19/6/18

PROBABILITY

Introduction :-

Probability and Statistics are two separate academic disciplines which are studied together.

Statistical analysis use probability distributions. where in probability theory contains more of mathematics. However, there are topics in statistics which are independent of probability.

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. Descriptive statistics involves methods of Organizing, presentating and summarizing information from data. Inferential statistics involves methods of using information from a sample to draw conclusions about the population.

Probability is a measure of how likely something will happen (or) that statement is true. If a high degree of probability is present, it is more likely that the event is to happen. Many events cannot be predicted with total certanity.

Probability :

The probability of an given event is an expression of likely hood of occurrence of an event.

Random experiment :-

A random experiment is an experiment or a proces for which the outcome cannot be predicted with certainity.

Experiment :-

An experiment is the process of making an observations / taking a measurement which results in different possible outcomes.

Ex:-

→ tossing a coin

→ throwing a dice

→ drawing a card from deck of cards

Trial and event :-

Any particular performance of a random experiment is called trial and outcomes of an experiment is called event.

Ex:-

→ throwing a dice is a trial and occurrence of 1, 2, 3, ..., 6 are called outcomes / events.

Sample Space :-

Set of all possible outcomes in an experiment is called Sample Space.

Ex:-

→ Flip a coin, the possible outcomes are head and tail

Outcome :-

A possible result of a probability experiment is called an outcome.

Ex :-

→ when we toss a coin, head is the possible result i.e outcome or may be tail is the outcome ∴ head, tail are the outcomes

Exhaustive events :-

The total no. of possible outcomes of a random experiment is known as an exhaustive event.

Ex:-

→ In tossing a coin, there are 2 exhaustive events namely head and tail.

→ In throwing of a dice, there are 6 exhaustive events

Favourable events :-

The no. of events favourable to an event in a trial is the no. of outcomes which favours the happening of the event.

Ex:-

→ In throwing of 2 dices, the no. of events favourable to get a some sum 5 in the two faces are 2,3 ; 3,2 ; 1,4 ; 4,1

Mutually exclusive events :-

Events are said to be mutually exclusive if they cannot occur simultaneously in the same trial (i.e the occurrence of one (1) prevents the occurrence of the other)

Ex:-

→ In tossing a coin, the head and tail are mutually exclusive events

Equally likely events :-

Events are said to be equally likely if any one of them cannot be expected to occur in preference to the others.

Ex:-

→ In tossing a coin, head and tail are equally likely events.

Independent event :-

Events are said to be independent if the occurrence of one (1) does not effect occurrence of the other.

Ex:-

→ An drawing a diamond card from a pack of well shuffled cards is $\frac{13}{52}$ and that of another diamond card after replacing the first one is also $\frac{13}{52}$. Here the result of Second trial not affected by the result of first trial since the card was replaced.

Dependent event :-

Events are said to be dependent, if the occurrence of one affects the occurrence of other

Ex:-

→ An drawing a diamond card from a pack of playing Cards is $\frac{13}{52}$ while that of another diamond card when first being not replaced these is $\frac{12}{51}$ Since the card was not replaced before second experiment.

21/6/18

Mathematical or Classical Probability :-

If a random experiment or a trial results in small n exhaustive, mutually exclusive and Equally likely outcomes out of which m are favourable outcomes to the occurrance of event, then the probability (p) of occurrance of E, usually denoted by P(E) is given by P(E) =

$$P(E) = \frac{\text{No. of favourable outcomes}}{\text{Total no. of outcomes}} = \frac{m}{n}$$

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

NOTE: • The probability always lies between 0 to 1.

• The universal equation of probabity $\underset{\text{success}}{P} + \underset{\text{failure}}{q} = 1$

- If $P(E) = 1$, then E is called a certain event.
- If $P(E) = 0$, then E is called impossible event.

Problems

Q. what is the change a leap year selected at random will contain 53 Sundays?

Sol: In a leap year, there are 366 days i.e 52 weeks + 2 days over. The following are the possible combinations for these 2 days are :- Sunday, monday,
- mon, tues
- tues, wed
- wed, thurs
- thurs, fri
- fri, sat
- Sat, sun

In order that a leap year selected at random should contain 53 Sundays. One of two over days must be Sunday since out of the seven possibilites two are favourable to this event.

$\therefore P(E) = \frac{2}{7}$

Q. A bag contain 3 red, 6 white and 7 blue balls. what is the probability that the 2 balls drawn are white and blue?

Sol: The total no. of balls are 16 from which two balls can be drawn in 16_{C_2} ways. i.e $n = 16_{C_2} = \frac{\overset{8}{\cancel{16}} \times 15}{\cancel{2} \times 1}$

$= 120$

$m = 6_{C_1} \times 7_{C_1} = 6 \times 7 = 42$

$P(E) = \frac{m}{n} = \frac{42}{120} = \frac{21}{60}$.

Q. Two cards are drawn at random from a well-shuffled pack of cards, show that the chance of drawing of two aces is $\frac{1}{221}$.

$\frac{^4C_2}{^{52}C_2}$

Sol: Total no. of cards = 52

Total no. of aces in pack of cards = 4

Total No. of outcomes = $^{52}C_2$ (n)

$^nC_r = \frac{n!}{r!(n-r)!}$

∵ Two cards are drawn

No. of favourable outcomes (m) = $^4C_2$

$\frac{4\times3}{2\times1}$

$\therefore\ P(E) = \frac{m}{n} = \frac{^4C_2}{^{52}C_2} = \frac{\cancel{4}\times\cancel{3}}{\underset{13}{\cancel{52}}\times\underset{17}{\cancel{51}}} = \frac{1}{221}.$

Q. From 25 ~~decord~~ tickets marked with the first 25 numbers, one is drawn at random. Find the chance that (i) it is a multiple of 5 or 7 (ii) it is a multiple of 3 or 7.

Sol: From 25 tickets, one ticket can be drawn in $^{25}C_1$

i.e $n = 25$.

(i) multiple of 5 or 7 :

multiples of 5 are 5, 10, 15, 20, 25 i.e = 5

7 are 7, 14, 21 i.e = 3

$P(E) = \frac{m}{n} = \frac{5+3}{25} = \frac{8}{25}$

(ii) multiple of 3 or 7 :

multiples of 3 are 3, 6, 9, 12, 15, 18, 21, 24 i.e = 8

7 are 7, 14, 21 i.e = 3

$P(E) = \frac{m}{n} = \frac{8+2}{25} = \frac{10}{25} = \frac{2}{5}$

Scanned by CamScanner

Q. 4 cards are drawn from a pack of cards. Find the probability that
(i) All are diamonds.
(ii) There is one card of each suite and (iii) there are
(iii) 2 spades & 2 hearts.

Sol: (i) Total no. of cards = 52 $\quad \dfrac{^{13}C_4}{^{52}C_4}$

Total no. of diamonds = 13

4 cards are drawn

$\therefore \quad n = {}^{52}C_4$

$m = {}^{13}C_4$

$\therefore \quad P(E) = \dfrac{m}{n} = \dfrac{^{13}C_4}{^{52}C_4} = \dfrac{11}{4165}$

(ii) Total no. of cards = 52

Total no. of cards in each suite = 13

One card is drawn from each suite

total 4 cards are drawn

$\therefore \quad n = {}^{52}C_4$

$m = {}^{13}C_1 \, {}^{13}C_1 \, {}^{13}C_1 \, {}^{13}C_1$

$\therefore \quad P(E) = \dfrac{m}{n} = \dfrac{^{13}C_1 \, {}^{13}C_1 \, {}^{13}C_1 \, {}^{13}C_1}{^{52}C_4} = \dfrac{2197}{20825}$

(iii) Total no. of cards = 52

No. of spades = 13

No. of hearts = 13

2 are drawn from spades, 2 are drawn from hearts

$\therefore \quad n = {}^{52}C_4$

$m = {}^{13}C_2 \, {}^{13}C_2$

Scanned by CamScanner

$$P(E) = \frac{m}{n} = \frac{{}^{13}C_2 \cdot {}^{13}C_2}{{}^{52}C_4} = \frac{468}{20825}$$

22/6/18

Q. Two dice are thrown, find the probability that (i) both the dice show the same number

(ii) first dice shows six

(iii) The total of the no's on the dice is 8

(iv) The sum of no's on the dice is greater than 9

(v) The total of the no's on the dice is 13.

(vi) The total of the no's on the dice is any number from 2 to 12. both inclusive.

Sol (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)

Total no. of outcomes $= 6^2 = 36$
$(n) = 36$

(i) $m = 6$

$$P(E) = \frac{m}{n} = \frac{6}{36} = \frac{1}{6}$$

(ii) $m = 6$

$$P(E) = \frac{m}{n} = \frac{6}{36} = \frac{1}{6}$$

(iii) $m = 5$

$$P(E) = \frac{m}{n} = \frac{5}{36}$$

(iv) $m = 6$

$$P(E) = \frac{m}{n} = \frac{6}{36} = \frac{1}{6}$$

(v) $m = 0$

$P(E) = \frac{m}{n} = \frac{0}{36} = 0$

(vi) $m = 36$

$P(E) = \frac{36}{36} = 1.$

Q. An urn contains 6 white, 4 red, 9 black balls. If 3 balls are drawn at random. Find the probability that (i) 2 of the balls are white.

(ii) one is of each color.

(iii) None is red

(iv) Atleast one is white.

$6+4+9 = 19$

${}^{19}C_3$

${}^6C_2 \cdot {}^{13}C_1$

Sol The total no. of balls are $6+4+9 = 19$.

3 balls can be drawn out of 19.

i.e the exhaustive no. of outcomes $(n) = {}^{19}C_3$
$= 969$

(i) If 2 balls of the 3 drawn balls are to be white, these 2 balls should be drawn out of 6 white balls which can be done in ${}^6C_2$ ways., the remaining balls i.e 13 can be drawn in ${}^{13}C_1$ ways.

$m = {}^6C_2 \times {}^{13}C_1$

$\therefore P(E) = \frac{m}{n} = \frac{{}^6C_2 \times {}^{13}C_1}{969} = \frac{65}{323}.$

(ii) $m = {}^6C_1 \times {}^4C_1 \times {}^9C_1$

$\because$ one is of each color

$\therefore P(E) = \frac{m}{n} = \frac{{}^6C_1 \times {}^4C_1 \times {}^9C_1}{969} = \frac{216}{969} = \frac{72}{323}.$

(iii) If none of the balls are red, then all the 3 balls must be white and black balls.

$6+9=15.$

$\therefore \; m = {}^{15}C_3$

$$P(E) = \frac{m}{n} = \frac{{}^{15}C_3}{969} = \frac{455}{969}.$$

(iv) $P(E) = 1 - p(\text{none of white})$

$$= 1 - \frac{{}^{13}C_3}{{}^{19}C_3}$$

$$= 1 - \frac{286}{969}$$

$$= \frac{683}{969}.$$

Q. 3 coins are tossed. what is the Probability of (i) getting atmost 2 heads?

Sol (H,H,H) (T,T,T) (H,T,T) (H,T,H) (H,H,T) (T,H,H) (T,H,T) (T,T,H)

$n = 8$

$m = 7$

$\therefore \; P(E) = \frac{7}{8}.$

(ii) getting atleast 2 heads?

$m = 4$

$n = 8$

$\therefore \; P(E) = \frac{4}{8} = 1/2.$

Q. A card is drawn at random from a deck of cards. Find the probability of getting King of heart.

$m = {}^4C_1$

$n = {}^{52}C_1$

$\therefore P(E) = \frac{{}^4C_1}{{}^{52}C_1} = \frac{4}{52} = \frac{1}{13}$.

The committe of 4 people is to be appointed from 3 officers of the production department, 4 officers of the purchase department, 2 officers of the sales department and 1 C.A. Find the probability of forming the committe in the following manner.

i) There must be 1 from each category

ii) It should have atleast one from the purchase department

iii) The C.A must be in the committe.

$3+2+4+1 = 10$

No. of exhaustive events (total) $(n) = {}^{10}C_4$

$= 210$

i) Since there must be 1 from each category,

$m = {}^3C_1 \times {}^4C_1 \times {}^2C_1 \times {}^1C_1$

$\therefore p(E) = \frac{m}{n} = \frac{\text{No. of favourable events}}{\text{No. of total events}}$

$= \frac{{}^3C_1 \times {}^4C_1 \times {}^2C_1 \times {}^1C_1}{{}^{10}C_4}$

$= \frac{24}{210} = \frac{4}{35}$

(ii) prob. of committee has atleast one purchase dept officer

$= 1 - p$ (Committee has no purchase officer)

$= 1 - \frac{{}^6C_4}{{}^{10}C_4}$

$= 1 - \frac{1}{14} = \frac{13}{14}$

Scanned by CamScanner

(iii) The C.A must be in committee $= {}^{1}C_1$.

remaining ${}^{9}C_3$

$\therefore \quad m = {}^{1}C_1 \times {}^{9}C_3$

$$P(E) = \frac{m}{n} = \frac{{}^{1}C_1 \times {}^{9}C_3}{{}^{10}C_4}$$

$$= \frac{2}{5}$$

Axioms of probability :-

1. The probability of an event is non-negative real number.

$P(A) \geq 0$

2. The sum of probability of all sample events is unity

$P(S) = 1$

3. A and B are mutually exclusive and exhaustive events, then $P(A \cup B) = P(A) + P(B)$ $\left(\because A \cap B = \phi,\ A \cup B = S\right)$

Theorem-1 :-

* The probability of impossible event is 0 i.e $P(\phi) = 0$

Proof :-

we know that an impossible event contains no sample point and hence certain event S and the impossible event ϕ are mutually exclusive

$$S \cup \phi = S$$

$$P(S \cup \phi) = P(S)$$

$$P(S) + P(\phi) = P(S)$$

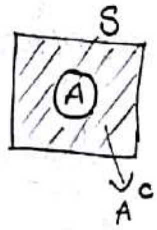
$$P(\phi) = P(S) - P(S)$$

$$P(\phi) = 0$$

Theorem-2 :-

* The probability of the complimentary event $\bar{A}$ (or) A^c (or) A' is given by $P(\bar{A}) = 1 - P(A)$.

Proof :-



disjoint
A and A^c are mutually designed events

$\therefore \quad A \cup A^c = S$

$P(A \cup A^c) = P(S) \quad \langle \because$ axioms of 2 & 3 $\rangle$

$P(A) + P(\bar{A}) = 1$

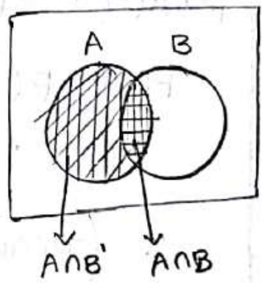
$P(\bar{A}) = 1 - P(A)$

Theorem-3 :-

* For any 2 events A & B is

(i) $P(A \cap B') = P(A) - P(A \cap B)$

(ii) $P(A' \cap B) = P(B) - P(A \cap B)$



Proof :-

(i) A & B are any two events

let $(A \cap B') \cup (A \cap B) = A$ &

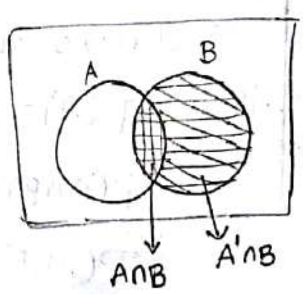
$(A \cap B') \cap (A \cap B) = \phi$

$P(A \cap B') \cup (A \cap B) = P(A)$

$P(A \cap B') + P(A \cap B) = P(A)$

$P(A \cap B') = P(A) - P(A \cap B)$

(ii) $P(A' \cap B) = P(B) - P(A \cap B)$



$(A' \cap B) \cup (A \cap B) = B$ &

$(A' \cap B) \cap (A \cap B) = \phi$

$P(A' \cap B) \cup (A \cap B) = P(B)$

$P(A' \cap B) + P(A \cap B) = P(B)$

$P(A' \cap B) = P(B) - P(A \cap B)$

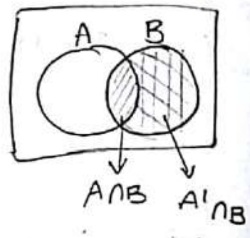
Addition theorem :-

* If A & B are any 2 events and they are not disjoint (mutually exclusive) then $P(A\cup B) = P(A) + P(B) - P(A\cap B)$

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

Proof :-

Given that A & B are any 2 events and they are not disjoint.



From the diagram,

$A\cap(A'\cap B) = \phi$ and

$A\cup(A'\cap B) = A\cup B$

$P(A\cup(A'\cap B)) = P(A\cup B)$

$P(A) + P(A'\cap B) = P(A\cup B)$ $\langle \because$ from 3^{rd} theorem$\rangle$

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

$\therefore \boxed{P(A\cup B) = P(A) + P(B) - P(A\cap B)}$

27/6/18

Problems :-

Q. The probability that the students pass physics test is $2/3$ and the probability that he pass both physics and english test is $14/45$. The probability that he pass atleast one test is $4/5$. What is the probability that he pass english test?

Sol: Let A denotes the student pass physics test and B denotes the student pass english test

then $P(A) = 2/3$

$P(A\cap B) = 14/45$

$P(A\cup B) = 4/5$

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

$4/5 = 2/3 + P(B) - 14/45$

$\frac{4}{5} + \frac{14}{45} - \frac{2}{3} = P(B)$

$\Rightarrow P(B) = \frac{4}{9}$ //

$\frac{36+14}{45}$

$\frac{50}{45} - \frac{2}{3}$

$\frac{10-6}{9} = \frac{4}{9}$

Q. If two dices are thrown, what is the probability that the sum is
(i) greater than 8.
(ii) Neither 7 nor 11.

Sol: Two dices are thrown, the exhaustive events $(n) = 36$

(i) $P(S>8) = P(S=9) + P(S=10) + P(S=11) + P(S=12)$

$S=9$;	(3,6) (4,5) (5,4) (6,3)	$m=4$
$S=10$;	(4,6) (5,5) (6,4)	$m=3$
$S=11$;	(5,6) (6,5)	$m=2$
$S=12$;	(6,6)	$m=1$

$P(S>8) = \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36}$

$= \frac{10}{36}$

let S represent the sum on the 2 ~~days~~ dices.

(ii) let A denotes 7 (the event of getting the sum 7)
B denotes 11 (the event of getting the sum 11)

$S=7$.

$S=7$; (4,3) (3,4) (2,5) (5,2) (1,6) (6,1) $m=6$.

$P(A) = \frac{6}{36} = \frac{1}{6}$

$S=11$; (5,6) (6,5) $m=2$

$P(B) = \frac{2}{36} = \frac{1}{18}$

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

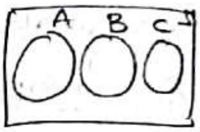
$= 1 - P(A \cup B)$

$= 1 - [P(A) + P(B)]$

$= 1 - \left[\frac{1}{6} + \frac{1}{18}\right] = 1 - \frac{4}{18} = 7/9$ //

Q. If A, B, C are mutually exclusive and exhaustive events in random experiment. Find P(A), P(B) & P(C). Given than $P(B) = \frac{3}{2} P(A)$ and $P(C) = \frac{1}{2} P(B)$

Sol Given that,
A, B, C are mutually exclusive and exhaustive events.



$P(S) = P(A) + P(B) + P(C)$

Given, $P(B) = \frac{3}{2} P(A) \rightarrow (1)$ $\because P(S) = 1$

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

from (1), $P(C) = \frac{1}{2}\left[\frac{3}{2} P(A)\right]$

$= \frac{3}{4} P(A)$

$\therefore \quad P(S) = P(A) + \frac{3}{2} P(A) + \frac{3}{4} P(A)$

$1 = \frac{4 P(A) + 6 P(A) + 3 P(A)}{4}$

$1 = \frac{13 P(A)}{4}$

$\Rightarrow 13 \frac{P(A)}{4} = 1$

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

$P(B) = \frac{3}{\cancel{2}} \times \frac{\cancel{4}^{2}}{13} = \frac{6}{13}$

$P(C) = \frac{3}{\cancel{4}} \times \frac{\cancel{4}}{13} = \frac{3}{13}$

Q. An integer is chosen at a random from 5 hundred digits, what is the probability that the integer is divisible by 5/50.

Sol Given,
Sample Space $S = \{1, 2, 3, \ldots, 500\}$
$n = 500$

Let A denotes the event that the integer chosen is divisible by 5.

B denotes the event that the integer chosen is divisible by 50.

No. of integers divisible by $5 = \frac{500}{5} = 100$

No. of integers divisible by $50 = \frac{500}{50} = 10.$

The no. of integers divisible by 5 & 50.

$P(A) = \frac{\cancel{100}}{\cancel{500}} = \frac{1}{5}$

$P(B) = \frac{10}{500} = \frac{1}{50}$

$P(A \cap B) = \frac{10}{500} = \frac{1}{50}$

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

$= \frac{1}{5} + \frac{1}{50} - \frac{1}{50}$

$= \frac{1}{5}$ //.

28/6/18

Q. 3 news papers A, B, C are published in a certain city. It is estimated from a survey of population 20% reads A, 16% reads B, 14% reads C, 8% read both A and B, 5% read both A and C, 4% read B and C. and 2% read all the news paper. Find the percentage of persons read atleast one of the newspaper.

Sol $P(A) = \frac{20}{100}$ $\qquad P(A \cap C) = \frac{5}{100}$

$P(B) = \frac{16}{100}$ $\qquad P(B \cap C) = \frac{4}{100}$

$P(C) = \frac{14}{100}$ $\qquad P(A \cap B \cap C) = \frac{2}{100}$

$P(A \cap B) = \frac{8}{100}$

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{4}{100} - \frac{5}{100} + \frac{2}{100}$$

$$= \frac{35}{100}$$

$$= 35\%$$

Q. If $P(A) = P_1$, $P(B) = P_2$, $P(A\cap B) = P_3$ then express the following in terms of P_1, P_2 and P_3

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

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

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

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

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

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

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

Sol

$\{\because P(A^c) = 1-p(A)$

$P(A^c\cap B) = P(B) - P(A\cap B)$

1. $P(\overline{A\cup B}) = 1-p(A\cup B)$

$$= 1- [p(A)+p(B) - p(A\cap B)]$$

$$= 1 - P(A) - p(B) + p(A\cap B) = 1-P_1-P_2+P_3$$

2. $P(\bar{A}\cap B) = P(B) - p(A\cap B)$

$$= P_2 - P_3$$

3. $P(\bar{A}\cup\bar{B}) = P(\overline{A\cap B}) = 1 - p(A\cap B) = 1-P_3$

$$= \cancel{1} - P(A) - P(B) + P(A\cup B)$$

$$\neq A$$

4. $P(\bar{A}\cup B) = P(\bar{A}\cup\bar{\bar{B}}) = P(\overline{A\cap\bar{B}}) = 1-P(A\cap\bar{B})$

$$= 1 - P(A) + P(A\cap B)$$

$$= 1 - P_1 + P_3$$

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

$= 1 - P(A\cup B) = 1 - P_1 - P_2 + P_3$

6. $P(A\cap\bar{B}) = P(A) - P(A\cap B)$

$= P_1 - P_3$

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

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

$= \cancel{1-}\cancel{P_1} + P_2 - P_3 - \cancel{P_1} - P_3. = P_2 - 2P_3$

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

$0 + P(B) - P(A\cap B)$

$= \boxed{P_2 - P_3}$

Conditional Probability :-

The probability of an event occuring given that another event has already occured is a conditional probability i.e the probability that event 'B' occurs given that event A has already occured is $P(B/A) = \cancel{P(A)\cdot} P(A\cap B)/P(A)$

Ily, $P(A/B) = \cancel{P(B)\cdot} P(A\cap B)/P(B)$

Multiplication theorem :-

Statement :- If A and B are any 2 events in a Sample Space (S) then $P(A\cap B) = P(A)\cdot P(B/A)$, $P(A) > 0$

$P(A\cap B) = P(B)\cdot P(A/B)$, $P(B) > 0$

Proof :-

Given that A and B are any 2 events in a sample (S) by using conditional probability definition.

$P(A/B) = \cancel{P(B)\cdot} P(A\cap B)/P(B)$

$\Rightarrow \frac{P(A\cap B)}{P(B)} = P(A/B)$

$P(A\cap B) = P(B)\cdot P(A/B)$

Similarly, $P(B/A) = \frac{P(A\cap B)}{P(A)}$

$$P(A\cap B) = P(A)\cdot P(B/A)$$

Hence Proved.

Case 1 :- If A and B are said to be independent, if the occurence of one event does not affect on the occurence of the other event i.e

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

29/6/18

Q. A problem in statistics is given to 3 students A, B, C. whose chance of solving it are $1/2$, $3/4$ and $1/4$ respectively. what is the probability that problem will be solved. If all of them try independently

Sol Let A, B, C be the events that three students A, B, C solve the problem independently.

Probability of A solving the problem is $P(A) = 1/2$

" " B solving the problem is $P(B) = 3/4$

" " C solving the problem is $P(C) = 1/4$

Probability that 3 students will solve the problem if they try independently?

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

$$= \frac{1}{2}+\frac{3}{4}+\frac{1}{4} - [P(A)\cdot P(B)] - [P(B)\cdot P(C)] - [P(C)\cdot P(A)] + [P(A)\cdot P(B)\cdot P(C)]$$

$$= 1/2 + 1 - \left[\frac{1}{2}\cdot\frac{3}{4}\right] - \left[\frac{3}{4}\cdot\frac{1}{4}\right] - \left[\frac{1}{4}\cdot\frac{1}{2}\right] + \left[\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{4}\right]$$

$$= \frac{3}{2} - \frac{3}{8} - \frac{3}{16} - \frac{1}{8} + \frac{3}{32} = \frac{48-12-6-4+3}{32} = \frac{29}{32}$$

Q. The odds in favour of A solving a problem are $3 \text{ to } 4$ and the odds against B solving a Problem are $5 \text{ to } 7$. Find the probability that the problem will be solved by atleast one of them.

Sol The odds in favour of A Solving a problem $P(A) = \frac{3}{7}$

The odds in against of B Solving the problem $P(B) = \frac{7}{12}$

P(A can solve the Problem) $P(A) = \frac{3}{7}$

P(B can Solve the problem) $P(B) = \frac{7}{12}$.

A and B can Solve independently

The probability that both can Solve problem

$$P(A \cap B) = P(A) \cdot P(B)$$

$$= \frac{3}{\cancel{7}} \cdot \frac{\cancel{7}}{12} = \frac{1}{4}.$$

probability that the problem will be Solved by atleast one of them $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{3}{7} + \frac{7}{12} - \frac{1}{4}$$

$$= \frac{36+49}{7 \cdot 12} - \frac{1}{4}$$

$$= \frac{85}{84} - \frac{1}{4} = \frac{16}{21}$$

** Baye's Theorem :-

Statement:- If $E_1, E_2, E_3 \ldots E_n$ are mutually disjoint events with $P(E_i) \neq 0$ for $(i = 1, 2, 3, \ldots n)$ then for any arbitary event A which is a subset of $\bigcup_{i=1}^{n} E_i$ with $P(A) > 0$ i.e

$$P\left(\frac{E_i}{A}\right) = \frac{P(E_i) \cdot P\left(\frac{A}{E_i}\right)}{\sum_{i=1}^{n} P(E_i)\, P\left(\frac{A}{E_i}\right)}$$

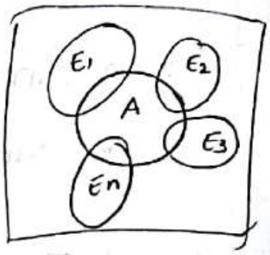
Proof :-

Given that $E_1, E_2, \dots E_n$ are mutually disjoint events and also it is given that $A \subset \bigcup_{i=1}^{n} E_i$

$$A = A \cap \left(\bigcup_{i=1}^{n} E_i\right)$$

Since $(A \cap E_i) \subset E_i$ $(i = 1, 2, \dots n)$ are mutually disjoint events by addition theorem of probability.

$$A = \bigcup_{i=1}^{n} (A \cap E_i)$$



$$P(A) = P\left[\bigcup_{i=1}^{n} (A \cap E_i)\right]$$

$$= \bigcup_{i=1}^{n} P(A \cap E_i)$$

$$= \sum_{i=1}^{n} P(A \cap E_i). \qquad \because P(A \cap E_i) = P(E_i) P\left(\frac{A}{E_i}\right)$$

$$= \sum_{i=1}^{n} P(E_i) P\left(\frac{A}{E_i}\right) \longrightarrow (1) \quad (\because \text{by multiplication theorem})$$

30/6/18

$$P\left(E_i/A\right) = \frac{P(E_i) \cdot P(A/E_i)}{P(A)}$$

$$= \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) P(A/E_i)}$$

Q. Three bags contain one white, 2 red, 3 green ; 2 white, 1 red, 1 green ; 4 white, 5 red, 3 green balls. Two balls are drawn from a bag choose at random. These are found to be 1 white and 1 green. Find the Probability that the balls drawn from the third bag.

Let B_1, B_2, B_3 be the 3 bags and A be the event that getting one white and one green ball from a bag.

$P(B_1) = 1/3$

$P(B_2) = 1/3$

$P(B_3) = 1/3$

$$P(A/B_1) = \frac{{}^1C_1\ {}^3C_1}{{}^6C_2} = 1/5$$

$$P(A/B_2) = \frac{{}^2C_1\ {}^1C_1}{{}^4C_2} = 1/3$$

$$P(A/B_3) = \frac{{}^4C_1\ {}^3C_1}{{}^{12}C_2} = \frac{2}{11}$$

Probability of 1 white and 1 green from third bag.

$$P(B_3/A) = \frac{P(B_3)\ P(A/B_3)}{\sum_{i=1}^{3} P(B_i)\ P(A/B_i)}$$

$$= \frac{1/3 \times \frac{{}^4C_1\ {}^3C_1}{{}^{12}C_2}}{\sum_{i=1}^{3} 1/3 \times \frac{{}^4C_1\ {}^3C_1}{{}^{12}C_2} + \frac{1}{3} \times \frac{{}^1C_1\ {}^3C_1}{{}^6C_2} + \frac{1}{3} \times \frac{{}^2C_1 \times {}^1C_1}{{}^4C_2}}$$

$\frac{12}{11}$

$\frac{118}{165}$ 15

$\cancel{2} \times \frac{165}{\cancel{118}}$ 59

$$= \frac{\frac{1}{3} \times \frac{2}{11}}{\frac{1}{3} \times \frac{2}{11} + \frac{1}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{3}} = \frac{\frac{1}{3} \times \frac{2}{11}}{\frac{1}{3}\left(\frac{2}{11} + \frac{1}{5} + \frac{1}{3}\right)} = \frac{15}{59}$$

$\frac{1}{3} \times \frac{165}{118} = \frac{33}{118}$

$\frac{55}{118}$

Q. A factory produce certain type of products by 3 types of machines. The respective daily productions as mission 1 3000 units machine 2 2500 units machine 3 4500 units. First experiment choose that 1% of the output produced by machine is defective. The corresponding defectives of other 2 machines are 1.2% and 2% resp. An item is drawn at random from the days production and is found to be defective. What is the probability that it comes from the o/p of 1^{st}, 2^{nd} and 3^{rd} machines.

Sol

		defective
M_1	3000	1 %
M_2	2500	1·2 %
M_3	4500	2 %
	10000	

$$P(M_1) = \frac{3000}{10,000} = 0.3$$

$$P(M_2) = \frac{2500}{10,000} = 0.25$$

$$P(M_3) = \frac{4500}{10,000} = 0.45$$

Let D be the defective item

Prob. that output defective from M_1 $P(D/M_1) = 0.01$

Prob. that output defective from M_2 $P(D/M_2) = 0.012$

Prob. that output defective from M_3 $P(D/M_3) = 0.02$

$$P(M_1/D) = \frac{P(M_1)\,P(D/M_1)}{\sum_{i=1}^{3} P(M_i)\,P(D/M_i)}$$

$$= \frac{0.3(0.01)}{0.3\times 0.01 + 0.25\times 0.012 + 0.45\times 0.02}$$

$$= \frac{3\times 10^{-3}}{3\times 10^{-3} + 3\times 10^{-3} + 9\times 10^{-3}} = \frac{3\times 10^{-3}}{3\times 10^{-3}(1+1+3)}$$

$$= \frac{1}{5}$$

$$P\left(\frac{M_2}{D}\right) = \frac{3\times 10^{-3}}{3\times 10^{-3}(1+1+3)} = \frac{1}{5}$$

$$P\left(\frac{M_3}{D}\right) = \frac{\overset{3}{\cancel{9}}\times \cancel{10^{-3}}}{\cancel{3}\times \cancel{10^{-3}}(1+1+3)} = \frac{3}{5}$$

2/7/18

Q In a ~~gold~~ Bolt factory, machine A, B, C manufacture 20%, 30%, 50% of the total of their output 6%, 3% and 2% of defective. A bolt is drawn at random and found to be defective. Find the probability that it is manufactured from machine A, B and C.

