KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)

COIMBATORE - 641 021

FACULTY OF ENGINEERING

DEPARTMENT OF SCIENCE AND HUMANITIES

I B.E COMPUTER SCIENCE AND ENGINEERING

LESSON PLAN

SUBJECT : PROBABILITY AND STATISTICS SUB.CODE : 19BECS201

S.No	Topics covered	No. of hours
	UNIT I - BASIC PROBABILITY	
1	Introduction of Probability and Applications	1
2	Probability spaces: Experiment, Events, Axioms and properties	1
3	Conditional probability – Problems	1
4	Concepts of Baye's rule – Problems	1
5	Problems based on Baye's theorem	1
6	Tutorial 1 – Problems based on probability and Baye's theorem	1
7	Idea of Discrete Random Variables	1
8	Problems based on discrete random variables	
9	Independent random variables – Problems	1
10	Concepts of multinomial distribution and sums of independent random variables	1
11	Expectation of Discrete Random Variables – Problems	1
12	Concept of Moments, Variance of a sum and Correlation coefficient	1
13	Chebyshev's Inequality	1
14	Tutorial 2 – Problems based on discrete random variables	1
	Total	**14**
	UNIT II - RANDOM VARIBALES	
15	Introduction to Continuous random variables and their properties	1
16	Continuous random variables – Normal distribution	1
17	Problems based on Normal distribution	1
18	Continuous random variables – Exponential distribution	1
19	Continuous random variables – Gamma distribution	1
20	Problems based on Exponential and Gamma distributions	1
21	Tutorial 3 - Problems based on Continuous random variables	1
22	Bivariate distributions and their properties	1
23	Bivariate Discrete random variables – Joint, marginal and conditional probability mass function	1
24	Problems based on bivariate Discrete random variables	1
25	Problems based on bivariate Discrete random variables	1
26	Bivariate continuous random variables – Joint, marginal and conditional probability density function	1
27	Problems based on bivariate continuous random variables	1
28	Tutorial 4 - Problems based on bivariate distributions	1
	Total	**14**
	UNIT III - BASIC STATISTICS	
29	Measures of Central tendency: Moments, Skewness and Kurtosis	1
30	Problems based on Moments, Skewness and Kurtosis	1

31	Probability distributions : Binomial and Poisson distributions	1
32	Problems based on Binomial distribution	1
33	Problems based on Binomial distribution	1
34	Problems based on Poisson distribution	1
35	Problems based on Poisson distribution	1
36	Tutorial 5 - Problems based on Binomial and Poisson distributions	1
37	Concepts of Correlation and Regression	1
38	Problems based on Karl Pearson's correlation coefficient	1
39	Problems based on Rank correlation coefficient	1
40	Problems based on lines of regression and regression coefficients	1
41	Problems based on lines of regression and regression coefficients	1
42	Tutorial 6 - Problems based on Correlation and Regression	1
	Total	**14**
	UNIT IV – APPLIED STATISTICS	
43	Introduction of Curve fitting by the method of least squares	1
44	Curve fitting by the method of least squares	1
45	Fitting of straight lines	1
46	Second degree parabolas and more general curves	1
47	Problems based on Curve fitting by the method of least squares	1
48	Problems based on Fitting of straight lines and Second degree parabolas	1
49	Tutorial 7 - Problems based on Curve fitting by the method of least squares	1
50	Concept of test of significance – Small and Large samples	1
51	Testing of significance for mean, variance, proportions and differences using large samples	1
52	Test of significance for single mean – Problems	1
53	Test of significance for difference means – Problems	1
54	Test of significance for single proportion – Problems	1
55	Test of significance for difference of proportions – Problems	1
56	Tutorial 8 - Problems based on test of significance for large samples	1
	Total	**14**
	UNIT V – SMALL SAMPLES	
57	Introduction to test of significance of small samples – t, F and Chi-square tests	1
58	Test for single mean - t test	1
59	Test for difference of means – t test	1
60	Problems based on t test	1
61	Test for ratio of variances – F test	1
62	Problems based on F test	1
63	Tutorial 9 - Problems based on t and F tests	1
64	Concepts of Chi-square test	1
65	Chi-square test for goodness of fit – Problems	1
66	Problems based on Chi-square test for goodness of fit	1
67	Chi-square test for independence of attributes – Problems	1
68	Problems based on Chi-square test for independence of attributes	1
69	Tutorial 10 - Problems based on chi-square test	1
70	Discussion of previous years ESE Questions	1
	Total	**14**
	GRAND TOTAL	**70**

Staff- Incharge **HoD / S&H**

Semester-II

19BECS201 Probability And Statistics 4H-4C

Course Objectives

- The objective of this course is to familiarize the students with statistical techniques.
- It aims to equip the students with standard concepts and tools at an intermediate to advanced level that will serve them well towards tackling various problems in the discipline.

Course Outcomes

The students will learn:

1 The ideas of probability and random variables and various discrete andcontinuous probability distributions and their properties.

2 The basic ideas of statistics including measures of central tendency, correlationand regression.

3 The statistical methods of studying data samples.

UNIT I - Basic Probability

Probability spaces, conditional probability, Bayes' rule, independence; Discrete random variables,Independent random variables, the multinomial distribution, sums of independent randomvariables; Expectation of Discrete Random Variables, Moments, Variance of a sum, Correlation coefficient, Chebyshev's Inequality.

UNIT II - Random Variables

Continuous random variables and their properties, distribution functions and densities,normal, exponential and gamma densities. Bivariate distributions and their properties, conditional densities,

UNIT III - Basic Statistics

Measures of Central tendency: Moments, skewness and Kurtosis - Probability distributions: Binomial, Poisson and Normal - evaluation of statistical parameters for these three distributions, Correlation and regression – Rank correlation.

UNIT IV - Applied Statistics

Curve fitting by the method of least squares- fitting of straight lines, second degree parabolas and more general curves. Test of significance: Large sample test for single proportion, difference of proportions, single mean, difference of means, and difference of standard deviations.

UNIT V - Small samples

Test for single mean, difference of means and correlation coefficients, test for ratio of variances - Chi-square test for goodness of fit and independence of attributes.

SUGGESTED READINGS

1. Erwin kreyszig, (2014), Advanced Engineering Mathematics, 9th Edition, John Wiley & Sons.
2. Bali N., Goyal M. (2010), A text book of Engineering Mathematics, 7th Edition, Firewall Media (An imprint of Lakshmi Publications Pvt., Ltd),
3. P.G.Hoel, S. C. Port and C. J. Stone,(2003) Introduction to Probability Theory, UniversalBook Stall
4. S. Ross, (2002)A First Course in Probability, 6th Edition, Pearson Education India
5. W. Feller, (1968)An Introduction to Probability Theory and its Applications, Vol. 1, 3rd Edition, Wiley
6. Veerarajan T, (2010)Engineering Mathematics (for semester III), Tata McGraw-Hill

Faculty Incharge **HoD**

UNIT - I

BASIC PROBABILITY

Introduction:

The word 'Probability or change' is very frequency used in day-to-day conversation. The Statistician I.J. Good, suggests in his "kinds of Probability" that "the theory of Probability is much older than the human species.

The concept and applications of probability, which is a formal term of the popular word "Change" while the ultimate objective is to facilitate calculation of probabilities in business and managerial, science and technology etc., the specific objectives are to understand the following terminology.

Random Experiment: The term experiment refers to describe, which can be repeated under some given conditions. The experiment whose result (outcomes) depends on change is called Random Experiment.

Example:

1. Tossing of a coin is a random experiment.
2. Throwing a die is a random experiment.
3. Calculation of he mean arterial blood pressure of a person under ideal environmental conditions, by using the formula, Blood pressure = $= \frac{Systoloic\ pressure}{Diastolic\ pressure}\ mm / Hg$ is a random experiment.

