin{cases} {}^nC_r\, p^r q^{n-r} & r = 0,1,2 \\ 0 & \text{otherwise} \end{cases}$ —— (i)

Substituting $n=10,\ p=\frac{1}{2},\ q=\frac{1}{2}$ in eq ① we get.

$$P(x=r)=P(r)={}^{10}C_r\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{10-r} \quad —②$$

i) atleast 7 heads i.e, $P(x\geq 7)$.

$$=P(x=7)+P(x=8)+P(x=9)+P(x=10)$$

$$={}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3+{}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2+{}^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)+{}^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0$$

$$=\left(\frac{1}{2}\right)^{10}\left[{}^{10}C_7+{}^{10}C_8+{}^{10}C_9+{}^{10}C_{10}\right]$$

$$=\frac{11}{64}=\frac{\cancel{176}\ 11}{\cancel{1024}\ 64}$$

2) atleast 6 heads i.e, $P(x\geq 6)$

$$=P(x=6)+P(x=7)+P(x=8)+P(x=9)+P(x=10)$$

$$={}^{10}C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4+{}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3+{}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2+{}^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)+{}^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0$$

$$=\left(\frac{1}{2}\right)^{10}\left({}^{10}C_6+{}^{10}C_7+{}^{10}C_8+{}^{10}C_9+{}^{10}C_{10}\right)$$

$$=\frac{193}{512}=\frac{\cancel{386}\ 193}{\cancel{1024}\ 512}$$

Determine the binomial distribution for which mean = 4 Variance = 3

Given mean $=4=np$ — ①

Variance $=npq=3$ — ②

eq $\frac{②}{①} \Rightarrow \frac{\cancel{np}q}{\cancel{np}}=\frac{3}{4}$

$q=\frac{3}{4}$.

$p=1-q=1-\frac{3}{4}=\frac{1}{4}$

$np=4$

$n\left(\frac{1}{4}\right)=4 \Rightarrow n=16$

$q=\frac{1}{3},\ p=\frac{2}{3},\ n=6,\ 0.9986$

1) The Mean & variance of binomial distribution are 4 & 4/3 respective find $P(X\geq 1)$ $\rightarrow P(x\geq 1)=1-P(x<1)=1-P(x=0)$

2) If the incidence of an occupational disease in an industry is such that the workers have 20% chance of suffering from it. what is the probability that out of 6 workers choosen at random, 4 are more will suffer from the disease

0.0175

$P=20=\frac{20}{100}=\frac{1}{5}=0.2 \quad q=1-p=$

$n=6$

$P(x\geq 4)=P(x=4)$

1) G

$\frac{②}{①}$

$\Rightarrow$

$P=$

np

$n($

$\Rightarrow P$

p

q

n

$P($

1) Given, Mean $= np = 4$, —(1)

Variance $= npq = \frac{4}{3}$ —(2)

$$\frac{(2)}{(1)} = \frac{\frac{4}{3}}{4} = \frac{4}{3\times 4} = \frac{1}{3} = \frac{npq}{np}$$

$\Rightarrow \boxed{q = \frac{1}{3}}$

$p = 1 - q = 1 - \frac{1}{3} = \frac{2}{3} \Rightarrow \boxed{p = \frac{2}{3}}$

$np = 4$

$n\left(\frac{2}{3}\right) = 4^2 \Rightarrow \boxed{n = 6}$

WKT

$P(x=r) = {}^nC_r\, p^r q^{n-r}$

$\Rightarrow P(x \geq 1) = 1 - P(x < 1) = 1 - P(x=0)$

$= 1 - \left[{}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^{6-0}\right] = 1 - [1(1)(0.0013717)]$

$= 0.9963$

2) $p = 20\% = \frac{20}{100} = \frac{1}{5} = 0.2$

$q = 1 - p = 1 - 0.2 = 0.8$

$n = 6$

$P(x \geq 4) = 1 - P(x < 4)$

$= 1 - [P(x=0) + P(x=1) + P(x=2) + P(x=3)]$

$= 1 - \left[{}^6C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^6 + {}^6C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^5 + {}^6C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4 + {}^6C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^3\right]$

$= 1 - [0.2621 + 0.3932 + 0.2458 + 0.0819]$

$= 1 - 0.9830$

$= 0.0170$

— * —

Q) A box contains a cards · numbered 1 to 9 if 4 cards are drawn with replacement what is the probability that none is one ·

Given, $n=4$

let p = probability of getting one $= \frac{1}{9}$

$Q = 1-p = 1-\frac{1}{9} = \frac{8}{9}$

WKT $p(x=r) = {}^nc_r\, p^r q^{n-r}$

Hence the probability of getting none is one is given by

$(x=0) = {}^4c_0 \left(\frac{1}{9}\right)^0 \left(\frac{8}{9}\right)^{4-0}$

$n=4,\ r=0,\ p=\frac{1}{9},\ q=\frac{8}{9}$

$P(x=0) = 1(1)\left(\frac{8}{9}\right)^4$

$= 0.624$

Q) the probability of that John hits a target is $\frac{1}{2}$ he fires Six times. Find the probability that he hits the target

1) exactly 2 times

2) more than 4 times

3) atleast once

The probability of getting hits a target $t = \frac{1}{2}$

$q = 1-p = 1-\frac{1}{2} = \frac{1}{2}$ = probability of failure target

$n=6,$

WKT $\boxed{p(x=r) = {}^nc_r\, p^r q^{n-r}}$

1) exactly 2 times

$P(x=2) = {}^6c_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2}$

$= {}^6c_2 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^4$

$= 0.2344$

2) more than 4 times $p(r>4) = p(r=5) + p(r=6)$

$= {}^6c_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{6-5} + {}^6c_5 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^{6-6}$

$= {}^6c_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^1 + {}^6c_5 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^0$

$= \left(\frac{1}{2}\right)^6 [{}^6c_5 + {}^6c_6]$

$= 0.1094$

ds 3) atleast once $p(x \geq 1) = 1 - p(x=0)$

$= 1 - {}^6C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{6-0}$

$= 1 - 1\left(\frac{1}{2}\right)^6 = 1 - \frac{1}{2^6}$

$= 0.9844$

— No Incomplete —

ires

t ✱ Derive Mean & variance of binomial distribution

A) Mean of the binomial Distribution: The binomial probability distribution is given by $P(r) = {}^nC_r p^r q^{n-r}$ $r = 0, 1, 2, 3 \dots n$
$q = 1 - p$.

$$\text{Mean} = \mu = E(x) = \sum_{r=0}^{n} r\,p(r) = \sum_{r=0}^{n} r\,{}^nC_r p^r q^{n-r}$$

$$= 0 + {}^nC_1 p^1 q^{n-1} + 2\,{}^nC_2 p^2 q^{n-2} + 3\,{}^nC_3 p^3 q^{n-3} + \dots + n\,{}^nC_n p^n q^{n-n}$$

$$= n p q^{n-1} + 2\,{}^nC_2 p^2 q^{n-2} + 3\,{}^nC_3 p^3 q^{n-3} + \dots + n p^n \quad \boxed{p^n = p^{n-1} p}$$

$$= n p q^{n-1} + \frac{2\,n(n-1)}{2!} p^2 q^{n-2} + \frac{3\,n(n-1)(n-2)}{3!} p^3 q^{n-3} + \dots + n p^{n-1} p$$

$$= n p q^{n-1} + n(n-1) p^2 q^{n-2} + \frac{n(n-1)(n-2)}{2} p^3 q^{n-3} + \dots + n p^{n-1} p.$$

$$= np\left[q^{n-1} + (n-1) p q^{n-2} + \frac{(n-1)(n-2)}{2} p^2 q^{n-3} + \dots + p^{n-1}\right]$$

$$= np(q+p)^{n-1}$$

$$\left[(q+p)^{n-1} = q^{n-1} + (n-1)pq^{n-2} + \frac{n(n-1)(n-}{2} p^2 q^{n-3} + \dots + p^{n-1}, \text{ Binomial theorem}\right]$$

$$= np(1)^{n-1}$$

$$= np(1) = np$$

$$\therefore \boxed{\text{Mean} = E(x) = \mu = np}$$

$$\text{Variance} = V(x)\,\sigma^2 = E(x^2) - (E(x))^2 = \sum_{r=0}^{n} r^2 p(r) - \mu^2$$

$$= \sum_{r=0}^{n} (r(r-1) + r)\,p(r) - \mu^2$$

$$= \sum^{n} r(r-1)p(r) + \sum^{n} r\,p(r) - \mu^2$$

$= \left(2(2-1)\,{}^nC_2\, p^2 q^{n-2} + 3(3-1)\,{}^nC_3\, p^3 q^{n-3} + \cdots + n(n-1)\,{}^nC_n\, p^n q^{n-n}\right) + \mu - \mu^2$

$= \left[\cancel{2} \cdot \frac{n(n-1)}{\cancel{2!}} p^2 q^{n-2} + \cancel{3 \cdot 2}\, \frac{n(n-1)(n-2)}{\cancel{3!}} p^3 q^{n-3} + \cdots + n(n-1)p^n\right) + \mu - \mu^2$

$= n(n-1)p^2\left[q^{n-2} + (n-2)p\, q^{n-3} + \cdots + p^{n-2}\right] + \mu - \mu^2$

formula $(q+p)^{n-2}$] Binomial theorem

$= n(n-1)p^2(q+p)^{n-2} + \mu - \mu^2$

w.k.T $q+p=1 \Rightarrow q = 1-p$, $\mu = np$

$= n(n-1)p^2(1)^{n-2} + np - n^2p^2$

$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$

$= np[1-p]$

$V(x) = \sigma^2 = npq$

$S.D = \sigma = \sqrt{npq}$

Poisson Distribution

Definition: A Random variable 'x' is said to be a poisson Distribution if it assumes only non-negative values & its probability density function is given by

$$P(x,\lambda) = P(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 0,1,2\ldots \\ 0 & \text{otherwise} \end{cases}$$

Derive Mean and Variance of poisson distribution

1) Mean of the poisson distribution:

$$\text{Mean} = E(x) = \sum_{x=0}^{\infty} x\,p(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{w.k.T } P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty} \frac{x\,\lambda^x}{x!}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty} \frac{\cancel{x}\,\lambda^x}{\cancel{x}(x-1)!}$$

Let $x-1=y$

$x = y+1$

$x=1$, $1-1=y$, $0=y$, $y=0$ | $\infty - 1 = y$, $\infty = y$

$$= e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}$$

$$= e^{-\lambda}\sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!}$$

$$= e^{-\lambda}\sum_{y=0}^{\infty} \frac{\lambda^y\lambda^1}{1!} = e^{-\lambda}\lambda\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} \qquad \lambda^0 = 0! = 1$$

$$= e^{-\lambda}\left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots + \cdots\right]$$

$$= \lambda e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right]$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$= \lambda e^{-\lambda}[e^{\lambda}] = \lambda e^{-\lambda+\lambda} = \lambda e^0 = \lambda.$$

2) Variance of poisson Distribution:

$V(x) = E(x^2) - (E(x))^2 = \sum_{x=0}^{\infty} x^2 p(x) - \lambda^2$ $\boxed{E(x) = \lambda}$

$= \sum_{x=0}^{\infty} [x(x-1)\cancel{p(x)} + x] \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2$

$= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2$

$= \sum_{x=0}^{\infty} \frac{\cancel{x(x-1)}\, e^{-\lambda}\lambda^x}{\cancel{x(x-1)}(x-2)!} + \sum_{x=0}^{\infty} \frac{\cancel{x}\, e^{-\lambda}\lambda^x}{\cancel{x}(x-1)!} - \lambda^2$

$= e^{-\lambda}\left[\sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}\right] - \lambda^2$

put $x-2=y$ $\boxed{x=y+2}$, $x-1=z$ $\boxed{x=z+1}$

$= e^{-\lambda}\left[\sum_{y=0}^{\infty} \frac{\lambda^{y+2}}{y!} + \sum_{z=0}^{\infty} \frac{\lambda^{z+1}}{z!}\right] - \lambda^2$

$x=2,\ y=0$; $x=1 \Rightarrow z=0$

$= e^{-\lambda}\left[\sum_{y=0}^{\infty} \frac{\lambda^y \lambda^2}{y!} + \sum_{z=0}^{\infty} \frac{\lambda^z \lambda^1}{z!}\right] - \lambda^2$

$x=\infty \Rightarrow y=\infty$; $x=\infty, z=\infty$

$\left[\sum \frac{y^\lambda}{y!} = \sum \frac{\lambda^z}{z!} = e^\lambda\right]$

$= e^{-\lambda}\cdot[e^\lambda \lambda^2 + e^\lambda \lambda] - \lambda^2$

$= e^{-\lambda}e^{\lambda}[\lambda^2 + \lambda] - \lambda^2$

$= e^0(\cancel{\lambda^2} + \lambda - \cancel{\lambda^2})$

$\boxed{V(x) = \lambda}$

$\therefore$ Mean $= E(x) = \lambda$
Variance $= V(x) = \lambda$ $\}$ P.D.