Sol Let A, B, C denotes the 3 machines &

$P(A) = 20\%$

$P(B) = 30\%$

$P(C) = 50\%$

Let D be the defective

$P(D/A) = 6\%$

$P(D/B) = 3\%$

$P(D/C) = 2\%$

If a bolt is defective then the prob. that it is from machine A.

i.e $$P(A/D) = \frac{P(A) \cdot P(D/A)}{\sum_{i=1}^{3} P(A) P(D/A)} = 12/31$$

If a bolt is defective then the prob. that it is from machine B

$$P(B/D) = \frac{P(B) \cdot P(D/B)}{\sum_{i=1}^{3} P(B) \cdot P(D/B)} = 9/31$$

if a bolt is defective that it is from C.

$$P(C/D) = \frac{P(C) \cdot P(D/C)}{\sum_{i=1}^{3} P(C) \cdot P(D/C)} = \frac{10}{31}.$$

Q. 3 of the men, the chances the politician, a bussiness man, an academician will be appointed as a vice chancellor (VC) of a university 0.5, 0.3, 0.2 respectively. The probability that research is promoted by these persons if they are appointed as VC are 0.3, 0.7, 0.8 resp. Determine the Probability that research is promoted.

Sol: let A, B, C be the events that a politician, bussiness man, an academician will be appointed as VC of the 3 men then.

$P(A) = 0.5$, $P(B) = 0.3$, $P(C) = 0.2$.

The probabilities that research is promoted if the 3 men are appointed as VC are $P\left(\frac{R}{A}\right) = 0.3$

$$P\left(\frac{R}{B}\right) = 0.7$$

$$P\left(\frac{R}{C}\right) = 0.8$$

The Probability that research is promoted when VC is a politician

$$P\left(\frac{A}{R}\right) = \frac{P(A) \cdot P\left(\frac{R}{A}\right)}{\sum_{i=1}^{3} P(A) \cdot P\left(\frac{R}{A}\right)} = \frac{0.5 \times 0.3}{0.52} = \frac{0.15}{0.52} = 0.28$$

The probability that research is promoted when VC is a bussiness man

$$P\left(\frac{B}{R}\right) = \frac{P(B) \cdot P(R/B)}{\sum_{i=1}^{3} P(B) \cdot P\left(\frac{R}{B}\right)} = \frac{0.21}{0.52} = 0.40$$

The probability that research is promoted when VC is an academician

$$P\left(\frac{C}{R}\right) = \frac{P(C) \cdot P(R/C)}{\sum_{i=1}^{3} P(C) \cdot P(R/C)} = \frac{0.16}{0.52} = 0.30.$$

The chances that doctor harshitha will diagnoise a disease x correctly is 60%. The chances that a patient will die by his treatment after correct diagnoise is 40% and the chance of death by wrong diagnosis is 70%. A Patent of doctor who had disease x died. What is the Probability that his disease was correctly diagnoised by him. corre

Let us define the following events.

E_1 denotes disease x is diagnoised correctly by doctor.

E_2 denotes disease x is not diagnoised correctly by doctor.

E denotes a patient who had disease x dised

100%

$P(E_1) = 0.6$

$P(E_2) = 0.4 = 1 - P(E_1) = 1 - 0.6 = 0.4$

$P(E/E_1) = 0.\cancel{4}$

$P(E/E_2) = 0.7$

$$P(E_1/E) = \frac{P(E_1) \cdot P(E/E_1)}{\sum_{i=1}^{2} P(E_i) \cdot P(E/E_i)} = \frac{0.24}{0.24 + 0.28} = \frac{0.24}{0.52} = \frac{6}{13}$$

Random Variables:-

A real variable 'x' whose value is determined by the outcome of a random experiment is called a random variable.

A random variable 'x' can also be regarded as a real value function defined on the sample 's' of a random experiment. Such that for each point of x of the sample Space. P(x) is probability of occurrence of the event represented by x.

The probability distribution of random variable

X = no. of heads in tossing 2 coins.

$S = \{HH, HT, TH, TT\}$

X = no. of heads

$X = \{2, 1, 1, 0\}$

X	0	1	2
$P(X=x)$	$1/4$	$2/4$	$1/4$
$F(x)$	$1/4$	$3/4$	$4/4 = 1$

They are two types of Random variables:

1. Descrete Random variables
2. continuous Random variables

1. Discrete Random Variables :- Variables If a random variable take atmost countable no. of values are called discrete random variables.

Ex:- → Marks obtained by a Student in a test
→ No. of Students in a college.

2. Continuous Random Variables :- If a random variable assumes any values within an interval or it takes all possible values between certain limit then the random variable is called continuous random variable

Ex:- → The height of Students between 5 to 6 feets in a city.

Probability mass function :- If is the random variable 'x' it takes the value $x_1, x_2, \dots x_n$ with corresponding Probabilities $P(x_1), P(x_2) \dots P(x_n)$. Then the probability mass function is defined as

$$P(x) = \sum_{i=1}^{n} p(x_i)$$

The probability mass function must satisfy

(i) $p(x) = \geq 0$

(ii) $\leq p(x) = 1$.

Let us consider a continuous random variable 'x' it takes the value $x_1, x_2, \ldots x_n$ with corresponding probabilities $p(x_1), p(x_2) \ldots P(x_n)$. Then the probability density function is defined as $f(x) = \int_{-\infty}^{\infty} f(x)\,dx$.

The probability density function must satisfy.

(i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Distribution function:- The probability distribution function associated with random variable 'x' is defined as the probability that the outcome of an experiment will be one of the outcomes.

The distribution function defined as

$$F_X(x) = f(x) = P(X \leq x)$$

$$= \leq p(x) \quad \text{(DRV)}$$

$$= \int_{-\infty}^{\infty} f(x)\,dx \quad \text{(CRV)}$$

Properties of distribution function :-

(1) If $f(x)$ is a distribution function of a random variable 'x' and $a < b$ then $p(a < x \leq b) = P(b) - p(a)$

(2) If $f(x)$ is distribution function of random variable 'x' then $0 \leq F(x) \leq 1$

(3) $F(x) \leq f(y)$ if $x \leq y$

(4) $f(-\infty) = \underset{x \to \infty}{\text{lt}} f(x) = 0$.

(5) $f(\infty) = \underset{x \to \infty}{\text{lt}} f(x) = 1$

Scanned by CamScanner

NOTE :- Mean $= \mu = \Sigma x \cdot p(x)$

Variance $\sigma^2 = \Sigma (x-\mu)^2 p(x)$

$= \Sigma x^2 \cdot p(x) - \mu^2$

Standard deviation $\sigma = \sqrt{\text{variance}} = \sqrt{\sigma^2}$

Q. A random variable 'x' has the following probability function.

$X = x$	0	1	2	3	4	5	6	7
$P(X=x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

Find (i) k (ii) $p(x \geq 6)$ (iii) $p(0 < x < 5)$ and (iv) Determine the distribution function of 'x'!

Sol: (i) $p(x) = \Sigma p(x) = 1$

$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$9k + 10k^2 = 1$$

$$10k^2 + 9k - 1 = 0.$$

$$10k^2 + 10k - k - 1 = 0$$

$$10k(k+1) - 1(k+1) = 0$$

$$(10k-1)(k+1) = 0$$

$$k = \frac{1}{10}, -1$$

(ii) $p(x \geq 6) = p(x=6) + p(x=7)$

$$= 2k^2 + 7k^2 + k$$

$$= 9k^2 + k = 9\left(\frac{1}{10}\right)^2 + \frac{1}{10}.$$

$$= \frac{19}{100}$$

(iii) $p(0 < x < 5) = p(x=1) + p(x=2) + p(x=3) + p(x=4)$

$$= k + 2k + 2k + 3k.$$

$$= 8k$$

$$= \frac{8}{10}.$$

(iv)

x	$P(X=x)$	$F(x)=P(X\le x)$		
0	0	0	$= 0$	$=0$
1	k	k	$= 1/10$	$=0.1$
2	$2k$	$3k$	$= 3/10$	$=0.3$
3	$2k$	$5k$	$= 5/10$	$=0.5$
4	$3k$	$8k$	$= 8/10$	$=0.8$
5	k^2	$8k+k^2$	$= \frac{8}{10}+\frac{1}{100} = \frac{81}{100}$	$=0.81$
6	$2k^2$	$8k+3k^2$	$= \frac{8}{10}+\frac{3}{100} = \frac{83}{100}$	$=0.83$
7	$7k^2+k$	$9k+10k^2$	$= \frac{9}{10}+\frac{10}{100} = \frac{10}{10}$	$=1$

5/7/18

Q. A random variable 'x' has the following probability distribution.

$X=x$	0	1	2	3	4	5	6	7	8
$P(X=x)$	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

Find (i) a (ii) $p(x<3)$ (iii) $P(x\ge 3)$ (iv) $p(0<x<5)$

(v) what is smallest value of x for which $P(X\le x)>0.5$

Sol (i) $p(x)=\Sigma p(x)=1$

$$a+3a+5a+7a+9a+11a+13a+15a+17a=1$$

$$81a=1$$

$$a=\frac{1}{81}$$

(ii) $p(x<3) = p(x=0)+p(x=1)+p(x=2)$

$$= a+3a+5a$$

$$= 9a = \frac{9}{81} = \frac{1}{9}$$

(iii) $P(x\ge 3) = p(x=3)+p(x=4)+p(x=5)+p(x=6)+p(x=7)+p(x=8)$

$$= 7a+9a+11a+13a+15a+17a$$

$$= 72a = \frac{72}{81} = \frac{8}{9}$$

(iv) $P(0<x<5) = p(x=1)+p(x=2)+p(x=3)+p(x=4)$

$$= 3a+5a+7a+9a$$

$$= 24a = \frac{24}{81} = \frac{8}{27}$$

(v) $P(X \le x) > 0.5$

x	$P(x)$	$f(x) = P(X \le x)$
0	a	$a = 1/81 = 0.012$
1	$3a$	$4a = 4/81 = 0.049$
2	$5a$	$9a = 9/81 = 0.11$
3	$7a$	$16a = 16/81 = 0.19$
4	$9a$	$25a = 25/81 = 0.30$
5	$11a$	$36a = 36/81 = 0.44$
6	$13a$	$49a = 49/81 = 0.60$ for $x=6$
7	$15a$	$64a = 64/81 = 0.79$
8	$17a$	$81a = 81/81 = 1$

(vi) Find mean, Variance & S.D

Mean $\mu = \sum x \cdot p(x)$

$$= 0\left(\frac{1}{81}\right) + 1\left(\frac{3}{81}\right) + 2\left(\frac{5}{81}\right) + 3\left(\frac{7}{81}\right) + 4\left(\frac{9}{81}\right) + 5\left(\frac{11}{81}\right) + 6\left(\frac{13}{81}\right) + 7\left(\frac{15}{81}\right) + 8\left(\frac{17}{81}\right)$$

$$= 0 + \frac{3}{81} + \frac{10}{81} + \frac{21}{81} + \frac{36}{81} + \frac{55}{81} + \frac{78}{81} + \frac{105}{81} + \frac{136}{81}$$

$$= 5.4938$$

Variance $\sigma^2 = \sum (x-\mu)^2 p(x)$

$$= \sum x^2 \cdot p(x) - \mu^2$$

$$= \left[(0)^2\left(\frac{1}{81}\right) + (1)^2\left(\frac{3}{81}\right) + (2)^2\left(\frac{5}{81}\right) + (3)^2\left(\frac{7}{81}\right) + (4)^2\left(\frac{9}{81}\right) + (5)^2\left(\frac{11}{81}\right) + (6)^2\left(\frac{13}{81}\right) + (7)^2\left(\frac{15}{81}\right) + (8)^2\left(\frac{17}{81}\right)\right] - \left(\frac{445}{81}\right)^2$$

$$= \left[0 + \frac{3}{81} + \frac{20}{81} + \frac{63}{81} + \frac{144}{81} + \frac{275}{81} + \frac{468}{81} + \frac{735}{81} + \frac{1088}{81}\right] - \left(\frac{445}{81}\right)^2$$

$$= \frac{2796}{81} - 30.1821$$

$$= 34.5185 - 30.1821$$

$$= 4.3364$$

Standard deviation = $\sqrt{\sigma^2}$

$= \sqrt{4.3364}$

$= 2.0824$

Q. $P(x) = \frac{x}{15}$, $x = 0, 1, 2, 3, 4, 5$

$= 0$, otherwise

x	0	1	2	3	4	5
$P(x) = x/15$	$0/15$	$1/15$	$2/15$	$3/15$	$4/15$	$5/15$

(i) $P(x=1)$ or $P(x=2)$ (ii) $P\left(\left(\frac{1}{2} < x < \frac{5}{2}\right) / (x > 1)\right)$

Sol (i) $P(x=1)$ or $(x=2) = P(x=1) + P(x=2) = \frac{1}{15} + \frac{2}{15} = \frac{3}{15} = \frac{1}{5}$

(ii) $P\left(\underbrace{\left(\frac{1}{2} < x < \frac{5}{2}\right)}_{A} / \underbrace{(x > 1)}_{B}\right) \Rightarrow P(A/B) = \frac{P(A \cap B)}{P(B)}$

$$= \frac{P(x=1) + P(x=2) \cap P(x=2) + P(x=3) + P(x=4) + P(x=5)}{P(x=2) + P(x=3) + P(x=4) + P(x=5)}$$

$$= \frac{P(x=2)}{P(x=2) + P(x=3) + P(x=4) + P(x=5)}$$

$$= \frac{2/15}{\frac{2}{15} + \frac{3}{15} + \frac{4}{15} + \frac{5}{15}} = \frac{2/15}{14/15} = 1/7.$$

Q. A random variable 'x' has the following probability function

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Find (i) k (ii) $P(x<4)$ (iii) $P(x \geq 5)$ (iv) $P(3 < x \leq 6)$ (v) what is the smallest value of 'x' for which $P(X \leq x) > 1/2$ (vi) find mean & variance.

Q. A random variable x has the following probability function

x	-2	-1	0	1	2	3
$P(x)$	0.1	k	0.2	$2k$	0.3	k

find (i) k (ii) $P(x<1)$ (iii) $P(x \geq 2)$ (iv) $P(-1 \leq x \leq 3)$ (v) find distribution function of 'x'.

Ans (i) $p(x) = \Sigma p(x) = 1$

$$k+3k+5k+7k+9k+11k+13k = 1$$

$$49k = 1$$

$$k = 1/49$$

(ii) $p(x<4) = p(x=0)+p(x=1)+p(x=2)+p(x=3)+p(x=4)$

$$= k+3k+5k+7k$$

$$= 16k = 16\left(\frac{1}{49}\right)$$

$$= \frac{16}{49}$$

(iii) $p(x\geq 5) = p(x=5)+p(x=6)$

$$= 11k+13k$$

$$= 24k = 24\left(\frac{1}{49}\right)$$

$$= \frac{24}{49}$$

(iv) $p(3<x\leq 6) = p(x=3) + p(x=4)+p(x=5)+p(x=6)$

$$= 9k+11k+13k$$

$$= 33k = 33(1/49)$$

$$= \frac{33}{49}$$

(v)

x	$P(x)$	$F(x) = p(X\leq x)$
0	k	$k = 1/49 = 0.0204$
1	$3k$	$4k = 4/49 = 0.0816$
2	$5k$	$9k = 9/49 = 0.1836$
3	$7k$	$16k = 16/49 = 0.3265$
4	$9k$	$25k = 25/49 = 0.5102$ for $x=4$
5	$11k$	$36k = 36/49 = 0.7346$
6	$13k$	$49k = 49/49 = 1$

(vi) Mean $\mu = \Sigma x \cdot p(x)$

$$= 0(1/49)+1(3/49)+2(5/49)+3\left(\frac{7}{49}\right)+4(9/49)+5\left(\frac{11}{49}\right)+6\left(\frac{13}{49}\right)$$

$$= 0+\frac{3}{49}+\frac{10}{49}+\frac{21}{49}+\frac{16}{49}+\frac{55}{49}+\frac{78}{49}$$

$= \frac{183}{49} = 3.7346 \quad \frac{29}{7}$

$\frac{\sqrt{6\times 4}}{14^4}$

varience $(\sigma^2) = \Sigma (x-\mu)^2 p(x)$

$= \Sigma x^2 \cdot p(x) - \mu^2$

$= \left[(0)^2\left(\frac{1}{49}\right) + (1)^2\left(\frac{3}{49}\right) + (2)^2\left(\frac{5}{49}\right) + (3)^2\left(\frac{7}{49}\right) + (4)^2\left(\frac{9}{49}\right) + (5)^2\left(\frac{11}{49}\right) + (6)^2\left(\frac{13}{49}\right)\right] - \left(\frac{183}{49}\right)^2$

$= \left[0 + \frac{3}{49} + \frac{20}{49} + \frac{63}{49} + \frac{144}{49} + \frac{275}{49} + \frac{468}{49}\right] - \left(\frac{29}{7}\right)^2$

$= \frac{893}{49} - 13.9479 = \frac{132}{49}$

$= 18.2244 - 13.9479 = 4.2765$

Ans (i) $p(x) = \Sigma p(x) = 1$

$0.1 + k + 0.2 + 2k + 0.3 + k = 1$

$0.6 + 4k = 1$

$4k = 1 - 0.6$

$k = 0.1$

(ii) $p(x<1) = p(x=-2) + p(x=-1) + p(x=0)$

$= 0.1 + k + 0.2$

$= 0.3 + k$

$= 0.3 + 0.1 = 0.4$

(iii) $p(x \geq 2) = p(x=2) + p(x=3)$

$= 0.3 + k$

$= 0.3 + 0.1 = 0.4$

(iv) $p(-1 \leq x \leq 3) = p(x=-1) + p(x=0) + p(x=1) + p(x=2) + p(x=3)$

$= k + 0.2 + 2k + 0.3 + k$

$= 4k + 0.5 = 0.4 + 0.5 = 0.9$

(v)

Scanned by CamScanner

(V)

x	$P(x)$	$F(x) = P(X \le x)$
-2	0.1	$0.1 = 0.1$
-1	k	$0.1+k = 0.2$
0	0.2	$0.3+k = 0.4$
1	$2k$	$0.3+3k = 0.6$
2	0.3	$0.6+3k = 0.9$
3	k	$0.6+4k = 1.$

7/7/18

Measure of central tendency for Continuous P.d.f :-

(1) Mean of a distribution is given by

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

if x is defined from $[a,b]$ then $\mu = \int_a^b x F(x) dr$

(2) Median is the point which divides the entire series into 2 equal parts

$$\text{Median} = \int_a^M f(x)\,dx = \int_M^b F(x)\,dx = \frac{1}{2}$$

(3) Mode is a value of x for which $f(x)$ is maximum, mode is given by

$$f'(x) = 0 \;\&\; F''(x) \le 0.$$

(4) Varience: It is given by $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$$= \int_{-\infty}^{\infty} x^2 F(x)\,dx - \int_{-\infty}^{\infty} \mu^2 f(x)\,dx.$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 \qquad (\because \int F(x)\,dx = 1$$

(5) Mean deviation (MD): MD about mean is given by

$$\int_{-\infty}^{\infty} (x-\mu) F(x)\, dx.$$

Q. Let x be a continuous random variable with P.d.f

$$f(x) = \begin{cases} ax & ; 0 \le x \le 1 \\ a & ; 1 \le x \le 2 \\ -ax+3a & ; 2 \le x \le 3. \\ = 0 & \text{Otherwise.} \end{cases}$$

find (i) Determine the value of a

(ii) compute $p(x < 1.5)$.

Sol. We have $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

$$\int_0^3 f(x)\,dx = 1$$

$$\Rightarrow \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx = 1$$

$$\int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 -ax+3a\,dx = 1$$

$$\int_0^1 ax\,dx = a\int_0^1 x\,dx = a\left[\frac{x^2}{2}\right]_0^1 = a\left[1/2\right] = \frac{a}{2}$$

$$\int_1^2 a\,dx = a\int_1^2 dx = a(x)_1^2 = a(1) = a$$

$$\int_2^3 -ax+3a\,dx = \int_2^3 a(3-x)\,dx = a\int_2^3 3\,dx - \int_2^3 x\,dx$$

$$= a\left[(3) - \left(\frac{x^2}{2}\right)_2^3\right]$$

$$= a\left[3 - \left(\frac{9}{2} - \frac{4}{2}\right)\right]$$

$$= a\left[3 - \frac{5}{2}\right] = \frac{a}{2}$$

$$\frac{a}{2} + a + \frac{a}{2} = 1 \; ; \; a = 1/2$$

(ii) $p(x < 1.5) = \int_0^1 f(x)\,dx + \int_1^{1.5} f(x)\,dx$

$$= \int_0^1 ax\,dx + \int_1^{1.5} a\,dx$$

$$= a\left(\frac{x^2}{2}\right)_0^1 + a(x)_1^{1.5}$$

$$= \frac{a}{2} + 0.5a = 0.5a + 0.5a = 1a = 1/2$$

Q. A continuous random variable x

$$f(x) = \begin{cases} 1/16\,(3+x)^2 & , -3 \le x \le -1 \\ 1/16\,(6-2x^2) & , -1 \le x \le 1 \\ 1/16\,(3-x)^2 & ; 1 \le x \le 3 \end{cases}$$

Find (i) Verify whether it is pdf or not.

Sol

$$\int_{-\infty}^{\infty} f(x)\,dx$$

$$\Rightarrow \int_{-3}^{3} f(x)\,dx$$

$$\Rightarrow \int_{-3}^{-1} f(x)\,dx + \int_{-1}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx = 1$$

$$\Rightarrow \int_{-3}^{-1} \frac{1}{16}(3+x)^2 dx + \int_{-1}^{1} \frac{1}{16}(6-2x^2)\,dx + \int_{1}^{3} \frac{1}{16}(3-x)^2 dx$$

$$= \frac{1}{16}\int_{-3}^{-1} 6x+x^2+9 + \frac{1}{16}\int_{-1}^{1} 36-2x^2-24x\,dx + \frac{1}{16}\int_{1}^{3} 9+x^2-6x\,dx$$

$$= \frac{1}{16}\left[6\left(\frac{x^2}{2}\right)_{-3}^{-1} + \left(\frac{x^3}{3}\right)_{-3}^{-1} + 9(x)_{-3}^{-1}\right] + \frac{1}{16}\left[36(x)_{-1}^{1} - 2\left(\frac{x^3}{3}\right)_{-1}^{1} - 24\left(\frac{x^2}{2}\right)_{-1}^{1}\right]$$

$$+ \frac{1}{16}\left[9\left(x\right)_{1}^{3} + \left(\frac{x^3}{3}\right)_{1}^{3} - 6\left(\frac{x^2}{2}\right)_{1}^{3}\right]$$

$$= \frac{1}{16}\left[6\left(\frac{1}{2}-\frac{9}{2}\right) + \left(\frac{-1}{3}+\frac{27}{3}\right) + 9(-1+3)\right] + \frac{1}{16}\left[36(1+1) - 2\left(\frac{1}{3}+\frac{1}{3}\right) - 24\left(\frac{1}{2}-\frac{1}{2}\right)\right]$$

$$+ \frac{1}{16}\left[9(3-1) + \left(\frac{27}{3}-\frac{1}{3}\right) - 6\left(\frac{9}{2}-\frac{1}{2}\right)\right]$$

$$= \frac{1}{16}\left[-24+\frac{26}{3}+18\right] + \frac{1}{16}\left(12-\frac{4}{3}\right) + \frac{1}{16}\left(18+\frac{26}{3}-24\right)$$

$$= \frac{1}{16}\left[-24+\frac{26}{3}+18+12-\frac{4}{3}+18+\frac{26}{3}-24\right]$$

$$= \frac{1}{16}(16) = 1$$

9/7/18

Q. Let x be a continous random variable with pdf $f(x) = k(1+x)$, $2 \le x \le 5$. Find (i) k (ii) $p(x<4)$

Sol

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{2}^{5} f(x)\,dx = 1 \Rightarrow \int_{2}^{5} k(1+x)\,dx = 1.$$

$$k\left(\int_{2}^{5} dx + \int_{2}^{5} x\,dx\right) = 1$$

$$k\left[(x)_2^5 + \left(\frac{x^2}{2}\right)_2^5\right] = 1$$

$$k\left[3 + \frac{25}{2} - \frac{4}{2}\right] = 1$$

Scanned by CamScanner

$$K\left[\frac{6+21}{2}\right] = 1$$

$$K = \frac{2}{27}$$

$$P(x<4) = \int_2^4 f(x)\,dx$$

$$= \int_2^4 K(1+x)\,dx$$

$$= K\left(x+\frac{x^2}{2}\right)_2^4$$

$$= K\left[(4-2)+\frac{(4)^2}{2}-\frac{(2)^2}{2}\right]$$

$$= K\left(2+\frac{16}{2}-\frac{\cancel{4}^{2}}{\cancel{2}}\right)$$

$$= \frac{2}{27}\times\frac{16}{2} = \frac{16}{27}$$

Q. A continuous random variable x has the following Pdf $f(x) = 3x^2$, $0 \le x \le 1$. Find a & b if (i) $p(x \le a) = P(x > a)$ (ii) $p(x > b) = 0.05$.

Sol: (i) $p(x \le a) = p(x > a)$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^a f(x)\,dx + \int_a^1 f(x)\,dx + \int_1^{\infty} f(x)\,dx$$

$$\int_0^a f(x)\,dx = \int_a^1 \cancel{3x^2} f(x)\,dx$$

$$\int_0^a 3x^2\,dx = \int_a^1 3x^2\,dx$$

$$3\left(\frac{x^3}{3}\right)_0^a = 3\left[\frac{x^3}{3}\right]_a^1$$

$$a^3 = 1-a^3$$

$$2a^3 = 1$$

$$a = \left(\tfrac{1}{2}\right)^{1/3}$$

(ii) $p(x \ge b) = 0.05$

$$\int_b^1 f(x)\,dx = 0.05$$

$$\int_b^1 3x^2\,dx = 0.05$$

$$3\left(\frac{x^3}{3}\right)_b^1 = 0.05$$

$$1-b^3 = 0.05$$

$$b^3 = 1-0.05 \Rightarrow b = (0.95)^{1/3}$$

Q. A continuous random variable x has the following Pdf $f(x) = cxe^{-x}$, $x>0$ find (i) c (ii) $p(x<2)$ (iii) $p(2<x<3)$

Sol (i) $\int_{-\infty}^{\infty} f(x)dx = 1$

$\int uv = u\int v - \int(u' \int v)$

$e^{-\infty} = 0$

$\int_0^{\infty} f(x)\,dx = 1.$

$\int_0^{\infty} cxe^{-x}dx = 1$

$c\int_0^{\infty} xe^{-x}dx = 1$

$c\left[-xe^{-x} - e^{-x}\right]_0^{\infty} = 1.$

$c\left[-0+1\right] = 1$

$c = 1$ //

(ii) $P(x<2) = \int_0^2 f(x)dx$

$= \int_0^2 cxe^{-x}dx$

$= c\left(-xe^{-x} - e^{-x}\right)_0^2$

$= \left(-2e^{-2} - e^{-2} + 1\right)$

$= 1 - 3e^{-2}.$

(iii) $p(2<x<3) = \int_0^2 f(x)dx + \int_2^3 f(x)dx$

$= c\left(-xe^{-x} - e^{-x}\right)_0^2 + \int_2^3 cxe^{-x}dx$

$= 1 - 3e^{-2} + \left(-xe^{-x} - e^{-x}\right)_2^3$

$= 1 - 3e^{-2} + \left(-3e^{-3} - e^{-3} + 2e^{-2} + e^{-2}\right)$

$= 1$ $\quad = 3e^{-2} - 4e^{-3}$

Q. A continuous random variable 'x' has the following pdf $f(x) = 3x^2$, $0 \le x \le 1$. find mean & variance.

Sol Mean $\mu = \int_{-\infty}^{+\infty} xf(x)\,dx$

$= \int_0^1 x\, 3x^2 dx$

$= \int_0^1 3x^3 dx$

$$= 3\left(\frac{x^4}{4}\right)_0^1$$

$$= 3(1/4) = 3/4$$

variance $(\sigma^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$= \int_0^1 x^2 (3x^2)\,dx - \left(\frac{3}{4}\right)^2$$

$$= \int_0^1 3x^4\,dx - \left(\frac{3}{4}\right)^2$$

$$= 3\left(\frac{x^5}{5}\right)_0^1 - \left(\frac{3}{4}\right)^2 (x)_0^1$$

$$= 3\left(\frac{1}{5}\right) - \frac{9}{16}$$

$$= \frac{3}{5} - \frac{9}{16}$$

Q. For the following destiny function $f(x) = 6x^2(1-x)$, $0 < x < 1$. Find the (i) constant c and (ii) mean, variance and standard deviation.

Q. A continuous random variable x has the following pdf $f(x) = 6x(1-x)$, $0 \leq x \leq 1$. Find (i) check whether is it a pdf (ii) determine b such that $P(x<b) = P(x>b)$ (iii) Mean, Variance and standard deviation.

Sol: Given $f(x) = cx^2(1-x)$, $0 < x < 1$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_0^1 cx^2(1-x)\,dx = 1$$

$$c\int_0^1 (x^2 - x^3)\,dx = 1$$

$$c\left[\left(\frac{x^3}{3}\right)_0^1 - \left(\frac{x^4}{4}\right)_0^1\right] = 1$$

$$c\left[\frac{1}{3} - \frac{1}{4}\right] = 1$$

$$c\left[\frac{1}{12}\right] = 1 \Rightarrow c = 12.$$

(ii) Mean $\mu = \int_{-\infty}^{\infty} x(cx^2(1-x)\,dx)$

$= \int_0^1 x(cx^2(1-x)\,dx)$

$= c\int_0^1 x^3(1-x)\,dx$

$= c\int_0^1 x^3 - x^4\,dx$

$= c\left[\left(\frac{x^4}{4}\right)_0^1 - \left(\frac{x^5}{5}\right)_0^1\right]$

$= c\left[\frac{1}{4} - \frac{1}{5}\right]$

$= 12 \times \frac{1}{20} = \frac{3}{5}$

Variance $\sigma^2 = \int_0^1 x^2\,[cx^2(1-x)]\,dx - \left(\frac{3}{5}\right)^2$

$= c\int_0^1 \left[(x^4 - x^5)\,dx - \left(\frac{3}{5}\right)^2\right]$

$= c\left[\left(\frac{x^5}{5}\right)_0^1 - \left(\frac{x^6}{6}\right)_0^1\right] - \left(\frac{3}{5}\right)^2$

$= 12\left[\frac{1}{30}\right] - \frac{9}{25}$

$= \frac{6}{15} - \frac{9}{25} = \frac{2}{5} - \frac{9}{25} = \frac{10-9}{25} = \frac{1}{25}$

Sol (i) $\int_{-\infty}^{\infty} f(x)\,dx$

$\int_0^1 6x(1-x)\,dx$

$= 6\int_0^1 x - x^2\,dx$

$= 6\left[\left(\frac{x^2}{2}\right)_0^1 - \left(\frac{x^3}{3}\right)_0^1\right] = 6\left[\frac{1}{2} - \frac{1}{3}\right] = \cancel{6} \times \frac{1}{\cancel{6}} = 1$

(ii) $P(x<b) = p(x>b)$

$\int_0^b 6x(1-x)\,dx = \int_b^1 6x(1-x)\,dx$

$6\left[\left(\frac{x^2}{2}\right)_0^b - \left(\frac{x^3}{3}\right)_0^b\right] = 6\left[\left(\frac{x^2}{2}\right)_b^1 - \left(\frac{x^3}{3}\right)_b^1\right]$

$$\frac{b^2}{2}-\frac{b^3}{3}=\frac{1}{2}-\frac{b^2}{2}-\frac{1}{3}+\frac{b^3}{3}$$

$$\frac{2b^2}{2}-\frac{2b^3}{3}=\frac{1}{6}.$$

$$6b^2-4b^3=1.$$

$$6b^2-4b^3-1=0$$

$$4b^3-6b^2+1=0 \Rightarrow b=\tfrac{1}{2}, \frac{1\pm\sqrt{3}}{2}$$

(iii) Mean $\mu=\int_{-\infty}^{\infty}x\,[6x(1-x)]\,dx$

$$=\int_0^1 6x^2(1-x)\,dx$$

$$=6\int_0^1(x^2-x^3)\,dx$$

$$=6\left[\left(\frac{x^3}{3}\right)_0^1-\left(\frac{x^4}{4}\right)_0^1\right]$$

$$=6\left[\frac{1}{3}-\frac{1}{4}\right]=\frac{6}{12}=\frac{1}{2}$$

Variance $\sigma^2=\int_{-\infty}^{\infty}x^2\,[6x(1-x)]\,dx-\left(\frac{1}{2}\right)^2 dx$

$$=\int_0^1(6x^3(1-x))\,dx-\left(\frac{1}{2}\right)^2 dx$$

$$=6\left[\left(\frac{x^4}{4}\right)_0^1-\left(\frac{x^5}{5}\right)_0^1\right]-\frac{1}{4}$$

$$=6\left(\frac{1}{20}\right)-\frac{1}{4}=\frac{3}{10}-\frac{1}{4}=\frac{12-10}{40}=\frac{1}{20}$$

Standard deviation $=\sqrt{\sigma^2}=\sqrt{1/20}$

11/7/18

Mathematical expectation :-

The expected value of a random variable is weighted avg. of all possible values of the random variable where the weights of probabilities. associated with the corresponding values. The mathematical expectation for computing expected values of the descrete random variable x with probability mass function is given below $E(x) = \sum_{x} x \cdot p(x)$.