Sample Space:

The totality of all possible outcomes of a random experiment is called a sample space and it is denoted by s and a possible outcome are element.

The no. of the coins in a sample space denoted by n(s).

Example:

Tossing a coin n(s)=2={H,T}

Event:

The output or result of a random experiment is called an event or result or outcome.

Example:

1. In tossing of a coin, getting head or tail is an event.
2. In throwing a die getting 1 or 2 or 3 or 4 or 5 or 6 is an event.

Events are generally denoted by capital letters A, B, C etc. The events can be of two types. One is simple event and the other is compound event

Favorable event:

The no. of events favorable to an event in a trail is the no.of outcomes which entire the happening of the event.

Mutually Exclusive Events:

Two or more events are said to be mutually exclusive events if the occurrence of one event precludes (excludes or prevents) the occurrence of others, i.e., both cannot happen simultaneously in a single trail.

Example:

1. In tossing of a coin, the events head and tail are mutually exclusive.
2. In throwing a die, all the six faces are mutually exclusive.

Equally Likely Events: Two or more events are said to be equally likely, if there is no reason to expect any one case (or any event) in preference to others. i.e., every outcome of the experiment has equal possibility of occurrence. These are equally likely events.

Exhaustive Number of Cases or Events: The total number of possible outcomes in an experiment is called exhaustive number of cases or events.

Dependent event:

Two events are said to be dependent if the occurance or non occurance of a event in any trail affect the occurance of the other event in other trail.

Classical Definition of Probability: Suppose that an event 'A' can happen in 'm' ways and fails to happen (or non-happen) in 'n' ways, all these 'm+n' ways are supposed equally likely. Then the probability of occurrence (or happening) of the event called its success is denoted by 'P(A)' or simply 'p' and is defined as $P(A) = \frac{m}{m+n}$ ...(1) and the probability of non-occurrence (or non-happening) of the event called its failure is denoted by $P(\overline{E})$ or simply 'q' and is defined as. $P(\overline{A}) = \frac{n}{m+n}$...(2)

From (1) and (2) we observe that the probability of an event can be defined as

$$P(event) = \frac{The\ number\ of\ favourable\ cases\ for\ the\ event}{Total\ number\ of\ possible\ cases}$$

Definition:

Let S be the sample space and A be the event associated with a random experiment. Let n(S) and n(A) be the no .of elements of S & A. Then the probability of the event A occurring denoted as P(A) is defined by

$$P(event) = \frac{The\ number\ of\ favourable\ cases\ for\ the\ event}{Total\ number\ of\ possible\ cases} = \frac{n(A)}{n(S)}$$

Note:

It follows that, $P(A) + P(\overline{A}) = 1$ or $p + q = 1$.

This implies that p=1-q or q=1-p.

Hence $0 \le P(A) \le 1$.

Axiomatic Definition of Probability: Let S be the sample space and A be an event associated with a random experiment. Then the probability of the event A, denoted by P(A), is defined as a real number satisfying he following axioms.

(i) $0 \le P(A) \le 1$

(ii) P(S)=1

(iii) If A and B are mutually exclusive events, $P(A \cup B) = P(A) + P(B)$

(iv) If $A_1, A_{2,\ldots,} A_{n,\ldots}$ are a set of mutually exclusive events, $P(A_1 \cup A_2 \cup \ldots \cup A_{n,\ldots}) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$

Theorem 1: The probability of the impossible event is zero, i.e., if ϕ is the subset (event) containing no sample point, P(ϕ)=0.

Proof: The certain event S and the impossible event ϕ are mutually exclusive.

Hence $P(S \cup \phi) = P(S) + P(\phi)$ [axiom (iii)]

But $S \cup \phi = S$.

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

Hence $P(\phi) = 0.$

Theorem 2: If $\overline{A}$ is the complementary event of A, $P(\overline{A}) = 1 - P(A) \le 1.$

Proof: A and $\overline{A}$ are mutually exclusive events, such that $A \cup \overline{A} = S$

Therefore, $P(A \cup \overline{A}) = P(S)$=1 (Since axiom (ii))

i.e., $P(A) + P(\overline{A}) = 1.$

Therefore, $P(\overline{A}) = 1 - P(A)$

Since $P(A) \ge 0,$it follows that $P(\overline{A}) \le 1.$

Theorem 3: If $B \subset A\, then\, P(B) \le P(A).$

Proof: B and $A\overline{B}$ are mutually exclusive events such that $B \cup A\overline{B} = A$.

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

i.e., $P(B) + P(A\overline{B}) = P(A)$ [axiom (iii)]

Therefore, $P(B) \le P(A).$

Theorem 4: Addition theorem of probability

Statement: For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B).$

Proof: Since $(A\cup B)=A\cup(A'\cap B)$ here A and $(A'\cap B)$ are mutually exclusive.

$$P(A\cup B)=P[A\cup(A'\cap B)]\ ...(1)$$

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

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

Here $(A\cap B)\&(A'\cap B)$ are mutually exclusive events.

$$P(B)=P[(A\cap B)\cup(A'\cap B)]\ ...(2)$$

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

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

From (1), $P(A\cup B)=P(A)+P(B)-P(A\cap B).$

Conditional Probability: The Conditional probability of an event B, assuming that the event A has happened, is denoted by P(B/A) and defined as, $P(B/A)=\dfrac{P(A\cap B)}{P(A)}$, provided $P(A)\neq 0.$

Rewriting the definition of conditional probability, we get $P(A\cap B)=P(A)\times P(B/A).$ [Product theorem of probability]

Properties:

1. If $A\subset B, P(B/A)=1, Since\ A\cap B=A.$
2. If $B\subset A, P(B/A)\geq P(B), Since\ A\cap B=B,$ and $\dfrac{P(B)}{P(A)}\geq P(B), as\ P(A)\leq P(S)=1.$
3. If A and B are mutually exclusive events, P(B/A)=0, since $P(A\cap B)=0$
4. If P(A)>P(B), P(A/B)>P(B/A).
5. If $A_1\subset A_2, P(A_1/B)\leq P(A_2/B).$

Independent Events: A set of events is said to be independent if the occurrence of any one of them does not depend on the occurrence or non-occurrence of the others.

The product theorem can be extended to any number of independent events: $A_1, A_{2,...,} A_n$ are n independent events. $P(A_1\cap A_2\cap ...\cap A_n)=P(A_1)\times P(A_2)\times ...\times P(A_n)$, when this condition is satisfied, the events $A_1, A_{2,...,} A_n$ are also said to be totally independent. A set of events $A_1, A_{2,...,} A_n$ is said to be mutually independent if the events are totally independent when considered in sets of 2,3,. . . n events.

Theorem 5: If the events A and B are independent, then so are $\overline{A}\&\overline{B}.$

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

$=1-[P(A)+P(B)-P(A\cap B)]$ (By addition theorem)

$=1-P(A)-P(B)+P(A)\times P(B)$ {since A and B are independent)

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

$$= P(\overline{A}) \times P(\overline{B})$$

Example 1: In how many different ways can the director of a research laboratory choose two chemists from among seven applicants and three physicists from among nine applicants?

Solution:

The two chemists can be chosen in ${}^{7}C_{2}$ =21 ways

The three physicists can be chosen in ${}^{9}C_{3}$ = 84 ways

Then these two things can be done in 21 x 84 = 1764 ways.

Example 2: What is the probability that a non-leap year contains 53 Sundays?

Solution:

A non-leap year consists of 365 days, of these there are 52 complete weeks and 1 extra day. That day may be any one of the 7 days. So already we have 52 Sundays. For one more Sunday, the probability that getting a one more Sunday is 1/7.

Hence the probability that a non-leap year contains 53 Sundays is 1/7.