Q) Find Recurrence Relation for the poisson Distribution (P.D)

W.K.T $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$p(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}$ $\quad (x+1)! = (x+1)x!$

$p(x+1) = \frac{e^{-\lambda}\lambda^x \lambda}{(x+1)x!} = \frac{\lambda}{x+1}\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)$

$= \frac{\lambda}{x+1}\cdot p(x)$

$p(x+1) = \frac{\lambda}{x+1}\cdot p(x)$

Q) The Average number of accidents on a national highway is $\lambda = 1.6$, $\lambda = 1.8$. Determine the probability that the no. of accidents are

a) W.K.T $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Mean $= \lambda = 1.8$

$P(x) = \frac{e^{-1.8}(1.8)^x}{x!}$

1) $P(\text{atleast one}) = P(x \geq 1)$

$= 1 - P(x < 1)$

$= 1 - P(x=0)$

$= 1 - \frac{e^{-1.8}(1.8)^0}{0!}$

$= 1 - e^{-1.8}$

$= 0.8347$

2) $P(\text{atmost one}) = P(x \leq 1)$

$= P(x=0) + P(x=1)$

$= \frac{e^{-1.8}(1.8)^0}{0!} + \frac{e^{(-1.8)}(1.8)^1}{1!}$

$= e^{-1.8} + e^{-1.8}(1.8)$

$= e^{-1.8}(1+1.8)$

$= e^{-1.8}(2.8)$

$= 0.4628$

b) W.K.T $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Mean $= \lambda = 1.6$

1) $P(\text{atleast one}) = P(x \geq 1)$

$= 1 - P(x<1) = 1 - P(x=0)$

$= 1 - \frac{e^{-1.6}(1.6)^0}{0!}$

$= 1 - e^{-1.6}$

$= 0.7981$

2) $P(\text{atmost one}) = P(x \leq 1)$

$= P(x=0) + P(x=1)$

$= \frac{e^{-1.6}(1.6)^0}{0!} + \frac{e^{(-1.6)}(1.6)^1}{1!}$

$= 0.2018 + 0.3230$

$= 0.5248$

Q) A hospital switch board receives an average of 4 emergency calls in a 10min interval. What is the probability that 1) there are atmost 2 emergency calls in a 10minute interval.

2) There are exactly '3' in a 10 min interval. -2010

W.K.T

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Given $\lambda = 4$

$P(x) = \frac{e^{-4}(4)^x}{x!}$

1) $P(\text{at most 2 calls}) = P(x \leq 2)$

$= P(x=0) + P(x=1) + P(x=2)$

$= \frac{e^{-4}(4)^0}{0!} + \frac{e^{-4}(4)^1}{1!} + \frac{e^{-4}(4)^2}{2!}$

$= 0.0183 + 0.0732 + 0.1465 = 0.2381$

2) $P(\text{exactly 3 calls}) = P(x=3)$

$= \frac{e^{-4}4^3}{3!}$

$= 0.1953$

If a Random variable has a p.d.f such that $p(1)=p(2)$ find

1) Mean of the distribution

2) $p(4)$

3) $p(x \geq 1)$ 4) $p(1 < x < 4)$

1) Given $p(1)=p(2) \Rightarrow \frac{p(2)}{p(1)} = 1$

W.K.T the regression relation

$$\boxed{p(x+1) = \frac{\lambda}{x+1} p(x)}$$

substituting $x=1$ in above eq.

$$p(2) = \frac{\lambda}{2} p(1)$$

$$\frac{p(2)}{p(1)} = \frac{\lambda}{2} \Rightarrow \frac{1}{1} = \frac{\lambda}{2} \Rightarrow \boxed{\lambda=2}$$

2) $p(4)$

WKT $p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$$= \frac{e^{-2}(2)^4}{4!} = 0.0902$$

3) $p(x \geq 1) = 1 - p(x=0)$

$$= 1 - \frac{e^{-\lambda}\lambda^{0}}{0!} = 1 - e^{-2} = 0.8646$$

4) $p(1 < x < 4) = p(x=2) + p(x=3)$

$$= \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!}$$

$$= \frac{e^{-2}(2)^2}{2!} + \frac{e^{-2}(2)^3}{3!}$$

$$= 0.4511.$$

H.W

Q) 2% of the items of a factory are defective. The items are packed in boxes. what is the probability there will be a) 2 defective items b) atleast 3 defective items in a box of hundred items

Given $n=100$, $p = 2\% = \frac{2}{100}$

Mean $\lambda = np = 100 \times \frac{2}{100} = 2$

w.K.T $p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

a) $p(x=2) = \frac{e^{-2}(2)^2}{2!}$

$$= 0.2706$$

b) $p(\text{atleast } 3) = p(x \geq 3)$

$$= 1 - p(x<3)$$

$$= 1 - [p(x=0)+p(x=1)+p(x=2)]$$

$$= 1 - \left[\frac{e^{-2}(2)^0}{0!} + \frac{e^{-2}(2)^1}{1!} + \frac{e^{-2}(2)^2}{2!}\right]$$

$$\cancel{= 1 - [0.1353 + 0.1831 + 0.2706]}$$

$$= 0.3233$$

Q) Suppose 2% of people on the average are lefthanded. find the probability 3 (or) more lefthanded

b) the probability of finding none (or) 1 left handed.

Given Mean $= \lambda = 2\% = \frac{2}{100} = 0.02$

WKT $p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$$p(x) = \frac{e^{-0.02}(0.02)^x}{x!}$$

a)

$p(x \geq 3) = 1 - p(x < 3)$

$$= 1 - [p(x=0) + p(x=1) + p(x=2)]$$

$$= 1 - \left[\frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!} + \frac{e^{-0.02}(0.02)^2}{2!}\right]$$

$$= 1.3077 \times 10^{-16} = 0.0000013$$

b) $p(x=0) + p(x=1)$

none or one

$$= \frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!}$$

$$= 0.9998$$

H.W

Q) The Average no. of phone calls per minute coming into a switch board b/w 2PM & 5PM is 2.5. Determine the probability during 1 particular minute there will be

1) 4 or fewer - 0.8912
2) more than 6 calls - 0.01416

Given $\lambda = 2.5$

WKT $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

1) $P(x \le 4) = P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$= \frac{e^{-2.5}(2.5)^0}{0!} + \frac{e^{-2.5}(2.5)^1}{1!} + \frac{e^{-2.5}(2.5)^2}{2!} + \frac{e^{-2.5}(2.5)^3}{3!} + \frac{e^{-2.5}(2.5)^4}{4!}$

$= e^{-2.5}\left[1 + 2.5 + \frac{(2.5)^2}{2!} + \frac{(2.5)^3}{3!} + \frac{(2.5)^4}{4!}\right]$

$= 0.8912$

2) $P(x \ge 6) = 1 - P(x < 6)$

Q) If a poisson distribution is such that $P(x=1)\cdot\frac{3}{2} = P(x=3)$ find 1) $P(x \ge 1)$ 2) $P(x \le 3)$ 3) $P(2 \le x \le 5)$

Given $P(x=1)\cdot\frac{3}{2} = P(x=3)$

We know that

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$\frac{e^{-\lambda}\lambda^1}{1!}\cdot\frac{3}{2} = \frac{e^{-\lambda}\lambda^3}{3!}$

$\frac{3}{2} = \frac{\lambda^2}{6} \Rightarrow 18 = 2\lambda^2$

$\lambda^2 = 9 \Rightarrow \lambda = \pm 3$

$\lambda = +3, -3$

$\therefore \boxed{\lambda = 3}$ [Take non negative values $\lambda > 0$]

1) $P(x \ge 1) = 1 - P(x < 1) = 1 - P(x=0)$

$= 1 - \frac{e^{-3}3^0}{0!}$

$= 1 - e^{-3} = 0.9502$

2) $P(x \le 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$

$= \frac{e^{-3}3^0}{0!} + \frac{e^{-3}(3^1)}{1!} + \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!}$

$= e^{-3}\left[1 + 3 + \frac{9}{2} + \frac{27}{6}\right]$

$= 0.6472$

3) $P(2 \le x \le 5) = P(x=2) + P(x=3) + P(x=4) + P(x=5)$

$= \frac{e^{-3}(3)^2}{2!} + \frac{e^{-3}(3)^3}{3!} + \frac{e^{-3}(3)^4}{4!} + \frac{e^{-3}(3)^5}{5!}$

$= e^{-3}\left[\frac{9}{2} + \frac{27}{6} + \frac{81}{24} + \frac{243}{120}\right]$

$= 0.7169$

H.W

If x is a poisson variate such that $3P(x=4) = \frac{1}{2}P(x=2) + P(x=0)$

find 1) The Mean of x - 2011

2) $P(x \le 2)$

Given

$3P(x=4) = \frac{1}{2}P(x=2) + P(x=0)$

WKT, $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$3\frac{e^{-\lambda}\lambda^4}{4!} = \frac{1}{2}\frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^0}{0!}$

$\frac{3e^{-\lambda}\lambda^4}{24} = \frac{1}{4}e^{-\lambda}\left[\frac{\lambda^2}{4} + 1\right]$

$$\frac{\lambda^4}{8} = \frac{\lambda^2}{4} + 1$$

$$\frac{\lambda^4}{8} - \frac{\lambda^2}{4} - 1 = 0$$

$$\frac{\lambda^4 - 2\lambda^2 - 8}{8} = 0 \Rightarrow \lambda^4 - 2\lambda^2 - 8 = 0$$

$\lambda^4 - 2\lambda^2 - 8 = 0$

$(\lambda^2 + 2)(\lambda^2 - 4) = 0$

$\lambda^2 + 2 = 0$ $\quad\quad$ $\lambda^2 - 4 = 0$

$\lambda^2 = -2 = i^2(\sqrt{2})^2$ $\quad\quad$ $\lambda^2 = 4$

$\lambda = \pm i\sqrt{2}$ $\quad\quad$ $\lambda = \pm 2$

$\lambda = i\sqrt{2}, -i\sqrt{2}$ $\quad\quad$ $\lambda = 2, -2$

$\boxed{\lambda = 2}$

2) $p(x \le 2) = p(x=0) + p(x=1) + p(x=2)$

$$= \frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}$$

$= e^{-2}[1 + 2 + 2] = 0.6767$

4. If 2% of light bulbs are defective find

1) atleast 1 is defective ($p(x \ge 1)$) 0.8646

2) exactly 7 are defective $p(x=7)$ 0.0034

3) $p(1 < x < 8)$ in a sample of 100 0.59

Sol: Given $n = 100$, $p = 2\% = \frac{2}{100}$

$\lambda = np = 100 \times \frac{2}{100} = 2$

$\boxed{\lambda = 2}$, WKT $= p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

1) $p(\text{atleast 1 defective}) = p(x \ge 1)$

$= 1 - p(x < 1) = 1 - p(x=0)$

$$= 1 - \frac{e^{2}(2)^0}{0!} = 1 - e^{2} = 0.8646$$

2) $p(\text{exactly } 7) = p(x=7)$

$$= \frac{e^{-2}(2)^7}{7!} = 0.00347$$

3) $p(1 < x < 8) = \cancel{p(x=1)} + p(x=2) + p(x=3) + p(x=4) + p(x=5) + p(x=6) + p(x=7)$

$$\Rightarrow \frac{e^{-2}(2)^2}{2!} + \frac{e^{-2}(2)^3}{3!} + \frac{e^{-2}(2)^4}{4!} + \frac{e^{-2}(2)^5}{5!} + \frac{e^{-2}(2)^6}{6!} + \frac{e^{-2}(2)^7}{7!}$$

$$\Rightarrow e^{-2}\left[2 + 1.3 + \frac{16}{24} + \frac{32}{120} + \frac{64}{720} + \frac{128}{5040}\right]$$

$= 0.5883$

Q) Fit a ~~posio~~ poisson distribution to the following data.