If the random variables are continuous

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx.$$

Properties of expectation :-

(1) Addition theorem of expectation :- If x and y are two random variables then expectation of

$$E(x+y) = E(x) + E(y)$$

Let x, y be continuous random variables with join probability density function.— —

$f(x,y)$, $f(x)$, $f(y)$ respectively then by definition

$$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx \quad , \quad E(y) = \int_{-\infty}^{\infty} y\, f(y)\, dy.$$

(2) $E(x_1 + x_2 + x_3 + \cdots + x_n) = E(x_1) + E(x_2) + \cdots + E(x_n)$

(3) $E[ax] = aE(x)$

(4) $E[ax+b] = aE(x) + b.$

Properties of variance :-

(1) If x is a random variable then $V(ax+b) = a^2 V(x)$

Proof :- let $y = (ax+b)$

then $E(y) = E(ax+b) = aE(x) + b$

$y - E(y) = ax - aE(x)$

squaring and taking expectation

$$[y - E(y)]^2 = a^2 [x - E(x)]^2$$

$E[Y - E(Y)]^2 = a^2 E[X - E(X)]^2$ $\quad [\because E[x - E(x)]^2 = E(x^2) - E(x))^2 = V(x)]$

$V(Y) = a^2 V(X)$

$V(aX + b) = a^2 V(X)$

② $V(cx)$: where c is constant then $V(cx) = c^2 V(x)$

③ $V(c) = 0$: where 'c' is constant

$\therefore$ variance of only constant is zero

④ $cov(ax, bY) = ab\, cov(x, y)$ $\quad$ cov → co-Varience.

⑤ $cov.(x+a, Y+b) = cov(x, Y)$

⑥ If x, Y are independent $E(X.Y) = E(X)\, E(Y)$

$\therefore cov(x, y) = 0$.

Prove that $V(x) = E(x^2) - [E(x)]^2$

L.H.S $\quad V(x) = E(x - \mu)^2$

$= E[x^2 + \mu^2 - 2x\mu]$

$= E(x^2) + E(\mu^2) - 2\mu E(x)$

$= E(x^2) + \mu^2 - 2\mu \cdot \mu$ $\quad (\because E(\mu^2) = \mu^2$, $E(x) = \mu)$

$= E(x^2) + \mu^2 - 2\mu^2$

$= E(x^2) - \mu^2$

$= E(x^2) - [E(x)]^2$

Q when 4 coins are tossed what is the expectation of no. of heads. ?

Sol If 4 coins are tossed, the exhaustive events $= 2^4 = 16$.

let 'x' denotes no. of heads.

4	1	2	3	0
HHHH	TTTH	TTHH	HHHT	TTTT
	TTHT	THHT	HHTH	
	THTT	HHTT	HTHH	
	HTTT	HTHT	THHH	
		HTTH		
		THTH		

1

Probability of getting 0 heads $P(X=0) = 1/16$

1 head $P(X=1) = 4/16$

2 heads $P(X=2) = 6/16$

3 heads $P(X=3) = \frac{4}{16}$

4 heads $P(X=4) = 1/16$.

$$E(x) = \sum_{x=0}^{4} x\,(p(x)$$

$$= 0\cdot\frac{1}{16} + 1\cdot\frac{4}{16} + 2\cdot\frac{6}{16} + 3\cdot\frac{4}{16} + 4\cdot\frac{1}{16}$$

$$= \frac{4}{16} + \frac{12}{16} + \frac{12}{16} + \frac{4}{16} = \frac{16+16}{16} = 2$$

Q. Find the expectation of the number on a die when it ~~rolled~~ thrown.

Sol: If 1 die is thrown, the exhaustive events $= 6^1 = 6$

$x =$	1	2	3	4	5	6
$P(x) =$	$1/6$	$1/6$	$1/16$	$1/6$	$1/6$	$1/6$

$$E(x) = \sum_{x=1}^{6} x\,p(x)$$

$$= 1\cdot 1/6 + 2\cdot 1/6 + 3\cdot 1/6 + 4\cdot 1/6 + 5\cdot 1/6 + 6\cdot 1/6$$

$$= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{\cancel{21}^{7}}{\cancel{6}_{2}} = \frac{7}{2}$$

6
6
9
21

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

$$= \sum_{x=1}^{6} x^2\,p(x) - \left(\frac{7}{2}\right)^2$$

$$= \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} - \frac{49}{4}$$

$$= \frac{91}{6} - \frac{49}{4} = \frac{364 - 294}{24} = \frac{70}{24} = \frac{35}{12}$$

36
25
61
16
77
9
86
4
90
91

Biased and unbiased dice are thrown. Find the expected values of the sum of numbers of points on dice

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
= 2	= 3	= 4	= 5	= 6	= 7
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
= 3	= 4	= 5	= 6	= 7	= 8
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
= 4	= 5	= 6	= 7	= 8	= 9
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
= 5	= 6	= 7	= 8	= 9	= 10
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
= 6	= 7	= 8	= 9	= 10	= 11
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)
= 7	= 8	= 9	= 10	= 11	= 12

Probability of getting sum

2 $P(x=2) = 1/36$

3 $P(x=3) = 2/36$

4 $P(x=4) = 3/36$

5 $P(x=5) = 4/36$

6 $P(x=6) = 5/36$

7 $P(x=7) = 6/36$

8 $P(x=8) = 5/36$

9 $P(x=9) = 4/36$

10 $P(x=10) = 3/36$

11 $P(x=11) = 2/36$

12 $P(x=12) = 1/36$

$$E(x) = \sum_{x=2}^{12} x \cdot p(x)$$

$$= 2 \cdot 1/36 + 3 \cdot \frac{2}{36} + 4 \cdot 3/36 + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36}$$

$$= \frac{252}{36} = 7.$$

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

$$= \sum_{x=2}^{12} x^2 \cdot p(x) - (7)^2$$

$$= \left[(2)^2 \cdot \frac{1}{36} + (3)^2 \cdot \frac{2}{36} + (4)^2 \cdot \frac{3}{36} + (5)^2 \cdot \frac{4}{36} + (6)^2 \cdot \frac{5}{36} + (7)^2 \cdot \frac{6}{36} + (8)^2 \cdot \frac{5}{36} + (9)^2 \cdot \frac{4}{36} + (10)^2 \cdot \frac{3}{36} + (11)^2 \cdot \frac{2}{36} + (12)^2 \cdot \frac{1}{36}\right] - (7)^2$$

$$= \left[\frac{4}{36} + \frac{18}{36} + \frac{48}{36} + \frac{100}{36} + \frac{180}{36} + \frac{294}{36} + \frac{320}{36} + \frac{324}{36} + \frac{300}{36} + \frac{242}{36} + \frac{144}{36}\right] - 49$$

$$= \frac{1974}{36} - 49 = 5.833.$$

Module-II Probability Distributions

Moment Generating Function (MGF) :- It is a tool used to calculate the higher moments. The moment generating function of a random variable 'x' about a origin whose probability density function $F(x)$ is given by $M_x(t) = E(e^{tx})$

$\sum_x x p(x) = \sum_x e^{tx} p(x) \rightarrow$ Discrete Distribution function.

$\int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \rightarrow$ continuous dist. function.

Since $M_x(t)$ is used to generate moments, it is known as MGF.

we have $e^{tx} = 1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots + \frac{t^r x^r}{r!}$

$$M_x(t) = E[e^{tx}]$$

$$= E\left[1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots + \frac{t^r x^r}{r!}\right]$$

$$= 1 + tE(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3}{3!}E(x^3) + \dots + \frac{t^r}{r!}E(x^r)$$

$$\Rightarrow M_x(t) = 1 + t\mu + \frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \dots + \frac{t^r}{r!}\mu_r$$

where μ_r is the r^{th} moment of x about the origin i.e $E(x^r) = \mu_r$ is the co-efficient of $\frac{t^r}{r!}$, in the expansion $M_x(t)$.

In order to find of μ_r we have to differentiate the moment generating function r times w.r.t t and substitute $t=0$ i.e $\mu_r = \frac{d^r}{dt^r}\left[M_x(t)\right]_{t=0}$

Properties of MGF :-

Let $Y = ax+b$, where x is a random variable with MGF $M_x(t)$ then $M_{ax}(y) = M_x(ax+b)$

Scanned by CamScanner

$$M_x(y) = M_x(ax+b)$$
$$= e^{tx}[M_x(at)]$$

2. $M_{kx}(t) = M_x(kt)$, k is constant

3. $M_{x+y}(t) = M_x(t) \cdot M_y(t)$ ($\because$ x, y are independent)

4. A random variable x may have no moments even if, its moment generating function exists.

5. A random variable x can have all (or) sum moments but moment generating function does not exists.. perhaps at one point.

Binomial distribution :-

If x is a discrete random variable with probability function $\boxed{p(X=x) = {}^nC_x\, p^x q^{n-x}}$, where $x = 0, 1, 2, \ldots n$ is called a binomial variant denoted by $x \sim B(n, p)$, where n, p are called parameters

NOTE :- $(p+q)^n = \sum_{x=0}^{n} {}^nC_x\, p^x q^{n-x}$

Conditions :-

1. If the random experiment containing only two cases i.e Success / Failure, True/Fase, Head/tail.
   probability of success p is same in each trial.
2. probability of Success p is same in each trial.
3. No. of repeatations are finite

Mean of B.D :-

$$E(x) = \sum_{x=0}^{n} x \cdot p(x)$$

$$= \sum_{x=0}^{n} x\, {}^nC_x\, p^x q^{n-x}$$

$$= \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)!\, x!} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} \cancel{x} \cdot \frac{n!}{(n-x)!\, \cancel{x}(x-1)!} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n(n-1)!}{(n-1)-(x-1)!\,(x-1)!} p^{x-1+1} q^{(n-1)-(x-1)}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-1)-(x-1)!\,(x-1)!} p^{(x-1)} q^{(n-1)-(x-1)}$$

$$= np \sum_{x=1}^{n} {}^{n-1}C_{x-1}\, p^{x-1} q^{(n-1)-(x-1)}$$

$$= np\,(p+q)^{n-1} \qquad [\because (p+q)^n = {}^nC_x\, p^x q^{n-x}]$$

$$= np\,(1)^{n-1} \qquad (\because p+q=1)$$

$$= np$$

13/7/18

Variance of B.D :-

$$V(x) = E(x^2) - [E(x)]^2$$

$$= \sum_{x=0}^{n} x^2 p(x) - (np)^2$$

$$= \sum_{x=0}^{n} [x(x-1)+x]\, {}^nC_x\, p^x q^{n-x} - n^2p^2$$

$$= \sum_{x=0}^{n} x(x-1)\, {}^nC_x\, p^x q^{n-x} + \sum_{x=0}^{n} x\, {}^nC_x\, p^x q^{n-x} - n^2p^2$$

$$= \sum_{x=2}^{n} x(x-1) \frac{n!}{(n-x)!\,x!} p^x q^{n-x} + \sum_{x=1}^{n} x \frac{n!}{(n-x)!\,x!} p^x q^{n-x} - n^2p^2$$

$$= \sum_{x=2}^{n} x(x-1) \frac{n!}{(n-x)!\,x(x-1)(x-2)!} p^x q^{n-x} + np - n^2p^2$$

$$= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(n-2)-(x-2)!\,(x-2)!} p^{x-2+2} q^{(n-2)-(x-2)} + np - n^2p^2$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-2)-(x-2)!\,(x-2)!} p^{(x-2)} q^{(n-2)-(x-2)} + np - n^2p^2$$

$$= n(n-1)p^2 \sum_{x=2}^{n} {}^{n-2}C_{x-2}\, p^{x-2} q^{(n-2)-(x-2)} + np - n^2p^2$$

$$= n(n-1)p^2 (p+q)^{n-2} + np - n^2p^2 \qquad (\because (p+q)^n = \Sigma\, {}^nC_x\, p^x q^{n-x})$$

$$(\because p+q=1)$$

$$= n(n-1)p^2 (1)^{n-2} + np - n^2p^2$$

$$= n^2p^2 - np^2 + np - n^2p^2 = -np^2 + np$$

$$= np - np^2$$

$= np(1-p)$ $\quad (\because p+q=1)$

$= npq$ $\quad (1-p=q)$

NOTE :- The mean of B.D is always greater than its variance $\boxed{np > npq}$

Recurrence Relation for the B.D :-

we know $P(X=x) = {}^nC_x p^x q^{n-x}$

$$P(X=x+1) = {}^nC_{x+1} p^{(x+1)} q^{n-x-1}$$

$$\frac{P(x+1)}{P(x)} = \frac{{}^nC_{x+1} p^{x+1} q^{n-x-1}}{{}^nC_x p^x q^{n-x}}$$

$$\frac{P(x+1)}{P(x)} = \frac{n-x}{x+1} \frac{p}{q}$$

$$P(x+1) = \frac{n-x}{x+1} \frac{p}{q} P(x).$$

M.G.F of BD :-

$$M_x(t) = E[e^{tx}]$$

$$= \sum_{x=0}^{n} e^{tx} p(x)$$

$$= \sum_{x=0}^{n} e^{tx}\, {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nC_x (pe^t)^x q^{n-x}$$

$$= (pe^t+q)^n.$$

Q. The mean and variance of binomial distribution are 4 and $4/3$ respectively. Find $p(x \geq 1)$

Sol: Mean $np=4$

Variance $npq = 4/3$

$$4q = 4/3$$

$$q = 1/3$$

Scanned by CamScanner

$np = 4$

$n(\cancel{2/3}) = \cancel{4}\,2$

$n = 6$

$p(x \geq 1) = p(x=1) + p(x=2) + p(x=3) + p(x=4) + p(x=5) + p(x=6)$

or

$p(x \geq 1) = 1 - p(x < 1)$

$= 1 - p(x=0)$

$= 1 - {}^nC_x\, p^x q^{n-x}$

$= 1 - {}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^{6-0}$

$= 0.998$

Find the parameters in binomial distribution given mean $= 2 \times$ variance and mean + variance $= 3$, also find $p(x \leq 3)$.

Mean $= 2 \times$ variance

$\cancel{n}\cancel{p} = 2 \times \cancel{n}\cancel{p}q$

$q = 1/2$

$p + q = 1$

$p + 1/2 = 1$

$p = 1/2$

$np + npq = 3$

$np(1+q) = 3$

$np\left(1+\frac{1}{2}\right) = 3$

$np\left(\frac{3}{2}\right) = \cancel{3}$

$np = \frac{2}{3} \cdot 2$

$n\left[1/2\right] = 2$

$n = 4$

$p(x \leq 3) = \cancel{p(x=} 1 - p(x > 3)$

$= 1 - p(x=4)$

$= 1 - {}^nC_x\, p^x q^{n-x}$

$= 1 - {}^4C_4 (1/2)^4 \left(\frac{1}{2}\right)^0$

$= 1 - \frac{1}{16} = 0.9375$

Q. The incidence of an occupational disease in an industry, the workers has 20% chance of suffering from teeth. what is the probability that out of 6 workers chosen at random, 4 or more will suffer from disease.

sol: Given number of workers $(n) = 6$.

$$P = \frac{20}{100} = \frac{1}{5}$$

$$P + q = 1$$

$$\frac{1}{5} + q = 1$$

$$q = 1 - \frac{1}{5}$$

$$q = \frac{4}{5}$$

$$\frac{6!}{(1)!\,(5)!}$$

$$p(x \geq 4) = p(x=4) + p(x=5) + p(x=6)$$

$$= {}^6C_4 \left(\frac{1}{5}\right)^4 \left(\frac{4}{5}\right)^2 + {}^6C_5 \left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right)^1 + {}^6C_6 \left(\frac{1}{5}\right)^6 \left(\frac{4}{5}\right)^0$$

$$= (15)\left(\frac{1}{625}\right)\left(\frac{16}{25}\right) + 6\left(\frac{1}{3125}\right)\left(\frac{4}{5}\right) + 1\left(\frac{1}{15625}\right)(1)$$

$$= \frac{15 \times 16}{15625} + \frac{24}{15625} + \frac{1}{15625}$$

$$= \frac{240 + 24 + 1}{15625} = \frac{265}{15625} = 0.01696.$$

14/7/18

Q. The probability of a man hitting a target is $\frac{1}{3}$. (i) If he fires 5 times what is the probability of his hitting the target atleast twice? (ii) How many times must be fired so that the probability of his hitting the target atleast once is more than 90%?

sol: The probability of a man hitting a target is $\frac{1}{3}$.

$$P = \frac{1}{3}.$$

$$P + q = 1$$

$$q = 1 - \frac{1}{3} = \frac{2}{3}$$

Scanned by CamScanner

$n=5$

(i) $p(x \geq 2) = 1 - p(x < 2)$

$$= 1 - [p(x=0) + p(x=1)]$$

$$= 1 - p(x=0) - p(x=1)$$

$$= 1 - {}^nC_x p^x q^{n-x} - {}^nC_x p^x q^{n-x}$$

$$= 1 - {}^5C_0 \left(\tfrac{1}{3}\right)^0 \left(\tfrac{2}{3}\right)^5 - {}^5C_1 \left(\tfrac{1}{3}\right)^1 \left(\tfrac{2}{3}\right)^{5-1}$$

$$= 1 - \frac{32}{243} - 5\left(\frac{1}{3}\right)\left(\frac{16}{81}\right)$$

$$= \frac{243-32-80}{243} = \frac{131}{243} = 0.539$$

(ii) $p = \frac{9\cancel{0}}{10\cancel{0}} = 0.9$.

$p(x \geq 1) > 90\%$

$1 - p(x<1) > 0.9$

$1 - p(x=0) > 0.9$

for $n=5$, $1 - {}^5C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{5-0} > 0.9$

$0.8683 > 0.9$

For $n=6$,

$1 - {}^6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 > 0.9$

$1 - \frac{64}{729} > 0.9$

$0.9122 > 0.9$

So, for $n=6$

6 times must be fired. so that the probability of his hitting the target atleast once is more than 90%.

Fitting of binomial distribution :-

Q. Fit a binomial distribution to the following frequency distribution.

x	0	1	2	3	4	5	6
$f(x)$	13	25	52	58	32	16	4

Sol:- $n=6$

Total frequency $N = \sum_{x=0}^{n} f(x) = 13+25+52+58+32+16+4$

$= 200$

Scanned by CamScanner

$$\text{Mean } \mu = \frac{\sum f_i x_i}{N}$$

$$\sum_{x=0}^{n} f_i x_i = 0\times13 + 1\times25 + 2\times52 + 3\times58 + 4\times32 + 5\times16 + 6\times4$$

$$= 25 + 104 + 174 + 128 + 80 + 24$$

$$= 535.$$

$$\mu = \frac{535}{200} = 2.675$$

Mean for B.D $= np$

$$np = 2.675$$

$$p = \frac{2.675}{6}$$

$$p = 0.4458.$$

$$p + q = 1$$

$$q = 1 - 0.4458$$

$$q = 0.5541$$

x	$P(X=x) = {}^nC_x\, p^x q^{n-x}$	$N \cdot p(X=x)$	
0	${}^6C_0 (0.4458)^0 (0.5541)^6$	200×0.0289	$= 5.78 \simeq 6$
1	${}^6C_1 (0.4458)^1 (0.5541)^5$	200×0.1397	$= 27.9422 \simeq 28$
2	${}^6C_2 (0.4458)^2 (0.5541)^4$	200×0.2807	$= 56.1526 \simeq 56$
3	${}^6C_3 (0.4458)^3 (0.5541)^3$	200×0.30138	$= 60.2766 \simeq 60$
4	${}^6C_4 (0.4458)^4 (0.5541)^2$	200×0.1814	$= 36.2874 \simeq 36$
5	${}^6C_5 (0.4458)^5 (0.5541)^1$	200×0.0585	$= 11.7076 \simeq 12$
6	${}^6C_6 (0.4458)^6 (0.5541)^0$	200×0.0078	$= 1.5698 \simeq 2$
			$= 199.7162$
			$\simeq 200$

0.0942×0.1987

0.1701×0.08859

0.0394×0.3070

Modified BD

x	0	1	2	3	4	5	6
$f(x)$	13	25	52	58	32	16	4
Modified Value	6	28	56	60	36	12	2

II method:-

$n=6,\ p=0.4458,\ q=0.5541$

$N[p+q]^n = 200[0.4458+0.5541]^6$

$= 200[{}^6C_0(0.4458)^0(0.5541)^{6-0} + {}^6C_1(0.4458)^1(0.5541)^{6-1}$

$+ {}^6C_2(0.4458)^2(0.5541)^{6-2} + {}^6C_3(0.4458)^3(0.5541)^3$

$+ {}^6C_4(0.4458)^4(0.5541)^2 + {}^6C_5(0.4458)^5(0.5541)^1$

$+ {}^6C_6(0.4458)^6(0.5541)^0]$

$= 200[0.0289 + 0.1397 + 0.2807 + 0.30138 + 0.1814 + 0.0585 + 0.0078]$

$= \simeq 200$

Q. Fit a B.D to the following data

x	0	1	2	3	4	5
$f(x)$	2	14	20	34	22	8

Sol

$n=5$

$N = \sum_{x=0}^{n} f(x) = 100$

Mean $\mu = \frac{\sum f_i x_i}{N}$

$\sum_{n=0}^{N} f_i x_i = 0\times2 + 1\times14 + 2\times20 + 3\times34 + 4\times22 + 5\times8$

$= 0+14+40+102+88+40$

$= 284.$

$\mu = \frac{284}{100} = 2.84$

Mean for B.D $= np$

$np = 2.84$

$p = \frac{2.84}{5}$

$= 0.568$

$p+q=1$

$q = 1-0.568$

$q = 0.432.$ (on solving we get solution)

16/7/18

Poisson distribution :-

Poisson distribution is the limiting case of a binomial distribution under the conditions that

(i) n is very large ($n \to \infty$)

(ii) P is very small ($p \to 0$)

(iii) $\lambda = np$

* Derivation of poisson distribution :-

$$P(x) = {}^nC_x\, p^x q^{n-x}$$

$$= \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{n-x}$$

$$= \frac{n(n-1)(n-2)\cdots\cdots[n-(x-1)]\cancel{(n-x)!}}{\cancel{(n-x)!}\,x!}\, p^x \frac{(1-p)^n}{(1-p)^x}$$

$\left\langle \because \lambda = \frac{n}{p} \Rightarrow n = \frac{\lambda}{p} \Rightarrow p = \frac{\lambda}{n} \right\rangle$

$$= \frac{\frac{\lambda}{p}\left(\frac{\lambda}{p}-1\right)\left[\frac{\lambda}{p}-2\right]\cdots\left[\frac{\lambda}{p}-(x-1)\right]}{x!}\, p^x \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}$$

$$= \frac{\lambda(\lambda-p)(\lambda-2p)\cdots\cdots(\lambda-(x-1)p)}{\cancel{p^x}\cdot x!}\,\cancel{p^x}\,\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}$$

$\left\langle \because n\to\infty, p \to 0 \right\rangle$

$$= \frac{\lambda(\lambda-0)(\lambda-0)\cdots\cdots[\lambda-[x-1]0]}{x!}\ \lim_{n\to\infty} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}$$

$$= \frac{\lambda^x}{x!}\left(\frac{e^{-\lambda}}{1}\right) \qquad \left\langle \because \lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n = e^{-x} \right\rangle$$

$$= \frac{e^{-\lambda}\lambda^x}{x!}$$

* Definition of poisson distribution :-

Let x be a random variable, the poisson distribution function is defined as

$$P(x,\lambda) = \begin{cases} \dfrac{e^{-\lambda}\cdot\lambda^x}{x!} & ;(\text{if } x \geq 0) \\ 0 & ;\ \text{Otherwise.} \end{cases}$$

NOTE:- $\Sigma p(x)$

$$= \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$ $\left\langle \because \sum_{x=0}^{\infty} \frac{x^n}{n!} = e^x \right\rangle$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \cdot e^{\lambda} = 1$$

$\therefore$ $p(x)$ is probability function.

Mean of a poisson distribution :-

*

$$\mu = \Sigma x p(x)$$

$$\mu = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!}$$

$$\mu = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\cancel{x} \cdot \lambda^x}{\cancel{x}(x-1)!}$$

$$\mu = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda \lambda^{x-1}}{(x-1)!}$$

for $x-1 = y$
$x = y+1$.

$$\mu = e^{-\lambda} \cdot \sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!}$$ $\left\langle \because \sum_{x=0}^{\infty} \frac{x^n}{n!} = e^x \right\rangle$

$$\mu = e^{-\lambda} \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

$$\mu = \cancel{e^{-\lambda}} \cdot \lambda \cdot \cancel{e^{\lambda}} = \lambda. \Rightarrow \boxed{\mu = \lambda}$$

**

Variance of a poisson distribution :-

$$\sigma^2 = \Sigma x^2 p(x) - \mu^2$$

$$= \Sigma x^2 \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} - \lambda^2$$

$$= \Sigma \frac{x^2 e^{-\lambda} \lambda^x}{\cancel{x}(x-1)!} - \lambda^2$$

$$= e^{-\lambda} \Sigma \frac{x \lambda^x}{(x-1)!} - \lambda^2$$

$$= e^{-\lambda} \Sigma \left[\frac{(x-1)+1}{(x-1)!} \right] \lambda^x - \lambda^2$$

$$= e^{-\lambda}\left[\sum \frac{(x-1)\lambda^x}{(x-1)!} + \sum \frac{\lambda^x}{(x-1)!}\right] - \lambda^2$$

let $x-1=y$
$x=y+1$

$$= e^{-\lambda}\left[\sum \frac{\cancel{(x-1)}\lambda^x}{\cancel{(x-1)}(x-2)!} + \sum \frac{\lambda^{y+1}}{y!}\right] - \lambda^2$$

let $x-2=t$
$x=t+2$

$$= e^{-\lambda}\left[\sum \frac{\lambda^{t+2}}{t!} + \lambda \sum \frac{\lambda^y}{y!}\right] - \lambda^2$$

$$= e^{-\lambda}\left[\lambda^2 \sum \frac{\lambda^t}{t!} + \lambda \sum \frac{\lambda^y}{y!}\right] - \lambda^2$$

$$= e^{-\lambda}\left[\lambda^2 e^{\lambda} + \lambda e^{\lambda}\right] - \lambda^2$$

$$= \cancel{e^{-\lambda}}\lambda^2\cancel{e^{\lambda}} + \cancel{e^{-\lambda}}\lambda\cancel{e^{\lambda}} - \lambda^2$$

$$\sigma^2 = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}$$

$$\sigma^2 = \lambda$$

$$\boxed{\sigma^2 = \lambda}$$

* Properties of poisson distribution:-

1. Poisson distribution is a limiting case of binomial distribution.
2. The mean of poisson distribution is a $\lambda = np$ $(\mu = \lambda)$
3. In poisson distribution, mean = variance.
4. λ is the parameter of the poisson distribution.

Problems:-

Q. A hospital switch board receives an average of 4 emergency calls in a 10 mins interval. what is the probability that (i) there are atmost 2 emergency calls in 10 mins interval. (ii) there are exactly 3 emergency calls in a 10 mins interval (iii) between 3 to 5 inclusive emergency calls in 10 mins interval

Sol: Hospital switch board receiving an average 4 calls in 10 mins interval i.e mean $\lambda = \mu = 4$

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-4} \cdot 4^x}{x!}$$

(i) $p(x \leq 2) = p(x=0) + p(x=1) + p(x=2)$

$$= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!}$$

$$= e^{-4} \left(0 + 4 + \frac{\cancel{16}\, 8}{\cancel{2}}\right)$$

$$= e^{-4} \cdot 12 = 0.2197$$

(ii) $p(x=3) = \frac{e^{-4} \cdot 4^3}{3!}$

$$= \frac{e^{-4} \cdot 64}{6} = e^{-4} (10.6666) = 0.1953$$

(iii) $p(3 \leq x \leq 5) = p(x=3) + p(x=4) + p(x=5)$

$$= \frac{e^{-4} \cdot 4^3}{3!} + \frac{e^{-4} \cdot 4^4}{4!} + \frac{e^{-4} \cdot 4^5}{5!}$$

$$= 0.1953 + \frac{e^{-4} \cdot 256}{24} + \frac{e^{-4} \cdot 1024}{120}$$

$$= 0.1953 + 0.19536 + 0.15629$$

$$= 0.5469.$$

18/7/18

Q. In a given city, 6% of all drivers get atleast 1 parking ticket per year. use the poisson to the binomial distribution to approximation determine the probability that among 80 drivers

(i) 4 will get atleast 1 parking ticket in any given year

(ii) Atleast 3 will get one parking ticket in any given year.

(iii) Any where from ~~2~~ 3 to 6 inclusive will get atleast one parking ticket in any given year.

Scanned by CamScanner

Sol
$p = 6\% = 0.06$

$n = 80$

$\lambda = np$

$= 80 \times 0.06$

$= \frac{48}{10\cancel{0}} \times 1\cancel{0} = 4.8$

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-4.8} (4.8)^x}{x!}$$

(i) $p(x=4) = \frac{e^{-4.8} (4.8)^4}{4!}$

(ii) $p(x \geq 3) = 1 - p(x < 3)$

$= 1 - [p(x=0) + p(x=1) + p(x=2)]$

$$= 1 - \left[\frac{e^{-4.8}(4.8)^0}{0!} + \frac{e^{-4.8}(4.8)}{1!} + \frac{e^{-4.8}(4.8)^2}{2!} \right]$$

(iii) $p(3 \leq x \leq 6) = p(x=3) + p(x=4) + p(x=5) + p(x=6)$

Q. Ships arrive in a harbour at a mean rate of 2 per hour. Suppose that this situation can be ~~decte~~ described by poisson distribution. Find the probability for a 30 mins period that

(i) No ships arrived

(ii) 3 ships arrived.