Example 3: A bag contains 7 white, 6 red and 5 black balls. Two balls are drawn at random. Find the probability that they will both the white?

Solution:

Given that Balls White(7), Red(6) & Black(5), total 18 balls.

Two balls are drawn at random from 18 balls in ${}^{18}C_{2}$ ways

Two white balls are drawn at random from 7 balls in ${}^{7}C_{2}$ ways.

Hence the required probability = $({}^{7}C_{2})/({}^{18}C_{2}) = 21/153.$

Example 4 : Determine the probability that for a non-defective bolt will be found if out of 600 bolts already examined 12 were defective.

Solution:

Given that out of 600 bolts 12 were defective.

Therefore, probability that a defective bolt will be found = $\frac{12}{600} = \frac{1}{50}$

Therefore, Probability of getting a non-defective bolt = $1 - \frac{1}{50} = \frac{49}{50}.$

Example 5: A fair coin is tossed 4 times. Define the sample space corresponding to this experiment. Also give the subsets corresponding to the following events and find the respective probabilities:

a).More heads than tails are obtained.

b).Tails occur on the even numbered tosses.

Solution:

S= {HHHH, HHHT, HHTH, HHTT, HTHH, HTHT, HTTH, HTTT, THHH, THHT, THTH, THTT, TTHH, TTHT, TTTH, TTTT}

a). Let A be the event is which more heads occur than tails

Then A= {HHHH, HHHT, HHTH. HTHH,THHH}

b).Let B be the event is which tails occur is the second and fourth tosses.

Then B= {HTHT, HTTT,TTHT,TTTT}

$$P(A)=\frac{n(A)}{n(S)}=\frac{5}{16};\ P(B)=\frac{n(B)}{n(S)}=\frac{1}{4}.$$

Example 6: A box contains 4 bad & 6 good tubes. Two are drawn out from the box at a time. One of them is tested and found to be good. What is probability that the other one is also good?

Solution:

Let A = one of the tubes drawn is good and B = the other tube is good .

$P(A\cap B)$ = P(both tubes drawn are good)

$$=\frac{{}^{6}C_{2}}{{}^{10}C_{2}}=\frac{1}{3}$$

Knowing that one tube is good, the conditional probability that the other tube is also good is required, i.e., P(B/A) is required.

By definition, $P(B/A)=\dfrac{P(A\cap B)}{P(A)}=\dfrac{1/3}{6/10}=\dfrac{5}{9}.$

Example 7: In a shooting test, the probability of hitting the target is ½ for A , 2/3 for B , 3/4 for C. If all of them five at the target, find the probability that

i). none of them hits the target.

ii). Atleast one of them hits the target.

Solution:

Let A = event of A hitting the target.

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

$$P(\overline{A}\cap\overline{B}\cap\overline{C})=P(\overline{A})\times P(\overline{B})\times P(\overline{C})\quad \text{(by independence)}$$

i.e., P(none hits the target) = $\dfrac{1}{2}\times\dfrac{1}{3}\times\dfrac{1}{4}=\dfrac{1}{24}$

P(atleast one hits the target) = 1 – P(none hits the target)

$$=1-\frac{1}{24}=\frac{23}{24}.$$

Example:8

Three coins are tossed together find they are exactly 2 head?

Solution:

Total no. of chances by throwing 3 coins are n(S)= 8.

The event A to get exactly 2 heads are A = {HHT, THH,HTH}

n(A)= 3

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$$

Example:9

A bag contains 4 red, 5 white and 6 black balls. What is the probability that 2 balls drawn are red and black?

Solution:

Given that Balls White(5), Red(4) & Black(6), total 15 balls.
Two balls are drawn at random from 15 balls in $15C_2$ ways

n(A)= $4C_1$X $6C_1$, Hence the required probability = $\frac{4C_1 X\, 6C_1}{15C_2} = \frac{8}{35}$

Example :10

A bag contains 3 red and 4 white balls. Two draws are made without replacement.
What is the probability that both balls are red

Solution:

Total no. of balls = 3Red + 4 White = 7 balls

P(Drawing a red ball in the first drawn is red) = $P(A) = \frac{3}{7}$

P(Drawing a red ball in the second drawn is red) = $P(B/A) = \frac{2}{6}$

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

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

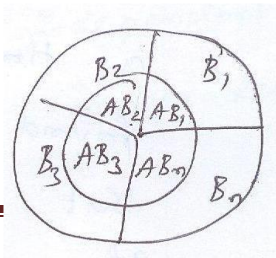
$$P(A \cap B) = P(A)P(B/A)$$

$$= \frac{1}{7}$$

Theorem of Total Probability

Statement: If $B_1, B_{2,...,} B_n$ be a set of exhaustive and mutually exclusive events, and A is another event associated with (or caused by) B_i, then $P(A) = \sum_{i=1}^{n} P(B_i)P(A/B_i)$

Proof. The inner circle represents the event A. A can occur along with (or due to) $B_1, B_{2,...,} B_n$ that are exhaustive and mutually exclusive.



Therefore, $AB_1, AB_{2,...,} AB_n$ are also mutually exclusive.

Therefore, $A = AB_1 + AB_2 + ... + AB_n$ (by addition theorem)

Hence $P(A)=P(\sum AB_i)$

$=\sum P(AB_i)$(since $AB_1, AB_2, \ldots, AB_n$ are mutually exclusive)

$$P(A)=\sum_{i=1}^{n} P(B_i)P(A/B_i)$$

Baye's theorem on Probability (or) Rule of inverse probability

Statement: If $B_1, B_2, \ldots, B_n$ be a set of exhaustive and mutually exclusive events associated with a random experiment and A is another event associated with (or caused by) B_i, then

$$P(B_i/A)=\frac{P(B_i)\times P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\times P(A/B_i)}, i=1,2,\ldots,n$$

Proof. Since by product theorem, $P(A\cap B_i)=P(B_i)\times P(A/B_i)$ . . . (1)

or $P(A\cap B_i)=P(A)\ P(B_i/A)$ . . . (2)

From (1) and (2), $P(A)P(B_i/A)=P(B_i)\ P(A/B_i)$

$$P(B_i/A)=\frac{P(B_i)\ P(A/B_i)}{P(A)} \quad . . . (3)$$

Therefore from total probability, $P(A)=\sum_{i=1}^{n} P(B_i)P(A/B_i)$ substitute in (3), we get

$$P(B_i/A)=\frac{P(B_i)\times P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\times P(A/B_i)}, i=1,2,\ldots,n$$

Example 11: A bag contains 5 balls and it is not known how many of them are white. Two balls are drawn at random from the bag & they are note to be white. What is the chance the all the balls in the bag are white?

Solution:

Since 2 white balls have been drawn out, the bag must have contained 2, 3, 4, or 5 white balls.

Let B_1 = Event of the bag containing 2 white balls.

B_2 = Event of the bag containing 3 white balls.

B_3 = Event of the bag containing 4 white balls.

B_4 = Event of the bag containing 5 white balls.

Let A = Event of drawing 2 white balls.

$$P(A/B_1)=\frac{^2C_2}{^5C_2}=\frac{1}{10},\ P(A/B_2)=\frac{^3C_2}{^5C_2}=\frac{3}{10}$$

$$P(A/B_3)=\frac{^4C_2}{^5C_2}=\frac{3}{5},\ P(A/B_4)=\frac{^5C_2}{^5C_2}=1$$

Since the number of white balls in the bag is not known, B_i's are equally likely.