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{0 + 156 + 138 + 81 + 20 + 5}{400}$$ $\boxed{\Sigma f = 400}$

Mean $= \lambda = 1$

Hence, the theoritical frequency for 'x' success is given by $\boxed{Np(x)}$

where $x = 0, 1, 2, 3, 4, 5$.

$N = \Sigma f = 400$

$$Np(x) = 400\frac{e^{-\lambda}\lambda^x}{x!}$$

where $x = 0, 1, 2, 3, 4, 5$ and $\boxed{\lambda = 1}$

$$\Rightarrow 400\frac{e^{-1}1^0}{0!}, 400\frac{e^{-1}1^1}{1!}, 400\frac{e^{-1}1^2}{2!}, 400\frac{e^{-1}1^3}{3!}, 400\frac{e^{-1}1^4}{4!}, 400\frac{e^{-1}1^5}{5!}$$

$\Rightarrow 400e^{-1}, 400e^{-1}, 200e^{-1}, 66.67e^{-1}, 16.67e^{-1}, 3.33e^{-1}$

$\Rightarrow 147.15, 147.15, 73.57, 24.52, 6.1325, 1.225$

The theoretical frequencies are

x	0	1	2	3	4	5
f	147	147	73	25	6	1

Fit a poisson distribution to the following data

x	0	1	2	3	4	5	6	7
f	305	365	210	80	28	9	2	1

Mean $= \frac{\Sigma fx}{\Sigma f}$

$= \frac{0+365+420+240+112+45+12+7}{1000}$

$= \frac{1201}{1000} = 1.201$

$\boxed{\text{Mean} = \lambda = 1.201}$

Here the theoritical fequency for "x" successes is given by $\boxed{Np(x)}$

where $x = 0,1,2,3,4,5,6,7$

$N = \Sigma f = 1000$

$Np(x) = 1000 \frac{e^{-\lambda}\lambda^x}{x!}$

where $x = 0,1,2,3,4,5,6,7$ & $\boxed{\lambda = 1.201}$

$\Rightarrow 1000 \frac{e^{1.201}(1.201)^0}{0!}, 1000\frac{e^{1.201}(1.201)^1}{1!}, 1000\frac{e^{1.201}(1.201)^2}{2!}, 1000\frac{e^{1.201}(1.201)^3}{3!}, 1000\frac{e^{1.201}(1.201)^4}{4!}, 1000\frac{e^{1.201}(1.201)^5}{5!}, 1000\frac{e^{1.201}(1.201)^6}{6!}, 1000\frac{e^{1.201}(1.201)^7}{7!}$

$\Rightarrow 300.89, 361.37, 271.00, 86.87, 26.08, 6.26$

The theortical frequencies are

x	
f	

Normal Distribution:

A Random variable x is said to have Normal distribution if its density of probability distribution is give by

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$

where μ is mean and σ is standard deviation of x

Derive Mean and variance of normal distribution

Mean of normal distribution

consider the normal distribution with b, σ as the parameters

w.k.T $f(x,b,\sigma) = \frac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$

Mean $= \mu = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$

$\mu = \int_{-\infty}^{\infty} x \frac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\,dx$

let $z = \frac{x-b}{\sigma}$ $dz = \frac{dx}{\sigma} \Rightarrow dx = \sigma dz$

$x = b + \sigma z$

$\mu = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (b+\sigma z) \frac{e^{-\frac{z^2}{2}}}{\cancel{\sigma}} \cancel{\sigma} dz$

$\mu = \frac{b}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} e^{-z^2/2}\,dz}_{\text{even}} + \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} z e^{-z^2/2}\,dz}_{\text{odd}}$

$\left[\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx \quad f(x) \text{ is even}\right]$

$\int_{-a}^{a} f(x)\,dx$. The function is odd it takes integral 0

$= \frac{b}{\sqrt{2\pi}} 2\int_0^{\infty} e^{-z^2/2}\,dz + 0$

wkT $\int_0^{\infty} e^{-z^2/2}\,dz = \sqrt{\frac{\pi}{2}}$

$= \frac{b}{\sqrt{2\pi}} 2\sqrt{\frac{\pi}{2}}$

$= \frac{2b\sqrt{\pi}}{\sqrt{2}\sqrt{\pi}\sqrt{2}}$

$\boxed{\mu = E(x) = b}$

Variance of normal distribution

WKT

$f(x,b,\sigma) = \frac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$

Variance $= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

let $z = \frac{x-b}{\sigma}$

Here $\mu = \text{mean} = b$

$\boxed{x - b = \sigma z}$

$= \int_{-\infty}^{\infty} (x-b)^2 \frac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\,dx$

$dz = \frac{dx}{\sigma}$

$dx = \sigma dz$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z^2) e^{-\frac{z^2}{2}} \frac{\cancel{\sigma}}{\cancel{\sigma}} dz$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{z^2 e^{-z^2/2}}_{\text{even}} dz$$

let $\frac{z^2}{2} = t$, $z^2 = 2t$, $2z\,dz = 2\,dt$, $dz = \frac{dt}{z}$, $\boxed{dz = \frac{dt}{\sqrt{2t}}}$

$$= 2\frac{\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} 2t e^{-t} \frac{dt}{\sqrt{2t}}$$

$$= 2\frac{\sigma^2}{\sqrt{2}\sqrt{\pi}} \int_0^{\infty} \cancel{2}t e^{-t} \frac{t^{-1/2}}{\sqrt{2}} dt$$

$$= 2\frac{\sigma^2}{\sqrt{\pi}} \int_0^{\infty} e^{-t} t^{1/2} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} e^{-t} t^{\frac{3}{2}-1} dt$$

$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$

$\Gamma(\frac{3}{2}) = \int_0^{\infty} e^{-t} t^{3/2-1} dt$

$\frac{1}{2} = \frac{3}{2} - 1$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$\Gamma(n+1) = n\Gamma(n)$

W.K.T $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$

$$= \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \frac{1}{\cancel{2}} \cancel{\sqrt{\pi}}$$

Variance $= \sigma^2$

Mode of normal distribution:

Mode is the value of x for which $f(x)$ is maximum. i.e., Mode is the solution of

$f'(x) = 0$ & $f''(x) < 0$

W.K.T

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \quad \text{--- (1)}$$

Differentiating on b.s w.r to x we get

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} \left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \frac{-1}{2} 2\left(\frac{x-\mu}{\sigma}\right)\right]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot -\left(\frac{x-\mu}{\sigma}\right)$$

$$f'(x) = -f(x)\left(\frac{x-\mu}{\sigma}\right)$$

$$f'(x) = 0 \quad -f(x)\left(\frac{x-\mu}{\sigma}\right) = 0$$

$$\frac{x-\mu}{\sigma} = \frac{0}{f(x)} = 0$$

$$x - \mu = \sigma(0) \Rightarrow x - \mu = 0 \Rightarrow \boxed{x = \mu}$$

$$\text{Ily } f''(x) = \frac{-f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

sub $x = \mu$ in $f''(x)$

Then $f''(x) = \frac{-f(x)}{\sigma^2}\left[1 - \frac{(\mu-\mu)^2}{\sigma^2}\right] = \frac{-f(x)}{\sigma^2}$

$$f''(x) = \frac{-f(x)}{\sigma^2} < 0$$

Hence $\boxed{x = \mu}$ is the mode of the normal distribution.

Median of normal distribution

If M is the median of the normal distribution, we have

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2}$$

i.e., $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$

i.e., $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$ --- (1)

consider $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

$x - \mu = \sigma z$,

put $\frac{x-\mu}{\sigma} = z$ then $dx = \sigma dz$

$$\therefore \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{z^2}{2}} dz$$

(by symmetry) $= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\pi}{2}} = \frac{1}{2}$ --- (2)

$\therefore$ from (1) & (2), we have

$$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0$$

$$\Rightarrow \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0$$

$\Rightarrow \mu = M$ [$\because$ if $\int_a^b f(x)\,dx = 0$ then $a = b$, where $f(x) > 0$]

Median $M = \mu$

$\therefore$ Mean = Median = Mode

Q) The Mean and standard deviation of the marks obtained by 1000 students in an examination or respectively 34.5 & 16.5. Assuming the normality of the distribution find the approximate no. of students expected to obtain marks between 30 and 60

Given mean $\mu = 34.5$

$\sigma = 16.5$

WKT $z = \frac{x-\mu}{\sigma}$

When $x = 30 \Rightarrow z = \frac{30-34.5}{16.5}$

$z = -0.27\cancel{27}$ z_1 say

when $x = 60$

$z = \frac{60-34.5}{16.5} = 1.54\cancel{54}$ (z_2 say)

$P(30 \le x \le 60) = p(z_1 \le x \le z_2)$

$= A(z_1) + A(z_2)$

$= A(-0.27) + A(1.54)$

$= A(0.27) + A(1.54)$

from Tables

$A(0.27) = 0.1084$ [Area is alway true]

$A(1.54) = 0.4382$ $= 0.1084 + 0.4382$

$= 0.5466$

The no. of students who get marks between 30 and 60

$= 0.5466 \times 1000$

$= 546.6$

$= 547$

Hence 547 Students get marks between 30 and 60.

Q) If x is a normal variate with Mean 30 & S.D 5 find 1) $P(26 \le x \le 40)$ 2) $P(x \ge 45)$

Given Mean $\mu = 30$,

S.D $\sigma = 5$.

WKT $z = \frac{x-\mu}{\sigma}$

1) When $x = 26 \Rightarrow z = \frac{26-30}{5}$

$z = -0.8$ z_1 say

when $x = 40 \Rightarrow z = \frac{40-30}{5} = 2$ z_2 say

$P(26 \le x \le 40) = P(z_1 \le z \le z_2)$

$= A(z_1) + A(z_2)$

$= A(-0.8) + A(2)$ $A(0.8) = 0.2881$ } from

$= A(0.8) + A(2)$ $A(2) = 0.4772$ } table

$= 0.2881 + 0.4772$

$= 0.7653$

2) $P(x \ge 45)$

when $x = 45$

$z = \frac{x-\mu}{\sigma} = \frac{45-30}{5} = \frac{15}{5} = 3$

$P(x \ge 45) = P(z_1 \ge 45)$

$= 0.5 - A(z_1)$

$= 0.5 - A(3)$ [$A(3) = 0.4987$ from Table]

$= 0.5 - 0.4987$

$= 0.0013$

Q) If x is a normal variant Find the area A.