Sol No avg. 2 ships arrived per hour i.e one ship for 30 mins period

$\lambda = 1$

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-1} \cdot 1^x}{x!} = \frac{e^{-1}}{x!}$$

(i) $p(x=0) = \frac{e^{-1}}{0!} = e^{-1} = 0.3678$

Scanned by CamScanner

(iii) $p(x=-3) = \frac{e^{-1}}{3!} = 0.061.$

Q. If a random variable x has a poisson distribution $p(x=1) = p(x=2)$. Find (i) mean of a poisson distribution (ii) $p(x=4)$ (iii) $p(x>1)$ (iv) $p(1<x\leq 4)$

Sol

$p(x=1) = p(x=2)$

$$p(x) = \frac{e^{-\lambda}.\lambda^x}{x!}$$

$$\frac{e^{-\lambda}.\cancel{\lambda^1}}{1!} = \frac{e^{-\lambda}.\lambda^{2}}{2!}$$

$$\cancel{e^{-\lambda}} = \frac{\cancel{e^{-\lambda}}.\lambda}{2}$$

$\lambda = 2.$

(i) mean $(\lambda) = 2$ $\Rightarrow p(x) = \frac{e^{-2}.2^x}{x!}$

(ii) $p(x=4) = \frac{e^{-\lambda}.\lambda^4}{4!} = \frac{e^{-2}.2^{\cancel{4}}}{\cancel{4}\times\cancel{2}\times 3} = \frac{e^{-2}.2}{3} = 0.0902$ //

(iii) $p(x>1) = 1-p(x\leq 1)$

$= 1-[p(x=0)+p(x=1)]$

$= 1-\left[\frac{e^{-2}.2^0}{0!} + \frac{e^{-2}.2^1}{1!}\right]$

$= 1-e^{-2}.3$

$= 1-0.270 = 0.\cancel{7}593$ //

(iv) $p(1<x\leq 4) = p(x=2)+p(x=3)+p(x=4)$

$= \frac{e^{-2}.2^2}{2!} + \frac{e^{-2}.2^3}{3!} + \frac{e^{-2}.2^4}{4!}$

$= 0.270+0.180+0.0902$

$= 0.540$ //

Q. If poisson distribution such that $\frac{3}{2}p(x=1) = p(x=3)$. Find (i) $p(x\geq 1)$ (ii) $p(x\leq 3)$ (iii) $p(2\leq x\leq 5)$

Sol

$$p(x) = \frac{e^{-\lambda}.\lambda^x}{x!}$$

$$p(x=1) = \frac{e^{-\lambda}.\lambda^x}{1!} = e^{-\lambda}.\lambda$$

$$P(x=3) = \frac{e^{-\lambda}\cdot\lambda^3}{3!}$$

$$\frac{3}{2}e^{-\lambda}\lambda = \frac{e^{-\lambda}\cdot\lambda^3}{6} \qquad \left(\frac{3}{2}\,p(x=1) = p(x=3)\right)$$

$$\lambda^2 = 9.$$

$$\lambda = \pm 3$$

$$\lambda = 3.$$

(i) $p(x \geq 1) = 1 - p(x<1)$

$= 1 - [p(x=0)]$

$= 1 - \dfrac{e^{-3}\cdot 3^0}{0!} = 1 - e^{-3} = 0.950.$

(ii) $p(x \leq 3) = p(x=0) + p(x=1) + p(x=2) + p(x=3)$

$= \dfrac{e^{-3}\cdot 3^0}{0!} + \dfrac{e^{-3}\cdot 3^1}{1!} + \dfrac{e^{-3}\cdot 3^2}{2!} + \dfrac{e^{-3}\cdot 3^3}{3!}$

$= e^{-3}\left(1 + 3 + \dfrac{9}{2} + \dfrac{\cancel{27}^{\,9}}{\cancel{3}\times 2}\right)$

$= e^{-3}\,[1+3+9]$

$= e^{-3}\cdot 13. = 0.647.$

(iii) $p(2 \leq x \leq 5) = p(x=2) + p(x=3) + p(x=4) + p(x=5)$

Q. The avg. number of radioactive particles passing through a counter during one milli second in an experiment is 4. what is the probability that (i) 6 particles enter the counter in a given milli second.

Sol Given $\lambda = 4$

$$P(x) = \frac{e^{-\lambda}\cdot\lambda^x}{x!}$$

$$= \frac{e^{-4}\cdot 4^x}{x!}$$

$$P(x=6) = \frac{e^{-4}\cdot 4^6}{6!} = 0.104$$

Q. It is known that 5% book bound at a certain bindary have defective binding. Find the probability that 2 of 100 books by this bindary will have defective binding, use the Poisson approximation to the binomial distribution.

Sol: $p = 5\% = 0.05$

$n = 100$

$\lambda = 0.05 \times 100 = 5$. $p(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-5} \cdot 5^x}{x!}$

$P(x=2) = \frac{e^{-5} \cdot 5^2}{2!}$

$= 0.084$

Q. 2% of the out of a machine is a defective a lot of 300 pieces will be produced. Determine the probability that exactly 4 pieces will be defective.

Q. If the probability that an individual suffers a bad reaction from a certain medicine is 0.001. Determine the probability that out of 2000 individuals more than 2 individuals will suffer a bad reaction.

Sol: $p = 0.001$

$n = 2000$

$\lambda = np$
$= 2000 \times \frac{1}{1000} = 2$

$p(x) = \frac{e^{-2} \cdot 2^x}{x!}$

$p(x>2) = 1 - p(x \leq 2)$

$= 1 - [p(x=0) + p(x=1) + p(x=2)]$

$= 1 - \left[\frac{e^{-2} \cdot 2^0}{0!} + \frac{e^{-2} \cdot 2^1}{1!} + \frac{e^{-2} \cdot 2^2}{2!}\right]$

$= 1 - e^{-2}(1+2+2)$

$= 1 - e^{-2}(5) = 1 - 0.6766 = 0.3234$

19/7/18

Q. Suppose 2% of the people on average are left handed. Find (i) the probability 3 or more left handed. (ii) the probability of none or one left handed.

Sol

$\lambda = 2\% = 0.02$

$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-0.02}(0.02)^x}{x!}$

(i) $p(x \geq 3) = 1 - p(x < 3)$

$= 1 - [p(x=0) + p(x=1) + p(x=2)]$

$= 1 - \left[\frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!} + \frac{e^{-0.02}(0.02)^2}{2!}\right]$

$= 0.003$

(ii) $p(x=0) + p(x=1) = \frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!}$

$= 0.999.$

Fitting poisson distribution:-

$\text{Mean} = \frac{\Sigma x f(x)}{\Sigma f(x) = N}$

Q. Fit a poisson distribution for the following data and also calculate expected frequencies.

x	0	1	2.	3	4
$f(x)$	109	65	22	3	1.

Sol $N = \Sigma f(x) = 109 + 65 + 22 + 3 + 1 = 200$

$\Sigma x f(x) = 0 \times 109 + 1 \times 65 + 2 \times 22 + 3 \times 3 + 4 \times 1$

$= 65 + 44 + 9 + 4$

$= 122$

$\text{Mean}(\lambda) = \frac{122}{200} = 0.61$

$N p(x) = N \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{200 \cdot e^{-0.61}(0.61)^x}{x!}$

$$P(x=0) = 200 \times \frac{e^{-0.61}(0.61)^0}{0!} = 108.670 .$$

$$P(x=1) = 200 \times \frac{e^{-0.61}(0.61)^1}{1!} = 66.288$$

$$P(x=2) = 200 \times \frac{e^{-0.61}(0.61)^2}{2!} = 20.218$$

$$P(x=3) = 200 \times \frac{e^{-0.61}(0.61)^3}{3!} = 4.111$$

$$P(x=4) = 200 \times \frac{e^{-0.61}(0.61)^4}{4!} = 0.626.$$

$$\underline{\underline{199.913}}$$

$\therefore$ the expected frequencies are 108.670, 66.288, 20.218, 4.111, 0.626.

Q. Fit a poisson distribution for the following data and also calculate expected frequencies

x	0	1	2	3	4	5
$f(x)$	125	95	49	20	8	3.

$N = \Sigma f(x) = 125+95+49+20+8+3 = 300$

$\Sigma x f(x) = 0\times125 + 1\times95 + 2\times49 + 3\times20 + 4\times8 + 5\times3$

$= 95+98+60+32+15$

$= 300$

$\lambda = \frac{300}{300} = 1$

the expected frequencies are given by

$$Np(x) = N \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} = 300 \cdot \frac{e^{-1}(1)^x}{x!}$$

$$P(x=0) = 300 \times \frac{e^{-1}(1)^0}{0!} = 110.3638$$

$$P(x=1) = 300 \times \frac{e^{-1}(1)^1}{1!} = 110.3638$$

$$P(x=2) = 300 \times \frac{e^{-1}(1)^2}{2!} = 55.18$$

$$P(x=3) = 300 \times \frac{e^{-1}(1)^3}{3!} = 18.393$$

$P(x=4) = 300 \times \frac{e^{-1}(1)^4}{4!} = 4.598$

$P(x=5) = 300 \times \frac{e^{-1}(1)^5}{5!} = 0.9196$

$\simeq 299.818$

$\therefore$ the expected frequencies are 110.3638, 110.3638, 55.18, 18.39, 4.598, 0.9196.

Q. Fit a poisson distribution for the following data and also calculate expected values

x	0	1	2	3	4	5	6	7
$f(x)$	305	365	210	80	28	9	2	1

Sol: mean $= \frac{\Sigma x f(x)}{\Sigma f(x) = N}$

$N = \Sigma f(x) = 1000$

$\Sigma x f(x) = 1201$

$\lambda = \frac{1201}{1000} = 1.201$

$Np(x) = 1000 \times \frac{e^{-1.201}(1.201)^x}{x!}$

$P(x=0) = 1000 \times \frac{e^{-1.201}(1.201)^0}{0!} = 300.893$

$P(x=1) = 361.372$

$P(x=2) = 217.0043$

$P(x=3) = 86.874$

$P(x=4) = 26.083$

$P(x=5) = 6.265$

$P(x=6) = 1.254$

$P(x=7) = 0.215$

$= 999.95$

the expected values are 300.893, 361.372, 217.0043, 86.874, 26.083, 6.265, 1.254, 0.215

Moment generating function for poisson distribution :-

$$M_x(t) = E[e^{tx}]$$
$$= \Sigma e^{tx} p(x)$$
$$= \Sigma e^{tx} \frac{e^{-\lambda}\lambda^x}{x!}$$
$$= e^{-\lambda} \Sigma \frac{(e^t\lambda)^x}{x!} \qquad \left\langle \because e^x = \Sigma \frac{x^n}{n!} \right\rangle$$
$$= e^{-\lambda} e^{e^t\lambda} = e^{e^t\lambda-\lambda} = \boxed{e^{\lambda(e^t-1)}}$$

Probability Generating function for poisson distribution :-

$$P_x(t) = E(t^x)$$
$$= \Sigma t^x p(x)$$
$$= \Sigma t^x \cdot \frac{e^{-\lambda}\cdot\lambda^x}{x!}$$
$$= e^{-\lambda} \Sigma \frac{t^x\cdot\lambda^x}{x!} \qquad \left\langle \because e^x = \frac{\Sigma x^n}{n!} \right\rangle$$
$$= e^{-\lambda} \Sigma \frac{(t\lambda)^x}{x!}$$
$$= e^{-\lambda} e^{t\lambda} = e^{t\lambda-\lambda} = \boxed{e^{\lambda(t-1)}}$$

20/7/18 **Normal distribution :-**

A random variable 'x' is said to having normal distribution if its density function is given by

$$\boxed{f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}, \quad -\infty \le x \le \infty \quad \text{where} \quad -\infty \le \mu \le \infty$$

μ = mean of the normal distribution and

σ = Standard deviation of the normal distribution.

μ, σ are the parameters of the normal distribution.

Mean of the normal distribution :-

The mean of the continuous probability distribution is $\mu = \int_{-\infty}^{\infty} x F(x) dx$

$$\boxed{\mu = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx}$$

Let $z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{dx}{\sigma}$

$\Rightarrow x = z\sigma + \mu$

if $x = -\infty \Rightarrow z = -\infty$
$x = \infty \Rightarrow z = \infty$

$$\int_{-\infty}^{\infty} (z\sigma + \mu) \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-\frac{z^2}{2}} (\cancel{\sigma} dz)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z\sigma + \mu) e^{-z^2/2} (dz)$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \underbrace{z\sigma e^{-z^2/2}}_{\text{odd function}} dz + \int_{-\infty}^{\infty} \mu e^{-z^2/2} dz \right]$$

$$= \frac{1}{\sqrt{2\pi}} \left[0 + 2\mu \int_{0}^{\infty} e^{-z^2/2} dz \right] \qquad \left(\because \int_{-a}^{a} f(x) dx = 2\int_0^a f(x) dx \right)$$

$$= \frac{1}{\sqrt{2\pi}} 2\mu \sqrt{\frac{\pi}{2}} = \frac{\cancel{2}\mu\sqrt{\pi}}{\cancel{2}\sqrt{\pi}} = \mu \qquad \frac{2\mu\sqrt{\pi}}{2\sqrt{\pi}}$$

Q. Assume that 50% of all engineering students good in mathematics. Determine the probability that among that 20 engineering students studied

(i) exactly 9

(ii) Atmost 8

(iii) Atleast 7

(iv) Atleast 2 and Atmost 4

Sol. $n = 20$

$P = 50\% = 0.5$

$q = 0.5$ $\qquad (\because p+q=1)$

$$P(n,x,p) = nC_x p^x q^{n-x}$$
$$= 20C_x (0.5)^x (0.5)^{20-x}$$

(i)
$$P(x=9) = 20C_9 (0.5)^9 (0.5)^{11}$$

(ii) $P(x \leq 8) = p(x=0) + p(x=1) + p(x=2) + p(x=3) + p(x=4) + p(x=5) + p(x=6) + p(x=7) + p(x=8)$

(iii) $p(x \geq 7) = 1 - p(x < 7)$
$$= 1 - [p(x=0) + p(x=1) + p(x=2) + \ldots\ldots + p(x=6)]$$

(iv) $p(2 \leq x \leq 4) = p(x=2) + p(x=3) + p(x=4)$

Q. A traffic control engineer reports 75% of vechicals passing through check point are from within the state. what is the probability that fewer than 4 of the next 9 vehicles are from outs of the state?

Sol:
$q = 75\% = 0.75$

$p = 0.25$

$n = 9$

$$P(x=x) = nC_x p^x q^{n-x}$$
$$= 9C_x (0.25)^x (0.25)^{9-x}$$

$$P(x<4) = p(x=0) + p(x=1) + p(x=2) + p(x=3)$$

Q. Find the parameters in binomial distribution, given that mean $= 2 \times$ variance and mean + variance $= 3$ also find $p(x \leq 3)$

26/7/18

Variance of normal distribution :-

$$\text{Variance} = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$= \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2$$

Let $z = \frac{x-\mu}{\sigma} \Rightarrow x = z\sigma + \mu$

$dx = \sigma dz$

$x = -\infty \Rightarrow z = -\infty$

$x = \infty \Rightarrow z = \infty$

$$= \int_{-\infty}^{\infty} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} (z\sigma+\mu)^2 e^{-\frac{z^2}{2}} \cdot \cancel{\sigma} dz - \mu^2$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (z\sigma+\mu)^2 e^{-z^2/2} dz - \mu^2$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} z^2 \sigma^2 e^{-z^2/2} dz + 2\sigma\mu \int_{-\infty}^{\infty} z e^{-z^2/2} dz + \mu^2 \int_{-\infty}^{\infty} e^{-z^2/2} dz \right] - \mu^2$$

let $\frac{z^2}{2} = t \Rightarrow 2t = z^2 \Rightarrow z = \sqrt{2t}$

$dz = \sqrt{2} \frac{1}{2\sqrt{t}} dt = \frac{1}{\sqrt{2t}} dt$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} 2t \cdot \sigma^2 e^{-t} \cdot \frac{1}{\sqrt{2t}} dt + 0 + \mu^2 2 \int_{0}^{\infty} e^{-z^2/2} dz \right] - \mu^2$$

$$= \frac{1}{\cancel{\sqrt{2\pi}}} \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{2}}} \int_{-\infty}^{\infty} t^{1/2} \cdot e^{-t} dt + \cancel{2}\mu^2 \frac{1}{\cancel{\sqrt{2\pi}}} \sqrt{\frac{\pi}{\cancel{2}}} - \mu^2$$

$$= \frac{\sigma^2}{\sqrt{\pi}} 2 \int_{0}^{\infty} t^{3/2-1} \cdot e^{-t} dt + \cancel{\mu^2} - \cancel{\mu^2}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \gamma(3/2) = \frac{2\sigma^2}{\sqrt{\pi}} \gamma\left(\frac{1}{2}+1\right)$$

$$= \frac{\cancel{2}\sigma^2}{\sqrt{\pi}} \frac{1}{\cancel{2}} \gamma(1/2)$$

$$= \frac{\sigma^2}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}}$$

$$= \sigma^2$$

$\gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt$

$\gamma(n+1) = n\gamma(n)$

$\gamma(1/2) = \sqrt{\pi}$

Mode of a normal distribution:-

Mode is the value of x for which $f(x)$ is max.
i.e mode is the solution of $f'(x)=0$ and
$f''(x) < 0$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot -\frac{\cancel{2}(x-\mu)}{\cancel{2}\sigma^2}$$

$$= -\frac{(x-\mu)}{\sigma^2} f(x)$$

$\langle$ if $x=\mu$, $f'(x)=0 \rangle$

$$f''(x) = -\frac{(x-\mu)}{\sigma^2} f'(x) - f(x)\frac{1}{\sigma^2}$$

$$= -\frac{(x-\mu)}{\sigma^2} \cdot -\frac{(x-\mu)}{\sigma^2} f(x) - \frac{1}{\sigma^2} f(x)$$

$$= \frac{f(x)}{\sigma^2}\left[\frac{(x-\mu)^2}{\sigma^2} - 1\right]$$

(if $x=\mu$)

$$f''(x) = -\frac{f(x)}{\sigma^2} < 0$$

Median of a normal distribution :-

27/7/18 Median of a normal distribution

If M is the median of the normal distribution then (integral $-\infty$ to M), $\boxed{\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2}}$

$$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2}$$

$$\int_{-\infty}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{2}$$

$$\int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{2}$$

Let $z = \frac{x-\mu}{\sigma} \Rightarrow dz = \frac{dx}{\sigma} \Rightarrow dx = \sigma dz$

$x = \infty \Rightarrow z = -\infty$

$x = \mu \Rightarrow z = 0$

$$\Rightarrow \int_{-\infty}^{0} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-z^2/2} \cancel{\sigma}\, dz + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1/2$$

$$\Rightarrow \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1/2$$

$$\frac{1}{\sqrt{2\cancel{\pi}}} \sqrt{\frac{\cancel{\pi}}{2}} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{2}.$$

$$\cancel{\frac{1}{2}} + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \cancel{\frac{1}{2}}$$

$$\Rightarrow \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 0.$$

$$\boxed{M = \mu}$$

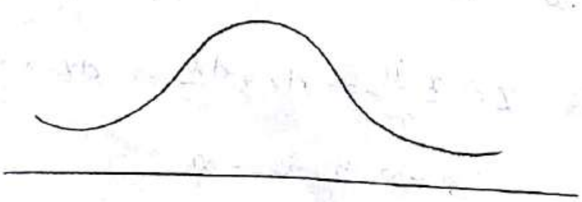
Standard normal distribution:-

If x is a normal variant with mean (μ) and standard deviation (σ) [s.d] then $z = \frac{x-\mu}{\sigma}$ is a standard normal variate with mean (0) and standard deviation (1) then the probability density function standard normal variate (z)

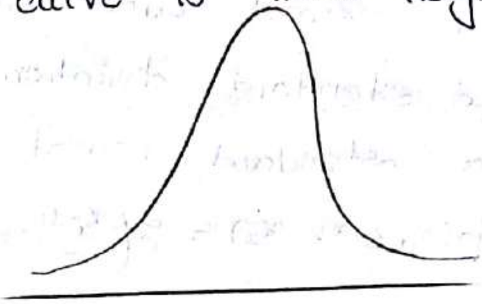
$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} \quad ; -\infty \le z \le \infty$$

Normal distribution curve:-

The graph of the normal distribution is depends on 2 factors mean and the standard deviation. The mean of the distribution determines the location of the graph and the standard deviation depends on the height and width of the graph. If the standard deviation is large then the curve is less height and more width.



If the standard deviation is small, then the shape of curve is more height and less width.



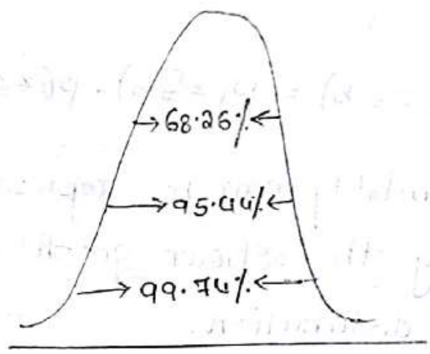
Characteristics of normal distribution :-

- An normal distribution, mean = mode = median
- The curve is bell shaped.
- The normal curve is symetrical about the line $x = \mu$
- As x increases numerically, $f(x)$ decreases rapidly. The maximum probability occurs at the point $x = \mu$ and given by $\boxed{[f(x)]_{max} = \frac{1}{\sigma\sqrt{2\pi}}}$
- The total area under the cure $\int_{-\infty}^{\infty} f(x)dx = 1$ is distributed as follows:

$\mu - \sigma \leq x \leq \mu + \sigma$ covers 68.26% of area.

$\mu - 2\sigma \leq x \leq \mu + 2\sigma$ covers 95.44 % of area

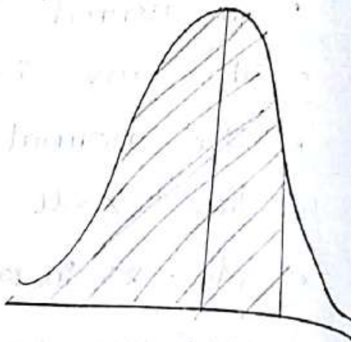
$\mu - 3\sigma \leq x \leq \mu + 3\sigma$ covers 99.74% of area.



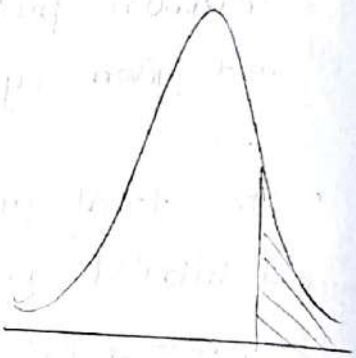
Formula:-

- If x is a normal curve variable with mean (μ) and standard deviation (σ) then $\boxed{z = \frac{x-\mu}{\sigma}}$ is a standard normal random variable and hence $p(x_1 \le x \le x_2) = p\left(\frac{x_1-\mu}{\sigma} \le z \le \frac{x_2-\mu}{\sigma}\right)$

- z is identical to $N(0;1)$ is standard normal variate by using the standard normal distribution area tables, we calculate various probabilites as follows



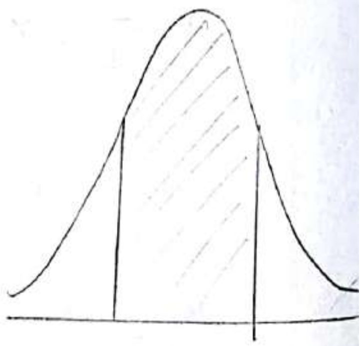
(i) $p(z \le a)$, The probability is directly from the table, we get the value and corresponding probabilities



(ii) $p(z \ge a)$

This probability can be represented by using the graph of standard normal distribution but can't directly obtained from table.

$p(z \ge a) = 1 - p(z < a)$



(iii) $p(a \le z \le b) = p(z \le b) - p(z \le a)$

this probability can be represented by using the shown graph of normal distribution.

23/7/18

NOTE:-

The normal approximation to the binomial distribution given by x is a random variable which follows the B.D with parameters n & p then the limiting form of the distribution function of a standard normal variate $Z = \frac{x-np}{\sqrt{npq}}$ where

$q = 1-p$.

If n is large and p is not closed to 0 or 1. If both np and nq are greater than 5 then approximation will be good.

Problems:-

Q. If x is a normal variate with mean 30 and standard deviation 5. Find the probability that
(i) $26 \leq x \leq 40$ (ii) $x \geq 45$

Sol Given,

mean $(\mu) = 30$

S.D $(\sigma) = 5$

Standard distribution $(z) = \frac{x-\mu}{\sigma} = \frac{x-30}{5}$

(i) $p(26 \leq x \leq 40) = p\left(\frac{26-30}{5} \leq z \leq \frac{40-30}{5}\right)$ $\langle \because p(a \leq z \leq b) = p(z \leq b) - p(z \leq a) \rangle$

$= p(-0.8 \leq z \leq 2)$

$= p(z \leq 2) - p(z \leq -0.8)$

$= 0.9772 - 0.2119$

$= 0.7653$

(ii) $p(x \geq 45) = p\left(z \geq \frac{45-30}{5}\right)$ $\langle \because p(z \geq a) = 1 - p(z < a) \rangle$

$= p(z \geq 3)$

$= 1 - p(z < 3) = 1 - 0.9987$

$= 0.0013$

Q. The marks obtained by 1000 students in an examination mean is 34.5 and S.D is 16.5. Find assuming in the normality of distribution. Find the approximate no. of students, expected to obtain marks (i) more than 65 (ii) marks under 37 (iii) marks btw 30 & 60.

Sol (i) $z = \frac{x-\mu}{\sigma} = \frac{x-34.5}{16.5}$

$p(x>65) = p\left(z > \frac{65-34.5}{16.5}\right)$

$= p(z \geq 1.84)$

$= 1 - p(z \leq 1.84)$

$= 1 - 0.9671 = 0.0329.$

N. of students get more than 65 marks $= 0.0329 \times 1000$

$= 32.9 \sim 33$

(ii) $p(x<37) = p\left(z < \frac{37-34.5}{16.5}\right)$

$= p(z < 0.15) = 0.5596$

Total no. of students got under 37 marks

$= 0.5596 \times 1000$

$= 559.6 \sim 560$

(iii) $p(30 \leq x \leq 60) = p\left(\frac{30-34.5}{16.5} \leq z \leq \frac{60-34.5}{16.5}\right)$

$= p(-0.27 \leq z \leq 1.54)$

$= p(z \leq 1.54) - p(z \leq -0.27)$

$= 0.9382 - 0.3936 = 0.5446$

Total no. of students who got btw 30 and 60 marks $= 0.5446 \times 1000$

$= 544.6 \simeq 545.$

Scanned by CamScanner

** Q. Given that the mean heights of the students in a class is 158 cm with standard deviation 20cm. Find how many students heights lies between 150 cm and 170 cm, more than 165 cm. There are 100 students in the class room.

sol. $\mu = 158$, $\sigma = 20$

$$z = \frac{x-\mu}{\sigma} = \frac{x-158}{20}$$

(i) $p(150 \le x \le 170) = p\left(\frac{150-158}{20} \le z \le \frac{170-158}{20}\right)$

$= p(-0.4 \le z \le 0.6)$

$= p(z \le 0.6) - p(z \le -0.4)$

$= 0.7257 - 0.3446$

$= 0.3811.$

The no. of students whose heights lies btw 150 cm and 170 cm $= 0.3811 \times 100$

$= 38.11. \sim 38$

(ii) $p(x \ge 165) = 1 - p(\cancel{z} < 165)$

$= 1 - p\left(z < \frac{165-158}{20}\right)$

$= 1 - p(z \le 0.35)$

$= 1 - 0.6368.$

$= 0.3632.$

The no. of students whose heights more than 165 cm $= 0.3632 \times 100$

$= 36.32 \sim 36.$

Q. If the weights of 300 students are normally distributed with mean 60 kg & standard deviation 3kg. How many students have (i) > 72 kg (ii) ≤ 64 kg (iii) 65 kg & 71 kg inclusive.

Q. If the mean & S.D of a normal variate are 8 & 4 resp. (i) Find the $p(5 \le x \le 10)$ (ii) $p(x \ge 5)$ (iii) $p(10 \le x \le 15)$ (iv) $p(x \le 15)$

Q. In a normal distribution 7% of the items under 35 and 89% are under 63. Determine the mean and variance of the distribution.

Sol $Z = \frac{x-\mu}{\sigma}$

$p(x < 35) = 7\%$

$p\left(z < \frac{35-\mu}{\sigma}\right) = 0.07$

$\frac{35-\mu}{\sigma} = -1.47$

$\Rightarrow \mu - 1.47\sigma = 35 \rightarrow (1)$

$p(x < 63) = 89\%$

$p\left(z - \frac{63-\mu}{\sigma}\right) = 0.89$

$\frac{63-\mu}{\sigma} = 1.23$

$\mu + 1.23\sigma = 63 \rightarrow (2)$

$1.47 - 1.23 = 0.24$

Solving (1) & (2)

$63 - 1.23\sigma - 1.47\sigma = 35$

$63 - 35 = 2.70\sigma$

$28 = 2.70\sigma$

$\sigma = \frac{28}{2.70}$

$\sigma = 10.37 \Rightarrow \boxed{\sigma = 10.37}$

From (1), $\mu - 1.47(10.37) = 35$

$\boxed{\mu = 50.24}$

30/7/18

Q. The S.Deviation and mean of the of the normal distribution are 70 and 160. Find (i) $p(38 < x < 16)$ (ii) $p(x > 152)$ (iii) $p(x < 140)$

Sol $\sigma = 70$, $\mu = 160$

$z = \frac{x-\mu}{\sigma} = \frac{x-160}{70}$

Scanned by CamScanner

(i) $p(38 < x < 46)$

$\Rightarrow p\left(\frac{38-160}{70} < z < \frac{46-160}{70}\right) = p(-1.74 < z < -1.62)$

$= p(z < -1.62) - p(z < -1.74)$

(ii)

$0.0526 - 0.0409$

(ii) $p(x > 152) = p\left(z > \frac{152-160}{70}\right)$

$= p\left(z > -\frac{80}{70}\right) = p(z > -0.11)$

$= 1 - p(z < 0.11)$

$= 1 - 0.4562 = 0.5438$

(iii) $p(x < 140) = p\left(z < \frac{140-160}{70}\right)$

$= p(z < -0.28) = 0.3897$

Q. The marks obtained in statistics in a certain examination found to be normal distribution. If 15% of the students more than or equal to 60 marks and 40% of the students less than 30 marks. Find the mean and standard deviation of the distribution.

Sol $p(x \geq 60) = 15\%$

$p\left(z \geq \frac{60-\mu}{\sigma}\right) = 0.15$

$1 - p\left(z < \frac{60-\mu}{\sigma}\right) = 0.15$

$p\left(z < \frac{60-\mu}{\sigma}\right) = 0.85$

$\frac{60-\mu}{\sigma} = +1.04$

$\mu + 1.04\sigma = 60 \rightarrow (1)$

(ii) $p(x < 30) = 40\%$

$p\left(z < \frac{30-\mu}{\sigma}\right) = 0.4$

$\frac{30-\mu}{6} = -0.25$

$\mu - \sigma 0.25 = 30 \rightarrow (2)$

① - ② $1.29\sigma = 30$

$$\sigma = \frac{30}{1.29} = 23.25$$

for ②, $\mu = 30 + 0.25 \times 23.25$

$= 30.82$.

** Q. The marks obtained by 1000 students is a N.D with mean 78% and standard deviation 11%. Determine (i) How many students got marks above 90% (ii) what was highest obtained by the lowest 10% of the students (iii) within what limits did the middle of 90% of the students lie.

Sol $\mu = 78\% = 0.78$

$\sigma = 11\% = 0.11$

$$z = \frac{x-\mu}{\sigma} = \frac{x-0.78}{0.11}$$

(i)

$$p(x > 0.9) = p\left(z > \frac{0.9-0.78}{0.11}\right)$$

$= p(z > 1.09)$

$= 1 - p(z < 1.09)$

$= 1 - 0.8621$

$= 0.1379$.

The no. of students got 90% marks are $= 0.1379 \times 1000$

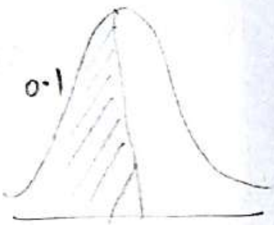
$= 137.9$

~ 138.

(ii) $p(x < x_1) = 0.1$

$$p\left(z < \frac{x_1 - 0.78}{0.11}\right) = 0.1$$

0·1



$$\frac{x_1 - 0.78}{0.11} = -1.28$$

$x_1 = -1.28 \times 0.11 + 0.78$.

$= 0.6392$

~~The no. of student got lowest 10% marks are~~

Highest obtained by lowest 10% of students $= 0.6392 \times 1000$

$= 639.2$

$= 64\%$.

① - ② $1.29\sigma = 30$

$$\sigma = \frac{30}{1.29} = 23.25$$

for ②, $\mu = 30 + 0.25 \times 23.25$

$= 30.82.$

** Q. The marks obtained by 1000 students is a N.D with mean 78% and standard deviation 11%. Determine (i) How many students got marks above 90% (ii) what was highest marks obtained by the lowest 10% of the students. (iii) within what limits did the middle of 90% of the students lie.

Sol: $\mu = 78\% = 0.78$

$\sigma = 11\% = 0.11$

$$z = \frac{x-\mu}{\sigma} = \frac{x-0.78}{0.11}$$

(i)

$$p(x>0.9) = p\left(z > \frac{0.9-0.78}{0.11}\right)$$

$= p(z>1.09)$

$= 1-p(z<1.09)$

$= 1-0.8621$

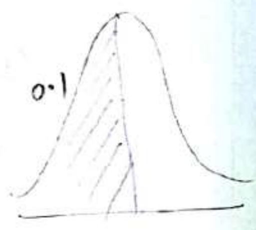
$= 0.1379.$

The no. of students got 90% marks are $= 0.1379 \times 1000$

$= 137.9$

$\sim 138.$

(ii) $p(x<x_1) = 0.1$



$$p\left(z < \frac{x_1-0.78}{0.11}\right) = 0.1$$

$$\frac{x_1-0.78}{0.11} = -1.28$$

$x_1 = -1.28 \times 0.11 + 0.78.$

$= 0.6392$

~~The no. of students got lowest 10% marks are~~

Highest obtained by lowest 10% of students $= 0.6392 \times 100$

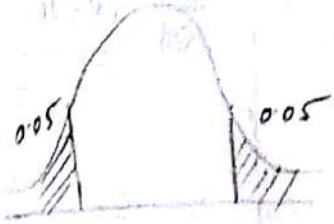
$= 63.9$

$= 64\%.$

Scanned by CamScanner

(iii) $P(x < x_1) = 0.05$

$P\left(z < \frac{x_1 - 0.78}{0.11}\right) = 0.05$



$\frac{x_1 - 0.78}{0.11} = -1.64$

$x_1 = -1.64 \times 0.11 + 0.78$

$= 0.599$

The percentage is $= 0.599 \times 100$
$= 59.9\% = 60\%$.