Therefore $P(B_1)=P(B_2)=P(B_3)=P(B_4)=\frac{1}{4}$

By Baye's theorem,

$$P(B_4/A)=\frac{P(B_4)\times P(A/B_4)}{\sum_{i=1}^{4}P(B_i)\times P(A/B_i)}=\frac{\frac{1}{4}\times 1}{\frac{1}{4}\times\left(\frac{1}{10}+\frac{3}{10}+\frac{3}{5}+1\right)}=\frac{1}{2}.$$

Example 12: There are 3 true coins and 1 false coin with 'head' on both sides. A coin is closer at random and tosses 4 times, If 'head' occurs are the 4 times, What is the probability that the false coin has been chosen and used?

Solution:

P(T) = P(the coin is a true coin) = 3/4

P(F) = P(the coin is a false coin) = 1/4

Let A = Event of getting all heads is 4 tosses,

Then, P(A/T) = ½ * ½ * ½ * ½ = 1/16 and P(A/F) = 1

By Baye's theorem, $P(F/A)=\frac{P(F)\times P(A/F)}{P(F)\times P(A/F)+P(T)\times P(A/T)}=\frac{\frac{1}{4}\times 1}{\frac{1}{4}\times 1+\frac{3}{4}\times\frac{1}{16}}=\frac{16}{19}.$

Example 13:

There are three bags , bag one contains 3 white balls , 2 red balls and 4 black balls. Bag two contains 2 white balls, 3 red balls and 5 black balls. Bag three contains 3 white balls, 4 red balls and 2 black balls. One bag is chosen at random and from it 3 balls were drawn out of which 2 balls were white and 1 is red. What is the probability that it is drawn from bag one, two and three?

Solution:

Selection of bags are mutually exclusive events. The selection of the 2 white and 1 red ball is an independent event.

$P(B_1)=P(B_2)=P(B_3)=1/3$

$P(A/B_1)=$ P(Bag 1 selected from 2W&1R ball chosen)

$= \frac{3C_2 X 2C_1}{9C_3}$

=0.07

$P(A/B_2)=$ P(Bag 2 selected from 2W&1R ball chosen)

$= \frac{2C_2 X 3C_1}{10C_3}$

= 0.025

$P(A/B_3)=$ P(Bag 3 selected from 2W&1R ball chosen)

$= \frac{3C_2 X 4C_1}{9C_3}$

= 0.14

By using Baye's theorem we have

$P(B_i)$	$P(A/B_i)$	$P(B_i)\ P(A/B_i)$
1/3	0.07	0.0233
1/3	0.025	0.0083
1/3	0.14	0.0466
	$\sum P(B_i)\ P(A/B_i)$	0.0782

$P(B_1/A)$ = P(The balls selected from the first bag)

$= \frac{0.0233}{0.0782}$

= 0.29

$P(B_2/A)$ = P(The balls selected from the second bag)

$= \frac{0.008}{0.0782}$

= 0.102

$P(B_3/A)$ = P(The balls selected from the third bag)

$= \frac{0.046}{0.0782}$

= 0.58

Exercise:

1. In a bolt factory machines A,B,C manufactures 25%,35% and 40% of the total respectively. Out of their output 5%,4% and 2% are defective bolts respectively. A bolt is drawn at random and is found to be defective. What are the probabilities that it was manufactured by the machines A,B and C respectively?

2. A bag contains five balls and it is not known how many of them are white. Two balls are drawn at random from the bag and they are found to be white. What is the probability that all the balls in the bag are white?

RANDOM VARIABLES

Definition: A real-valued function defined on the outcome of a probability experiment is called a random variable. A Random variable (RV) is a rule that assigns a numerical value to each possible outcome of an experiment.

1. Discrete Random Variables.
2. Continuous Random Variables

Probability distribution function of X: If X is a random variable, then the function F(x) defined by $F(x) = P\{X \le x\}$ is called the distribution function of X.

1. **Discrete Random Variable:** A random variable whose set of possible values is either finite or countable infinite is called discrete random variable.

Probability Mass Function (pmf): If X is a discrete variable, then the function $p(x) = P[X = x]$ is called the pmf of X. It satisfies two conditions

i) $p(x_i) \geq 0$

ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

Cumulative distribution [discrete R.V] or distribution function of X: The cumulative distribution F(x) of discrete random variable X with probability f(x) is given by

$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ for } -\infty < x < \infty$

Properties of distribution function:

1. $F(-\infty) = 0$
2. $F(\infty) = 1$
3. $0 \leq F(x) \leq 1$
4. $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
5. $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) + P[X = x_1]$
6. $P(x_1 < X < x_2) = F(x_2) - F(x_1) - P[X = x_2]$
7. $P(x_1 \leq X < x_2) = F(x_2) - F(x_1) - P[X = x_2] + P[X = x_1]$

Results:

1. $P(X \leq \infty) = 1$
2. $P(X \leq -\infty) = 0$
3. $P(X > x) = 1 - P[X \leq x]$
4. $P(X \leq x) = 1 - P[X > x]$

Example 14: A R.V X has the following probability distribution.

x:	-2	-1	0	1	2	3
p(x):	0.1	k	0.2	2k	0.3	3k

Find (1) The value of k, (2) Evaluate P(X<2) and P(-2<X<2) .

Solution:

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

0.1+k+0.2+2k+0.3+3k = 1

K = 1/15.

(2) P[X<2] = P[x=-2] + P[x=-1] + P[x=0] + P[x=1]

= 0.1 + 1/15 + 0.2 + 2/15

= ½

P[-2 < X < 2] = P[x=-1]+ P[x=0]+ P[x=1]

= 1/15 + 0.2 + 2/15 = 2/5

Example 15:

A random variable X has the following probability function

Values of x	0	1	2	3	4	5	6	7	8
Probability P(x)	a	3a	5a	7a	9a	11a	13a	15a	17a

i) Determine the value of 'a'.
ii) Find P(X<3), P(X ≥ 3) and P (0 < X< 5).
iii) Find the distribution function of X.

Solution:

i) To find 'a' value:

Given discrete random variable, $\sum_{i=1}^{\infty} p(x_i) = 1$

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

81a=1

a=1/81

ii) To find P(X<3):

P(X<3) = P(X=0)+P(X=1)+P(X=2)

=a+3a+5a

=9a

=1/9

iii) To find $P(X \geq 3)$:

$P(X \geq 3) = 1 - P(X < 3)$

=1-1/9 =8/9

iv) To find $P(0 < X < 5)$:

$P(0 < X < 5) = P(X = 1) + ...P(X = 4)$

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

= 24/81

v) To find the distribution function of X:

Value of x	0	1	2	3	4	5	6	7	8
P(x)	a	3a	5a	7a	9a	11a	13a	15a	17a
P(x)	1/81	3/81	5/81	7/81	9/81	11/81	13/81	15/81	17/81
F(x)	1/81	4/81	9/81	16/81	25/81	36/81	49/81	64/81	1

Example 16: A R.V X has the following function:

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

(a) find k (b) Evaluate P[X<6], P[x≥6], (c) Evaluate P[1.5<X<4.5 / X>2] (d) Find P[X<2], P[X>3], P[1<X<5].

Solution:

(a). Since $\sum_{i=1}^{n} p(x_i) = 1$

i.e., $0+k+2k+2k+3k+k^2+2k^2+7k^2+k = 1$

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

K = -1 or 1/10 (since k=-1 is not permissible, $P(X) \geq 0$)

Hence k = 1/10.