1) To the left of $z = -1.78$ (−)

2) $z = -1.45$ to the right (+)

3) corresponding to $-0.8 \le z \le 1.53$

4) To the left of $z = -2.52$ & to the right of $z = 1.83$

A) 1) Required Area

$= 0.5 - \text{Area}(0 \text{ To } -1.78)$

$= 0.5 - \text{Area}(0 \text{ To } -1.78)$ (area is true)

$= 0.5 - 0.4625$ (from Tables)

$= 0.0375$

2) Required Area A

$= 0.5 + \text{Area}(0 \text{ TO } -1.45)$

$= 0.5 + \text{Area}(0 \text{ to } 1.45)$

$= 0.5 + 0.4265$ (from Table)

$= 0.9265$.

3) Required Area $= A(z_2) + A(z_1)$

$= A(1.53) + A(-0.8)$

$= A(1.53) + A(0.8)$

$= 0.437 + 0.2881 = 0.7251$

4) Required area $= [0.5 - \text{Area from } 0 \text{ to } 2.52] + [0.5 + \text{Area from } 0 \text{ to } 1.83]$

$= (0.5 - 0.4941) + (0.5 + 0.4664)$

$= 0.0059 + 0.9664$

$= 0.9723$

Q) In a sample of 1000 cases the mean of a certain test 14 and S.D is 2.5 assuming the distribution to be normal find

i) how many student score b/w 12 & 15.

ii) how many score above 18.

iii) how many score below 18.

Given, $\mu = 14$

S.D $\sigma = 2.5$

WKT $z = \frac{x-\mu}{\sigma}$

when $x = 12$, $z = \frac{12-14}{2.5} = (-0.8)$ (z_1 say)

$x = 15$, $z = \frac{15-14}{2.5} = 0.4$ (z_2 say)

$P(12 < x < 15) = P(z_1 < z < z_2)$

$= A(z_1) + A(z_2)$

$= A(-0.8) + A(0.4)$

$= A(0.8) + A(0.4)$

$= 0.2881 + 0.1554$

$= 0.4435$

Hence the no. of students score b/w 12 and 15

$= 1000 \times 0.4435$.

$= (443.5) = (444)$ or (443)

2) when $x = 18$ $z = \frac{x-\mu}{\sigma}$

$= \frac{18-14}{2.5} = 1.6$

$P(x > 18) = P(z > 1.6)$

$= 0.5 - A(1.6)$

$= 0.5 - 0.4452$

$= 0.0548$

Hence the no. of students scored above $18 = 1000 \times 0.0548$

$= 54.8$

$= 54$ (or) 55

3) when $x = 18$ $z = \frac{18-14}{2.5} = 1.6$

$P(x < 18) = P(z < 1.6)$ (less than (+))

$= 0.5 + A(1.6)$

$= 0.5 + 0.4452$

$= 0.9452$

Hence the no. of students scored below 18 is 1000×0.9452

$= 945.2$

Q) In a normal distribution, 7% of the items are under 35 & 89% are under 63. Determine the mean & variance of normal distribution.

let μ be the mean and σ is the SD of the normal curve.

7% of items are under 35 and 89% are under 63.

Given $P(x < 35) = 0.07$ ($7\% = \frac{7}{100} = 0.07$)

$P(x < 63) = 0.89$

$\therefore P(X>63) = 1-P(X<63) = 1-0.89 = 0.11$

when $z = \frac{x-\mu}{\sigma}$

when $x = 35.2 = \frac{35-\mu}{\sigma} = z_1 \rightarrow (1)$

$x = 63 \quad z = \frac{x-\mu}{\sigma} = \frac{63-\mu}{\sigma} = z_2 \rightarrow (2)$

from the figure we have

$\left.\begin{array}{l} P(0<z<z_2) = 0.39 \quad z_2 = 1.23 \\ P(0<z<z_1) = 0.43 \quad z_1 = 1.48 \end{array}\right\}$ from Tables

from (1)

$\frac{35-\mu}{\sigma} = -1.48 \rightarrow (3)$

from (2)

$\frac{63-\mu}{\sigma} = 1.23 \rightarrow (4)$

$(4)-(3) \Rightarrow \frac{63-\cancel{\mu}}{\sigma} - \frac{35-\cancel{\mu}}{\sigma} = 1.23+1.48$

$\frac{63-35}{\sigma} = 2.71$

$\frac{28}{\sigma} = 2.71 \Rightarrow \frac{28}{2.71} = \sigma$

$\boxed{\sigma = 10.332}$

from eq (3)

$\frac{35-\mu}{10.332} = (-1.48)$

$35-\mu = (-1.48)(10.332)$

$= -15.29$

$= 35+15.29 = \mu$

$\boxed{\mu = 50.29}$

Mean $= \mu = 50.29$

S.D $= \sigma = 10.332$

Variance $= \sigma^2 = (10.332)^2$

$= 106.75$

Normal approximation of the binomial distribution (or) Approximation of the binomial distribution to Normal

Find the probability of getting an even number on face 3 to 5 times in throwing 10 dies together.

Let $x_1 = 3$

$x_2 = 5, n = 10$

p = probability of getting an even number on face

$= (2,4,6)$

$P(2,4,6) = \frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

$q = 1-p = 1-\frac{1}{2} = \frac{1}{2}$

Mean $= \mu = np = 10 \times \frac{1}{2} = 5$

Variance $= npq = 10 \times \frac{1}{2} \times \frac{1}{2} = \frac{5}{2}$

S.D $= \sigma = \sqrt{5/2} = 1.58$

WKT

$z_1 = \frac{(x_1-\frac{1}{2})-\mu}{\sigma} = \frac{3-\frac{1}{2}-5}{1.58}$

$= -1.50$

$z_2 = \frac{(x_2+\frac{1}{2})-\mu}{\sigma}$

$= \frac{5+\frac{1}{2}-5}{1.58} = 0.32$

Hence the required probability

$= \int_{x_1}^{x_2} \phi(z)\,dz = \int_{-1.58}^{0.32} \phi(z)\,dz$

$= P(-1.58 \le z \le 0.32)$

$= A(0.58) + A(0.32)$

$= 0.4429 + 0.1256$

$= 0.5689$

19/2/25

Estimation and Testing of Hypothesis (Large Sample test)

Estimate: An estimate is a statement made to find an unknown population parameters.

Estimator: The procedure or rule to determine an unknown population parameter is called an estimator.

Types of estimation:

Basically there are two kinds of estimates to determine the statistic of the population parameters namely
1) point estimation 2) Interval estimation

1) point estimation: If an estimate of population parameter is given by a single value then the estimate is called a point estimation of the parameter.

2) Interval estimation: The estimate of the population parameter is given by 2 different values between which the parameter may be considered to lie. Then the estimate is called an interval estimation of the parameter.

Ex: If the height of a student is measured as 162 cm then the measurement gives a point estimation. But if the height is given as (16.3 ± 3.5) cm then the height lies between 159 and 166.5 cm and the measurement gives an interval Estimation.

Statistical Estimation:

W.A Spur and C.P Bonini defined statistical infurence has the process by which we draw a conclusion about some measure of a population based on a sample value. The measure might a variable such as the mean S.D etc. The purpose of sampling is to estimate some characteristics for the population from which the sample is selected.

Unbiased estimator: A statistic or point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if $E(\hat{\theta}) = \theta$

Biased estimator: The statistic 't' is a biased estimator of 'θ' and the biased estimator is $E(t) - \theta$

2m Confidence interval: If an interval estimation of the population parameter θ. If we can find two quantities the populations such that the ... parameter θ is included in the interval $[t_1, t_2]$ in a specified percentage of cases then these interval is called a confidence interval for the parameter 'θ'.

Formulae:

*Confidence limits for 95% for the population mean 'μ' are respectively $\bar{x} \pm 1.96$

* 99% confidence limits are $\bar{x} \pm 2.58$

Testing of Hypothesis for large Sample test

Testing of Hypothesis: we need to decide whether to accept or reject a statement about the parameter. This statement is called a hypothesis and the decision making procedure about the hypothesis is called hypothesis testing.

Procedure of testing a hypothesis:

There are two types of hypothesis

1) Null Hypothesis

2) Alternative Hypothesis

1) Null Hypothesis: For applying the test of significance we first set up a hypothesis a definite statement about the population parameter such a hypothesis is usually a hypothesis of no difference is called Null Hypothesis

*It is in the form $H_0: \mu = \mu_0$

2) Alternative Hypothesis: Any hypothesis which contradicts the null Hypothesis is called an alternative Hypothesis.

*It is denoted by H_1

*It is in the form $H_1: \mu \neq \mu_0$

$H_1: \mu > \mu_0$

$H_1: \mu < \mu_0$

2m Level of Significance: The level of significance is denoted by α is the confidence with which we reject or accept the null Hypothesis H_0 i.e., It is maximum possible Probability with which we are willing to risk an error is rejecting H_0 when it is true.

* Errors of Sampling

There are two types of Errors

1) Type 1 Error

2) Type 2 Error

1) Type 1 Error: Reject H_0 when it is true.

If the null Hypothesis H_0 is true. But it rejected by test procedure. when the error made is called Type 1 Error

2) Type 2 Error: Accept H_0 when it is False

i.e., accept H_0 when H_1 is true. If the null Hypothesis is false but it is accepted by Test then the Error commited as called Type 2 Error

Critical Region: A region corresponding to a statistic 't' in the Sample Space 'S' when leads to the rejection of H_0 is called critical Region / rejection region

Right tailed test: $H_1: \mu > \mu_0$

left tailed test: $H_1: \mu < \mu_0$

One Tailed Test:

If Right tailed Test $H_1: \mu > \mu_0$ (or) Left tailed Test $H_1: \mu < \mu_0$

Two-Tailed test:

If $H_1: \mu \neq \mu_0$ is called Two Tailed test.

Sampling distribution:

Population and sample: population is the aggregate or totality of statistical data forming a subject of investigation. for example,

1) The population of the heights of Indians.

Classification of samples:

Samples are classified into two ways.

1) Large sample: If the size of the Sample $n \geq 30$ then the sample is called large sample.

2) Small sample: If the size of the sample $n < 30$ then the sample is called small sample.

Sampling distribution

Sampling theory is a study of relationship b/w a population & samples are drawn from the population and it is applicable to random sample.

Sampling theory also useful in testing of hypothesis & significance which is important in the theory of decisions.

Central limit theorem:

If $\bar{x}$ be the mean of a sample size n drawn from a population with mean 'μ' and S.D 'σ' then the standardized sample mean

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

problems:

Find the values of the finite population correction factor
1) for $n=10$ & $N=1000$ 2) for $n=5$ & $N=200$

1) Given $N=1000$
$n=10$

correction factor $= \frac{N-n}{N-1} = \frac{1000-10}{1000-1} = 0.991$

2) Given $N=200$, $n=5$

correction factor $= \frac{N-n}{N-1} = \frac{\cancel{2000-10}}{} = \frac{200-5}{200-1} = 0.979$

Q) How many different samples size 2 can be choosen from the finite population of size 25

we can take N_{C_n} samples of size

Given $n=2$, $N=25$

$\therefore N_{C_n} = {}^{25}C_2 = 300$ samples of size to from the finite population of size 25

Q) A population consists of 5 numbers 2, 3, 6, 8 & 11 consider all the possible samples of size two which can be drawn:
-2009

a) with replacement

b) without replacement

from this population, find

i) The Mean of the population

ii) The S.D of population

iii) the Mean of the sampling distribution of mean

i) Mean of the population is given by

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6.$$

ii) variance of the population σ^2 is given by

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5} = \frac{54}{5} = 10.8$$

with & without replacement

$$S.D = \sigma = \sqrt{10.8} = 3.286 = 3.29$$

iii) Sampling with replacement (finite population).