$P(x \geq x_2) = 0.05$

$P\left(z > \frac{x_2 - 0.78}{0.11}\right) = 0.05$

$1 - P\left(z < \frac{x_2 - 0.78}{0.11}\right) = 0.05$

$P\left(z < \frac{x_2 - 0.78}{0.11}\right) = 0.95$

$\frac{x_2 - 0.78}{0.11} = 1.64$

$x_2 = 0.9604$.

The percentage is $= 0.9604 \times 100$
$= 96.04\%$.

Mean deviation of a normal distribution :-

$\int_{-\infty}^{\infty} |x - \mu| f(x) dx$

$\int_{-\infty}^{\infty} |x - \mu| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$ let $\frac{x-\mu}{\sigma} = z \Rightarrow z\sigma = x - \mu$
$\sigma dz = dx$

$\Rightarrow \int_{-\infty}^{\infty} |z\sigma| \, e^{-\frac{z^2}{2}} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \cancel{\sigma} dz$

$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-z^2/2} dz$

$= \frac{\sigma}{\sqrt{2\pi}} 2 \int_{-\infty}^{\infty} z \, e^{-z^2/2} dz$ let $\frac{z^2}{2} = t \Rightarrow z dz = dt$

$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t} dt$

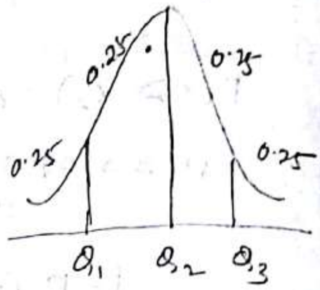
Scanned by CamScanner

$$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \left(e^{-t}\right)_0^{\infty} = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \left[-e^{-\infty} + e^{0}\right]$$

$$= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} = \frac{1.414\sigma}{\sqrt{3.14}} = 0.795\sigma$$

$$\simeq 0.8\sigma = \frac{4}{5}\sigma$$

* Prove that in normal deviation, the quaretial deviation and the standard deviations are approximately 10:12:15.



Standard deviation of normal distribution is σ.

The mean deviation above the mean for normal distribution is $\frac{4}{5}\sigma$.

The quaretial deviation is $z = \frac{Q_3 - Q_1}{2}$

Let x ~~acts~~ follows the normal distribution. if Q_1 and Q_3 are the first and 3^{rd} quaretials respectively then by defination

$P(z < Q_1) = 0.25$

$P(z > Q_3) = 0.25$

$P(0 < z < z_1) = 0.25$

if $x = Q_3 \Rightarrow z = \frac{Q_3 - \mu}{\sigma} = z_1 \rightarrow (1)$

$x = Q_1 \Rightarrow z = \frac{Q_1 - \mu}{\sigma} = -z_1 \rightarrow (2)$

(1) - (2) $\frac{Q_3 - \mu}{\sigma} - \frac{Q_1 - \mu}{\sigma} = 2z_1$

$$\frac{Q_3 - Q_1}{\sigma} = 2z_1$$

$$\Rightarrow z_1\sigma = \frac{Q_3 - Q_1}{2}$$

from the graph,

$$P(0 < z < z_1) = 0.25$$

$$P(z < z_1) - P(z < 0) = 0.25$$

$$P(z < z_1) - 0.5 = 0.25$$

$$P(z < z_1) = 0.75$$

$$z_1 = 0.68\sigma = 2/3\,\sigma$$

Q.D : M.D : S.D

$\frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma$

Multiply with $\frac{15}{\sigma}$

$\frac{2}{\cancel{3}}\cancel{\sigma} \times \frac{\cancel{15}^{5}}{\cancel{\sigma}} : \frac{4}{\cancel{5}}\cancel{\sigma} \times \frac{\cancel{15}^{3}}{\cancel{\sigma}} : \cancel{\sigma} \times \frac{15}{\cancel{\sigma}}$

10 : 12 : 15

1/8/18 Fitting of a normal distribution :-

$$\mu = \frac{\Sigma x f(x)}{\Sigma f(x)}$$

$$\sigma^2 = \frac{\Sigma (x-\bar{x})^2 f(x)}{\Sigma f(x)}$$

Step 1 :-

In order to fit a normal distribution to given data, first we calculate mean and standard deviation of

Step 2 :- To calculate expected normal frequencies, we first find the standard normal variable, to the corresponding lower units of each of the class interval to compute $z = \frac{x_i - \mu}{\sigma}$, where x_i is the lower unit of i class interval.

Step 3: The area under the normal curve to the left of the ordinate at $z = z_i$ say $p(z_i) \leq p(z < z_i)$ are computed from the table.

Step 4:- Finally, area for the class intervals are applying by subtracting $p(z_{i+1}) - P(z_i)$ and on multiplying an area by total frequency, we get the expected normal frequency.

Q. Fit a normal distribution curve for the following data and also expected normal frequencies

C.I	60-65	65-70	70-75	75-80	80-85	85-90	90-9[5]	95-[100]
f(x)	3	21	150	335	326	135	26	4

Sol: $N = \Sigma f(x) = 1000$

Let x be the mean value of class interval

62.5, 67.5, 72.5, 77.5, 82.5, 87.5, 92.5, 97.5

$$\Sigma x f(x) = (62.5)(3) + (67.5)(21) + (72.5)(150) + (77.5)(335) + (82.5)(326) + (87.5)(135) + (92.5)(26) + (97.5)(4).$$

$$= 79945$$

$$\mu = \frac{\Sigma x f(x)}{\Sigma f(x)} = \frac{\cancel{80740}\ 79945}{1000} = 79.945$$

$$\sigma^2 = \frac{\Sigma (x-\bar{x})^2 f(x)}{\Sigma f(x)}$$

$$= \frac{[62.5-79.9]^2 3}{1000} + \frac{[67.5-79.9]^2 21}{1000} + \frac{[72.5-79.9]^2 150}{1000} + \frac{[77.5-79.9]^2 335}{1000} + \frac{[82.5-79.9]^2 326}{1000} + \frac{(87.5-79.9)^2 135}{1000} + \frac{(92.5-79.9)^2 26}{1000} + \frac{(97.5-79.9)^2 4}{1000}$$

~~= -52.2 -260.4 -1110 -804 +847.6 +1028~~

$$= \frac{908.28 + 3228.96 + 8214 + 1929.6 + 2203.76 + 7797.6 + 4127.76 + 1239.04}{1000}$$

$$= \frac{29649}{1000} = 29.649$$

$$\sigma = \sqrt{29.649} = 5.445$$

C.I	LCI	$Z = \frac{x-\mu}{\sigma} = \frac{x-79.945}{5.44}$	P(z)	$\Delta p(z)$
	<60	∞	0	0
60-65	60	-3.66	0	0.0031
65-70	65	-2.747	0.0031	0.0313
70-75	70	-1.828	0.0344	0.1467
75-80	75	-0.909	0.1811	0.3229
80-85	80	0.010	0.5040	0.3172
85-90	85	0.929	0.8212	0.1459
90-95	90	1.848	0.9671	0.03
95-100	95	2.76	0.9971	-0.9971
	>100		0	

Np(z)
0
3.1
31.3
146.7
322.9
317.2
145.9
30
997.1

997.1 ~ 1000.

∴ The expected frequencies are 3.1, 31.3, 146.7, 322.9, 317.2, 145.9, 30.

1/8/18 # 3. Correlation - Regression Sampling Distribution and Estimation.

2/8/18 ## Correlation :

When there are 2 continous variables which are Concerned a joint distribution is known as bivariate distribution.

If there are more than 2 such variables, they joint distribution is known as multivariate distribution.

In case of bivariate/ multi-variate distribution we may be interested in discovering and measuring a magnitude and direction of the relationship between two/ more variables. For this purpose, we use the statistical tool known as correlation

Definition : If the change in one variable effect in the change in the other variable, then the 2 variables are said to be correlated and the degree of relationship is known as correlation.

Types of correlation :-

For a given bivariate data (x, y), the correlation can be deviated into 3 types.

① positive correlation : If 2 variables deviate in the Same direction i.e if the increasing/ decreasing in one variable results in corresponding increasing/ decreasing in the other variable then the correlation is said to be positive correlation

Ex:- Heights and weights

② Negative correlation : If the 2 variables constantly deviate in the opp. direction i.e if the increasing/ decreasing in one variable results in corresponding decreasing/increasing in other variable then the correlation is said to be negative correlation

Ex:- price and demand of items

③ Zero Correlation :- If there is no relationship btw 2 variables such that the value of one variable change and the other variable remains constant is called zero correlation.

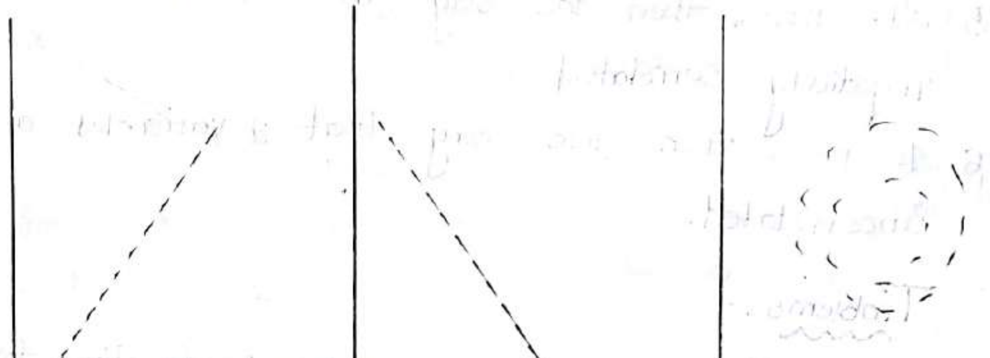
Ex:- weather conditions.

Correlation co-efficient (methods for finding correlation co-efficient):-

1. scatter diagrams
2. Karl pearson's co-efficient of correlation.
3. Spearman Rank. correlation.

1. Scatter diagrams :-

It is the simplest way of the diagramatic representation of bivariate data. Thus for the bivariate distribution. $(x_i, y_i)\ \forall\ i = 1, 2, \ldots, n$. If the values of the variables x and y be plotted along the x-axis and y-axis resp. in xy-plane. The diagram of dots so obtained is known as a scatter diagram.



(+ve) correlation (-ve) (zero)

② Karl pearson's co-efficient of correlation :-

It is useful for measuring the degree of linear relationship b/w the 2 variables x & y and it is usually denoted as r (or) r' (or) r_{xy}

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\ \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2}\ \sqrt{\Sigma (y-\bar{y})^2}}$$

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

$\bar{y} = \frac{\Sigma y}{n}$.

$x-\bar{x}$ means deviation of mean from $\bar{x}$

where n is the no. of observations.

3/8/18

NOTE :-

1. The correlation co-efficient value always lies b/w -1 and 1 i.e $-1 \leq r \leq 1$
2. If $r = +1$ then we say that the two random variables are perfectly positively correlated
3. If $r = -1$ then we say that the two random variables are perfectly negatively correlated
4. If $r > 0$, then we say that the 2 random variables are positively correlated
5. If $r < 0$, then we say that the 2 variables are negatively correlated
6. If $r = 0$, then we say that 2 variables are uncorrelated.

Problems :-

1. Calculate the correlation co-efficient for the following marks obtained by a class of students in english and mathematics.

Marks in English	44	42	40	52	39	32	24	46	41	50
Marks in Maths	24	25	28	29	32	35	36	41	45	50.

Sol:

x	y	x^2	y^2	xy
44	24	1936	576	1056
42	25	1764	625	1050
40	28	1600	784	1120
52	29	2704	841	1508
39	32	1521	1024	1248
32	35	1024	1225	1120
24	36	576	1296	864
46	41	2116	1681	1886
41	45	1681	2025	1845
50	50	2500	2500	2500
410	345	17422	12577	14197.

$\Sigma x = 410$

$\Sigma y = 345$

$\Sigma x^2 = 17422$

$\Sigma y^2 = 12577$

$\Sigma xy = 14197$

$n = 10$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\ \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{10 \times 14197 - 410 \times 345}{\sqrt{10 \times 17422 - (410)^2}\ \sqrt{10 \times 12577 - (345)^2}} = 0.0809 //$$

$\therefore$ Marks in English and Marks in maths are correlated.

2. Calculate the correlation co-efficient for the following heights in inches of fathers and their sons

x	65	66	67	67	68	69	70	72	$= 544$.
y	67	68	65	68	72	72	69	71	

Scanned by CamScanner

Sol

x	y	x^2	y^2	xy
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	1512
544	552	37028	38132	37560

$\Sigma x = 544$
$\Sigma y = 552$
$\Sigma x^2 = 37028$
$\Sigma y^2 = 38132$
$xy = 37560$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\ \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{8 \times 37560 - 544 \times 552}{\sqrt{8 \times 37028 - (544)^2}\ \sqrt{8 \times 38132 - (552)^2}} = 0.603$$

$\therefore$ x and y are positively correlated

4/8/18

3. A computer while calculating correlation co-efficient btw 2 variables x and y from 25 pairs of observations obtained by the following results $n = 25$, $\Sigma x = 125$, $\Sigma y = 100$, $\Sigma x^2 = 650$, $\Sigma y^2 = 460$, $\Sigma xy = 508$. However later discovered the time checking that he had copied down two pairs are

x	6	8
y	14	6

but the values

x	8	6
y	12	8

obtained the calculate the correct correlation. coefficient

Sol The correct values are $n = 25$

$\Sigma x = 125 - 6 - 8 + 8 + 6 = 125$

$\Sigma y = 100 - 14 - 6 + 12 + 8 = 100$

$\Sigma x^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$

$\Sigma y^2 = 460 - 14^2 - 16^2 + 12^2 + 8^2 = 436$

$\Sigma xy = 508 - (6 \times 14) - (8 \times 6) + (8 \times 12) + (6 \times 8)$

$= 520.$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{25 \times 520 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}}$$

$$= 0.66$$

$\therefore$ x and y are positively correlated

4\. For the following data compute the correlation btw x and y. No. of items is equal to 15; Sum of squares of deviation from the mean x and y are 136 and 138. Summation of product of deviation of x and y from the respective arithmetic is 122.

Sol $n = 15$

$\Sigma (x-\bar{x})^2 = 136$

$\Sigma (y-\bar{y})^2 = 138$

$\Sigma (x-\bar{x})(y-\bar{y}) = 122$

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2}\sqrt{\Sigma (y-\bar{y})^2}} = \frac{122}{\sqrt{136}\sqrt{138}} = 0.89$$

$\therefore$ x and y are positively correlated.

5\. Calculate co-relation co-efficient btw x and y from the following data

x	1	3	4	5	7	8	10
y	2	6	8	10	14	16	20

6\. Calculate the co-relation co-efficient for the following marks in Statistics 1st and 2nd papers

Marks in Paper 1	18	45	55	56	58	60	65	68	70	75	85
Marks in Paper 2	82	56	50	48	60	62	64	65	70	74	90.

Model-II

Spearman Rank Correlation :- (Rank correlation)

$\gamma\ \rho$

$$\boxed{r = 1 - \frac{6\sum d^2}{n(n^2-1)}},$$

where d is the diff. b/w the two ranks and n is the no. of observations.

properties :-

1. The rank correlation value always lies b/w -1 and +1 i.e $-1 \le r \le +1$
2. If $r = 1$ then we say that the 2 variables are perfectly positively correlated if $r = -1$ then we say that the 2 variables perfectly negatively correlated
3. If $r > 0$ then we say that, the 2 variables are positively correlated if $r < 0$ then we say that the 2 variables are negatively correlated
4. If $r = 0$ then we say that, the 2 variables are uncorrelated

Formula for tide ranks / repeated ranks :-

- In some cases there is tie b/w 2 or more cases, in such cases each item have ranks 2nd and 3rd respectively then they are given by $\left(\frac{2+3}{2}\right)^{th}$ rank $= 2.5^{th}$ rank. If 3 items have equal ranks 3^{rd}, 4^{th} and 5^{th} resp then they are given by $\frac{3+4+5}{3} = 4^{th}$ rank to each one.
- Let T_x and T_y be the repeated ranks of x and y. The factor $\frac{1}{12}(T_x + T_y)$ is added to $\sum d^2$. Therefore, rank correlation $\boxed{r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}[T_x + T_y]\right]}{n(n^2-1)}}$

where $T_x = \Sigma m_i (m_i^2 - 1)$

$T_y = \Sigma m_j (m_j^2 - 1)$

7. Calculate the rank correlation from the following data

No. of students	1	2	3	4	5	6	7	8	9	10
Ranks in Maths	1	3	5	4	6	2	10	9	8	7
Ranks in stats	3	1	4	5	6	9	7	8	10	2

Sol:

R_1	1	3	5	4	6	2	10	9	8	7
R_2	3	1	4	5	6	9	7	8	10	2
$d = R_1 - R_2$	-2	2	1	-1	0	-7	3	1	-2	5
d^2	4	4	1	1	0	49	9	1	4	25

$\Sigma d^2 = 98$, $n = 10$

$$r = 1 - \frac{6 \Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 98}{10(10^2-1)} = 0.4060$$

∴ Ranks in maths and ranks in stats are positively correlated

8. calculate the rank correlation from the following data.

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
10	30	9	9.	0	0
15	42	5	3	2	4
12	45	8	2	6	36
17	46	3	1	2	4
13	33	7	8	-1	1
16	34	4	7	-3	9
24	40	1	4	-3	9
14	35	5	6	0	0
22	39.	2	5	-3	9
					72.

$\Sigma d^2 = 72$

$n = 9.$

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

$$= 1 - \frac{\cancel{6}^{2} (\cancel{72})^{\cancel{24}\,12}}{\cancel{9}_{3} (\cancel{80})^{\cancel{40}\,20}}$$

$$= \frac{\cancel{8}^{6.0}}{\cancel{20}} = 0.\cancel{40}$$

$= 0.4$

∴ x and y are positively correlated

8/8/18

9. Calculate rank correlation of 6 students from the following data.

Marks in statistics	40	42	45	35	36	35
Marks in maths	46	43	44	39	40	43

Sol:

x	R_1	y	R_2	$d = R_1 - R_2$	d^2
40	3	46	1	2	4
42	2	43	3.5	-1.5	2.25
45	1	44	2	-1	1
35	5.5	39	6	-0.5	0.25
36	4	40	5	-1	1
35	5.5	43	3.5	+2	4
					12.50

$n = 6$

$\Sigma d^2 = 12.50$

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(T_x + T_y)\right]}{n(n^2-1)}$$

$T_x = m(m^2-1)$ $\langle m=2 \rangle$

$= 2(2^2-1)$

$= 6$

$T_y = m(m^2-1)$, $\langle m=2 \rangle$

$= 2(2^2-1)$

$= 6.$

$$\rho = 1 - \frac{6\left[12.50 + \frac{1}{\cancel{12}}(\cancel{12})\right]}{6(36-1)}$$

$$= 1 - \frac{\cancel{6}(13.50)}{\cancel{6} \times 35}$$

$$= 1 - \frac{\cancel{13.5}\ 2.7}{\cancel{35}\ 7}$$

$$= \frac{4.3}{7} = 0.614$$

$\therefore$ marks in statistics and marks in maths are +vely correlated

10. Calculate the rank correlation co-efficient data for the following data.

x	65	66	67	67	68	68	70	72
y	67	68	65	68	72	72	68	71

Sol:

X	R_1	Y	R_2	$d = R_1 - R_2$	d^2
65	8	67	7	1	1
66	7	68	5	2	4
67	5.5	65	8	-2.5	6.25
67	5.5	68	5	0.5	0.25
68	3.5	72	1.5	2	4
68	3.5	72	1.5	2	4
70	2	68	5	-3	9
72	1	71	3	-2	4
					32.5

$n=8$

$\Sigma d^2 = 32.5$

$$r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(T_x + T_y)\right]}{n(n^2-1)}$$

$$T_x = m_1(m_1^2 - 1) + m_2(m_2^2 - 1)$$

$$m_1 = 2, \quad m_2 = 2$$

$$= 2(2^2-1) + 2(2^2-1) = 12$$

$$T_y = m_1(m_1^2-1) + m_2(m_2^2-1)$$

$$m_1 = 2, \quad m_2 = 3$$

$$= 2(2^2-1) + 3(3^2-1) = 30$$

$$r = 1 - \frac{6\left[32.5 + \frac{1}{12}(12+30)\right]}{8(8^2-1)}$$

$$= 0.57$$

$\therefore$ x and y are positively correlated

11. 10 compitetors in musical test are ranked by the 3 judges A, B and C in the following order

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

using rank correlation for which pair of judges has a nearest approach to common liking in music

Sol | The possible pairs of judges are three (A,B) (B,C) (C,A)
At is enough to calculate rank correlation for pair of judges (A,B) (B,C) (C,A)

Ranks of A	Ranks of B	Ranks of C	$d_1=A-B$	d_1^2	$d_2=B-C$	d_2^2	$d_3=A-C$	d_3^2
1	3	6	-2	4	-3	9	-5	25
6	5	4	1	1	1	1	2	4
5	8	9	-3	9	-1	1	-4	16
10	4	8	6	36	-4	16	2	4
3	7	1	-4	16	6	36	2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	1	1
9	1	10	8	64	-9	81	-1	1
7	6	5	1	1	1	1	2	4
8	9	7	-1	1	2	4	1	1
				200		214		60

$n=10,\ \Sigma d_1^2=200,\ \Sigma d_2^2=214,\ \Sigma d_3^2=60$

$$r_1 = 1-\frac{6\Sigma d_1^2}{n(n^2-1)} = 1-\frac{6\times 200}{10(10^2-1)} = 1-\frac{1200}{990}$$
$$= -0.21$$

$$r_2 = 1-\frac{6\Sigma d_2^2}{n(n^2-1)} = 1-\frac{6\times 214}{10(10^2-1)} = 1-\frac{1284}{990}$$
$$= -0.29$$

$$r_3 = 1-\frac{6\Sigma d_3^2}{n(n^2-1)} = 1-\frac{6\times 60}{990} = 0.63.$$

$\therefore$ The pair of (A,B) (B,C) have the nearest approach to common liking in music.

12. The value of sphere maths rank correlation co-efficient for a certain number of pair of observations was found to be $\frac{2}{3}$. The sum of the squares of difference between the corresponding ranks were 55. Find the number of pairs.

sol) $r = \frac{2}{3}$

$\Sigma d^2 = 55$

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

$\frac{2}{3} = 1 - \frac{6 \times 55}{n(n^2-1)}$

$\frac{330}{n(n^2-1)} = 1 - \frac{2}{3} = \frac{1}{3}$

$n(n^2-1) = 990. = 10(10^2-1)$

$n = 10$ //

13. Find the rank correlation for the following data

X	56	42	72	36	63	47	55	49	38	42	68	60
Y	147	125	116	118	149	128	150	145	115	140	152	150

14. The ranking of 10 students in a subjects A and B are given below

A	3	5	8	4	7	10	2	1	6	9
B	6	4	9	8	1	2	3	10	5	7.

Find the rank correlation.

* Regressions:-

The regression equations are two type.

(i) Regression equation y on x

(ii) " " x on y

9/8/18

Regression can be defined as a method that estimates the value of one variable when that of another variable known.

Line regression :-

Line regressions are two types.

(i) line regression equation y on x

(ii) " " " x on y

Scanned by CamScanner

Direct method of regression equation y on x :-

$\boxed{y = a + b_{yx} x}$, where x is independent variable and y is dependent variable, a is intercept, b_{yx} is regression coefficient y on x. The constants a & b_{yx} can be estimate with by applying least square method. $\boxed{a = \bar{y} - b_{yx}\bar{x}}$ where $\bar{x}$ & $\bar{y}$ be the mean of x and mean of y

$$a = \bar{y} - b_{yx}\bar{x}$$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}$$

$$\boxed{b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}}$$

$\boxed{x = a + b_{xy} y}$

$$a = \bar{x} - b_{xy}\bar{y}$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

Regression co-efficient line x on y is of the form $\boxed{x = a + b_{xy} y}$, y is independent variable x is dependent variable

Properties of regression co-efficient :-

- Correlation co-efficient is the geometric mean of the 2 regression co-efficient
- An one of the regression co-efficient is more than unity then another regression co-efficient is less than unity.
- Arithmetic mean of regression co-efficient is greater than correlation co-efficient

NOTE :- * Both the lines regression passes through the points x, y or $\bar{x}, \bar{y}$ in other hands, the mean values of x, y can be obtained as the point of intersection of 2 lines.

* If $r = 0$, the 2 variables are uncorrelated, the lines of regression becomes $\perp$'er to each other

* If $r = \pm 1$, then the 2 lines of regression either coincide / they are parallel to each other.

* If two regression co-efficients are positive then correlation co-efficient is positive

* If two regression co-efficients are negative then correlation co-efficient is negative

* correlation co-efficient btw x and y is given by

$$\boxed{r = \pm\sqrt{b_{yx}\, b_{xy}}} \quad ; \quad b_{yx} = \frac{r\,\sigma_y}{r\,\sigma_x}$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

where r is the correlation co-efficient, σ_x & σ_y be the standard deviations of x and y.

* formulas for two regression equations by using deviation method.

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

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

15. Compare the co-efficient of correlation and 2 lines of regression for the following data.

Price (x)	Demand (y)	x^2	y^2	xy
14	84	196	7056	1176.
16	78	256	6084	1248
17	70	289	4900	1190
18	75	324	5625	1350
19	66	361	4356	1254
20	67	400	4489	1340
21	62	441	3844	1302
22	58	484	3364	1276
23	60	529	3600	1380
170	620	3280	43318	11516.

Regression equation y on x:

$$y = a + b_{yx}\,x$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{170}{9} = 18.86$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{620}{9} = 68.8$$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}} = \frac{11516 - \frac{170 \times 620}{9}}{\sqrt{3280 - \frac{(170)^2}{9}}} = -2.832$$

$$= \cancel{-23.5078}$$

$$a = \bar{y} - b_{yx}\bar{x}$$

$$= 68.8 - (\cancel{-2.83324})(18.86)$$

$$= \cancel{512.15} \quad = 122.211$$

$$y = \overset{122.211}{\cancel{512.15}} + (\cancel{-23.832})x$$

<u>x on y</u>

$$x = a + b_{xy}y$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}} = \frac{11516 - \frac{170 \times 620}{9}}{\sqrt{43318 - \frac{(620)^2}{9}}} = -0.3214$$

$$= \cancel{-7.920}$$

$$a = \bar{x} - b_{xy}\bar{y}$$

$$= (18.86) - (\overset{0.3214}{\cancel{-7.920}})\,68.8$$

$$= \cancel{563.756} = 40.71232$$

$$x = 40.712 + (-0.3214)y.$$

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

$$= \pm\sqrt{(-2.832)(-0.32)}$$

$$= -0.95$$

Both regretion co-efficients are negative then correlation coefficient (r) is negative

16. ~~If r = −0.2~~

10/8/18

16. Production(x) 55 56 58 59 60 60 62

Export (y) 35 38 38 39 44 43 45.

Find (i) Karl pearson's correlation co-efficient

(ii) regression line x on y

(iii) regression line y on x.

(iv) Estimate export when production is 50

Sol:

x	y	x^2	y^2	xy
55	35	3025	1225	1925
56	38	3136	1444	2128
58	38	3364	1444	2204
59	39	3481	1521	2301
60	44	3600	1936	2640
60	43	3600	1849	2580
62	45	3844	2025	2790
410	282	24050	11444	16568

$n = 7.$

$\Sigma x = 410$

$\Sigma y = 282$

$\Sigma x^2 = 24050$

$\Sigma y^2 = 11444$

$\Sigma xy = 16568$

$\bar{x} = \frac{410}{7} = 58.57$

$\bar{y} = \frac{282}{7} = 40.28$

(ii) x on y

$x = a + b_{xy} y$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}} = \frac{16568 - \frac{410 \times 282}{7}}{11444 - \frac{(282)^2}{7}} = 0.609$$

$$a = \bar{x} - b_{xy}\bar{y} = 58.57 - (0.609)(40.28)$$
$$= 34.03$$

$$x = 34.03 + (0.609)y$$

(iii) y on x

$y = a + b_{yx} x$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{16568 - \frac{410 \times 282}{7}}{24050 - \frac{(410)^2}{7}} = 1.4238$$

$a = \bar{y} - b_{yx}\bar{x}$

$= 40.28 - (1.42)(58.5)$

$= -42.79.$

$y = (-42.79) + 1.42x.$

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

$= \pm\sqrt{1.42 \times 0.609}$

$= \pm 0.929 = 0.929$

Both regression co-efficients are positive therefore correlation co-efficient ~~are~~ is positive.

(iv) Given that $x = 50$

$y = -42.79 + (1.42)(50)$

$= 28.21$

17 2 lines of regressive are given by $x + 2y - 5 = 0$, $2x + 3y - 8 = 0$ and variance of x is 12. Calucculate

(i) mean values of x & y

(ii) correlation co-efficient (iii) variance

Sol $x + 2y - 5 = 0 \rightarrow (1)$

$2x + 3y - 8 = 0 \rightarrow (2)$

On solving (1) & (2)

$y - 2 = 0$

$y = 2$

$x = 1.$

$x = 5 - 2y$

$y = \frac{8}{3} - \frac{2}{3}x$

$b_{xy} = -2$, $b_{yx} = \frac{-2}{3}$

$r = \pm\sqrt{b_{yx}\, b_{xy}} = \pm\sqrt{(-2)(-2/3)}$

$= 1.15$ ✗

$x = 4 - 3/2\, y$

$y = \frac{5}{2} - \frac{1}{2}x$

$b_{xy} = \frac{-3}{2}$, $b_{yx} = \frac{-1}{2}$

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

$= \pm\sqrt{\frac{-3}{2} \times \frac{1}{2}} = \sqrt{0.75}$

$= \pm 0.86$ $r = -0.86$ ✓

Scanned by CamScanner

(iii) $\sigma_x^2 = 12$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$b_{yx}^2 = r^2\frac{\sigma_y^2}{\sigma_x^2}$$

$$\left(\frac{-1}{2}\right)^2 = (-0.86)^2\frac{\sigma_y^2}{12}$$

$$\sigma_y^2 = 4.056$$

18. The equations of 2 regression lines obtained in a correlation analysis are as follows $3x+12y=19$, $3y+9x=46$. Find the values of x and y. The value of correlation co-efficient.