(b). $P[x \geq 6] = P[X=6] + P[X=7]$

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

$= 2/100 + 7/100 + 1/10 = 19/100$

$P[X<6] = 1 - P[x \geq 6]$

$= 1 - 19/100$

$= 81/100$

(c). $P[1.5<X<4.5 / X>2] = \dfrac{p[(1.5<x<4.5) \cap x>2]}{p(x>2)}$ (by conditional probability)

$$= \frac{p[2<x<4.5]}{1-p(x \leq 2)}$$

$$= \frac{p(3)+p(4)}{1-[p(0)+p(1)+p(2)]}$$

$$= \frac{\frac{2}{10}+\frac{3}{10}}{1-\left[0+\frac{1}{10}+\frac{2}{10}\right]} = \frac{\frac{5}{10}}{\frac{7}{10}} = \frac{5}{7}$$

(d). $p(X<2) = p[x=0] + p[x=1]$

$= 0 + k = k = 1/10$

$P(X>3) = 1 - p(x \leq 3)$

$= 1 - [p(x=0)+p(x=1)+p(x=2)+p(x=3)]$

$= 1 - [0+k+2k+2k]$

$= ½$

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

$= 2k + 2k + 3k$

$= 7/10$

Example 17: If the R.V. X takes the values 1,2,3 and 4 such that 2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4). Find the probability distribution and cumulative distribution function of X.

Solution:

Since X is a discrete random variable.

Let 2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4) = k

2P(X = 1) = k implies that P(X = 1) = k/2

3P(X = 2) = k implies that P(X = 2) = k/3

P(X = 3) = k

5P(X = 4) = k implies that P(X = 4) = k/5

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

i.e., k/2 + k/3 + k + k/3 = 1

k[1/2 + 1/3 + 1 + 1/5] = 1

Therefore k = 30/61

x_i	$p(x_i)$	F(X)
1	P(1) = k/2 = 15/61	F(1) = p(1) = 15/61
2	P(2) = k/3 = 10/61	F(2) = F(1) + p(2) = 15/61 + 10/61 = 25/61
3	P(3) = k = 30/61	F(3) = F(2)+p(3) = 25/61 + 30/61 = 55/61
4	P(4) = k/5 = 6/61	F(4) = F(3)+p(4) = 55/61 + 6/61 = 61/61 = 1

Example 18: A discrete random variable X has the following probability mass function:

X	0	1	2	3	4	5	6	7
P(X)	0	a	2a	2a	3a	a^2	$2a^2$	$7a^2+a$

Find (i) the value of 'a' (ii) $P(X<6)$, $P(X \geq 6)$ (iii) $P(0<X<5)$ (iv) the distribution function of X (v) If $P(X \leq x) > 1/2$, find the minimum value of X.

Solution:

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

i.e., $0+a+2a+2a+3a+a^2+2a^2+7a^2+a = 1$

$10a^2 + 9a - 1 = 0$

a = -1 or 1/10 (since a=-1 is not permissible, P(X)≥0)

Hence a = 1/10.

(ii). P[x≥6] = P[X=6] + P[X=7]

$= 2a^2+7a^2+a$

= 2/100 + 7/100 + 1/10 = 19/100

(iii). P[X<6] = 1 - P[x≥6]

= 1 – 19/100

= 81/100

(iv). To find P(0<X<5):

P(0<X<5) = P(X=1)+…..P(X=4)

= a+2a+2a+3a

= 8a = 8/10

(v). To find distribution function of X :

x	0	1	2	3	4	5	6	7
P(x)	0	a	2a	2a	3a	a^2	2 a^2	7 a^2+a
F(x)	0	1/10	3/10	5/10	8/10	81/100	83/100	1

Minimum value of X:

$P(X \le x) > 1/2$

The minimum value of X for which $P(X \le x) > 0.5$, is the x value is 4.

Example 19: A RV X has the following distribution

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

(a) find k (b) Evaluate P(X<2) & P(-2<X<2)

Solution:

(a) $\sum$ P(X)=1

6K+0.6=1

K=1/15

Since the distribution is

X	-2	-1	0	1	2	3
P(X)	1/10	1/15	1/5	2/15	3/10	1/5

(b) P(X<2) = P(X=-2) + P(X=-1) + P(X=0) + P(X=1)

= 1/10 + 1/15 + 1/5 + 2/15 =1/2

& P(-2<X<2) = P(X=-1) + P(X=0) + P(X=1)

= 1/15 + 1/5 + 2/15 = 2/5.

Moments

The moment generating function (MGF) of a random variable X (about origin) whose probability function f(x) is given by

$$M_x(t) = E(e^{tx}) \quad = \sum_{x=\infty}^{\infty} e^{tx} P(x), \; \textit{for a discrete probability distribution}$$

where t is real parameter and the integration or summation being extended to the entire range of x.

Example 20

The probability function of an infinite discrete distribution is given by $P(X = x) = \frac{1}{2^x}, x = 1, 2, ...\infty$. Find the mean and variance of the distribution. Also find P(X is even).

Solution

We know that

$$M_x(t) = \sum_{x=1}^{\infty} e^{tx} p(x)$$

$$= \sum_{x=1}^{\infty} e^{tx} \frac{1}{2^x}$$

$$= \sum_{x=1}^{\infty} \left(\frac{e^t}{2}\right)^x$$

$$= \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + ...$$

$$= \frac{e^t}{2}\left[1 + \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + ...\right]$$

$$= \frac{e^t}{2}\left[1 - \frac{e^t}{2}\right]^{-1} \qquad \text{[Using } (1-x)^{-1} = 1 + x + x^2 + ...\text{]}$$

$$= \frac{e^t}{2}\left[\frac{(2-e^t)^{-1}}{2^{-1}}\right]$$

$$M_x(t) = \frac{e^t}{2-e^t} = (2-e^t)^{-1}e^t$$

$$M_x'(t) = -e^t(2-e^t)^{-2}(-e^t) + (2-e^t)^{-1}e^t$$

$$= e^{2t}(2-e^t)^{-2} + (2-e^t)^{-1}e^t$$

$$M_x''(t) = 2(2-e^t)^{-2}e^{2t} + e^{2t}(-2)(2-e^t)^{-3}(-e^t) + (2-e^t)^{-1}e^t + e^t(-1) + (2-e^t)^{-2}(-e^t)$$

Now E(X) = Mean $= M_x'(0) = 1 + 1 = 2$

$E(X^2) = M_x''(0) = 6$

Mean $\mu_1' = 2$

Variance = $E(X^2) - [E(X)]^2$

= 6-4 =2

Now p(X = even) = p(x = 2) + p(x = 4)+……

$$= \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4 + ...$$

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

$$= \frac{1/4}{1-1/4} = \frac{1}{4} \text{ x } \frac{4}{4-1} = \frac{1}{3}$$

MGF	**Mean**	**Variance**	**p(x=even)**
$e^t(2-e^t)^{-1}$	2	2	$1/3$

UNIT - II

RANDOM VARIABLES

Introduction:

In the last chapter, we introduced the concept of a single random variable. We observed that the various statistical averages or moments of the random variable like mean, variance, standard derivation, skewness give an idea about the characteristics of the random variable.

But in many practical problems several random variables interact with each other and frequently we are interested in the joint behavior of the health conditions of a person, doctors measure many parameters like height, weight, blood pressure, sugar level etc. we should now introduce techniques that help us to determine the joint statistical properties of several random variables.

The concepts like distribution function, density function and moments that we defined for single random variable can be extended to multiple random variables also.

Continuous Random Variables: A random variable X is said to be continuous if it takes all possible values between certain limits say from real number ‘a’ to real number ‘b’.

Example: The length time during which a vacuum tube installed in a circuit functions is a continuous random variable, number of scratches on a surface, proportion of defective parts among 1000 testes, number of transmitted in error.

Probability density function (pdf): For a continuous R.V X, a probability density function is a function such that (1) $f(x) \geq 0$ (2) $\int_{-\infty}^{\infty} f(x)dx = 1$ (3)

$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx =$ area under f(x) from a to b for any a and b.

Cumulative distribution function: The Cumulative distribution function of a continuous R.V. X is

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \ \ for -\infty < x < \infty.$

Mean and variance of the Continuous R.V. X: Suppose X is continuous variable with pdf f(x). The mean or expected value of X, denoted as μ or E(X)

$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$. And the variance of X, denoted as V(X) or σ^2 is $E[X^2] - [E(X)]^2$

Example: 1

A continuous random variable ‘X’ has a probability density function f(x) = K, $0 \leq x \leq 1$. Find ‘K’.