The total no. of samples with replacement is

$N^n = 5^2 = 25$ Here, $N=5, n=2$

with replacement we get 25 samples

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

The sample means are

$$\begin{Bmatrix} \frac{2+2}{2}=2, & \frac{2+3}{2}=2.5 & \frac{2+6}{2}=4 & \frac{2+8}{2}=5 & \frac{2+11}{2}=6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{2+2.5+4+5+6.5+2.5+3+4.5+5.5+7+4+4.5+6+7+8.5+5+5.5+7+8+9.5+6.5+7+8.5+9.5+11}{25}$$

$$\mu_{\bar{x}} = \frac{150}{25} = 6$$

$$\boxed{\mu_{\bar{x}} = 6 = \mu.}$$

without Replacement

$$\begin{Bmatrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ (8,11) & & & \end{Bmatrix}$$

$\mu_{\bar{x}} = 2.5$

$$\begin{Bmatrix} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$\mu_{\bar{x}} = \frac{60}{10} = 6$$

$$\boxed{\mu_{\bar{x}} = 6 = \mu}$$

(iv) with replacement :

From eq(1) squaring the result adding the 25 members thus obtain and dividing by 25

$$\sigma_{\bar{x}}^2 = \frac{(2-6)^2+(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+\dots+(11-6)^2}{25}$$

$$= \frac{135}{25} = 5.40$$

$$\sigma = \sqrt{5.40} = 2.32$$

(or)

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$$

without replacement :

$$\sigma_{\bar{x}}^2 = \frac{(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2}{10}$$

$$= 4.05$$

$$\sigma_{\bar{x}} = \sqrt{4.05} = 2.01$$

H.W

Q) Sample of size 2 are taken from the population 1,2,3,4,5,6 find

a) with replacement b) without replacement

1) find mean of the population

2) S.D of the population

3) The mean of the sampling Distribution of means

4) S.D of the sampling Distribution of means -

Given $n=2$

i) Mean of the population is given by $\mu = \bar{x} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6}$

$= 3.5$

ii) Variance of the population σ^2 is given by

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$= \frac{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2}{6}$$

$$= \frac{6.25+2.25+0.25+0.25+2.25+6.25}{6}$$

$$= \frac{17.5}{6} = 2.9167$$

$\sigma^2 = 2.9167$

$\sigma = \sqrt{2.9167} = 1.7078$

iii) Sampling with replacement (infinite population)

The total no. of samples with replacement is

$N = 6, n = 2$

$N^n = 6^2 = 36$

with replacement we get 36 samples

$$\begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

The Sample means are

$$\begin{Bmatrix} 1 & 1.5 & 2 & 2.5 & 3 & 3.5 \\ 1.5 & 2 & 2.5 & 3 & 3.5 & 4 \\ 2 & 2.5 & 3 & 3.5 & 4 & 4.5 \\ 2.5 & 3 & 3.5 & 4 & 4.5 & 5 \\ 3 & 3.5 & 4 & 4.5 & 5 & 5.5 \\ 3.5 & 4 & 4.5 & 5 & 5.5 & 6 \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{1+1.5+2+2.5+3+3.5+1.5+2+2.5+3+3.5+4+4.5+5+3+3.5+4+4.5+5+5.5+3.5+4+4.5+5+5.5+6}{36}$$

$$= \frac{126}{36} = 3.5$$

without replacement

$$\begin{Bmatrix} (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,3) & (2,4) & (2,5) & (2,6) \\ (3,4) & (3,5) & (3,6) \\ (4,5) & (4,6) \\ (5,6) \end{Bmatrix}$$

The sample means are

$$\begin{Bmatrix} 1.5 & 2 & 2.5 & 3 & 3.5 \\ 2.5 & 3 & 3.5 & 4 \\ 3.5 & 4 & 4.5 \\ 4.5 & 5 \\ 5.5 \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{1.5+2+2.5+3+3.5+2.5+3+3.5+4+3.5+4+4.5+4.5+5+5.5}{15}$$

iv) with replacement

$$\sigma_{\bar{x}}^2 = \frac{(1-3.5)^2+(1.5-3.5)^2+\cdots+(6-3.5)^2}{36}$$

$$\sigma_{\bar{x}^2} = \frac{\sigma^2}{n} = \frac{\sigma}{\sqrt{n}} = \frac{1.7078}{\sqrt{2}} = 1.2076$$

with replacement

$$\sigma_{\bar{x}}^2 = \frac{(1.5-3.5)^2+(2-3.5)^2+(2.5-3.5)^2+(3-3.5)^2+(3.5-3.5)^2+(2.5-3.5)^2+(3-3.5)^2+(4-3.5)^2+(3.5-4.5)^2+(4-3.5)^2+(4.5-3.5)^2+(4.5-3.5)^2+(5-3.5)^2+(5.5-3.5)^2}{15}$$

$$\sigma_{\bar{x}^2} = \frac{17.5}{15} = 1.1667$$

$\sigma_{\bar{x}} = \sqrt{1.1667} = 1.0801$

Q) The mean height of Students in a college is 155cm and S.D is 15. what is the probability that the mean height of 36 Students is less than 157cm.

Sol:- Given

$\mu = 155cm$

$\sigma = 15$

$n = 36$

$\bar{x}$ = mean of Sample = 157cm.

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{157-155}{\frac{15}{\sqrt{36}}} = 0.8$$

$$P(\bar{x} \leq 157) = P(z < 0.8)$$
$$= 0.5 + P(0 \leq z \leq 0.8)$$
$$= 0.5 + 0.2881$$
$$= 0.7881$$

$\therefore$ The probability that the mean height of 36 Students is less than 157cm is 0.7881

Q) A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $\sigma^2 = 256$. what is the probability that $\bar{x}$ will be between 75 and 78.

Sol:- Given

$n = 100$

$\mu = 76$

$\sigma^2 = 256$

$\sigma = \sqrt{256} = 60$

Let $\bar{x}_1 = 75$ and $\bar{x}_2 = 78$

$$z_1 = \frac{\bar{x}_1-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{75-76}{\frac{60}{\sqrt{100}}} = -0.625$$

$$z_2 = \frac{\bar{x}_2-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78-76}{\frac{60}{\sqrt{100}}} = 1.25$$

$P(75 \le \bar{x} \le 78) = P(Z_1 \le \bar{x} \le Z_2)$

$= P(-0.625 \le Z \le 1.25)$

$= P(-0.625 \le Z \le 0) + P(0 \le Z \le 1.25)$

$= 0.2334 + 0.3944$

$= 0.628$

Q) A normal population has a mean of 0.1 and S.D of 2.1. find the probability that mean of a Sample of Size 900 will be negative.

Sol:- Given $\mu = 0.1$ $\sigma = 2.1$ and $n = 900$

Standard normal variate

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 0.1}{2.1/\sqrt{900}} = \frac{\bar{x} - 0.1}{2.1/30} = \frac{\bar{x} - 0.1}{0.07}$$

$\bar{x} = 0.1 + 0.007 Z$ where $Z \sim N(0,1)$

The required probability, that the sample mean is negative is given by

$P(\bar{x} < 0) = P(0.1 + 0.07 Z < 0)$

$= P(0.07 Z < -0.1)$

$= P\left(Z < \frac{-0.1}{0.07}\right)$

$= P(Z < -1.43)$

$= 0.50 - P(0 < Z < 1.43) = 0.50 - 0.4236 = 0.0764$

Q) If a 1 gallon can of paint covers, on an average 513 Square feet standard deviation of 31.5 Square feet, what is the probability that the mean area covered people of these 1-gallon cans will be anywhere from 510 to 520 Square feet

Given $n = 40$, $\mu = 513$ and $\sigma = 31.5$ sq.ft

The test statistics is $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

when $\bar{x} = 510$, $Z = \frac{510 - 513}{\frac{31.5}{\sqrt{40}}} = -0.6$

when $\bar{x} = 520$, $Z = \frac{520 - 513}{\frac{31.5}{\sqrt{40}}} = 1.4$

Required probability $= P(-0.6 < Z < 1.4)$

$= P(-0.6 < Z < 0) + P(0 < Z < 1.4)$

$= P(0 < Z < 0.6) + P(0 < Z < 1.4)$

Test of significance of a single mean-large samples:

A sample of 64 students have a mean weight of 70kgs. Can this be recorded as a sample from a population with a mean weight of 56 kgs & SD 25 kgs with los 0.05

Given $\bar{x} = 70$kgs.

$\mu = 56$kgs.

$S.D = \sigma = 25, n = 64$.

Null Hypothesis $H_0: \mu = 56$

Alternative Hypothesis $H_1: \mu \neq 56$

L.O.S $\alpha = 0.05$

The test statistic $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{70-56}{\frac{25}{\sqrt{64}}} = 4.48$

The tabulated value at 5% los is 1.96

The calculated value is greater than the tabulated value.

We reject null Hypothesis H_0 at 5% los.

Q, It is claimed that a random sample of 49 tyres has a mean life of 15200kms. This sample was drawn from a population whose mean is 15,150kms and a S.D of 1200 kms. Test the significance at 0.05 level.

Given $n = 49$.

$\bar{x} = 15,200$

$\mu = 15,150$, $S.D = \sigma = 1200$

Null Hypothesis $H_0: \mu = 15,150$

Alternative Hypothesis $H_1: \mu \neq 15,150$

LOS $\alpha = 0.05$

The test statistic $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{15,200 - 15,150}{\frac{1200}{\sqrt{49}}} = 0.2916$

The tabulated value at 5% L.O.S is 1.96

The calculated value is less than the tabulated value.

We accept H_0 at 5% los.

Q, An Ambulance service claim that it takes of the average less than 10min to reach its destination Emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 los.

Given $n = 36$

$\bar{x} = 11$

$\mu = 10$, $\sigma^2 = 16 \Rightarrow \sigma = 4$.

Null Hypothesis $H_0: \mu = 10$

Alternative Hypothesis $H_1: \mu < 10$

LOS $\alpha = 0.05$

The test statistic $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11 - 10}{\frac{4}{\sqrt{36}}} = 1.5$

The tabulated value at 5% LOS is 1.645

The calculated value is less than the tabulated value

We accept H_0 at 5% LOS.

Q) A trucking RM suspects the claim that average life of certain tyres is atleast 28,000 miles. To check the claim the RM puts of 40 of this tyres on its trucks and gets a mean life time of 27463 Miles with a S.D of 1348 miles can the claim be true.

Given $\bar{x} = 27463$

$\mu = 28000$, $\sigma = 1348$, $n = 40$

Null Hypothesis $H_0: \mu = 28000$

Alternative Hypothesis $H_1: \mu \neq 28000$

LOS $\alpha = 0.05$

The test statistic $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{27463 - 28000}{\frac{1348}{\sqrt{40}}} = -2.52$

$|z| = 2.52$

The tabulated value at 5% LOS is 1.96

The calculated value is greater than the tabulated value

We reject the null Hypothesis H_0 at 5% LOS.

Q) A sample of 400 items is taken from a population whos S.D is 10. The mean of the sample. Test whether the sample has come from a population with mean 38. also calculate 95% confident interval for the population.

Given $n = 400$

$\bar{x} = 40$

$\mu = 38$

$\sigma = 10$

null Hypothesis $H_0: \mu = 38$

Alternative Hypothesis $H_1: \mu \neq 38$

L.O.S $\alpha = 0.05$

The test statistic $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$z = \frac{40-38}{\frac{10}{\sqrt{400}}} = 4$$

The tabulated value at 5% LOS is 1.96

The calculated value is greater than the tabulated value.

We reject the null hypothesis H_0 at 5% LOS.

95% confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

$$= \left(40 - 1.96\frac{10}{\sqrt{400}}, 40 + 1.96\frac{10}{\sqrt{400}}\right)$$

95% confidence interval $= (39.02, 40.98)$

Q) A sample of 900 members as a mean of 3.4 cms and S.D 2.61 cms in is this sample has been taken from a large population of mean 3.25 cms. If the population is normal and its mean is unknown. Find 95% confidence limits of true mean

Given

$\bar{x} = 3.4$

$\mu = 3.25$

$S.D = \sigma = 2.61$

$n = 900$.