Sol

$3x+12y=19 \rightarrow (1)$

$3y+9x=46 \rightarrow (2)$

$(1)\times 3-(2)$ $\quad x=5$ & $y=1/3$

from (1) $\quad y = \frac{19}{2} - \frac{3}{12}x$

from (2) $\quad x = \frac{46}{9} - \frac{3}{9}y$

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

$$b_{xy} = \frac{-3}{9} = -1/3$$

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

$$= \pm\sqrt{\frac{1}{12}} = -0.289.$$

19. In the partially destroyed laboratory records only the lines of regretion y on x and x on y are available as $4x-5y+330=0$, $20x-9y-107=0$. calculate the mean of x on y, correlation co-efficient (r)

20. If the 2 regression equations of variables x and y are $x = 19.13 - 0.87y$, $y = 11.64 - 0.5x$. Calculate mean of x & y, correlation co-efficient (r). calculate the

Q1. (i) correlation co-efficient (r)

(ii) regression values x on y & y on x

(iii) Obtain the x for $y = 70$

x	65	66	67	67	68	69	70	72
y	67	68	65	68	72	69^{72}	$\cancel{71}$ 69	71

* Find the angle btw 2 regression lines, $y-\bar{y} = b_{yx}(x-\bar{x})$ and $x-\bar{x} = b_{xy}(y-\bar{y})$

Sol

$$m_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$m_2 = b_{xy} = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$$

$$\tan\theta = \left|\frac{m_1 - m_2}{1+m_1 m_2}\right|$$

$$= \left|\frac{\frac{r\sigma_y}{\sigma_x} - \frac{1}{r}\frac{\sigma_y}{\sigma_x}}{1 + \frac{r\sigma_y}{\sigma_x}\left(\frac{1}{r}\frac{\sigma_y}{\sigma_x}\right)}\right|$$

$$= \left|\frac{\left(r-\frac{1}{r}\right)\left(\frac{\sigma_y}{\sigma_x}\right)}{1+\frac{\sigma_y^2}{\sigma_x^2}}\right| = \left|\frac{\left(\frac{r^2-1}{r}\right)\left(\frac{\sigma_y}{\sigma_x}\right)}{\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}}\right|$$

$$= \left|\frac{\left(\frac{r^2-1}{r}\right)\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right|$$

$$\boxed{\theta = \tan^{-1}\left|\frac{\left(\frac{r^2-1}{r}\right)\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right|}$$

* **least Squares :-**

Principle of least square method is sum of squares of residuals is minimum. The straight line is of the form $y = a+bx$, the residual or derivations is denoted by 'E' and is given by $\boxed{E = y-a-bx}$

* Method of least Squares :-

This is method which is used for obtaining the equations of curve which fits best to a given set of observations

* Fitting of Straight line (or) linear curve fitting :-

→ Equation of the Straight line of the form $y=a+bx$ and residual $E=y-a-bx$. According to the principle of least square method the sum of squares of residuals is minimum, $E=\Sigma(y-a-bx)^2 \rightarrow ①$

Differentiate eq① partially w.r.t a and equating to zero.

$$\frac{\partial E}{\partial a}=2\Sigma(y-a-bx)(-1)=0$$

$$\Rightarrow \Sigma y-a-bx=0$$

$$\Sigma y-\Sigma a-\Sigma bx=0$$

$$\Sigma y=\Sigma a+\Sigma bx$$

$$\boxed{\Sigma y=na+\Sigma bx} \rightarrow ②$$

differentiate eq② partially w.r.t b and equating to zero

$$\frac{\partial E}{\partial b}=2\Sigma(y-a-bx)(-x)=0$$

$$-2\Sigma yx-ax-bx^2=0$$

$$\Sigma yx=\Sigma ax+\Sigma bx^2$$

$$\boxed{\Sigma xy=a\Sigma x+b\Sigma x^2} \rightarrow ③$$

eq② & ③ are called normal equations of a st. line equation $y=a+bx$. Solve this two equations we get values of a & b, substitute a & b values in st. line equations, we are getting the required st. line equation.

Another form of Straight line equation is $x=a+by$. The normal eq's of the Straight line equation are,

$$\boxed{\Sigma x=na+\Sigma by \quad \text{and} \quad \Sigma xy=a\Sigma y+b\Sigma y^2}$$

Q. predict y at $x=5$ by fitting a least square st. line to the following data;

x	2	4	6	8	10	12
y	18	15	14	11	11	9

Scanned by CamScanner

Sol

x	y	xy	x^2
2	18	36	4
4	15	60	16
6	14	84	36
8	11	88	64
10	11	110	100
12	9	108	144
42	78	486	364

$n=6$

$\Sigma x=42$

$\Sigma y=78$

$\Sigma x^2=364$

$\Sigma xy=486$

we predict y value that is enough to fit a st. line equation of form $y=a+bx$ i.e the normal eq. are $\Sigma y=na+b\Sigma x$

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

$\Sigma y=na+b\Sigma x$

$78=6a+b(42) \rightarrow (1)$

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

$486=42a+364b \rightarrow (2)$

solving (1) & (2)

$6a+42b=78$

$42a+364b=486$

$a=19, \; b=-0.8571$

∴ Required straight line is,

$y=19-0.8571\times 5$

$y=14.7145$

Q Fit a straight line for the following data;

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

Sol

x	y	xy	x^2
1	2.4	2.4	1
2	3	6	4
3	3.6	10.8	9
4	4	16	16
6	5	30	36
8	6	48	64
24	24	113.2	130

$n=6$

$\Sigma x=24$

$\Sigma y=24$

$\Sigma x^2=130$

$\Sigma xy=113.2$

$\Sigma y = na + b\Sigma x$

$24 = 6a + b(24)$

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

$113.2 = a(24) + b(130)$

$6a + 24b = 24$

$24a + 130b = 113.2$

$a = 1.97$, $b = 0.505$

$\therefore$ Required straight line is,

$y = 1.97 + 0.5x$

Q. Fit a straight line for following data:

x	40	50	60	70	80	90
y	2.5	5	5.5	6	7	8.5

Also predict y at $x = 65$

x	y	xy	x^2
40	2.5	100	1600
50	5	250	2500
60	5.5	330	3600
70	6	420	4900
80	7	560	6400
90	8.5	765	8100

Q. Estimate the production for the year 2010 by fitting a straight line to the following data,

Year	2003	2004	2005	2006	2007
Production (in lakhs)	5	8	14	12	13

Sol

x	y	xy	x^2
2003	5	10015	4012009
2004	8	16032	4016016
2005	14	28070	4020025
2006	12	24072	4024036
2007	13	26091	4028049
10025	52	104280	20100135

$n = 5$

$\Sigma x = 10025$

$\Sigma y = 52$

$\Sigma xy = 104280$

$\Sigma x^2 = 20100135$

$y = a + bx$

$\Sigma y = an + b\Sigma x$

$52 = 5a + b(10025)$

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

$104280 = a(10025) + b(20100135)$

$a = -3999.6$

$b = 2$

$y = a + bx$

at $x = 2010$,

$y = -3999.6 + 2(2010)$

$y = 20.4$

$y = a + bx$

$\Sigma y = na + b\Sigma x$

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

* Fitting of a parabola (or) 2nd degree polynomial :-

Non-linear curve :-

⇒ The second degree non-linear curve is of the form:-

$\boxed{y = a + bx + cx^2} \rightarrow (1)$

⇒ The normal equations of 2nd degree polynomial are,

$\Sigma y = na + b\Sigma x + c\Sigma x^2$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$

⇒ Solving these normal equations we get the constant a, b, c. Substituting the constants in eq (1) we get the required 2nd degree polynomial.

Q. Fit a 2nd degree parabola for a following data:-

x	0	1	2	3	4
y	1	5	10	22	38

Sol

x	y	xy	x^2	x^3	x^4	x^2y
0	1	0	0	0	0	0
1	5	5	1	1	1	5
2	10	20	4	8	16	40
3	22	66	9	27	81	198
4	38	152	16	64	256	608
10	76	243	30	354	100	851

$\Sigma x = 10$, $\Sigma y = 76$, $\Sigma xy = 243$, $\Sigma x^2 = 30$, $\Sigma x^3 = 100$,
$\Sigma x^4 = 354$, $\Sigma x^2 y = 851$, $n = 5$

$\Sigma y = na + b\Sigma x + c\Sigma x^2$

$76 = 5a + 10b + 30c$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$

$243 = 10a + 30b + 100c$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$

$851 = 30a + 100b + 354$

$a = 1.42$, $b = 0.24$, $c = 2.21$.

The required equation is $y = 1.42 + 0.24x + 2.21x^2$

Q. Using the least square method, $y = ax^2 + bx + c$ to the following data:

x	1	3	5	6
y	12	18	25	35

Sol:

x	y	xy	x^2	x^3	x^4	x^2y
1	12	12	1	1	1	12
3	18	54	9	27	81	162
5	25	125	25	125	625	625
6	35	210	36	216	1296	1260
15	90	401	71	369	2003	2059

$\Sigma x = 15$, $\Sigma y = 90$, $\Sigma xy = 401$, $\Sigma x^2 = 71$,

$\Sigma x^3 = 369$, $\Sigma x^4 = 2003$, $\Sigma x^2 y = 2059$, $n = 4$

$\Sigma y = a\Sigma x^2 + b\Sigma x + cn$

$90 = 71a + 15b + 4c$

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$

$401 = 369a + 71b + 15c$

$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$

$2059 = 2003a + 369b + 71c$

$a = 0.709$, $b = -0.63$, $c = 12.29$

The required equation is,

$$y = 0.709 - 0.63x + 12.29x^2$$

Q. Fit a parabola curve for the following data

x	1	2	3	4	5	6	7	8
y	19	22	23	25	26	28	17	20.

Scanned by CamScanner

Q. Using the least square method, $y = ax^2 + bx + c$ to the following data:

x	1	3	5	6
y	12	18	25	35

Sol

x	y	xy	x^2	x^3	x^4	x^2y
1	12	12	1	1	1	12
3	18	54	9	27	81	162
5	25	125	25	125	625	625
6	35	210	36	216	1296	1260
15	90	401	71	369	2003	2059

$\Sigma x = 15$, $\Sigma y = 90$, $\Sigma xy = 401$, $\Sigma x^2 = 71$,

$\Sigma x^3 = 369$, $\Sigma x^4 = 2003$, $\Sigma x^2 y = 2059$, $n = 4$

$\Sigma y = a\Sigma x^2 + b\Sigma x + cn$

$90 = 71a + 15b + 4c$

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x$

$401 = 369a + 71b + 15c$

$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2$

$2059 = 2003a + 369b + 71c$

$a = 0.709$, $b = -0.63$, $c = 12.29$

The required equation is,

$$y = 0.709\,\cancel{x} - 0.63x + 12.29x^2$$

Q. Fit a parabola curve for the following data

x	1	2	3	4	5	6	7	8
y	19	22	23	25	26	28	17	20.

* **Fitting of an exponential curve :-**

→ The form of exponential curve is, $y = ae^{bx}$ (or) $y = ab^{x}$

(i) $y = ae^{bx}$

$$\log y = \log ae^{bx}$$

$$= \log a + \log e^{bx}$$

$$\log y = \log a + bx$$

let $Y = \log y$, $A = \log a$, $B = b$, $X = x$

$$Y = A + BX$$

$$\Sigma Y = \Sigma A + \Sigma BX$$

$$\boxed{\Sigma Y = nA + B\Sigma X}$$

$$\boxed{\Sigma XY = \Sigma XA + B\Sigma X^2}$$

(ii) $y = ab^{x}$

$$\log y = \log ab^{x}$$

$$\log y = \log a + \log b^{x}$$

$$\log y = \log a + x\log b$$

let $Y = \log y$, $A = \log a$, $B = \log b$, $X = x$

$\langle a = 10^{A} \rangle$

$$Y = A + BX$$

$$\Sigma Y = An + B\Sigma X$$

$$\Sigma XY = \Sigma XA + B\Sigma X^2$$

Q. Find the curve for the best fit of curve to data by method least squares.

x	1	5	7	9	12
y	10	15	12	15	21

$y = ae^{bx}$

$\log y = \log ae^{bx}$

$\log y = \log a + bx \log e$

$\log y = \log a + bx$

$Y = \log y$

$A = \log a$

$B = b$

$X = x$

$Y = A + BX$

$\Sigma Y = \Sigma A + B \Sigma X$

$\Sigma XY = \Sigma AX + B \Sigma X^2$

$x = X$	y	$Y = \log y$	X^2	XY
1	10	1	1	1
5	15	1.17	25	5.85
7	12	1.079	49	7.49
9	15	1.17	81	10.53
12	21	1.322	144	15.86
34	73	5.741	300	40.75

$n = 5$, $\Sigma X = 34$, $\Sigma Y = 5.741$, $\Sigma XY = 40.75$, $\Sigma X^2 = 300$

$\Sigma Y = nA + B \Sigma X$

$5.741 = 5A + B(34)$

$\Sigma XY = A \Sigma X + B \Sigma X^2$

$40.75 = 34A + 300B$

$A = \log a$

$a = e^A$

$A = 0.978$, $B = 0.024$

$b = B = 0.024$

$a = 10^A = 10^{0.978}$ [illegible]

$y = ae^{bx}$

$y = 10^{0.978}(e^{0.024x})$

Scanned by CamScanner

16/8/18

Q. Fit an exponential curve of form $y = ae^{bx}$ for the following data

$X = x$	y	$Y = \log y$	X^2	XY
1	19	1.27	1	1.27
2	22	1.34	4	2.68
3	23	1.36	9	4.08
4	25	1.39	16	5.56
5	26	1.41	25	7.05
6	28	1.44	36	8.64
7	17	1.23	49	8.61
8	20	1.30	64	10.4
36		10.74	204	48.29

$\Sigma n = 8$

$\Sigma X = 36$

$\Sigma Y = 10.74$

$\Sigma X^2 = 204$

$\Sigma XY = 48.29$

$y = ae^{bx}$

$\log y = \log a + \log e^{bx}$

$\log y = \log a + bx$

$Y = A + BX$

$\Sigma Y = An + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$\log y = Y$, $\log a = A$

$b = B$, $x = X$

$10.74 = 8A + 36B$

$48.29 = 36A + 204B$

$A = 1.34$

$B = -0.0009$

$a = 10^A = 10^{1.34} = 21.87$

$b = B = -0.0009$

$y = 218 \cdot e^{-0.0009x}$

Q. Fit an exponential curve $y = ab^x$ for the following data.

$X = x$	y	$Y = \log y$	X^2	XY
40	30	1.47	1600	58.8
65	20	1.30	4225	84.5
90	10	1	8100	90
5	80	1.90	25	9.5
30	40	1.60	900	48
10	65	1.81	100	18.1
80	150	1.17	6400	93.6
85	150	1.17	7225	99.45
70	20	1.30	4900	91.3
25	50	1.69	625	42.25
500		14.41	34100	635.2

$\Sigma n = 10$

$\Sigma X = 500$

$\Sigma Y = 14.41$

$\Sigma X^2 = 34100$

$\Sigma XY = 635.2$

$y = ab^x$

$\log y = \log a + x \log b$

Scanned by CamScanner

$\log y = Y$, $\log a = A$

$\log b = B$, $x = X$

$Y = A + BX$

$\Sigma Y = An + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$14.41 = 10A + 500B$

$636.901 = 500A + 34100B$

$A = 1.90$

$B = -0.009$.

$\left\{ \begin{array}{l} A = \log a \\ \Rightarrow a = 10^A \end{array} \right.$

$\left\{ \begin{array}{l} B = \log b \\ b = 10^B \end{array} \right.$

$a = 10^A = 10^{1.90} = 79.4$

$b = 10^B = 10^{-0.009} = 0.979$

$y = ab^x$

$y = 79.4\,(0.979)^x$

Q. Fit an exponential curve $y = ab^x$ for the following data

$X = x$	y	$Y = \log y$	X^2	XY
1	1	·0	1	0
2	1.2	0.079	4	0.158
3	1.8	0.255	9	0.765
4	2.5	0.397	16	1.588
5	3.6	0.556	25	2.78
6	4.7	0.672	36	4.032
7	6.6	0.819	49	5.733
8	9.1	0.959	64	7.672
36		3.737	204	22.728

$\Sigma n = 8$

$\Sigma X = 36$

$\Sigma Y = 3.737$

$\Sigma X^2 = 204$

$\Sigma XY = 22.728$

$\Sigma Y = An + B\Sigma X$

$3.737 = 8A + 36B$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$22.728 = 36A + 204B$

$A = 0.682$

$B = 1.380$

$A = -0.166$

$B = 0.14$

$a = 10^A = 10^{-0.166} = 0.682$

$b = 10^B = 10^{0.14} = 1.380$

$y = (0.682)(1.380)^x$.

Fitting of a power curves:- (square curve)

$y = ax^b$

The form of the power curve is $y = ax^b$

$$y = ax^b$$

$$\Rightarrow \log y = \log ax^b$$

$$\log y = \log a + \log x^b$$

$$\log y = \log a + b \log x$$

$$\log y = Y, \quad \log a = A.$$

$$b = B, \quad \log x = X$$

$$Y = A + BX$$

$$\Sigma Y = An + B \Sigma X$$

$$\Sigma XY = A \Sigma X + B \Sigma X^2.$$

Fit a square geometric curve $y = ax^b$

x	y	$X = \log x$	$Y = \log y$	X^2	XY
1	0.5	0	-0.30	0	+0.00
2	2	0.30	0.30	0.09	0.09
3	4.5	0.47	0.65	0.22	0.30
4	8	0.60	0.90	0.36	0.54
5	12.5	0.69	1.09	0.47	0.75
15		2.06	2.64	1.14	1.38

$\Sigma n = 5$

$\Sigma X = 2.06$

$\Sigma Y = 2.64$

$\Sigma X^2 = 1.14$

$\Sigma XY = 1.38$

$$2.64 = 5A + 2.06B.$$

$$1.38 = 2.06A + 1.14B$$

$A = ?$

$B = ?$

$$Y = A + BX$$

$$\Sigma Y = An + B \Sigma X$$

$$\Sigma XY = A \Sigma X + B \Sigma X^2$$

$a = ?$

$b = ?$

$$y = ax^b = ?$$

18/8/18

Q. Fit a power curve $y = ax^b$ for the following data and estimate by when $x = 4.6$

x	y	$X = \log x$	$Y = \log y$	X^2	XY
1	20	0	1.301	0	0
2	30	0.301	1.477	0.090	0.444
3	52	0.477	1.716	0.227	0.818
4	77	0.602	1.886	0.362	1.135
5	135	0.698	2.130	0.487	1.486
6	211	0.778	2.324	0.605	1.808
7	326	0.845	2.513	0.714	2.110
8	550	0.903	2.740	0.815	2.474
9	1052	0.954	3.022	0.910	2.882
		5.558	9.109	4.15	13.107

$y = ax^b$

$\log y = \log a + \log x^b$

$= \log a + b \log x$

$n = 9$

$\Sigma X = 5.558$

$\Sigma Y = 9.109$

$\Sigma X^2 = 4.15$

$\Sigma XY = 13.107$

$Y = A + BX$

$\Sigma Y = An + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$19.109 = 9A + 5.558B$

$13.17 = 5.558A + 4.15B$

$A = 0.94$ $\quad a = 10^A = 10^{0.94} = 8.70$

$B = 1.907$ $\quad b = B = 1.907$

$y = 8.70x^{1.907}$

$x = 4.6 \Rightarrow y = 8.70(4.6)^{1.907}$

$= 159.73$

Q. Fit a power curve is of the form $y = ax^b$ to the following data

Scanned by CamScanner

x	y	$X=\log x$	$Y=\log y$	X^2	XY
1	8	0	0.9030	0	0
2	15	0.3010	1.176	0.090	0.353
3	30	0.477	1.477	0.2275	0.704
4	60	0.6020	1.778	0.3624	1.076
5	125	0.698	2.096	0.484	1.463
		2.078	7.43	1.16	3.596

$n=5$

$\Sigma X = 2.078$

$\Sigma Y = 7.43$

$\Sigma X^2 = 1.16$

$\Sigma XY = 3.59$

$Y = A + BX$

$\Sigma Y = An + B\Sigma X$

$7.43 = 5A + 2.078B$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$3.594 = 2.078A + 1.16B$

$A = 0.793$

$B = 1.664$

$a = 10^A = 0.793 = 6.20$

$b = B = 1.664$

$y = 6.20\,(x)^{1.664}$

Scanned by CamScanner

MODULE-IV

20/8/18

Testing of Hypothesis

Statistical Hypothesis :-

A statistical hypothesis is an assumption/guess about the parameters of population distribution.

Ex:- The avg. of life of electrical goods is 10 years

Null Hypothesis :-

It is a statistical hypothesis which is to be actually listed for acceptance/regection and usually denoted as H_0.

Alternative Hypothesis :-

An a hypothesis, which is complimentary to the null hypothesis is called an alternative hypothesis and is usually denoted by H_1.

Ex:- The alternative hypothesis of Null hypothesis is $(\mu = \mu_0)$ that implies $\mu \neq \mu_0$ i.e $\mu < \mu_0$ (or) $\mu > \mu_0$

Two Tailed Alternative Hypothesis :-

The alternative hypothesis $H_1 : \mu \neq \mu_0$ is known as two tailed alternative hypothesis as it contains both $\mu < \mu_0$ and $\mu > \mu_0$ otherwise i.e either $\mu < \mu_0$ or $\mu > \mu_0$ then it is called one tailed alternative hypothesis. Here $\mu < \mu_0$ is a left tailed test and $\mu > \mu_0$ is a right tailed test ∵ the test is based on sample observations and the decision of acceptance/rejection of null hypothesis is always subjected to some year errors

Error in Sampling :-

The main object in sampling theory is to draw valid infereance about the population parameter on the basis of the sampling results in practice, we decide to accept / reject the large of an examination A sample from it as such we have two types of errors.

(i) Reject null hypothesis when it is true

(ii) Accept null hypothesis when it is false. i.e accept H_0 when H_1 is true. i.e propability of reject H_0 when it is true is equal to type I error that is equal to α.

Probability of acceptance of H_0 when it is wrong $p(\text{type II error}) = \beta$

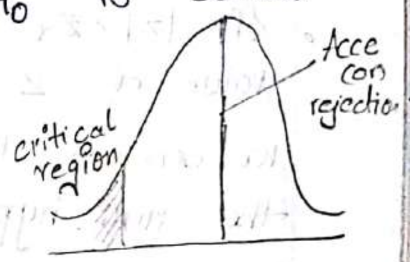
Then α and β are called type I and type II error

Level of significance:-

The level of significance of test is denoted by α, indicates the probability of vomiting type I error. Generally we fix α, before sampling information. In general, we use α as 5% or 1%. when level = 5% means 5 chances in 100. are null hypo-thesis is true (or) 95% confidence level, that the correct decision made. Level of significance is also denotes the size of test

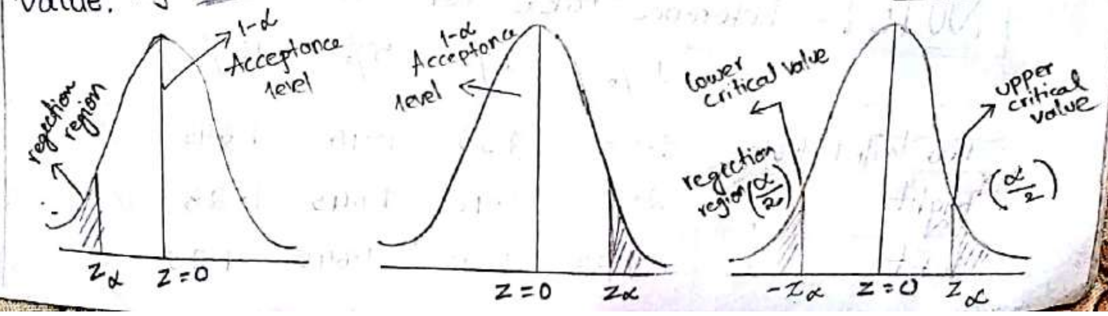
critical region :-

A region in the sample space which corresponds to the area of rejection of H_0 is called critical region.



Critical value :-

The critical value is the value of test statistics which separates the critical region and acceptance region is called critical value (or) significance value.



23/8/18

Step 1:- NULL Hypothesis (H_0)

Define or set up a null hypothesis (H_0) in clear terms

Step 2:- Alternative Hypothesis (H_1)

Set up the alternative hypothesis (H_1). So, that we could decide whether we should use one-tailed/ two-tailed test

Step 3:- level of Significance

Select the appropriate level of significance (α) depending on the reliability of the estimates and permicable risk i.e a suitable α is selected in advance

Step 4:- Test Statistics

compute the test statistic by using formula

$$z = \frac{t - E(t)}{S \cdot E(t)}$$

Step 5:- we Compare the computed value of the test statistic (z) with the critical value (z_α) at given level of Significance (α)

- If $|z| < z_\alpha$ i.e if the absolute value of the calculated value of z is less than the critical value z_α we conclude that it is not significant., we accept the null hypothesis.
- If $|z| > z_\alpha$ i.e then the diff. is significant, we reject the Null hypothesis, at the given level of Significance.

NOTE:- Reference table for z_α values

	1%	2%	5%	10%
Two tailed test	2.58	2.33	1.96	1.645
Right	2.33	1.96	1.645	1.28
left	-2.33	1.96	-1.645	-1.28

Test law Significance for large Samples:

If the sample size $n>30$, we usually take sample as large sample. If n is large, the distributions such as binomial, poission, Chy square and so on are closely approximated by ~~to~~ normal distribution.
$\therefore$ For large samples, we apply normal tests assuming the populations are normal.

① Testing of significance for single proportions
② " " " " difference of proportions
③ " " " " Single mean
④ " " " " difference of means.

I-model:- Testing of Significance for Single proportions:-

① Suppose a large sample (n) is taken from a normal population; to test the significance diff btw sample proportion P_0 and the population proportion P. we use the static $z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}}$, where $Q = 1-p$.

The confidence limits for single proportion

$P \pm Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$ (or) $P \pm 3\sqrt{\frac{PQ}{n}}$

Q. A coin tossed 10,000 times and it turns up head 5195 times, can the coin may be regarded as an unbiased coin? Test a 5% level of Significance.

sol Sample size $(n) = 10,000$ times

$x = 5195$

$P_0 = \frac{5195}{10,000} = \frac{x}{n} = 0.5195$

$P = 0.5$, $Q = 0.5$

1) Null hypothesis H_0: coins are unbiased i.e $p = 0.5$

2) Alternative hypothesis :- (H_1) : $P \neq 0.5$ (two-tailed test)

3) level of significance :- $\alpha = 5\%$ (At 5% level of Significance) $Z_\alpha = 1.96$

4) Test statistics :- $Z = \dfrac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.5195 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{10,000}}}$ $\frac{195}{5}$

$= 3.9$

5) Conclusion :- $|z| > Z_\alpha$

$(3.9 > 1.96)$

$\therefore$ we reject the Null hypothesis at 5% level of Significance

Q. A Sample of 1000 people in Karnataka State. 540 are rice eaters and the rest are wheat eaters. Can we assume that both the rice and wheat eaters are equally populated in this state at 1% level of significance.

Sol

$n = 1000$

$x = 540$

$P_0 = \dfrac{x}{n} = \dfrac{540}{1000} = 0.54$

$P = 0.5$, $Q = 0.5$

1) Null hypothesis (H_0) :- $P = 0.5$ (Both rice & wheat eaters are equally populated)

2) Alternative hypothesis (H_1) :- $P \neq 0.5$ (two tailed test)

3) level of significance :- 1%

$Z_\alpha = 2.58$

4) Test statistics :- $Z = \dfrac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \dfrac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}} = 2.52$

5) Conclusion :- $|z| < Z_\alpha$

$\therefore$ we accept the Null hypothesis at 1% level of Significance.

24/8/18

Q. A random sample of 125 cool drinkers 68 set they prefer thumbsup to pepsi. Testing of Null hypothesis $P=0.5$ against of alternative hypothesis $P>0.5$

Sol

$n=125$

$x=68$

$P_0 = \frac{68}{125} = 0.544$

$P=0.5, Q=0.5$

1) Null hypothesis: $(H_0) = P=0.5$
2) Alternative hypothesis $(H_1): P>0.5$ (Right tailed test)
3) level of significance $\alpha=5\%$

at 5% $Z_\alpha = 1.645$

4) Test statistics: $z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544-0.5}{\sqrt{\frac{0.5\times0.5}{125}}} = 0.983.$

5) Conclusion: $|z| < Z_\alpha$

∴ we accept the Null hypothesis at 5% level of significance.

Q. Experience had shown that 20% of manufactured products is of the top quality in one day production of 400 articles only 50 are top quality. Test the Hypothesis at 0.05 level of Significance.

Sol

$n=400$

$x=50$

$P_0 = \frac{\cancel{50}}{\cancel{400}_{8}} = 0.125$

$P=0.2, Q=1-p=0.8$

1) Null hypothesis $(H_0) = p=0.2$
2) Alternative hypothesis (H_1): $p\neq 0.2$ (two tailed test)
3) level of significance $\alpha = 5\% = 0.05$

$Z_\alpha = 1.96$

Scanned by CamScanner

4) Test hypothesis $z = \frac{\hat{p} - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125 - 0.2}{\sqrt{\frac{0.2 \times 0.8}{400}}} = -3.75$

5) Conclusion: $|z| > z_\alpha$

$\therefore$ we ~~accept~~ reject the Null hypothesis at 5% level of significance.

Q. A social worker believes that less than 25% of the couples in a certain area used any form of breathe controllers. A random sample of 120 couples was contacted, 20 of them said, they used controllers. Test the belief of the social worker at 0.05 level.

Sol

$n = 120$

$x = 20$

$P_0 = \frac{20}{120} = 0.16$

6

$P = 0.25$, $Q = 1 - 0.25 = 0.75$

1) N.H: $(H_0) = p = 0.25$

2) A.N.H: $(H_1) = p < 0.25$ (L.T.T)

3) L.S: $\alpha = 0.05$

at 5% of L.S $z_\alpha = -1.645$

4) Test hypothesis: $z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.16 - 0.25}{\sqrt{\frac{0.25 \times 0.75}{120}}} = -2.27$

5) Conclusion: $|z| > |z_\alpha|$

$\therefore$ we reject the Null hypothesis at 5% level of significance.

Q. 20 People are attacked by a disease and only 18 are survived, will you reject the Null hypothesis that the survival rate are attacked by this disease is 85%. In favour of the hypothesis i.e more than at 5% level

Sol $n = 20$

$x = 18$

$P_0 = \frac{18}{20} = 0.9$

$P = 0.85,\ Q = 1 - P = 0.15$

1) NH: $(H_0) = P = 0.85$

2) A.N.H: (H_1) : $P > 0.85$ (R.T.T)

3) L.S: $\alpha = 0.05$

at 5% L.S $Z_\alpha = 1.645$

4) Test Statistics: $Z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.9 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = 0.626$

5) Conclusion: $Z \not< Z_\alpha$

$\therefore$ we accept the Null hypothesis at 0.05 level of Significance.

Q. A manufacturer claims that atleast 95% of the equipment, which is supplied to a factory confirmed to specifications and examination of a 200 pieces of equipment revealed that 18 are defective Test of hypothesis is at 5% level of Significance.

Sol $n = 200$

$x = 18$

$P_0 = \frac{18}{200} = 0.09$

$Q = 0.95,\ P = 0.05$

1) NH: $(H_0) = P = 0.05$

2) ANH: $(H_1) = P > 0.05$ (R.T.T)

3) LS: $\alpha = 0.05$

at 5% L.S $Z_\alpha = 1.645$

4) Test statistics: $Z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.09 - 0.05}{\sqrt{\frac{0.95 \times 0.05}{200}}} = 2.595$

5) conclusion: $|Z| > Z_\alpha$

$\therefore$ we reject the Null hypothesis at 0.05 level of significance.

Scanned by CamScanner

Q. A die was thrown 9000 times and of these 3220 yielded 3 (or) 4 in this consistance with the hypothesis that the die was unbiased.

Sol:

$n = 9000$

$x = 3220$

$P_0 = \frac{3220}{9000} = 0.357$

$P = \frac{\cancel{2}^1}{\cancel{6}_3} = \frac{1}{3} = 0.33$ (population proportion) $\left[\frac{1}{6}+\frac{1}{6}\right]$

$Q = \frac{2}{3} = 0.67$

1) N.H (H_0) : $p = 0.33$

2) A.N.H (H_1) : $P \neq 0.33$ (T.T.T)

3) L.S: $\alpha = 0.05$

at 5% L.S $Z_\alpha = 1.96$

4) T.S :- $Z = \frac{P_0 - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.357 - 0.33}{\sqrt{\frac{0.33 \times 0.67}{9000}}} = 5.44$

5) Conclusion: $Z > Z_\alpha$

$\therefore$ we reject the Null hypothesis at 5% level of significance.

Q. A random sample of 500 apples, was taken from a large consignment, 60 are found to be bad within 98% of confidence level.

Q. In a big city, 325 men out of 600 are found to be smokers, does this info. support the conclusion on that the majority of men in this city are smokers.

Q. A manufacturer claims that only 5% of his products are defective, a random sample of 500 are taken among which 100 are defective. Test the hypothesis at 0.01 level of significance.

25/8/18 Test of Significance for II - model.

Difference of proportions :-

Suppose two large samples of size n_1 and n_2 are taken respectively from two different Populations to test the Significance diff. btw the Sample proportions P_{01} and P_{02} then $z = \frac{P_{01} - P_{02}}{\sqrt{pq\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$;

where $p = \frac{n_1 P_{01} + n_2 P_{02}}{n_1+n_2}$, $q = 1-p$

Suppose the population Proportions P_1 and P_2 are given and $P_1 \neq P_2$. If we want to test the hypothesis, that the diff. $P_1 - P_2$ in population is likely to be hidden in Sample. Samples of size n_1 and n_2 from the two Populations respectively, then

$$z = \frac{(P_{01} - P_{02}) - (P_1 - P_2)}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)}}$$

Q Before increasing in excise duty on tea 800 persons out of a Sample of 1000 persons are found to be tea drinkers. After an increasing excise duty, 800 persons are tea drinkers in a sample of 1200 Persons. Stay whether there is a Significance in the consumption of tea after increasing the excise duty, test at 5% level of Significance.