Solution:

Given $f(x) = k, 0 \leq x \leq 1$

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

$$\int_0^\infty k\,dx = 1$$

k=1

Example 2: Given that the pdf of a R.V X is f(x)=kx, 0<x<1. Find k and P(X>0.5)

Solution:

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

$$\int_0^1 kx\,dx = 1$$

$$k\left[\frac{x^2}{2}\right]_0^1 = 1$$

K = 2

$$P(X > 0.5) = \int_{0.5}^{\infty} f(x)\,dx$$

$$= \int_{1/2}^{1} 2x\,dx$$

$$= 2\left[\frac{x^2}{2}\right]_{1/2}^{1}$$

= 3/4

Example 3: If $f(x)=\begin{cases} kxe^{-x}, & x>0 \\ 0, & elsewhere \end{cases}$ is the pdf of a R.V. X. Find k.

Solution:

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

Here $\int_0^\infty kxe^{-x}\,dx = 1$ [since x>0]

$$k\left[x\left(\frac{e^{-x}}{-1}\right) - 1\left(\frac{e^{-x}}{-1}\right)\right]_0^\infty = 1$$

K =1

Example 4: A continuous R.V. X has he density function $f(x) = \frac{k}{1+x^2}, -\infty < x < \infty$. find the value of k and the distribution function.

Solution:

Given is a pdf $\int_{-\infty}^{\infty} f(x)dx=1$, $f(x)=\dfrac{k}{1+x^2}, -\infty<x<\infty.$

$$k\int_{-\infty}^{\infty}\frac{1}{1+x^2}dx=1$$

$$2k\int_{0}^{\infty}\frac{1}{1+x^2}dx=1$$

$$2k\left[\tan^{-1}x\right]_0^{\infty}=1$$

$$2k\left[\frac{\pi}{2}-0\right]=1$$

$$\pi k=1; k=\frac{1}{\pi}$$

$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{x}\frac{1}{\pi}\left(\frac{1}{1+x^2}\right)dx$$

$$=\frac{1}{\pi}\left[\tan^{-1}x\right]_{-\infty}^{x}=\frac{1}{\pi}\left[\tan^{-1}x-\left(\frac{-\pi}{2}\right)\right]$$

$$=\frac{1}{\pi}\left[\tan^{-1}x+\left(\frac{\pi}{2}\right)\right] \text{ for } -\infty<x<\infty$$

Example:5

A continuous random variable X has a pdf $f(x)=3x^2,\ 0\le x\le 1.$ Find a and b such that (i) $P(X\le a)=P(X>a)$ and (ii) $P(X>b)=0.05.$

Solution:

A continuous random variable X has a pdf $f(x)=3x^2,\ 0\le x\le 1.$

i) To find $P(X\le a)=P(X\succ a)$

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

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

Since $P(X\le a)=P(X\succ a)$, $P(X\le a)=\dfrac{1}{2}=0.5$

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

a = 0.7937

ii) To find $P(X\succ b)=0.05$

$$\int_b^1 f(x)dx=0.05,\quad \int_b^1 3x^2dx=1-b^3=0.05$$

$$b^3=0.95$$

$b = (0.95)^{1/3}$

Example 6: If the density function of a continuous R.V. X is given by $f(x) = \begin{cases} ax, & 0 \le x \le 1 \\ a, & 1 \le x \le 2 \\ 3a - ax, & 2 \le x \le 3 \\ 0, & otherwise \end{cases}$

(1) Find the value of a.

(2) The cumulative distribution function of X.

(3) If x_1, x_2, x_3 are 3 independent observations of X. What is the probability that exactly one of these 3 is greater than1.5?

Solution:

(1) Since f(x) is a pdf, then $\int_{-\infty}^{\infty} f(x)dx = 1$

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

i.e., $\int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (3a - ax)dx = 1$

a = ½

(2). (i) If x<0 then F(x) = 0

(ii) If $0 \le x \le 1 \, then \, F(x) = \int_0^x ax\,dx = \int_0^x \frac{x}{2} dx$

$= \frac{x^2}{4}$

$(iii)\ If\ 1 \le x \le 2\ then\ F(x) = \int_{-\infty}^{x} f(x)dx$

$= \int_0^1 ax\,dx + \int_1^x adx$

$= \frac{x}{2} - \frac{1}{4}$

$(iv)\ If\ 2 \le x \le 3\ then\ F(x) = \int_{-\infty}^{x} f(x)dx$

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

$= \frac{3x}{2} - \frac{x^2}{4} - \frac{5}{4}$

(i) If x>3, then $F(x) = \int_{-\infty}^{x} f(x)dx$

$$= \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (3a - ax)\,dx + \int_3^x f(x)\,dx$$

$= 1$

(3). $P(X > 1.5) = \int_{1.5}^{3} f(x)\,dx = \int_{1.5}^{2} \frac{1}{2}\,dx + \int_2^3 \left(\frac{3}{2} - \frac{x}{2}\right)dx$

$= ½$

Choosing an X and observing its value can be considered as a trail and X>1.5 can be considered as a success.

Therefore, p=1/2, q=1/2.

As we choose 3 independent observation of X, n = 3.

By Bernoulli's theorem, P(exactly one value > 1.5) = P(1 success)

$$= {}^3C_1 \times (p)^1 \times (q)^2 = \frac{3}{8}.$$

Example:7

A continuous random variable X is having the probability density function

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the cumulative distribution function of x.

Solution:

Given $f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$

To find cumulative distribution function of x:

i) If $0 < x < 1$ F(x) = $\int_{-\infty}^{x} f(x)dx$

$$= \int_0^x x\,dx = \frac{x^2}{2}$$

ii) If $1 < x < 2$, F(x) = $\int_{-\infty}^{x} f(x)dx$

$$= \int_0^1 x\,dx + \int_1^x (2 - x)dx$$

$$= 2x - \frac{x^2}{2} - 1$$

iii) If $x > 2$, F(x) = $\int_{-\infty}^{x} f(x)dx$

$$= \int_0^1 x\,dx + \int_1^2 (2-x)\,dx$$

$= 1$

The cumulative distribution function of x is $F(x) = \begin{cases} \dfrac{x^2}{2}, & 0 < x < 1 \\ 2x - \dfrac{x^2}{2} - 1, & 1 < x < 2 \\ 1, & x > 2 \end{cases}$

CONTINUOUS RANDOM VARIABLE DISTRIBUTIONS

Normal distribution:

Definition:

A continuous random variable X is said to follow a normal distribution with mean μ and variance σ^2, if its density function is given by the probability law

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,\ \sigma > 0,\ -\infty < \mu < \infty.$$

If X follows normal distribution with mean μ and standard deviation σ, then it is denoted by $N \sim (\mu, \sigma)$ sometimes $N(\mu, \sigma^2)$ can also be used.