Null Hypothesis

$H_0 : \mu = 3.25$

Alternative Hypothesis

$H_1 : \mu \neq 3.25$

L.O.S $\alpha = 0.05$

The test statistics $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$z = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}}$$

$$z = 1.724$$

The tabulated value at 5% LOS is 1.96

The calculated value is less than tabulated value.

We accept at H_0 5% LOS

95% confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

$$\left(3.4 - 1.96\frac{2.4}{\sqrt{900}}, 3.4 + 1.96\frac{2.61}{\sqrt{900}}\right)$$

95% confidence interval $= (3.2295, 3.57)$

No Incomplete

Test of significance for ~~recurrence~~ difference of Means

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

problems:

The Mean of two large Samples of sizes 1000 and 2000 members are 67.5 inches & 68.0 inches respectively. Can the Samples be regarded as drawn from the same population of S.D 2.5 inches

Given $n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$

S.D $\sigma = 2.5$ inches

null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

The test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Here, $\sigma_1 = \sigma_2 = \sigma = 2.5$

$$z = \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.16$$

$|z| = 5.16$

The tabulated value at 5% los is 1.96

If the cal value is greater than tabulated value we reject null Hypothesis H_0 at 5% LOS.

Q, The Mean life of sample of 10 Electric bulbs (or motors) was found to be 1456 hours with S.D of 423 hours. A Second Sample of 17 bulbs (or motors) choosen from a different batch show a mean life of 1280 hours with S.D of 398 hours. Is there significance difference between the Means of two batches.

Given $n_1 = 10$, $n_2 = 17$

$\bar{x}_1 = 1456$, $\bar{x}_2 = 1280$

$\sigma_1 = 423$, $\sigma_2 = 398$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

LOS $\alpha = 0.05$

The test Statistics $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$z = \frac{1456 - 1280}{\sqrt{\dfrac{(423)^2}{10} + \dfrac{(398)^2}{17}}} = 1.067$$

The tabulated value at 5% LOS is 1.96.

$Z_{cal} < Z_{Tab}$. we accept H_0 at 5% LOS

Q, The Mean height of 50 male students who participated in Sports is 68.2 inches with a S.D of 2.5. The Mean height of 50 male students who have not participated in sports in 67.2 inches with a S.D of 2.8. Test the hypothesis that the height of students who participated in Sports is more than the students who have not participated in Sports

Given $n_1 = 50$, $n_2 = 50$

$\bar{x}_1 = 68.2$, $\bar{x}_2 = 67.2$

$\sigma_1 = 2.5$, $\sigma_2 = 2.8$

Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

LOS $\alpha = 5\% = 0.05$

The test Statistics $z = \frac{x_1 - x_2}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}}$

~~$z = \frac{50-50}{}$~~ $z = \frac{68.2 - 67.2}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}} = -5.16$. $|z| = 5.16$

The tabulated value at 5% LOS is 1.96

If $z_{cal} < z_{Tab}$. We accept H_0 at 5% LOS

Q) The Research investigator is interested in studying whether there is a significance difference in the salaries of MBA grades 2 metro politian cities. A random sample of size 100 from mumbai yields on average income of 20150. Another random sample of 60 chennai results in an average income of ₹20,250. If the variance of the both ~~32,400 respectively~~ populations are given as $\sigma_1^2 = 40,000$ $\sigma_2^2 = 32,400$ respectively ($z = 3.26$)

Given $n_1 = 100$, $n_2 = 60$

$\bar{x}_1 = 20150$, $\bar{x}_2 = 20,250$

$\sigma_1^2 = 40,000 \Rightarrow \sigma_1 = 200$

$\sigma_2^2 = 32,400 \Rightarrow \sigma_2 = 180$

Null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$.

L.O.S $\alpha = 5\% = 0.05$

the test Statistics $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}} = \frac{20150 - 20250}{\sqrt{\frac{40,000}{100} + \frac{32,400}{60}}}$

$z = 3.26$

The tabulated value at 5% LOS is 1.96

$z_{cal} > z_{tab}$. We reject H_0 at 5% L.O.S

Q) The nicotine in milligrams of 2 samples of tobacco were found to be as follows. ~~Hou~~ find the Standard error & Confidence limits for the difference between the means at 0.05 level

Sample A	24	27	26	23	25	-
Sample B	29	30	30	~~34~~ 31	34	36

Sample A			Sample B		
x_1	$x_1 - \bar{x}_1$ $x_1 - 25$	$(x_1 - \bar{x})^2$	x_2	$(x_2 - \bar{x}_2)$ $x_2 - 30$	$(x_2 - \bar{x}_2)^2$
24	-1	1	29	-1	1
27	2	4	30	0	0
26	1	1	30	0	0
23	-2	4	31	1	1
25	0	0	24	-6	36
-	-	-	36	6	36
		$\Sigma(x_i - \bar{x})^2 = 10$			$\Sigma(x_2 - \bar{x})^2 = 74$

$$\bar{x}_1 = \frac{24+27+26+23+25}{5} = \frac{125}{5} = 25$$

$$\bar{x}_2 = \frac{29+30+30+31+24+36}{6} = \frac{180}{6} = 30$$

$$\Sigma(x_i - \bar{x})^2 = 10, \quad \Sigma(x_2 - \bar{x})^2 = 74$$

$$S_1^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{10}{5-1} = 2.5$$

$$S_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{74}{6-1} = 14.8$$

Standard Error of $(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{2.5}{5} + \frac{14.8}{6}} = 1.7$

Hence 95% confidence limits are

$(\bar{x}_1 - \bar{x}_2) \pm 1.96(\text{S.E of } \bar{x}_1 - \bar{x}_2)$

$(25 - 30) \pm 1.96(1.72)$

-5 ± 3.3712

$(-5 - 3.37, -5 + 3.37)$

$= (-8.37, -1.63)$

Test of significance for single proportion:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

A manufacturer claimed that atleast 95% of the equipment which is supplied to a factory conformed to specifications. An examination of a sample of 200 pieces equipments revealed that 18 were faulty. Test is claimed at 5% LOS.

- 2011

A) Given Sample Size = 200

The no. of pieces conforming to Specification = 200 - 18 = 182

P = proportion of pieces conforming Specifications:

$$p = \frac{182}{200} = 0.91$$

P = population proportion = $95\% = \frac{95}{100} = 0.95$, $Q = 1 - P$

Null Hypothesis $H_0 : P = 95\%$ $Q = 1 - 0.95$

Alternative Hypothesis $H_1 : P > 0.25$ $Q = 0.05$

The test statistics $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{0.95(0.05)}{200}}} = -2.59$

$|z| = 2.59$

The tabulated value at right tail test at 5% los is 1.645

The calculated value is greater than tabulated value.

We reject H_0 at 5% LOS.

Q) In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in the state & 1% los

Sol: Given $n = 1000$

$$p = \frac{540}{1000} = 0.54$$

$$P = \frac{1}{2}, Q = 1 - P = 1 - \frac{1}{2} = 0.5$$

Null Hypothesis $H_0 : P = 0.5$

Alternative Hypothesis $H_1 : P \neq 0.5$ (two-tailed test)

The test statistics $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.529$

The tabulated value at 1% los is 2.58

The calculated value is less than the tabulated value

we accept H_0 at 1% LOS.

Q) In a big city 325 out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in the city are smokers.

Given $n = 600$

No. of smokers = 325

$$p = \frac{325}{600} = 0.5417$$

$P = \frac{1}{2} = 0.5$, $Q = 1 - P = 1 - 0.5 = 0.5$

Null Hypothesis $H_0: P = 0.5$

Alternative Hypothesis $H_1: P > 0.5$ (Right tailed test) (majority)

The test statistics $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.5417 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{600}}} = 2.04$

The tabulated value at 5% l.o.s is 1.645

The calculated value is greater than tabulated value.

So, we reject H_0 at 5% los.

Q) 20 people were attack by a disease only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease in 85% in favour of the hypothesis i.e. more at 5% level. -2012

Given $n = 20$, x = no. of survive people $= 18$

p = proportion of survived people $= \dfrac{x}{n} = \dfrac{18}{20} = 0.9$

$P = 85\% = \dfrac{85}{100} = 0.85$, $Q = 1 - P = 1 - 0.85 = 0.15$

Null Hypothesis $H_0: P = 0.85$

Alternative Hypothesis $H_1: P > 0.85$

LOS $\alpha = 0.5$

The test statistics $z = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.9 - 0.85}{\sqrt{\dfrac{0.85(0.15)}{20}}} = 0.625$

The tabulated value at 5% in right tailed test is 1.645

Since calculated value is less than the tabulated level. We accept null hypothesis H_0 at 5% LOS.

Q) A random sample of 500 apples was taken from a large consignment of 60 were found to be bad, obtain 98% confidence limits for the percentage limit no. of bad apples in the consignment. - 2017

Given $n = 500$

x = no. of bad apples $= 60$

p = proportion of bad apples $= \dfrac{x}{n} = \dfrac{60}{500} = 0.12$

$q = 1 - p = 1 - 0.12 = 0.88$

The 98% confidence limits ~~were~~ ~~per~~ for p are

$\left[p - z \sqrt{\dfrac{pq}{}} , p + z \sqrt{\dfrac{pq}{}}\right]$

where $z_{\alpha/2} = 2.33$

$$=\left[0.12 - 2.33\sqrt{\frac{(0.12)(0.88)}{500}},\ 0.12 + 2.33\sqrt{\frac{(0.12)(0.88)}{500}}\right]$$

$= [0.0861, 0.1539]$

Thus, the confidence limits for the percentage of bad apples $(0.087\times100, 0.153\times100) = (8.7, 15.3)$

Q. In a study design to investigate whether certain detonators used with explosives in coal mining meet the requirement that atleast 90% will ignite the explosive when charged. It is found that 174 of 200 detonators function properly. Test the null Hypothesis $\overset{capital}{P}=0.9$ against the alternative hypothesis $P\neq0.9$ and 0.05 los.

The null Hypothesis $H_0: P=0.9$

Alternative Hypothesis $H_1: P\neq0.9$

l.o.s $\alpha=0.05$.

Given $n=200$, $p=\frac{174}{200}=0.87$,

$P=0.9$; $Q=1-P=1-0.9=0.1$

The test statistics $z=\frac{p-P}{\sqrt{\frac{PQ}{n}}}=\frac{0.87-0.9}{\sqrt{\frac{0.9\times0.1}{200}}}=-1.413$

$|z|=1.43$

The tabulated value at 5% Los is 1.645

$Z_{cal} < Z_{Tab}$

we accept the H_0 at 5% los.

Q. A die was thrown 9000 times and of this 3220 yielded 3 or 4. Is these consistent with the hypothesis that the die was unbiased.

Given $n=9000$

p = proportion of success of getting 3 or 4 in 9000 throws $=\frac{3220}{9000}=0.3578$

$P=(\text{getting } 3 \text{ or } 4)=\frac{1}{6}+\frac{1}{6}=\frac{2}{6}=\frac{1}{3}=0.3333$

$Q=1-P=1-\frac{1}{3}=\frac{2}{3}=0.6666$

Alternative Hypothesis $H_1: P \neq \frac{1}{3}$

The test statistics $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.3578 - 0.3333}{\sqrt{\frac{(0.3333)(0.6666)}{9000}}}$

$= 4.93$

The tabulated value at 5% LOS is 1.96

If the calculated value is greater than tabulated value we reject null Hypothesis H_0 at 5% LOS.

Test of significance of difference between two Sample proportions - Large Samples.

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}} = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Q) Among the items produced by a factory out of 500, 15 were defective in another sample out of 400, 20 were defective. Test the significance between the differences of two proportions at 5% los.