Sol $n_1 = 1000$ $n_2 = 1200$

$x_1 = 800$ $x_2 = 800$

$P_{01} = \frac{x_1}{n_1} = \frac{800}{1000} = 0.8$

$P_{02} = \frac{x_2}{n_2} = \frac{800}{1200} = 0.67.$

1) N.H (H_0) : There is a Significance diff. in the consumption of tea after increasing the excise duty i.e $P_1 = P_2$

$P_1 - P_2 = 0$

2) A.N.H. (H_1) : $P_1 - P_2 \neq 0$

3) L.S; $\alpha = 5\%$

$z_\alpha = 1.96$

4) Test statistics, $z = \dfrac{P_{01} - P_{02}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$p = \frac{n_1 P_{01} + n_2 P_{02}}{n_1 + n_2}$$

$$= \frac{1000 \times 0.8 + 1200 \times 0.67}{1000 + 1200}$$

$$= 0.72$$

$$q = 1 - p = 0.28.$$

$$z = \frac{0.8 - 0.67}{\sqrt{0.72 \times 0.28\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.7$$

5) conclusion : $z > z_\alpha$

∴ we reject the null hypothesis at 5% of significance

30/8/18

Q. A machine puts out 16 imperfect articles in sample of 500. After machine is overheld, it puts out of 3 imperfect articles in a batch of 100 has the machine improved.

Sol: $x_1 = 16$ $n_1 = 500$

$x_2 = 3$ $n_2 = 100$

$$P_{01} = \frac{x_1}{n_1} = \frac{16}{500} = 0.032$$

$$P_{02} = \frac{x_2}{n_2} = \frac{3}{100} = 0.03$$

1) Null hypothesis $H_0 : P_1 = P_2$

2) Alternative hypothesis $H_1 : P_1 \neq P_2$. (Two tailed test)

3) level of significance

$\alpha = 5\%$

$z_\alpha = 1.96$

4) Test statistics, $z = \dfrac{P_{01} - P_{02}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Scanned by CamScanner

$$P = \frac{n_1 P_{01} + n_2 P_{02}}{n_1+n_2}$$

$$= \frac{16+3}{500+100} = \frac{19}{600} = 0.031$$

$$Q = 1-p = 0.969$$

$$Z = \frac{0.032 - 0.03}{\sqrt{0.969 \times 0.31\left(\frac{1}{500}+\frac{1}{100}\right)}} = 0.105$$

5) conclusion :- $Z < Z_\alpha$

$\therefore$ we accept the null hypothesis at 5% level of significance

Q. A random Sample of 715 men in a town contained 500 smokers. After an increasing in the price of tobacco, a sample of 550 men in another town contains 415 smokers, test whether there is any diff. btw the smokers in the 2 town

Sol

$x_1 = 500$ $n_1 = 750$

$x_2 = 415$ $n_2 = 550$

$$P_{01} = \frac{x_1}{n_1} = \frac{500}{750} = 0.67$$

$$P_{02} = \frac{x_2}{n_2} = \frac{415}{550} = 0.754$$

1) Null hypothesis $H_0 : P_1 = P_2$

2) Alternative hypothesis $H_1 : P_1 \neq P_2$ (two tailed test)

3) Level of significance $\alpha = 5\%$

$Z_\alpha = 1.96$

4) T.S, $Z = \dfrac{P_{01} - P_{02}}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$P = \frac{n_1 P_{01} + n_2 P_{02}}{n_1+n_2} = \frac{500+415}{750+550} = \frac{915}{1300}$$

$$= 0.703$$

$$q = 1-p = 0.297$$

$$Z = \frac{0.67 - 0.754}{\sqrt{0.703 \times 0.297\left(\frac{1}{750}+\frac{1}{550}\right)}} = -3.274$$

5) conclusion: , $|Z| > Z_\alpha$

Scanned by CamScanner

∴ we reject the null hypothesis at 5% level of Significance

Q. In 2 large ~~operations~~ populations there are ~~1~~30% and 25% resp. into fair faired people. As this diff. likely to be hidden in sample of 1200 and 900 resp. from the two populations

Sol:- $n_1 = 1200$

$n_2 = 900$

The sample proportion $P_{01} = 30\% = 0.3$

$P_{02} = 25\% = 0.25$

1) N.H :- $(H_0) : P_1 = P_2$

2) A.H :- $(H_1) : P_1 \neq P_2$ (2-Tailed)

3) L.S :- $\alpha = 5\%$

$z_\alpha = 1.96$

4) T.S :- $(z) = \dfrac{P_{01} - P_{02}}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$P = \frac{n_1 P_{01} + n_2 P_{02}}{n_1 + n_2} = \frac{1200 \times 0.3 + 900 \times 0.25}{1200 + 900}$$

$$= 0.278$$

$$q = 1 - p = 0.722$$

$$z = \frac{0.3 - 0.25}{\sqrt{(0.278)(0.722)\left(\frac{1}{1200} + \frac{1}{900}\right)}} = 2.53$$

5) Conclusion:- $z > z_\alpha$

∴ we reject the Null hypothesis at 5% level of Significance.

Q. A ciggratte manufacturing forum claims that its brand A line of ciggrattes out sales its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B. Test whether the 8% diff. is a

Scanned by CamScanner

valid claim.

Sol:
$n_1 = 200$ $x_1 = 42$

$n_2 = 100$ $x_2 = 18$

$P_{01} = \frac{x_1}{n_1} = \frac{42}{200} = 0.21$

$P_{02} = \frac{x_2}{n_2} = \frac{18}{100} = 0.18$

1) N.H :- (H_0) : $P_1 - P_2 = 8\% = 0.08$

2) A.H : (H_1) : $P_1 - P_2 \neq 8\%$ (two tailed test)

3) L.S : $\alpha = 5\%$

$Z_\alpha = 1.96$

4) T.S : $(z) = \frac{(P_{01} - P_{02}) - (P_1 - P_2)}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$P = \frac{n_1 P_{01} + n_2 P_{02}}{n_1 + n_2} = \frac{60}{300} = \frac{1}{5} = 0.2$$

$$q = 1 - p = 0.8$$

$$Z = \frac{(0.21 - 0.18) - 0.08}{\sqrt{(0.2)(0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}} = -1.02$$

5) Conclusion :- $|z| < Z_\alpha$

∴ we accept the null hypothesis at 5% level of significance

∴ 8% claim is Valid

Q. A manufacturer of electronic equipment subject to the samples upto completing brands of transistors to an accelerated performance test, if 45 of 180 transistors of the first kind and 34 of 120 transistors of the Second kind. Find the test what can he conclude at the level of significance $(\alpha) = 0.05$ about the difference btw the corresponding sample proportions

Sol:
$n_1 = 180$ $x_1 = 45$

$n_2 = 120$ $x_2 = 34$

Scanned by CamScanner

$P_{01} = \frac{x_1}{n_1} = \frac{45}{180} = 0.25$

$P_{02} = \frac{x_2}{n_2} = \frac{34}{120} = 0.283$

1) Null hypothesis:- $(H_0) : P_1 - P_2 = 0$

2) A.H:- $(H_1) : P_1 - P_2 \neq 0$

3) LS:- $\alpha = 0.05 = 5\%$

$Z_\alpha = 1.96$

(4) TS:- $(z) = \frac{P_{01} - P_{02}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$P = \frac{n_1 P_{01} + n_2 P_{02}}{n_1 + n_2} = \frac{45+34}{180+120} = \frac{79}{300} = 0.263$

$q = 1 - 0.263 = 0.737$

$Z = \frac{0.25 - 0.283}{\sqrt{(0.263)(0.737)\left(\frac{1}{180} + \frac{1}{120}\right)}} = -0.636$

5) Conclusion:- $|z| < z_\alpha$

$\therefore$ we accept the null hypothesis at 5% level of Significance

31/8/18 A random samples of 400 men and 600 women are asked whether they would like to have flyover near their residence. 200 men and 325 women are in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level

Sol $n_1 = 400$, $n_2 = 600$

$x_1 = 200$ $x_2 = 325$

$P_{01} = \frac{200}{400} = 0.5$

$P_{02} = \frac{325}{600} = 0.541$

1) NH (H_0):- $P_1 = P_2$

2) ANH $H_1: P_1 \neq P_2$ (Two- tailed test)

3) L.S $\alpha = 5\%$

$Z_\alpha = 1.96$

4) TS :- $$z = \frac{P_{01} - P_{02}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p = \frac{n_1 P_{01} + n_2 P_{02}}{n_1 + n_2} = \frac{200 + 325}{400 + 600} = 0.525$$

$p = 1 - q$

$\Rightarrow q = 1 - p = 0.475$

$$z = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.475\left(\frac{1}{400} + \frac{1}{600}\right)}} = -1.27$$

5) Conclusion:- $|z| < Z_\alpha$

$\therefore$ we accept the null hypothesis at 5% level of significance

Q. On the basis of their total scores 200 candidates of civil service examination are divided into two groups the upper 30% and the remaining 70% consider the first ques. of the examination among the first grp 40 had the correct answer where as among the second grp 80 are the correct answer on the basis of these results can one conclude that the first ques. is not good. At discrimination ability of time being examined here

Q. In a city A 20% of the random samples of 90 school buses had a certain slight physical defect in another city be 18.5% of a random sample 600 school buses had the same defect in the diff btw proportion significant at 0.05 level of significance

Q. In a sample of 600 students of a certain clg 400 are found to use ball point pens. In another clg from a sample of 900 students 450 used ball point pens. Test whether 2 clgs are significantly different w.r.t the habit of using ball pens

III model

Test of Significance for Single mean:—

Suppose, we want to test whether the given sample of size (n) has been drawn from a population with mean (μ). We set up null hypothesis that there is no diff. btw sample mean ($\bar{x}$) and population mean (μ) then the test of statistics $\boxed{(z) = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}$; where σ is population standard deviation.

If the population standard deviation is not known, then we use the statistics $\boxed{z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}}$; where s is the sample standard deviation.

The confidence intervals are

$$\boxed{CI = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Q. A random sample of 400 students is found to have a mean height 171.38 cm. Can it be reasonable regarded as a sample from a large population with mean height 171.17 cm and standard deviation 3.30 cm. Test 5% level of significance and also find confidence limits.

Sol $n = 400$

$\bar{x} = 171.38$ cm (sample mean)

$\mu = 171.17$ cm (population mean)

$\sigma = 3.30$ cm

1) N·H (Ho) $\mu = 171.17$

2) A·N·H (H_1) $\mu \neq 171.17$ (Two-tailed test)

3) L·S $\alpha = 5\%$

$Z_\alpha = 1.96$

4) T·S $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{171.38 - 171.17}{\frac{3.30}{\sqrt{400}}} = 1.57$$

5) Conclusion:- $z < Z_\alpha$ (we accept the N·H)

$$CI = \left[171.38 - 1.96\left(\frac{3.30}{\sqrt{400}}\right), 171.38 + 1.96\left(\frac{3.30}{\sqrt{400}}\right)\right]$$

$$= (171.38 - 0.3234, 171.38 + 0.3234)$$

$$= (171.0566, 171.7034)$$

Q A sample of 900 members has a mean of 3.4 cm and S.D 2.61 cm is the sample from a large population of mean 3.25 cm and S.D 2.61 cm. If the population is normal and its mean is unknown, find the 95% confidence limits of true mean (sample mean) and also test the hypothesis.

Sol $n = 900$ $\mu = 3.25$ cm

$\bar{x} = 3.4$ $\sigma = 2.61$ cm

$\sigma = 2.61$ cm

1) N·H (Ho) $\mu = 3.25$

2) A·N·H (H_1) $\mu \neq 3.25$ (T·T·T)

3) L·S $\alpha = 5\%$

$Z_\alpha = 1.96$

4) T·S $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = \frac{0.15 \times 30}{2.61} = 1.72$$

5) Conclusion:- $z < Z_\alpha$

we accept the N·H at 5% level of significance

$$CI = \left(3.4 - 1.96\left(\frac{2.61}{\sqrt{900}}\right), 3.4 + 1.96\left(\frac{2.61}{\sqrt{900}}\right)\right)$$

$$= (3.4 - 0.17052, 3.4 + 0.17052)$$

$$= (3.22, 3.57)$$

Q. An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 has had previously recorded what can he conclude at the level of significance $\alpha =$ 0.05. If reading taken at 40 random locations in the given region yielded in a mean of 59.1 with S.D of 5.2

Sol
$n = 40$
$\mu = 57.4$
$\bar{x} = 59.1$, $\sigma = 5.2$

1) NH (H_0): $\mu = 57.4$

2) ANH (H_1): $\mu \neq 57.4$ (T.T.T)

3) L.S $\alpha = 0.05 = 5\%$

$z_\alpha = 1.96$

4) Test statistics $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{59.1 - 57.4}{\frac{5.2}{40}} = 2.06$

5) Conclusion $|z| > z_\alpha$

1/9/18

Q. According to the norms, establishing a mechanical aptitude test, person's who are 18 yrs old and avg. height of 73.2 and S.D 8.6. If 45 randomly selected persons of that age 76.7. Test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance

Sol
$\mu = 73.2$ $\sigma = 8.6$
$n = 45$ $\bar{x} = 76.7$

1) N.H. (H_0): $\mu = 73.2$

2) A.N.H (H_1): $\mu > 73.2$ (R.T.T)

3) level of significance $(\alpha) = 0.01$

$$z_{\alpha} = 2.33$$

4) Test statistics $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{76.7 - 73.2}{\frac{8.6}{\sqrt{45}}} = 2.730$

5) conclusion:- $z > z_{\alpha}$

$\therefore$ we reject the null hypothesis at 1% level of significance.

Q An a random sample of 60 workers, the avg. time taken by them to get to work is 33.8 min with S.D of 6.1 min. can we reject the null hypothesis $\mu = 32.6$ min in favour of A.H $\mu > 32.6$ at $\alpha = 0.05$ level of significance.

Sol: $\bar{x} = 33.8$
$s = 6.1$
$\mu = 32.6$
$n = 60$

1) N.H (H_0): $\mu = 32.6$
2) A.N.H (H_1): $\mu > 32.6$ (R.T.T)
3) level of significance $\alpha = 0.05$
$z_{\alpha} = 1.645$

4) T.S $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{33.8 - 32.6}{\frac{6.1}{\sqrt{60}}} = 1.52$

5) conclusion: $z < z_{\alpha}$

Q. A sample of 400 items from a population whose SD is 10. The mean of the sample is 40. Test whether the sample has came from a population with mean 38. Test at 5% level of significance and also calculate confidence intreval for the population.

Sol: $\bar{x} = 40$ $\quad s = 10$
$\mu = 38$ $\quad n = 400$

1) Null hypothesis (H_0) :- $\mu = 38$

2) Alternative hypothesis (H_1) :- $\mu \neq 38$ (T.T.T)

3) level of significance (α) = 0.05

$z_\alpha = 1.96$

4) Test statistics $z = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$$= \frac{40-38}{\frac{10}{\sqrt{400}}} = \frac{2 \times 20}{10} = 4$$

5) conclusion :- $z > z_\alpha$

$\therefore$ we reject the null hypothesis

$$C.I = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \ \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

$$= \left(40 - 1.96\left(\frac{10}{\sqrt{400}}\right), \ 40 + 1.96\left(\frac{10}{\sqrt{400}}\right)\right)$$

Q. A ladies stenographer, claims that she can take distinction at the rate of 120 watts/min. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 watts with a S.D of 15 watts.

$n = 100$
$\bar{x} = 116$
$\mu = 120$
$s = 15$

1) Null hypothesis (H_0) : $\mu = 120$

2) A.N (H_1) : $\mu \neq 120$ (T.T.T)

3) Level of significance $\alpha = 0.05$

$z_\alpha = 1.96$

4) T.S $z = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{116 - 120}{\frac{15}{\sqrt{100}}} = \dfrac{-4}{1.5} = -2.60$

5) conclusion :- $|z| > z_\alpha$

$\therefore$ we reject the N.H at 5% L.S

Scanned by CamScanner

Q. A length of life of a certain computer is app. normally distributed with mean 800 hrs and S.D 40 hrs. If a random sample of 30 computes has an average, Test the null hypothesis $\mu = 800$ hrs against the A.N.H $\mu \neq 800$ hrs and mean of the sample is 730 hrs. Test the hypothesis at 0.01 level of significance.

Q. An ambulance service claims that, it talks on the average, less than 10 min to reach its destination in emergency calls. A sample of 36 calls, a mean of 30 mins and a varience of 16 min. Test the hypothesis at 0.01 level of significance.

Q. It is claimed that a random sample of 39 tyres has a mean life of 15200 km. This sample was drawn from a population whose mean is 15150 km and S.D of 1200 km. Test the hypothesis at 0.05 level of significance

Q. An insurance agent has claimed that an avg age of Policy holders who issues through him is less than the avg per all agents which is 30.5 years. A random sample of 100 Policy holders who had issued through him given the following age distributions

Age	16-20	21-25	26-30	31-35	36-40
Persons	12	22	20	30	16

calculate the A.M and S.D of this distribution. Use these values to test hypothesis at 5% level of significance.

Sol The given class intervals are discontinous. First we convert this to continous class intervals.

The diff. of upper limit of 1^{st} class interval and lower limit of next class interval. ∴

lower limit − 0.5 and upper limit + 0.5.

Age	15.5-20.5	20.5-25.5	25.5-30.5	30.5-35.5	35.5-40.5
f	12	22	20	30	16
x = mid term C.I	18	23	28	33	38
f(x)	216	506	560	990	608

$\bar{x} = \frac{\Sigma fx}{N} = \frac{2880}{100} = 28.80$

$$s^2 = \frac{\Sigma f(x-\bar{x})^2}{N} = \frac{12 \times (18-28.8)^2 + 22.(23-28.8)^2 + 20(28-28.8)^2 + 30(33-28.8)^2 + 16(38-28.8)^2}{100}$$

$S = 6.32$

IV - model

5/9/18 **Test of Significance for difference of means:-**

Let $\bar{x}_1$ be the mean of sample of size n_1 from a population mean μ_1 and variance σ_1^2, let $\bar{x}_2$ be the mean of a sample of size n_2 from a population with mean μ_2 and variance σ_2^2 to test whether there is any significant difference btw $\bar{x}_1$ and $\bar{x}_2$, we have to use the static $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

If the samples have been drawn from same population then $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

The difference btw the two population mean is not equal to zero then $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

If two population variances are unknown and sample variances are known then test statistics $z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Q. A random sample of 1200 house holders from one town gives the mean income as 500 rupees per month with standard deviation of Rs. 70/- and a random sample of 1000 house holders from another town gives the mean income of 600 Rs per month with S.D 90/- . Test whether the mean income of house holders from two towns differs significantly or not. Test at 5% level of significance.

Sol

$n_1 = 1200$ $n_2 = 1000$

$\bar{x}_1 = 500$ $\bar{x}_2 = 600$

$S_1 = 70$ $S_2 = 90$

① Null hypothesis (H_0) : $\mu_1 = \mu_2$

② A.N.H (H_1) : $\mu_1 \neq \mu_2$ (T.T.T)

③ level of significance $\alpha = 5\%$

$Z_\alpha = 1.96$

④ Test statistics $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{500 - 600}{\sqrt{\dfrac{70^2}{1200} + \dfrac{90^2}{1000}}} = -28.64$

⑤ Conclusion: $|z| > Z_\alpha$

∴ we reject the null hypothesis at 5% level of significance.

Q. A researcher wants to know the intelligence of students in a school, he selected 2 groups of students. In the first group, there are 150 students having mean iQ having of 75 with S.D of 15. In the second group, there are 250 students having mean iQ of 70 with S.D of 20. Test at 1% level of significance.

Sol

$n_1 = 150$ $n_2 = 250$

$\bar{x}_1 = 75$ $\bar{x}_2 = 70$

$S_1 = 15$ $S_2 = 20$

① Null hypothesis (H_0) : $\mu_1 = \mu_2$

② A.N.H (H_1) : $\mu_1 \neq \mu_2$ (T.T.T)

③ level of significance $\alpha = 1\%$

$Z_\alpha = 2.58$

④ Test Statistics $z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{15 - 70}{\sqrt{\frac{15^2}{150} + \frac{20^2}{250}}} = 2.839$

⑤ Conclusion: $z > z_\alpha$

$\therefore$ we reject the null hypothesis at 1% level of significance.

Q The mean heights of 2 large samples of sizes 1000 and 2000 members are 67.5 inches & 68 inches resp. Can the sample be regarded as drawn from the same sample of S.D 2.5 inches. Test the hypothesis.

Sol: $n_1 = 1000 \quad n_2 = 2000$

$\overline{x_1} = 67.5 \quad \overline{x_2} = 68$

$\sigma = 2.5$

1) Null hypothesis (H_0): $\mu_1 = \mu_2$

2) A.N.H (H_1): $\mu_1 \neq \mu_2$ (T.T.T)

3) Level of significance $\alpha = 5\%$

$z_\alpha = 1.96$

4) Test statistics $(z) = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{67.5 - 68}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}}$

$= -5.16$

5) Conclusion: $|z| > z_\alpha$

$\therefore$ we reject the null hypothesis at 5% level of significance.

Q The mean yield of wheat from a district A was 210 pounds of S.D 2.5 pounds from a sample of 100 plots. In another district, the mean yield was 220 pounds of S.D 12 pounds from a sample of 150 plots. Assuming that the S.D of yield in the entire state was 11 pounds. Test whether there was any significance diff btw any mean yield of plots in the 2 districts.

Sol: $n_1 = 100 \quad n_2 = 150$

$\overline{x_1} = 210 \quad \overline{x_2} = 220$

$S.D = 2.5 \quad S_2 = 12$

$\sigma = 11$

① Null hypothesis $(H_0): \mu_1 = \mu_2$

② A.N.H $(H_1): \mu_1 \neq \mu_2$ (T.T.T)

③ level of significance $\alpha = 5\%$

$Z_\alpha = 1.96$

④ Test statistics $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{210 - 220}{\sqrt{11^2\left(\frac{1}{100} + \frac{1}{200}\right)}} = -7.04$

⑤ conclusion: $|z| > z_\alpha$

∴ we reject the Null hypothesis at 5% level of significance

Q An a survey of buying habits 400 women shoppers are chosen at random in a super market A located in a certain section of the city, their average weekly food expenditure is Rs.250/- with S.D of Rs. 40/- . For 400 women shoppers chosen at random in super market B in another section of the city, the avg. weekly food expenditure is Rs. 220/- with S.D of Rs. 55/-. Test at 10% level of significance whether the avg. weekly food expenditure of the 2 populations of shoppers are equal.

sol

$n_1 = 400 \quad n_2 = 400$

$\overline{x_1} = 250 \quad \overline{x_2} = 220$

$s_1 = 40 \quad s_2 = 55$

① Null hypothesis $(H_0): \mu_1 = \mu_2$

② A.N.H $(H_1): \mu_1 \neq \mu_2$ (T.T.T)

③ level of significance $\alpha = 10\%$

$Z_\alpha = 1.645$

④ Test statistics $(Z) = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \dfrac{250 - 220}{\sqrt{\frac{40^2}{400} + \frac{55^2}{400}}} = 8.82$

⑤ conclusion: $Z > z_\alpha$

∴ we reject the Null hypothesis at 10% level of significance

Q. A research investigator is interested in studying whether there is a significant difference in the salaries of B.tech graduates in 2 cities a random sample of size 100 from Bangalore on avg income of 20150 Rs/- another random sample of 60 from chennai results in an avg income of Rs. 20250/-. If the variances of both the populations are given as $\sigma_1^2 = 40000$ & $\sigma_2^2 = 32400$ Rs respectively. Test the hypothesis at 2% level of hypothesis

6/9/18

Q. Two types of new cars producing in USA are tested per milage one sample is consisting of 42 cars averaged 15 km/lit. while the other sample consisting of 80 cars averaged 11.5 km/lit with population variances $\sigma_1^2 = 2$, $\sigma_2^2 = 1.5$ resp. Test whether there is any significance difference in the petrol consumption of this two types of cars (use $\alpha = 0.01$)

Sol: $n_1 = 42$ $\quad \bar{x}_1 = 15$ $\quad \sigma_1^2 = 2$

$n_2 = 80$ $\quad \bar{x}_2 = 11.5$ $\quad \sigma_2^2 = 1.5$

1) Null hypothesis (H_0): $\mu_1 = \mu_2$

2) Alternative hypothesis (H_1): $\mu_1 \neq \mu_2$ (T.T.T)

3) level of significance $(\alpha) = 0.01$

$Z_\alpha = 2.58$

4) Test statistics

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{15 - 11.5}{\sqrt{\frac{2^2}{42} + \frac{(1.5)^2}{80}}}$$

$= 13.58$

5) conclusion:- $Z > Z_\alpha$

$\therefore$ we reject the Null hypothesis at 1% level of significance

Q. Sample students are drawn from 2 universities and from their weights in kg mean and S.D's are calculated and shown below. Make a large sample test to test the significance of the difference btw the means

	Mean	S.D	sample size
University - I	55	10	400
- II	57.	15	100.

Sol

$n_1 = 400$ $\quad \bar{x}_1 = 55$ $\quad S_1 = 400$

$n_2 = 100$ $\quad \bar{x}_2 = 57$ $\quad S_2 = 15$

1) Null hypothesis (H_0) : $\mu_1 = \mu_2$

2) Alternative hypothesis (H_1) : $\mu_1 \neq \mu_2$ (T.T.T)

3) level of significance $(\alpha) = 5\%$.

$Z_\alpha = 1.96$

4) Test statistics

$$(z) = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{55-57}{\sqrt{\frac{10^2}{400} + \frac{15^2}{100}}}$$

$$= -1.26$$

5) conclusion : $|z| < Z_\alpha$

$\therefore$ we accept the Null hypothesis at 5% level of significance

6/9/18 Module-V

Test Of Significance for Small Samples

Small Samples :-

A every imp. concept of the sampling theory is the study of test of significance which enable to decide on the basis of the sample results.

(i) the deviation btw the observed sampled statistics & the hypothetical parameter value is significant

(ii) The deviation btw a sample statistics is significant

I-mod

Test of significance for small samples:-

when the size of the sample is less than 30 then that sample is called as small samples.

Some imp. tests for small samples :-

(i) t-test / students t-test / t-distribution:-

(ii) F-distribution / F-test

(iii) χ^2-distribution

(i) t-distribution :-

The test of statistics for $t = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$; where $\bar{x}$ is sample mean, μ is population mean, s is sample standard deviation and n is sample size

- If sample mean & sample standard deviation are not given then we calculate mean at standard deviations by using following formulas:

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

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

- For t-distribution, degree of freedom $\nu = n-1$

Assumptions for students t-test

(i) sample size must be less than 30 $(n<30)$

(ii) The parent population from which sample is drawn is normal.

(iii) The population S.D is unknown.

(iv) The sample observations are independent i.e Sample is random.

Use of t-test :- • To test for a specified mean

• To test for equality of 2 means of 2 independent samples drawn from 2 normal populations. S.D of the populations being unknown.

• To test the significance of difference btw the means of paired data.

Q. The average length of time for students to register for failed classes at a certain college has been 50 mins with a S.D of 10 mins. A new registration procedure using ~~mechanism~~ modern computing mechanism is being tried if random sample of 12 students has an avg. registration time of 42 mins with S.D of 11.9 mins under the new system. Test the hypothesis that the population mean now less than 50 using a level of significance $\alpha = 0.05$. Assume the population of times to be normal

Sol. $n = 12$, $\bar{x} = 42$, $s = 11.9$, $\mu = 50$, $\sigma = 10$

1) Null hypothesis (H0) :- $\mu_1 = 50$

2) Alternative hypothesis (H1) : $\mu < 50$ (L.T.T)

3) L.S $\alpha = 0.05$

$\nu = \cancel{df} = n-1 = 12-1 = 11$

$t_{11, 0.05} = -1.796$

4) (T.S) $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{42 - 50}{\frac{11.9}{\sqrt{12}}} = -2.328$

5) conclusion :- $|t| > |t_{11, 0.05}|$

$\therefore$ we reject the null hypothesis at 5% level of significance

7/9/18

Q. A sample of 26 bulbs gives a mean life of 990 hrs with S.D of 20 hrs. The manufacturer claims that the mean life of bulbs is 1000 hrs is the sample not up to the standard

Sol. $n = 26$, $\bar{x} = 990$, $s = 20$ hrs , $\mu = 1000$

Scanned by CamScanner

① Null hypothesis (H_0): Sample is upto the Standard i.e $\mu = 1000$

② A.N.H (H_1):- The sample is below standard i.e $\mu < 1000$ (L.T.)

③ level of significance $(\alpha) = 5\%$

$$\nu = df = n-1 = 26-1 = 25$$

$$t_{25,\,0.05} = -1.708$$

4) T.S $\quad (t) = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \dfrac{990-1000}{\frac{20}{\sqrt{26}}} = -2.52$

5) conclusion: $\quad |t| > |t_{25,0.05}|$

∴ we reject the null hypothesis at 5% level of significance. i.e sample is below standard.

Q. A mechanist is making engine parts with axle diameter of 0.70 inches. A random sample of 10 parts shows a mean diameter of 0.742 inches with S.D of 0.4 inches. Compute the statistic you would use to test whether the work is significant.

Sol: $n = 10$, $\bar{x} = 0.742$, $S = 0.4$, $\mu = 0.70$

(i) N.H (H_0): $\mu = 0.70$

(ii) A.N.H (H_1): $\mu \neq 0.70$ (T.T.T)

(iii) L.S $\quad (\alpha) = 5\%$

$$\nu = df = n-1 = 10-1 = 9$$

$$t_{9,\,0.05} = 2.262$$

④ T.S $\quad (t) = \dfrac{\bar{x}-\mu}{s/\sqrt{n}} = \dfrac{0.742-0.70}{0.4/\sqrt{10}} = 0.332$

⑤ conclusion:- $\quad t_{cal} < t_{tab}$

$$t < t_{9,0.05}$$

Q. A random sample of 6 steel beams has a mean comprasive strength of 5839 PSI with a S.D of 648 PSI use this information at the level of significance $\alpha = 0.05$ to test whether the two avg. compressive strength of the steel beam from which this sample come from a population whose mean is ... PSI. Assume normality.

$n=6$, $\bar{x}=5839$, $s=648$, $\mu=5800$

sol: ① N.H (H_0) : $\mu=5800$

② A.N.H (H_1) : $\mu \neq 5800$ (T.T.T)

③ L.S $(\alpha)=5\%$

$\nu = df = n-1 = 6-1 = 5$

$t_{5,0.05} = 2.571$

④ T.S $(t) = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{5839-5800}{\frac{548}{\sqrt{6}}} = 0.147$

⑤ conclusion: $t < t_{5,0.05}$

$\therefore$ we accept the null hypothesis at 5% level of significance

Q. A mechanic is designed to produce washes for electrical devices of avg thickness of 0.025 cm. A random sample of 10 washes was found to have thickness of 0.024 cm with S.D of 0.002 cm. Test the significance of the deviation value of t for 9 degrees of freedom at 5% level is 2.262

sol: $n=10$, $\bar{x}=0.024$, $s=0.002$, $\mu=0.025$.

① N.H (H_0) : $\mu=0.025$

② A.N.H (H_1) : $\mu \neq 0.025$

③ L.S $\alpha = 5\%$

$\nu = df = n-1 = 10-1 = 9$

$t_{9,0.05} = 2.262$

④ T.S $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{0.024-0.025}{\frac{0.002}{\sqrt{10}}} = -1.581$

⑤ conclusion: $|t| < t_{9,0.05}$

$|t_{cal}| < t_{tab}$

$\therefore$ we accept the null hypothesis at 5% level of significance.

Q. A mean life time of a sample of 25 florecent light bulbs produced by a company is computed to 1570 hrs with S.D of 120 hrs. The company claims that the avg. life of the bulb produced by the company is 1600 hrs. Using the level of significance. $\alpha = 0.05$ is the claim is accepted.

:- $n = 25$, $\bar{x} = 1570$, $s = 120$, $\mu = 1600$

(i) N.H $H_0: \mu = 1600$

(ii) A.N.H $H_1: \mu \neq 1600$ (T.T.T)

(iii) L.S $\alpha = 5\%$

$$\nu = df = n-1 = 25-1 = 24$$

$$t_{24, 0.05} = 2.064$$

(iv) (T.S) $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{1570 - 1600}{\frac{120}{\sqrt{25}}} = -1.25$

(v) conclusion $|t_{cal}| < t_{tab}$

$\therefore$ we accept the N.H at 5% level of significance

Q. The avg. breaking strength of steel rod is specified to be 18.5 thousand pounds to test this samples of 14 rods was tested. The mean & S.D are applied 17.85 and 1.955 resp. is the results of experiment significant.