Solution:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty} e^{tx}.\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

$$. \qquad = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

put $z = \dfrac{x-\mu}{\sigma}$ If $x = -\infty,\ z = -\infty$

$\sigma dz = dx$ If $x = \infty,\ z = \infty$

$$. \qquad = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)}.e^{\frac{-z^2}{2}} \sigma dz$$

. $= \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(z^2-2t\sigma z)}{2}} dz \qquad \because \frac{-(z^2-2t\sigma z)}{2} = \frac{-1}{2}\left[(z-\sigma t)^2 - \sigma^2 t^2 - \frac{(z-\sigma t)^2}{2} + \frac{\sigma^2 t^2}{2}\right]$

$$= e^{\mu t} . \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1}{2}(z-\sigma t)^2 + \frac{\sigma^2 t^2}{2}} dz$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}} . \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1}{2}(z-\sigma t)^2} dz$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}} . \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-u^2}{2}} du$$

$u = z - \sigma t \qquad z = \infty, \ u = \infty$

$du = dz \qquad z = -\infty, u = -\infty$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}} . \frac{1}{\sqrt{2\pi}} . \sqrt{2\pi} \qquad \left[\int_{0}^{\infty} e^{\frac{-u^2}{2}} du = \sqrt{2\pi}\right]$$

$$\therefore M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} .$$

Example: 8

A normal distribution has mean $\mu = 20$ and S.D $\sigma = 10$. Find $P(15 \le X \le 40)$.

Solution:

Given $\mu = 20$, $\sigma = 10$

The normal variate $Z = \frac{X-\mu}{\sigma} = \frac{X-20}{10}$

When $X = 15$, $Z = \frac{X-20}{10} = \frac{15-20}{10} = -0.5$

$X = 40$, $Z = \frac{40-20}{10} = 2$

$\therefore P(15 \le X \le 40) = P(-0.5 \le Z \le 2)$

$= P(-0.5 \le Z \le 0) + P(0 \le Z \le 2)$

$= P(0 \le Z \le 0.5) + P(0 \le Z \le 2)$

$= 0.1915 + 0.4772 \qquad$ [Using normal table]

$= 0.6687$

Example 9

If X is a normal variate with mean 1 and variance 4. Y is another normal variate independent of X with mean 2 and variance 3. What is the distribution of X+2Y.

Solution:

Given X and Y are independent normal variates.

X+2Y is also a normal variate by additive property.

$\therefore$ Mean of (X+2Y) = E(X+2y)

$= E(X) + E(2Y)$
$= E(X) + 2E(Y)$
$= 1+2\times 2 \quad [E(X)=1, E(Y)=2]$
$= 5$

$Var(X+2Y) = Var(X) + Var(2Y)$
$= 1^2 Var(X) + 2^2 Var(Y)$
$= 1 \times 4 + 4 \times 3 = 16$

$\therefore$ X+2Y follows normal distribution with mean 5 and variance 16.

Gamma Distribution:

The continuous random variable X is said to follow a Gamma distribution with parameter λ if its probability function is given by,

$$f(x) = \begin{cases} \dfrac{e^{-x} x^{\lambda-1}}{\Gamma(\lambda)}, & \lambda > 0, 0 < x < \infty \\ 0, & otherwise \end{cases}$$

Note: 1

A continuous random variable X whose probability density function is

$f(x) = \dfrac{a^{\lambda} e^{-ax} x^{\lambda-1}}{\Gamma(\lambda)}$, $a > 0,\ \lambda > 0,\ 0 < x < \infty$ is called a Gamma distribution with two parameters a and λ.

Note: 2

When $a = 1$

$f(x) = \dfrac{e^{-x} x^{\lambda-1}}{\Gamma(\lambda)}$, which is called the sample Gamma distribution or standard Gamma distribution.

Note: 3

Sometimes the definition of Gamma distribution is given by taking

$$a = \frac{1}{\beta},\quad f(x) = \frac{1}{\beta^{\lambda}} \cdot \frac{e^{\frac{-x}{\beta}} x^{\lambda-1}}{\Gamma(\lambda)}, x \geq 0$$

Find the moment generating function of Gamma distribution:

Solution:

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} f(x)dx = \int_0^{\infty} e^{tx} \cdot \frac{e^{-x} . x^{\lambda-1}}{\Gamma(\lambda)} dx$$

$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{tx}.e^{-x}.x^{\lambda-1}dx = \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{-(1-t)}.x^{\lambda-1}dx$$

put $(1-t)x = u$ *If* $x = 0, \quad u = 0$

$(1-t)dx = du$ *If* $x = \infty, \quad u = \infty$

$$= \frac{1}{\Gamma(\lambda)} \int_0^\infty e^{-u}.\left(\frac{u}{1-t}\right)^{\lambda-1}\left(\frac{du}{1-t}\right) = \frac{1}{\Gamma(\lambda)} \int_0^\infty \frac{u^{\lambda-1}e^{-u}}{(1-t)^\lambda} du$$

$$= \frac{1}{\Gamma(\lambda).(1-t)^\lambda}.\Gamma(\lambda) = \frac{1}{(1-t)^\lambda} \qquad \left[\Gamma(n) \int_0^\infty x^{n-1}e^{-x}dx\right]$$

$$M_X(t) = (1-t)^{-\lambda}, \quad |t| < 1.$$

Find the mean and variance of Gamma distribution:

Solution:

$M_X(t) = (1-t)^{-\lambda}$

$M'_X(t) = -\lambda(1-t)^{-\lambda-1}(-1)$ $\qquad \mu'_1 = M'_X(0) = \lambda$ $\qquad$ 1

$M''_X(t) = \lambda(-\lambda-1)(1-t)^{-\lambda-2}(-1)$ $\qquad \mu'_1 = M''_X(0) = \lambda(\lambda+1)$ $\qquad$ 2

Variance $\mu_2 = \mu'_2 - \mu'^2_1 = \lambda(\lambda+1) - \lambda^2 = \lambda^2 + \lambda - \lambda^2$

$\therefore$ Variance $= \lambda$.

Hence mean and variance of Gamma distribution $= \lambda$

Gamma			
p.d.f	MGF	Mean	Variance
$\frac{e^{-x}x^{\lambda-1}}{\Gamma(\lambda)}, \lambda > 0, \quad 0 < x < \infty$	$(1-t)^{-\lambda}$	λ	λ

Exponential distribution:

A continuous random variable X is said to follow an exponential distribution with parameter $\lambda > 0$ if its probability density function is given by,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & otherwise \end{cases}$$

Find the moment generating function of exponential distribution:

Solution:

$$M_X(t) = \int_0^\infty e^{tx} f(x)dx \qquad \left[Here \quad f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & otherwise \end{cases} \right]$$

$$= \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - t)x} dx$$

$$= \lambda \left[\frac{e^{-(\lambda - t)x}}{-(\lambda - t)} \right]_0^\infty \qquad \left[\because \int e^{ax} dx = \frac{e^{ax}}{a} \right]$$

$$= \frac{\lambda}{-(\lambda - t)} \left[e^{-\infty} - e^{-0} \right]$$

$$= \frac{\lambda}{(\lambda - t)} \qquad \left[\therefore e^{-\infty} = 0, \; e^0 = 1 \right]$$

$\therefore$ The MGF $= \dfrac{\lambda}{\lambda - t}, \; \lambda > t$

Find the mean and variance of exponential distribution:

Solution:

We know that MGF is,

$$M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - \frac{t}{\lambda}}$$

$$= \left(1 - \frac{t}{\lambda}\right)^{-1} = 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \dots\dots\dots\dots + \frac{t^r}{\lambda^r} + \dots\dots$$

$$= 1 + \frac{t}{\lambda} + \frac{t^2}{2!}\left(\frac{2!}{\lambda^2}\right) + \dots\dots\dots\dots + \frac{t^r}{\lambda^r}\left(\frac{r!}{\lambda^2}\right)$$

$$M_X(t) = \sum_{r=0}^{\infty} \left(\frac{t}{\lambda}\right)^r$$

$\therefore$ Mean $\mu_1' = \; coefficient \; of \; \dfrac{t'}{1!} = \dfrac{1}{\lambda}$

$$\mu_2' = \; coefficient \; of \; \frac{t^2}{2!} = \frac{2}{\lambda^2}$$

Now , variance $\mu_2 = \mu_2' - \mu_1'^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Variance $= \frac{1}{\lambda^2} = 1/\lambda^2$

Exponential			
p.d.f	MGF	Mean	Variance
$\lambda e^{-\lambda x}, x > 0$	$\frac{\lambda}{\lambda - t}, \lambda > t$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Memoryless property of the Exponential distribution:

If X is exponentially distributed , then $P\left(X > s+t \middle/ X > s\right) = P(X > t)$ for any $s, t > 0$

Proof:

$$P(X > k) = \int_k^{\infty} \lambda e^{-\lambda x} dx$$

$$= \lambda \left[\frac{e^{-\lambda x}}{\lambda} \right]_k^{\infty} = -e^{-\infty} + e^{-\lambda k} = e^{-\lambda k}$$

Also, $P\left(X > s+t \middle/ X > s\right) = \frac{P(X > s+t \text{ and } X > s)}{P(X > s)}$

$$= \frac{P(X > s+t)}{P(X > s)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}$$

$$= \frac{e^{-\lambda s}.e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

$\therefore P\left(X > s+t \middle/ X > s\right) = P(X > t)$

Thus $P(X > t) = e^{-\lambda t}$.