Imp

let P_1 & P_2 be the proportions of defective items in the populations of 2 sample items produced by the factory.

Null Hypothesis $H_0: P_1 = P_2$

Alternative Hypothesis $H_1: P_1 \neq P_2$

LOS $\alpha = 5\% = 0.05$

Given $n_1 = 500$, $n_2 = 400$

$x_1 = 15$, $x_2 = 20$

$P_1 = \frac{x_1}{n_1} = \frac{15}{500} = 0.03$, $P_2 = \frac{x_2}{n_2} = \frac{20}{400} = 0.05$

and $p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{500(0.03) + (400)(0.05)}{500 + 400} = \frac{35}{900}$

$p = 0.039$

$q = 1 - p = 1 - 0.039 = 0.961$

The test statistics $z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.03 - 0.05}{\sqrt{(0.039)(0.961)\left(\frac{1}{500} + \frac{1}{400}\right)}}$

$= -1.54$

$|z| = |-1.54| = 1.54$

The tabulated value at 5% los is 1.96.
The calculated value is less than tabulated value we accept null Hypothesis H_0 at 5% los.

Q) A sample poll of 300 voters from district A and 200 voters from district B. show that 56% and 48% respectively were in favour of a given candidate at 0.05 los. Test the hypothesis that where is as difference in the district.

Given

$n_1 = 300$, $n_2 = 200$

$P_1 = 56\% = \frac{56}{100} = 0.56$

$P_2 = 48\% = \frac{48}{100} = 0.48$

Null Hypothesis $H_0: P_1 = P_2$
Alternative Hypothesis $H_1: P_1 \neq P_2$
The los $\alpha = 0.05$
The test statistic $z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

where

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{300(0.56) + (200)(0.48)}{300 + 200} = 0.528$$

$q = 1 - p = 1 - 0.528 = 0.472$

$$z = \frac{0.56 - 0.48}{\sqrt{(0.528)(0.472)\left(\frac{1}{300} + \frac{1}{200}\right)}} = 1.755 = 1.76$$

The tabulated value at 5% los is 1.96.
If the calculated value is less than tabulated value. we accept null Hypothesis at 5% los.

Q) In two large populations, there are 30% & 25% of fair haired people. Is these difference likely to be hitten in sample of 1200 and 900 respectively

– 2013

Given $n_1 = 1200$, $n_2 = 900$

$P_1 = \frac{30}{100} = 0.3$, $P_2 = \frac{25}{100} = 0.25$

Null Hypothesis $H_0: P_1 = P_2$

l.o.s $\alpha = 0.05$

$$P = \frac{(1200)(0.3) + (900)(0.25)}{1200 + 900} = 0.2786.$$

$q = 1 - P = 0.7214$

The test statistics $z = \dfrac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$z = \frac{0.3 - 0.25}{\sqrt{(0.2786)(0.7214)\left(\frac{1}{1200} + \frac{1}{900}\right)}} = 2.529$$

Tabulated value at 5% l.o.s is 1.96

The calculated value is greater than tabulated value we reject H_0 at 5% l.o.s.

Q) Among the items produced by a factory, out of 800 65 were defective. in another sample out of 300, 40 were defective. Test the significance between the difference two proportions at 1% level.

Null Hypothesis $H_0 : P_1 = P_2$

Alternative Hypothesis $H_1 : P_1 \neq P_2$

Given $n_1 = 800$; $n_2 = 300$

$x_1 = 65$ $x_2 = 40$

$$P_1 = \frac{x_1}{n_1} = 0.081\left(\frac{65}{800}\right), \quad P_2 = \frac{x_2}{n_2} = \frac{40}{300} = 0.1333$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{800(0.81) + 300(0.1333)}{800 + 300} = 0.095$$

$q = 1 - P = 1 - 0.95 = 0.904$

The test statistics $z = \dfrac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.081 - 0.1333}{\sqrt{(0.095)(0.904)\left(\frac{1}{800} + \frac{1}{300}\right)}}$

$$z = -2.620$$

$|z| = |-2.620| = 2.620$

z_{tab} at 1% l.o.s is ~~1.645~~ 2.58

$z_{cal} > z_{tab}$ we reject H_0 at 1% l.o.s

Q) In a random sample of 1000 persons from Town A, 400 are found to be consumers ~~are~~ of wheat, ~~400~~ In a sample of 800 from Town B, 400 are found to be consumers from Town A & Town B. So ~~that~~ for as the proportions of wheat consumers is concerned.

-2012

Given $n_1 = 1000$ $n_2 = 800$

$x_1 = 400$, $x_2 = 400$

$P_1 = \frac{x_1}{n_1} = \frac{400}{1000} = 0.4$ $P_2 = \frac{x_2}{n_2} = \frac{400}{800} = \frac{1}{2} = 0.5$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{1000(0.4) + 800(0.5)}{1000 + 800} = \frac{800}{1800} = 0.4444$$

$q = 1 - P = 1 - 0.4444 = 0.5556$.

Null Hypothesis $H_0 : P_1 = P_2$

Alternative Hypothesis $H_1 : P_1 \neq P_2$

The test statistics $z = \frac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.4 - 0.5}{\sqrt{(0.4444)(0.5556)\left(\frac{1}{1000} + \frac{1}{800}\right)}}$

$z = -4.2426$

$|z| = |-4.2426| = 4.2426$

The tabulated value at 5% L.O.S is 1.96.

If the calculated value is greater than tabulated value. we reject the null Hypothesis H_0 at 5% L.O.S.

Q) During a country wide investigation the incidence of tuberculosis was found to be 5%. In a college of 400 students 3 reported to be affected, where as in another college of 1200 students 10 were affected. Thus, this indicate any significance difference.

Given $n_1 = 400$, $n_2 = 1200$

$x_1 = 3$, $x_2 = 10$

$P_1 = \frac{x_1}{n_1} = \frac{3}{400} = 0.0075$, $P_2 = \frac{x_2}{n_2} = \frac{10}{1200} = 0.0083$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(400)(0.0075) + 1200(0.0083)}{400 + 1200} = 0.0081$$

$q = 1 - P = 0.9919$

Null Hypothesis

Alternative Hypothesis.

The test statistics $z = \frac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.0075 - 0.0083}{\sqrt{(0.0081)(0.9919)\left(\frac{1}{400} + \frac{1}{1200}\right)}}$

The tabulated value at 5% l.o.s is 1.96

The tabulated value is ~~les~~ greater than calculated value.

So we accept the H_0 at 5% los

Q) If 120 out of 200 patients suffering from a certain disease are cured by allopathy & 240 out of 500 patients are cured by Homeopathy, is there reason enough to believe that allopathy is <u>better than</u> Homeopathy in curing the disease use $\alpha = 0.05$ l.o.s

Given $n_1 = 200$, $n_2 = 500$

$x_1 = 120$, $x_2 = 240$

$P_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.6$, $P_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48$

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(200)(0.6) + (500)(0.48)}{200 + 500} = 0.5142$$

$q = 1 - p = 1 - 0.5142 = 0.4858$

Null Hypothesis $H_0 : P_1 = P_2$

Alternative Hypothesis $H_1 : P_1 \cancel{\neq} > P_2$

The test statistic $z = \frac{P_1 - P_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.6 - 0.48}{\sqrt{(0.5142)(0.4858)\left(\frac{1}{200} + \frac{1}{500}\right)}}$

$z = 2.869$

l.o.s $\alpha = 0.05$

Z_{tab} at 5% l.o.s is $Z_{\alpha/2} = Z_{\frac{0.05}{2}} = Z_{0.025}$ is 1.645

$Z_{cal} > Z_{tab}$ we reject H_0 at 5% l.o.s.

Q) Fit a second degree polynomial to the following data by the method of least squares.

x	1	2	3	4	5
y	12	25	40	50	65

x	y	x^2	x^3	x^4	xy	x^2y
1	12	1	1	1	12	12
2	25	4	8	16	50	100
3	40	9	27	81	120	360
4	50	16	64	256	200	800
5	65	25	125	625	325	1625
$\Sigma x =$ 15	$\Sigma y =$ 192	$\Sigma x^2 =$ 55	$\Sigma x^3 =$ 225	$\Sigma x^4 =$ 979	$\Sigma xy =$ 707	$\Sigma x^2y =$ 2897

The normal equations are

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

$n=5$

$192 = 5a + 15b + 55c$ — (1)

$707 = 15a + 55b + 225c$ — (2)

$2897 = 55a + 225b + 979c$ — (3)

Solving (1) & (2), (1) × 3, (2) × 1

$$576 = \cancel{15a} + 45b + 165c$$

$$\underline{\overset{(-)}{707} = \overset{(-)}{\cancel{15a}} + \overset{(-)}{55b} + \overset{(-)}{225c}}$$

$$-131 = -10b - 60c$$

$$\boxed{10b + 60c = 131} \text{ — (4)}$$

Solving (2) & (3), (2) × 55, (3) × 15.

$$38885 = \cancel{825a} + 3025b + 12375c$$

$$\underline{\overset{(-)}{43455} = \overset{(-)}{\cancel{825a}} + \overset{(-)}{3375b} + \overset{(-)}{14685c}}$$

$$-4570 = -350b - 2310c$$

$$\div -5 \text{ O.B.S}$$

$$\boxed{914 = 70b + 462c} \text{ — (5)}$$

Solving (4) & (5), we get

(4) × 7, (5) × 1.

$$\cancel{70b} + 420c = 917$$

$$\underline{\cancel{70b} + 462c = 914}$$

$$-\cancel{42}c = 3$$
$$14$$

$$-14c = 1 \Rightarrow c = -\frac{1}{4}$$

$$c = -0.0714$$

Sub $c = -0.0714$ in eq (4) we get

$$10b + 60(-0.0714) = 131$$

$$10b - 4.284 = 131$$

$$10b = 131 + 4.284 = 135.284$$

$$b = \frac{135.284}{10} = 13.5284.$$

$b = 13.5284$, $c = 0.0714$ sub b & c values in (1), we get

$$192 = 5a + 15(13.5284) + 55(-0.0714)$$

$$5a = 192 - 202.926 + 3.927$$

$$a = \frac{192 - 202.96 + 8.927}{5} = -1.3998.$$

$\therefore y = a + bx + cx^2 \Rightarrow y = -1.3998 + (13.5284)x + (-0.0714)x^2$

Required parabola are second degree polynomial.

UNIT-5

Test of Significance (small Sa

Degrees of freedom (d.f)

The number of independent variates which make up the Statistic is known as the degrees of freedom. and is denoted by 'v'

Student's "t-test": Let $\bar{x}$ = Mean of a sample,

n = Size of the sample

σ = S.D of the sample

μ = Mean of the population

The student t-test is defined by the Statistic

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$$

Student t-test for Single Mean

Let a random sample of size n ($n < 30$) has a sample mean $\bar{x}$. To test the hypothesis that the

problem

A Mechanis

engine parts

eters of 0.70

Sample of 10 p

Mean diamete

with a S.D of

Compute the

Would use to

the work is n

Specification

level of Sig

Given, Sampl

$\bar{x}$ = 0.742

μ = 0.700

S.D $S = 0.04$

We use Stud

Null Hypothesi

is conforming

Alternative

Cal > Tab H_0 is rejected

Cal < Tab H_0 is accepted.

Q) A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacture claims that the Mean life of bulbs is 1000 hours. Is the Sample not upto the Standard.

Given $n = 26 < 30$

$\bar{x} = 990$

$\mu = 1000$

S.D $s = 20$

Degrees of freedom $= (n-1)$

$= (26-1)$

$= 25$

Q) The
Sample
light
a co
be 15
of 12
Claim
the b
compa
the l
0.05
Degree

Q) A random sample of 6 steel beams has a mean compressive strength of 58,392 (pounds per square inch) with a S.D of 648 P.S.I use these information & the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which this sample came is 58,000 P.S.I. Assume normality.