Sol. $n = 14$, $\bar{x} = 17.85$, $S = 1.955$, $\mu = 18.5$

(i) N.H (H_0): $\mu = 18.5$

(ii) A.N.H (H_1): $\mu \neq 18.5$

(iii) L.S $\alpha = 5\%$

$$\nu = df = n-1 = 14-1 = 13$$

$$t_{13, 0.05} = 2.160$$

(iv) T.S $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{17.85 - 18.5}{1.955/\sqrt{14}} = -1.245$

(v) Conclusion :- $|t| < t_{13, 0.05}$.

$\therefore$ we accept the N.H at 5% level of significance

Q. A random sample of size of 16 values from a normal Population gives a mean of 53 and sum of squares of deviation from the mean is 150 can this Sample be regarded as taken from the population having 56 as mean obtained 95% confidence limits for the mean of the population and also test the hypothesis.

sol $n=16$, $\bar{x}=53$.
sum of the squares deviation from the mean $\Sigma(x-\bar{x})^2=150$

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}=\sqrt{\frac{150}{15}}=\sqrt{10}=3.16. \quad , \mu=56.$$

① N.H (H_0) : $\mu=56$
② A.N.H (H_1) : $\mu\neq 56$
③ L.S $\alpha=5\%$

$\nu=df=n-1=16-1=15$

$t_{15,0.05}=2.131$

④ T.S $(t)=\dfrac{\bar{x}-\mu}{s/\sqrt{n}}=\dfrac{53-56}{3.16/\sqrt{16}}=-3.79$

⑤ conclusion $|t|>t_{15,0.05}$
∴ we reject the N.H at 5% level of significance.

Q. A random sample of 10 boys at point of following iq 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do this data support the assumption of the population mean iq of 100. Find a religion range in which most of the mean iq values of sample of 10 boys

sol $n=10$, $\bar{x}=\dfrac{70+120+110+101+88+83+95+98+107+100}{10}=97.2$

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}=\sqrt{\frac{1833.6}{9}}=14.27. \quad , \mu=100$$

① N.H (H_0) : $\mu=100$
② A.N.H (H_1) : $\mu\neq 100$
③ L.S $\alpha=5\%$

$\nu=df=n-1=10-1=9$

$t_{9,0.05}=2.262$

④ T.S $(t)=\dfrac{\bar{x}-\mu}{s/\sqrt{n}}=\dfrac{97.2-100}{14.27/\sqrt{10}}=-0.62$

⑤ conclusion $|t|<t_{9,0.05}$
∴ we accept the N.H at 5% level of significance

Q. The heights of 10 students of a given locality are found to be 70, 67, 62, 68, 61, 68, 64, 64, 66, 70 inches is a reasonable to beleive that the avg. height is greater than 64 inches, test at 5% level of significance for a degree of freedom. $(t=1.833)$

Sol $n = 10$, $\mu = 64$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 3.162$

$\bar{x} = 66$

① N.H (H_0) : $\mu = 64$

② A.N.H (H_1) : $\mu > 64$ (R.T.T)

③ L.S $\alpha = 5\%$

$\nu = df = n-1 = 10-1 = 9$

$t_{9,0.05} = 1.833$

④ T.S $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{66-64}{3.162/\sqrt{10}} = 2$

⑤ Conclusion $t > t_{9,0.05}$

∴ we reject the N.H at 5% level of significance

Q. A random sample from a company very expensive files shows that the orders for a certain kind of machinery where filled resp in 10, 12, 19, 14, 15, 18, 11, 13. Use the L.S $\alpha = 0.01$ to test the client that on the avg such orders are filled in 10.5 days. Assume normality.

Sol $n = 8$, $\bar{x} = 14$, $s = 3.207$, $\mu = 10.5$

$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 3.207$

$\bar{x} = 14$

① N.H (H_0) : $\mu = 10.5$

② A.N.H (H_1) : $\mu \neq 10.5$ (T.T.T)

③ L.S $\alpha = 0.01$

$\nu = df = n-1 = 8-1 = 7$

$t_{7,0.01} = 3.499$

④ T.S $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{14-10.5}{\frac{3.207}{\sqrt{8}}} = 3.089$

⑤ conclusion : $t < t_{7,0.01}$

∴ we accept the N.H at 1% level of significance.

8/9/18

Q The life time of electric bulb from a random sample of 10 from a large confinement of the following data

Item	1	2	3	4	5	6	7	8	9	10
life time in 1000 hrs	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 avg life time of bulbs is 4000 hrs.

Sol $n=10$

$\bar{x} = \frac{1.2+4.6+3.9+4.1+5.2+3.8+3.9+4.3+4.4+5.6}{10} = 4.1 \times 1000$
$= 4100$

$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = 3.47 \times 1000 = 3470$ hrs

$\mu = 4000$

① N.H (H_0): $\mu = 4000$

② A.N.H (H_1): $\mu \neq 4000$ (T.T.T)

③ L.S $\alpha = 0.05$

$\nu = df = n-1 = 10-1 = 9.$

$t_{9,0.05} = 2.262$

④ T.S $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4100-4000}{\frac{3470}{\sqrt{10}}} = \frac{\cancel{100 \times 3}}{\cancel{1170}\ 39} = 0.27$

⑤ conclusion : $t_{cal} < t_{9,0.05}$

$\therefore$ we accept the N.H at 5% level of significance.

II - model

Test of Significance for difference of mean :- (for small samples)

Test of significance for difference of mean for small samples (or) student

To test the significance b/w $\bar{x}_1$ & $\bar{x}_2$ are samples of size n_1 and n_2. we use the statistics $t = \frac{\bar{x}_1-\bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$;

where $S = \sqrt{\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2-2}}$ (or) $S = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2+\Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$ where s_1 and s_2 are the sample standard deviations. In this case $\mu_1 = \mu_2$ i.e $\mu_1 - \mu_2 = 0$.

If $\mu_1 \neq \mu_2$ then test statistics $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ and degree of freedom $\mu = n_1+n_2-2$

we assumin

Q. we assuming specimens Nylon yarn taken from the 2 machines, it was found that 8 specimens from first machine had a mean diameter 9.67 with standard deviation 1.81 while 10 specimens from a second machine had a mean diameter of 7.43 with standard deviation of 1.48. Assuming that the Populations are normal to test the Null hypothesis $\mu_1-\mu_2=1.5$ against the alternative hypothesis $\mu_1-\mu_2>1.5$ at 0.05 level of significance.

Sol:

$n_1=8 \quad \bar{x}_1=1.967 \quad s_1=1.81$

$n_2=10 \quad \bar{x}_2=7.43 \quad s_2=1.48$

① N.H (H_0) : $\mu_1-\mu_2=1.5$

② A.N.H (H_1) : $\mu_1-\mu_2>1.5$ (R.T.T)

③ L.S $\quad \alpha=0.05$

$\nu=df=n_1+n_2-2=8+10-2=16$

$t_{16,0.05}=1.746$

④ T.S $\quad t=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\dfrac{(1.967-7.43)-(1.5)}{S\sqrt{\frac{1}{8}+\frac{1}{10}}}=0.899$

$S=\sqrt{\dfrac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}}$

$=\sqrt{\dfrac{8(1.81)^2+10(1.48)^2}{8+10-2}}=1.734$

⑤ conclusion:- $t_{cal}<t_{16,0.05}$

∴ we accept the N.H at 5% level of significance.

Q. Find the max. diff that we can expect with probability 0.95 btw the means of samples of sizes 10 & 12 from a normal population if their standard deviations are found to be 2 & 3 resp.

Sol:

$n_1=10 \quad s_1=2$

$n_2=12 \quad s_2=3$

① N.H (H_0) : $\mu_1-\mu_2=0$

② A.N.H (H_1) : $\mu_1-\mu_2\neq0$ (T.T.T)

③ L.S $\quad \alpha=0.05$

$\nu=df=n_1+n_2-2=20$

$t_{20,0.05}=2.086$

Scanned by CamScanner

④ T.S $t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}} = \sqrt{\frac{10\times 2^2 + 12\times 3^2}{10+12-2}} = \sqrt{\frac{40+108}{20}} = 2.72$$

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{\left|s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right|} \Rightarrow |\bar{x}_1 - \bar{x}_2| = |t| \left|s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right|$$

$$= (2.086)\left|2.72\sqrt{\frac{1}{10}+\frac{1}{12}}\right|$$

$$= 2.41$$

⑤ conclusion.

Hence the max. diff btw the mean is 2.41.

Q. The means of 2 random samples of sizes 9 and 7 are ~~t and~~ 196.42 and ~~1 2~~198.82 resp. The sum of the squares of the deviations from the means are 26.94 and 18.73 resp. Can the sample be considered to have been drawn from the same Population.

Sol

$n_1 = 9 \quad \bar{x}_1 = 196.42 \quad \Sigma(x_1 - \bar{x}_1)^2 = 26.94$

$n_2 = 7 \quad \bar{x}_2 = 198.82 \quad \Sigma(x_2 - \bar{x}_2)^2 = 18.73$

$$s_1 = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2}{n_1 - 1}} = \sqrt{\frac{26.94}{8}}$$

$$s = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}} = \sqrt{\frac{26.94+18.73}{9+7-2}} = 1.80$$

① N.H (H_0): $\mu_1 - \mu_2 = 0$

② A.N.H (H_1): $\mu_1 - \mu_2 \neq 0$ (T.T.T)

③ L.S $\alpha = 0.05$

$\nu = df = n_1 + n_2 - 2 = 14$

$t_{14, 0.05} =$

④ T.S $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{196.42 - 198.82}{1.80\sqrt{\frac{1}{9}+\frac{1}{7}}}$

⑤ conclusion. $t \quad t_{14, 0.05}$

∴ we the N.H at 5% level of significance.

17/9/18

Q. The iQ's of 16 students from one area of a city showed a mean of 107 with S.D of 10 while the iQ's of 14 students from another city showed a mean of 112 with S.D 8 is the significant difference btw the iQ's of the 2 groups at 0.05 level of significance.

Sol
$n_1 = 16$, $\bar{x}_1 = 107$, $s_1 = 10$
$n_2 = 14$ $\bar{x}_2 = 112$ $s_2 = 8$

1) Null hypothesis (H_0): $\mu_1 = \mu_2$
2) A.N.H (H_1): $\mu_1 \neq \mu_2$ (T.T.T)
3) L.S $\alpha = 0.05$
$\nu = n_1 + n_2 - 2 = 16 + 14 - 2 = 28$
$t_{28, 0.05} = 2.048$
4) T.S $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$s = \sqrt{\frac{s_1^2 n_1 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{16 \times 10^2 + 14 \times 8^2}{16 + 14 - 2}} = 9.44$$

$$t = \frac{107 - 112}{9.44\sqrt{\frac{1}{16} + \frac{1}{14}}} = -1.44.$$

5) Conclusion: $|t_{cal}| < t_{28, 0.05}$
∴ we accept the N.H at 5% level of significance.

Q. A group of boys and girls are given an intelligence test, the mean scores standard deviation and members in each group are as follows:

	Boys	Girls
mean	124	121
S.D	12	10
members	18	14

is the mean scores of boys significantly different from that of girls.

Sol
$s_1 = 12$ $n_1 = 18$ $\bar{x}_1 = 124$
$s_2 = 10$ $n_2 = 14$ $\bar{x}_2 = 121$

1) N.H (H_0): $\mu_1 = \mu_2$
2) A.N.H (H_1): $\mu_1 \neq \mu_2$ (T.T.T)
3) L.S $\alpha = 0.05$
$\nu = n_1 + n_2 - 2 = 18 + 14 - 2 = 30$
$t_{30, 0.05} = 2.042$

4) T.S $t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{124-121}{11.535\sqrt{\frac{1}{18}+\frac{1}{14}}} = 0.729$ t_{cal}

$S = \sqrt{\frac{S_1^2 n_1 + S_2^2 n_2}{n_1+n_2-2}} = \sqrt{\frac{12^2 \times 18 + 10^2 \times 14}{18+14-2}} = 11.535$

5) conclusion :- $t_{cal} < t_{30,0.05}$

∴ we accept the N.H at 5% level of significance

Q. Samples of 2 types of electric light bulbs are tested for length of the life for following data are obtained.

	Type I	type II
	1234	1036
Mean	36	40
S.D	8	7
members		

is the difference in the means significant to warrant that type I is superior to type II regarding length of life.

Q. Two independent samples 8 and 7 items resp. had the following values

sample-I	9	11	13	11	15	9	12	4
sample-II	10	12	10	14	9	8	10	

sol: $\bar{x}_1 = \frac{9+11+13+11+15+9+12+4}{8} = 10.5$

$\bar{x}_2 = \frac{10+12+10+14+9+8+10}{7} = 10.42$

$S = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{n_1+n_2-2}}$

$= \sqrt{\frac{(9-10.5)^2+(11-10.5)^2+\dots+(4-10.5)^2+(10-10.42)^2+(12-10.42)^2+\dots(10-(10.42))^2}{8+7-2}}$

$= 2.769$

1) N.H (H_0) : $\mu_1 = \mu_2$

2) A.N.H (H_1) : $\mu_2 \neq \mu_2$ (T.T.T)

3) L.S $\alpha = 0.05$

$\nu = n_1+n_2-2 = 8+7-2 = 13$

$t_{13,0.05} = 2.160$

4) T.S $t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{10.5-10.42}{2.769\sqrt{\frac{1}{8}+\frac{1}{7}}} = 0.055$

5) conclusion: $t_{cal} < t_{13,0.05}$

∴ we accept the N.H at 5% level of significance

Q. 2 horses A & B are tested according to the time in sec to run a particular track with the following results.

Horse A 28 30 32 33 33 29 34

Horse B 29 30 30 24 27 29 . Test the 5% level of significance

Q. The following are the sample of skills. Test the significance difference btw the means at 0.05 level

Sample-I 74.1 77.7 74.4 74 73.8

Sample-II 70.8 74.9 74.2 70.4 69.2 72.2

19/9/18

Difference of means for paired observation :-

The test of statistic for paired observation $\boxed{t = \frac{\bar{d} - \mu}{\frac{s}{\sqrt{n}}}}$

if $\mu = 0$ then $\boxed{t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}}$, where $\bar{d}$ is the mean of the differences.

and $\boxed{s = \sqrt{\frac{n\sum d^2 - (\sum d)^2}{n(n-1)}}}$, where n is the no. of paired observations.

and degree of freedom (d.f) $\nu = n - 1$

** Q. 10 soldiers participated in a shooting competition in the first week after intensive training they participated in the competition. In the second week before and after training are given as follows:

Scores before training 67 24 57 55 63 54 56 68 33 43

Scores after training 70 38 58 58 56 67 68 75 42 38

Do the data indicate that the soldiers have been benefited by the train.

Sol

	67	24	57	55	63	54	56	68	33	43
	70	38	58	58	56	67	68	75	42	38
d :	-3	-14	-1	-3	7	-13	-12	-7	-9	5
d^2 :	9	196	1	9	49	169	144	49	81	25

$\sum d = -50$, $\sum d^2 = 732$.

$\bar{d} = \frac{\sum d}{10} = \frac{-50}{10} = -5$

$s = \sqrt{\frac{10(732) - (-50)^2}{10(9)}} = 7.318$ $\left(s = \sqrt{\frac{n\sum d^2 - (\sum d)^2}{n(n-1)}} \right)$

$t = \frac{-5}{\frac{7.318}{\sqrt{10}}} = -2.16$

i) N.H (H_0) : There is no diff. btw the scores of soldiers after & before training i.e $\mu = 0$ $(\mu_1 - \mu_2 = 0)$

2) A.N.H (H_1) : Soldiers are benefited after training
$\mu > 0$ (R.T.T)

3) L.S $\alpha = 0.05$
d.f $\nu = n-1 = 10-1 = 9$.
$t_{9,0.05} = 1.833$.

4) T.S $t = \dfrac{\bar{d}}{\frac{S}{\sqrt{n}}}$, $(\mu = 0)$

$t = \dfrac{-5}{\frac{7.318}{\sqrt{10}}} = -2.16$

5) conclusion :- $|t_{cal}| > t_{tab}$
$|t| > t_{9,0.05}$ at 5% level of significance

$\therefore$ we reject the N.H
i.e soldiers are benefited after training.

Q. Two kinds of bumper guards, 6 of each kind are maintained on a variante and then the vehicle was done into a concrete wall. The following are the cost of repair

Gaurd 1	107	148	123	165	102	119
Gaurd 2	134	115	112	151	133	129

Use the 0.01 L.S to test whether the diff. b/w two samples means is significant.

Q. To examine the hypothesis that the husbands are more intelligent than wives. An investigator took a sample of 10 couples and administrated any test which measures the iq of, the results are as follows

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	90	108	85

Test the hypothesis with a reasonable test at level of significance 0.05.

Sol:

	117	105	97	105	123	109	86	78	103	107
	106	98	87	104	116	95	90	90	108	85
d :	11	7	10	1	7	14	-4	-12	-5	22
d^2 :	121	49	100	1	49	196	16	144	25	484

$\Sigma d = 51$ $\bar{d} = \dfrac{\Sigma d}{n} = \dfrac{51}{10} = 5.1$

$\Sigma d^2 = 1185$

$$s = \sqrt{\frac{n\Sigma d^2 - (\Sigma d)^2}{n(n-1)}} = \sqrt{\frac{10 \times 1185 - (51)^2}{10(10-1)}} = 10.137$$

1) N.H (H_0) : $\mu = 0$ (or) $\mu_1 - \mu_2 = 0$

2) A.N.H (H_1) : $\mu > 0$ (or) $\mu_1 - \mu_2 > 0$

3) L.S $\alpha = 0.05$

$\nu = df = n - 1 = 10 - 1 = 9$

$t_{9, 0.05} = 1.833$

4) T.S $t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{5.1}{\frac{10.137}{\sqrt{10}}} = 1.59$

5) conclusion:- $t_{cal} < t_{9, 0.05}$

∴ we accept the N.H at 5% level of hypothesis.

*Q. The avg. losses of certain workers before and after certain prgm are given below use 0.05 level of significance to test whether the prgm is effective. (40 & 35), (70 & 65), (45 & 42), (120 & 116), (35 & 33), (55 & 50), (77 & 73)

Sol:-

	40	70	45	120	35	55	77
	35	65	42	116	33	50	73
d	5	5	3	4	2	5	4
d^2	25	25	9	16	4	25	16

$\Sigma d = 28$ $\bar{d} = \frac{\Sigma d}{n} = \frac{28}{7} = 4$

$\Sigma d^2 = 120$

$$s = \sqrt{\frac{n\Sigma d^2 - (\Sigma d)^2}{n(n-1)}} = \sqrt{\frac{7(120) - (28)^2}{7(7-1)}} = 1.154$$

1) N.H (H_0) : $\mu = 0$ (or) $\mu_1 - \mu_2 = 0$

2) A.N.H (H_1) : $\mu \neq 0$ (or) $\mu_1 - \mu_2 \neq 0$

3) L.S $\alpha = 0.05$

$\nu = df = n - 1 = 7 - 1 = 6$

$t_{6, 0.05} = 2.447$

4) T.S $t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{4}{\frac{1.154}{\sqrt{7}}} = 9.17$

5) conclusion:- $t_{cal} > t_{6, 0.05}$

∴ we ~~accept~~ reject the N.H at 5% level of hypothesis

F-distribution :-

To test whether there is any significance diff. b/w two estimates of population variances are to test, if the two samples come from the same population. We use F-test. In this case, we setup N.H (H_0): $\sigma_1^2 = \sigma_2^2$ i.e population variances are same under the test of statistics

$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$; if $F = \frac{s_1^2}{s_2^2}$ then $s_1^2 > s_2^2$

if $F = \frac{s_2^2}{s_1^2}$ then $s_2^2 > s_1^2$

the degree of freedom $\nu_1 = n_1 - 1$

$\nu_2 = n_2 - 1$

If the sample variance, s^2 is given, we apply the population variance (σ^2) by the equation $n\sigma^2 = (n-1)s^2$

If F is closed to 1 then the 2 samples variances are approximately same.

Q. The measurements of the output of 2 units have given the following results. Assuming that both samples have been obtained from the normal population at 5% level of significance. Test whether 2 populations have the same variances.

unit A	14·1	10·1	14·7	13·7	14
unit B	14	14·5	13·7	12·3	14·1

Sol:

$\bar{x}_1 = \frac{14\cdot1 + 10\cdot1 + 14\cdot7 + 13\cdot7 + 14}{5} = 13\cdot3$

$\bar{x}_2 = \frac{14 + 14\cdot5 + 13\cdot7 + 12\cdot3 + 14\cdot1}{5} = 13\cdot7$

$s_1^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{(14\cdot1 - 13\cdot3)^2 + (10\cdot1 - 13\cdot3)^2 + \dots + (14 - 13\cdot3)^2}{5-1} = 3\cdot37$

$s_2^2 = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{(14 - 13\cdot7)^2 + (14\cdot5 - 13\cdot7)^2 + \dots + (14\cdot1 - 13\cdot7)^2}{5 - 1} = 0\cdot712$

1) N·H (H_0) : $s_1^2 = s_2^2$

2) A·N·H (H_1) : $s_1^2 \neq s_2^2$

3) L·S $\alpha = 0\cdot05$

df $\nu_1 = n_1 - 1 = 5 - 1 = 4$

$\nu_2 = n_2 - 1 = 5 - 1 = 4$

$F = \frac{\text{Greater variance}}{\text{smaller variance}} = \frac{3\cdot37}{0\cdot712} = 4\cdot733$

$F(4, 4) = 6\cdot39$

21/9/18

Q. In one sample of 8 observations, the sum of squares of deviation of the sample values from the sample mean was 84·4 and in the another sample of 10 observation it was 102·6 ∴ Test whether this diff. is significant at 5% level.

Sol
$n_1 = 8$ $\quad \Sigma(x_1 - \bar{x}_1)^2 = 84{\cdot}4$
$n_2 = 10$ $\quad \Sigma(x_2 - \bar{x}_2)^2 = 102{\cdot}6$

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = \frac{84{\cdot}4}{8-1} = 12{\cdot}057$$

$$S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = \frac{102{\cdot}6}{10-1} = 11{\cdot}4$$

1) N·H (H_0): $S_1^2 = S_2^2$

2) A·N·H (H_1): $S_1^2 \neq S_2^2$

3) T·S $F = \frac{S_1^2}{S_2^2} = \frac{12{\cdot}057}{11{\cdot}4} = 1{\cdot}057$

4) L·S $\alpha = 0{\cdot}05$
$v = df = v_1 = n-1 = 7$
$v_2 = n_2 - 1 = 9$
$F_{(7,9),0{\cdot}05} = 3{\cdot}29$

5) conclusion:- $F_{cal} < F_{Tab}$

∴ we accept the N·H at 5% level of significance.

Q. An one sample of 10 observations from a normal population, the sum of squares of deviations of the sample values from the sample mean is 102·4 and in another sample of 12 observations from the another normal population. The sum of squares of the deviation of the sample values from the sample mean is 120·5. Examine whether the normal population have the same variance.

Sol
$n_1 = 10$ $\quad \Sigma(x_1-\bar{x}_1)^2 = 102{\cdot}4$
$n_2 = 12$ $\quad \Sigma(x_2-\bar{x}_2)^2 = 120{\cdot}5$

(same as above problem) Solve it !!

Q. The nicotine contains a mg's in 2 samples of tobacco are found to be as follows

SampleA 24 27 26 21 25
SampleB 27 30 28 31 22 36. Can it be said that 2 Samples came from the same normal population.

Sol
$$\bar{x}_1 = \frac{24+27+26+21+25}{5} = 24{\cdot}66 \qquad \bar{x}_2 = \frac{27+30+28+31+22+36}{6} = 29$$

$$S_1^2 = \frac{\Sigma(x_1-\bar{x}_1)^2}{n_1-1} = 5{\cdot}3$$

$$S_2^2 = \frac{\Sigma(x_2-\bar{x}_2)^2}{n_2-1} = 21.6$$

1) N.H (H₀) : $S_1^2 = S_2^2$

2) A.N.H (H_1) : $S_1^2 \neq S_2^2$

3) T.S $F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3} = 4.07$

4) L.S $\alpha = 0.01$

$\nu = df = \nu_1 = n_1 - 1 = 4$

$\nu_2 = n_2 - 1 = 5$

$F_{(5,4),0.01} = 15.52$

5) conclusion: $F_{cal} > F_{tab}$

∴ we reject the N.H at 1% level of significance

19/11

Q. If 2 independent random samples of size $n_1 = 13$, $n_2 = 7$ are taken from a normal population, what is the probability that the variance of the first sample will be atleast 4 times as large as that of the second sample.

Sol: i.e the variance of sample 1 is atleast 4 times of sample

$S_1^2 \geq 4S_2^2$

let $S_1^2 = 4S_2^2$

$$F = \frac{S_1^2}{S_2^2} = \frac{4\cancel{S_2^2}}{\cancel{S_2^2}} = 4$$

$df = \nu_1 = n_1 - 1 = 12$

$\nu_2 = n_2 - 1 = 6$

for the degree of freedom (12,6) at 5% level, the tabular value is 4.

∴ The req. probability is 0.95

Q. The random samples given the following results:

Sample	Size	Sample mean	sum of squares of deviations from mean
I	10	15	90
II	12	14	108

Test whether the samples came from the same normal populations.

Sol

Chi-square (χ^2) - distribution:-

If set of events $A_1, A_2 \ldots A_n$ are observed to occur with frequencies $O_1, O_2, \ldots O_n$ respectively according to the Probability $A_1, A_2, \ldots A_n$ are expected to occur with $O_1, O_2, \ldots O_n$ frequencies $E_1, E_2, \ldots E_n$ resp., with $O_1, O_2, \ldots O_n$ are called observed frequencies and $E_1, E_2, \ldots E_n$ are called expected frequencies then χ^2 is defined as

$$\boxed{\chi^2 = \frac{\Sigma(O_i - E_i)^2}{E_i}}$$

* χ^2 test for goodness of fit :- Suppose we are given a set of observed frequencies obtained under some experiments and we want to set the experiment results support particular hypothesis (or) theory. Karl Pearson developed a test for testing the significance btw experimental values and theoretical value obtained under some theory (or) hypothesis. This test is known as χ^2 test for goodness of fit.

• χ^2 is used to test whether difference btw observed and expected frequencies are significant.

NOTE :- If data is given in a series of n numbers, then degree of freedom (df) = n-1.

In case of binomial df = n-1,

" " Poisson distribution df = n-2,

" " normal distribution df = n-3,

* Conditions of validity of χ^2-test :- 1) The sample observations should be independent, the total frequency is large. i.e N > 50

2) The constraints of the cell frequency if any are linear.

3) No theoritical are expected frequency should be less than 10. If small theoretical frequencies occur, the difficulties overcome from two or more classes to get the before calculating is $(O_i - E_i)$

Q. A pair of dice are thrown 360 times and the frequencies of each sum indicated below.

Sum	2	3	4	5	6	7	8	9	10	11	12
frequency	8	24	35	37	44	65	51	42	26	14	14

. would you say on the basis of the χ^2-test that the dice are fair at 0.05 level of significance.

Sol The probability of getting of sum

SUM	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

1) Null hypothesis (H_0) : The dice are fair

2) A.N.H (H_1) : The dice are not fair

3)

sum	observed freq. (O)	Expected $E_i = 360\,p(x)$	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
2	8	10	-2	4	0.4
3	24	20	4	16	0.8
4	35	30	5	25	0.83
5	37	40	-3	9	0.225
6	44	50	-6	36	0.72
7	65	60	5	25	0.416
8	51	50	1	1	0.02

9	42	40	2	4	0·1
10	26	30	4	16	0·53
11	14	20	-6	36	1·8
12	14	10	4	16.	1·6
					7·441

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 7{\cdot}441$

(accept)

$df = v = n-1 = 11-1 = 10$ at 5%

χ^2- tabular value for 10 df, is 18·307 $\therefore \chi^2_{cal} < \chi^2_{Tab}$

Q. 4 coins are tossed 160 times and the following results are obtained

No. of heads 0 1 2 3 4

observed freq. 17 52 54 31 6 under the asumption that coins are balanced. Find the expected frequencies of 0, 1, 2, 3 & 4 and test the goodness of it at 5% level.

Sol N.H (H_0): coins are balanced

A.N.H (H_1): coins are unbalanced.

Probability of getting head = 1/2
tail = 1/2

Given Total frequency (N) = 160

No. of coins (n) = 4

$\therefore$ expected frequencies 0, 1, 2, 3 & 4 are $N(p+q)^n = 160\left[\frac{1}{2}+\frac{1}{2}\right]^4$

$= 160\left(4C_0\left(\frac{1}{2}\right)^4 + 4C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 + 4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + 4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) + 4C_4\left(\frac{1}{2}\right)^4\right]$

$= \left[160 \times \frac{1}{16} + 160 \times 4 \times \frac{1}{16} + 160 \times 6 \times \frac{1}{16} + 160 \times 4 \times \frac{1}{16} + 160 \times \frac{1}{16}\right]$

$= (10+40+60+40+10)$

$\therefore$ The expected frequencies of 0, 1, 2, 3 & 4 are 10, 40, 60, 40, 10.

observed freq (O)	Expected freq (E)	(O-E)	$(O-E)^2$	$\left(\frac{(O-E)^2}{E}\right)$
17	10	7	49	4·9
52	40	12	144	3·6
54	60	-6	36	0·6
31	40	-9	81	2·025
6	10	-4	16	1·6
				12·725

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 12{\cdot}725$

$\therefore$ No. of observations = 5

$df = n-1 = 5-1 = 4$

for 4 degree of freedom, the tabular value at 5% level of significance. $\chi^2_{tab} = 9{\cdot}488$ $\therefore \chi^2_{cal} > \chi^2_{Tab}$ (reject)

Q. The no. of automobile accidents for week in a certain community are as follows 12, 8, 20, 3, 14, 10, 15, 6, 8, 4 are these frequencies are in aggrement with the belief that the accidents conditions are same during this 10 week period.

Sol. N.H (H_0): accident conditions are same during 10 week period
A.N.H (H_1): " " " not " " " "

∴ The expected freq. of 10 period are $\frac{12+8+20+3+14+10+15+6+8+4}{10} = \frac{100}{10} = 10$

O	E	O-E	$\frac{(O-E)^2}{E}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10
3	10	-7	4.9
14	10	4	1.6
10	10	0	0
15	10	-5	2.5
6	10	-4	1.6
8	10	-2	0.4
4	10	-6	3.6

$\chi^2 = \frac{\Sigma(O-E)^2}{E} = 25.4$

$df = \nu = 10-1 = 9$

for 9 df at 0.05 level of significance

$\chi^2_{tab} = 16.919$

∴ $\chi^2_{cal} > \chi^2_{tab}$ (reject) i.e accidents not occured in same condition.

Q. A sample analysis of examination results of 500 students was made, it was found that 220 students had failed 170 has secured 3rd class and 90 placed in 2nd class and 20 placed in 1st class. Do this results (figures) compensate with the general exam results within the ratios of 4:3:2:1 for the various categories resp.

Sol. N.H (H_0): The observed results compensates with the general exam results
A.N.H (H_1): " " " not " " " "

Total freq = 500

Expected freq. are in the ratio of 4:3:2:1

If we divided the total freq. 500 in the ratio 4:3:2:1, we get the expected freq 200 : 150 : 100 : 50

O	E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
220	200	20	400	2
170	150	20	400	2.66
90	100	-10	100	1
20	50	-30	900	18

$\chi^2 = \frac{\Sigma(O-E)^2}{E} = 23.66$

$df = \nu = 4-1 = 3$

for 3 df at 0.05 level of significance $\chi^2_{tab} = 7.815$

∴ $\chi^2_{cal} > \chi^2_{tab}$ (reject)

Q. A die is thrown 272 times with the following results. Show that the dice is unbiased

No. appeared on dice	1	2	3	4	5	6
	40	32	28	58	54	60

Sol: N.H (H_0): The die is biased
A.N.H (H_1): The die is unbiased
The expected freq. of 1, 2, 3, 5 & 6 is $\frac{276}{6} = 45.00 = 46$

O	E	$O-E$	$\frac{(O-E)^2}{E}$
40	46	-6	
32	46	-14	
28	46	-18	
58	46	12	
54	46	8	
64	46	18	

$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i}$

d.f = $\nu = 6 - 1 = 5$

for 5 d.f at 0.05 l.s

$\chi^2_{tab} =$

Scanned by CamScanner