~~Null~~

Given $n = 6$

$\bar{x} = 58,392$

$\mu = 58,000$

S.D σ = 648

$d.f = n-1 = 6-1 = 5$

Level of significance $\alpha = 0.05$.

Null Hypothesis $H_0 : \mu = 58,000$

were teste

Obtained

respective

experimen

Given n:

μ=

d.f = (n-

l.o.s =

Null Hy

H_0: The

not S

Alternat

$H_1 : \mu \neq 1$

The test

$t = \frac{17.8}{\frac{1}{\sqrt{}}}$

$t = |t| =$

The test statistics $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}}$

$$t = \frac{67 - 70}{\frac{5.2}{\sqrt{154}}} = -7.16$$

$|t| = 7.16$

$t_{tab} = 3$

$t_{cal} > t_{tab}$

∴ Reject the null hypothesis H_0

conclude that the sample has not been taken from a large ~~Pro~~ population

Problems related to Student t test (when S.D of the sample is not given directly)

A Random Sample of 10 boys has the following I.Q's 70, 120, 110, 101, 88, 83, 95, 98, 107 & 100

a) Do these data support the assumption of a population

x	x
70	70-
120	+
110	1
101	3
88	-
83	-1
95	[illegible]
98	0
107	9
100	0

$\Sigma(x_i - \bar{x})^2$

$S = \sqrt{\frac{1833-6}{9}}$

Null Hyp

Support

a popul

100.

Alternati

$H_1 : \mu \neq 1$

$= 97.2 \pm 2.26 \frac{14.27}{\sqrt{10}}$

$= 97.2 + 2.26 \frac{14.27}{\sqrt{10}}, 97.2 - 2.6 \frac{14.27}{\sqrt{10}}$

$= 107.39, 87$

The 95% confidence limits with in which the mean I.Q values of sample of 10 boys will lie is (87, 107.39)

Q) The lifetime of electric bulbs in a random sample of 10. For a large consignment gave the following data.

Item	1	2	3	4	5	6	7	8	9	10
Life in hours	1.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average lifetime of

$\bar{x} = \frac{660}{10} = 66$

$s^2 = \sum_{i=1}^{10} \frac{1}{n-1}(x_i - \bar{x})^2$

$= \frac{1}{9}\big[(70-66)^2 + (67-66)^2 + (62-66)^2 + (68-66)^2 + (61-66)^2 + (68-66)^2 + (70-66)^2 + (64-66)^2 + (64-66)^2 + (66-66)^2\big]$

$s^2 = 10 \Rightarrow s = \sqrt{10} = 3.16$

$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{66-64}{\frac{3.16}{\sqrt{10}}} = 2.0014$

given $t_{tab} = 1.833$

$t_{cal} > t_{tab}$

$\therefore$ we reject the null Hypothesis H_0 at 5% l.o.s and conclude that the average height is not greater than the 64 inches.

$\bar{X} = 50$

$\bar{x} = 4$

$s^2 = \sum_{i=1}^{10}$

$= \frac{1}{9}[$

$+$

$+$

$n_1 = 7,\ n_2 = 6$

$$\bar{x} = \frac{\sum x}{n_1} = \frac{28+30+32+33+33+29+34}{7}$$

$$= \frac{219}{7} = 31.285$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{29+30+30+24+27+29}{6}$$

$$= \frac{169}{6} = 28.16$$

x	$x-\bar{x}$ $x-31.285$	$(x-\bar{x})^2$	y	$y-\bar{y}$ $y-28.16$	$(y-\bar{y})^2$
28	-3.285	1.0791	29	0.84	0.7056
30	-1.285	1.651	30	1.84	3.3856
32	0.715	0.511	30	1.84	3.3856
33	1.715	2.941	24	-4.16	17.3056
33	1.715	2.941	27	-1.16	1.3456
29	-2.285	5.221	29	0.84	0.7056
34	2.715	7.371	-	-	-
		31.427			26.833

The
gua
valu
we
Ho
two
sam
2) A
trea
wei
2^{nd}
from
with
~~ke~~
Do
cla
incr

$$s^2 = \frac{1}{10}(1216) = 121.6$$

$$S = \sqrt{121.6} = 11.03$$

l.o.s $\alpha = 0.05$

The test statistics

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{46 - 57}{11.03\sqrt{\frac{1}{5} + \frac{1}{7}}}$$

$t = -1.7$

$|t| = |-1.7|$

$t = 1.7$

The tabulated value of t $n_1 + n_2 - 2 = 5 + 7 - 2 = 10$ df at 0.05 is 2.228

x	x
117	
105	
97	
105	
123	
109	
86	
78	
103	
107	

Null

H_0: H

Alte

H_1:

l.o.s

The

	$(y-\bar{y})^2$
0.2	104.04
.2	4.84
8.8	77.44
8.2	67.24
0.2	408.04
0.8	0.64
5.8	33.64
6.8	718.24
.2	148.84
0.8	116.64
	$\Sigma(y-\bar{y})^2 = 1679.6$

Q, To compare two kinds of bumper guards 6 of each ki[nd] were mounted on a car an[d] then the car was run int[o a] concrete wall. The following [are] the cost of repairs

Bumper guard 1	107	148	123	165	102	[illegible]
Bumper guard 2	134	115	112	151	133	[illegible]

Use the 0.01 L.O.S to test whether the difference between the means of these two sam[ples] is significant. $t_{\alpha/2} = t_{\frac{0.01}{2}} = t_{0.005} = 3.16$[illegible]

s

Given $n_1 = 6$, $n_2 = 6$

$$\bar{x} = \frac{107+148+123+165+102+119}{6}$$

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$\bar{x} = 127.3$

$$\bar{y} = \frac{134+115+112+151+133+129}{6}$$

$+\Sigma(y_i-\bar{y})^2]$

$\bar{y} = 129$

L.O.S $\alpha = 0.05$

The test statistics $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Here $S^2 = \dfrac{1}{n_1 + n_2 - 2}\left[\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2\right]$

$= \dfrac{1}{8 + 7 - 2}\left[26 + 16\right]$

$= \dfrac{1}{13}\left[42\right] = 3.23$

$S = 1.797$

$t = \dfrac{12 - 10}{1.797\sqrt{\frac{1}{8} + \frac{1}{7}}} = 4.168$

$t = 4.168$

$n_1 + n_2 - 2 = 13$ df at 0.025 is 2.160

$t_{cal} > t_{tab}$

So we reject the null Hypothesis H_0

where $\bar{d}$

$S^2 = \Sigma$

Womens	Bl be of
1	
2	
3	
4	
5	

$n = 5$

$\bar{d} = \dfrac{\Sigma d}{n}$

$S^2 = \Sigma$

$= 14$

$S = \sqrt{3}$

$x - y$

$)^2$

$— \times n$

$-x$	d^2
	100
	4
	4
	16
	16
10	$\Sigma d^2 = 140$

Null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis $H_1 : \mu_1$

$1.0.S \quad \alpha = 5\%$

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} \text{ where } \bar{d} = \frac{\Sigma d}{n} \text{ \&}$$

$d = x - y,$

$$s^2 = \frac{\Sigma (d - \bar{d})^2}{n-1} = \frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}$$

Before training (x)	After training (y)	$d = y - x$	d^2
12	15	3	9
14	16	2	4
11	10	-1	1
8	7	-1	1
7	5	-2	4
10	12	2	4
3	10	7	49
0	2	2	4
5	3	-2	4
6	8	2	4
		$\Sigma d = 12$	$\Sigma d^2 =$

differ significantly.

let the null Hypothesis $H_0 : \sigma_1^2 = \sigma_1^2$

Alternative Hypothesis

$H_1 : \sigma_1^2 \neq \sigma_2^2$

x	$x-\bar{x}$ $x-22.3$	$(x-\bar{x})^2$	y	$(y-\bar{y})$ $y-34.4$	$(y-\bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
–	–	–	38	3.6	12.96
$\Sigma x = 134$		81.34			133.72

$\bar{x} = \frac{\Sigma x}{n} = \frac{134}{6} = 22.3$

$\bar{y} = \frac{\Sigma y}{n} = \frac{241}{7} = 34.4$

$n_1 = 6, \; n_2 = 7$

$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{\ldots} = \frac{81.34}{\ldots} = 16.26$

let the

$H_0 : \sigma_1$

Alterna

$H_1 : \sigma_1^2$

$\bar{x} = \frac{\Sigma x}{n_1}$

$= 24.$

$\bar{y} = \frac{\Sigma y}{n_2}$

$= 29$

x	$x-\bar{x}$
24	-0.6
27	2.4
26	1.4
21	-3.6
25	0.4
–	

$n_1 = 5$

$S_1^2 = \frac{\Sigma(\ldots}{n \ldots}$

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with } (n-1) \text{ df}$$

Q) The no. of automobile accidents per week in a certain community are as follows 12, 8, 20, 2, 14, 10, 15, 16, 9, 4. All these frequencies in agreement with the belief that accident conditions were same during these 10 week period

~~Excepte~~

Expected frequency of accidents each week $= \frac{12+8+20+2+14+10+15+6+9+4}{10}$

$= \frac{100}{10} = 10$, Ho: The accident condition was same during this 10 week period

observed frequency O_i	Expected frequency E_i	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	$\frac{4}{10} = 0.4$
8	10	-2	0.4
20	10	10	10
2	10	-8	6.4
14	10	4	1.6

x	0	1	2	3	4	Total
f	404.9	366.02	165.4	49.8	11.8	997.4

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i}$$

$$= \frac{(419-404.9)^2}{404.9} + \frac{(352-366)^2}{366} + \frac{(154-165.4)^2}{165.4} + \frac{(56-49.8)^2}{49.8} + \frac{(19-12.6)^2}{12.6}$$

$$\chi^2 = 5.948$$

$D.f = n-2 = 5-2 = 3$

$\chi^2_{0.05}$ 3 df $= 7.82$

$\chi^2 cal < \chi^2_{Tab}$

We accept the Null Hypothesis H_0.

Binomial Distribution.

Q)

x	4
$f(x)$	1

O_i
10
55
105
58
12

$\chi^2 =$

$D.f = n$

χ^2_{Tab}

χ^2_{cal}

We a

at 5

χ^2 test

$\frac{(O_i-E_i)^2}{E_i}$
1.66
0.41
2.5
0.06
0.6

23.

ypothesis

attributes

the

workers

Calculation of x^2

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
40	30	100	100/30 = 3.333
20	30	100	3.333
10	20	100	5
30	20	100	5
			16.666

$$x^2 = \frac{\Sigma(O_i-E_i)^2}{E_i} = 16.666$$

Tabulated value of x^2 ~~for~~ (2-1)

(2-1) = 1 df at 5%

l.o.s is 3.84.

$x^2_{cal} > x^2_{Tab}$. we reject H_0 at 5% l.o.s

6. from the following data find whether there is any significant liking in the habit of taking soft drinks among the categories of employees. Use x^2 distribution with the level of significance 0.05%.

Table of Expected frequencies

$\frac{700\times600}{1000}=420$	$\frac{300\times600}{1000}=180$	600
$\frac{700\times400}{1000}=280$	$\frac{300\times400}{1000}=120$	400
700	300	1000

calculations for x^2

O_i	E_i	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
460	420	1600	3.809
140	180	1600	8.880
240	280	1600	5.714
160	120	1600	13.33
			31.733

$x^2_{cal} = \frac{\Sigma(O_i-E_i)^2}{E_i} = 31.733$

x^2_{tab} $(2-1)(2-1) = 1$ df is at 5%.

L.O.S 3.84

$x^2_{cal} > x^2_{tab}$

2) Altern

$P_1 \neq P_2$

3) L.O.S

4) comp

Super conductor

Failures

Total

Table

$\frac{50\times120}{200}=3$

$\frac{50\times80}{200}=2$

50

Calcu