Example: 10

The time (in hours) required to repair a machine is exponentially distributed with parameter $\lambda = \frac{1}{2}$.

(a) What is the probability that the repair time exceeds $2h$?

(b) What is the conditional probability that a repair takes atleast $11h$ given that its duration exceeds $8h$?

Solution:

Let X be the random variable which represents the time to repair the machine then the density function of X is given by,

$$f(x) = \lambda e^{-\lambda x} = \frac{1}{2} e^{-\frac{1}{2}x}, \quad x > 0$$

(a) $P(X > k) = e^{-\lambda k}$

$$P(X > 2) = e^{-\frac{1}{2}\times 2} = e^{-1}$$

(b) $P\left(X \geq 11 / X > 8\right) = P\left(X \geq 8 + 3 / X > 8\right) = P(X > 3)$ $\left[\because P\left(X > s + t / X > s\right) = P(X > s)\right]$

by memoryless property

$$P(X > t) = e^{-\lambda t} = e^{-\frac{1}{2}\times 3} = e^{-1.5}$$

$\therefore P(X > 3) = e^{-1.5}$.

BIVARITE RANDOM VARIABLES

Definition:

Let S be the sample space. Let X=X(S) and Y=Y(S) be two functions each assigning a real no. to each outcome $s \in S$. Then (X,Y) is a two dimensional random variable.

Types of random variables:

1. Discrete random variables
2. Continuous random variables

Two dimensional discrete random variables:

If the possible values of (X,Y) are finite or countably infinite then (X,Y) is called a two dimensional discrete random variables when (X,Y) is a two dimensional discrete random variable the possible values of (X,Y) may be represented as (x_i, y_j) i = 1,2,…..n, j=1,2,….m.

Two dimensional continuous random variables:

If (X,Y) can assume all values in a specified region R in the XY plane (X,Y) is called a two dimensional continuous random variables.

Joint distributions – Marginal and conditional distributions:

(i) Joint Probability Distribution:

The probabilities of two events $A = \{X \leq x\}$ and $B = \{Y \leq y\}$ have defined as functions of x and y respectively called probability distribution function.

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

$F_Y(y) = P(Y \leq y)$

Discrete random variable important terms:

i) Joint probability function (or) Joint probability mass function:

For two discrete random variables x and y write the probability that X will take the value of x_i, Y will take the value of y_j as, $P(x,y) = P(X = x_i, Y = y_j)$

ie) $P(X = x_i, Y = y_j)$ is the probability of intersection of events $X = x_i \,\&\, Y = y_j$.

$P(X = x_i, Y = y_j) = P(X = x_i \cap Y = y_j)$, The function $P(X = x_i, Y = y_j) = P(x_i, y_j)$ is called a joint probability function for discrete random variables X,Y and it is denoted by P_{ij}.

P_{ij} satisfies the following conditions

(i) $P_{ij} > 0$, for every i,j

(ii) $\sum_j \sum_i P_{ij} = 1$

Continuous random variable (or) Joint Probability Density Function:

Definition:

The joint probability density function if (x,y) be the two dimensional continuous random variable then f(x,y) is called the joint probability density function of (x,y) the following conditions are satisfied.

(i) $f(x,y) \geq 0, \forall x, y \in R$

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$. Where R is a sample space.

Note: $P(a \leq x \leq b, c \leq y \leq d) = \int_a^b \int_c^d f(x,y)dydx$

Joint cumulative distributive function:

If (x,y) is a two dimensional random variable then $F(X,Y) = P(X \leq x, Y \leq y)$ is called a cumulative distributive function of (x,y) the discrete case $F(X,Y) = \sum_j \sum_i P_{ij} = 1$, $y_i \leq y, x_i \leq x$.

In the continuous case $F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(x,y)dxdy$

Properties of Joint Probability Distribution function:

1. $O \leq P(x_i, y_j) \leq 1$
2. $\sum_i \sum_j P(X_i, Y_j) = 1$
3. $P(X_i) = \sum_j P(X_i, Y_j)$
4. $P(y_i) = \sum_j P(X_i, Y_j)$
5. $P(x_i) \geq P(x_i, y_j)$ for any j
6. $P(y_j) \geq P(x_i, y_j)$ for any i

Properties:

1. The joint probability distribution function F xy (X, Y) of two random variable X and Y have the following properties. They are very similar to those of the distribution function of a single random variable.
2. $0 \leq f_{XY}(x,y) \leq 1$

3. $f_{XY}(\infty,\infty)=1$

4. $f_{XY}(x,y)$ is non decreasing

5. $f_{XY}(-\infty,y)=F_{xy}(x_1,\infty)=0$

6. For $x_1<x_2$ and $y_1<y_2$, $P(x_1<X\le x_2, Y\le y_1)=F(x_2,y_1)-F(x_1,y_1)$

7. $P(X\le x_1, y_1<Y\le y_2)=F(x_1,y_2)-F(x_1,y_1)$

8. $P(x_1<X\le x_2, y_1<Y\le y_2)=F(x_2,y_2)-F(x_1,y_2)-f(x_2,y_1)+f(x_1,y_1)$

9. $F_Y(y)=F_{XY}(\infty,y)=P(X\le\infty, y\le y)=P(y\le y)$

10. $F_X(x)+F_y(y)-1\le F_{XY}(x,y)\le\sqrt{F_X(x)F_Y(y)}$ for all x and y.

These properties can also be easily extended to multi dimensional random variables.

Marginal Probability Distribution function:

(i) Discrete case:

- Let (x,y) be a two dimensional discrete random variable, $P_{ij}=P[X=x_i, Y=y_j]$ then $P(X=x_i)=P_i*$ is called a marginal probability of the function X. Then the collection of the pair $\{x_i, P_i^*\}$ is called a marginal probability of X.
- If $P(Y=y_j)=P_{*j}$ is called a marginal probability of the function Y. Then the collection of the pair $\{y_i, P_{*j}\}$ is called a marginal probability of Y.

(ii) Continuous case:

- The marginal density function of X is defined as $f_x(x)=g(x)=\int_{-\infty}^{\infty} f(x,y)dy$ and
- The marginal density function of Y is defined as $f_y(y)=h(y)=\int_{-\infty}^{\infty} f(x,y)dx$

Conditional distributions:

(i) Discrete case:

- The conditional probability function of X given Y=y $_j$ is given by

$$P[X=x_i/Y=Y_j]=P[X=x_i, Y=y_j]/P[Y=y]=P_{ij}/P*j$$

The set $\{X=x_i, P_{ij}/P*j\}$, I =1,2,3,…is called the conditional probability distribution of X given $Y=y_j$

- The conditional probability function of Y given X=xi is given by

$$P=[Y=y_j\ /\ X=x_i]=\ P[Y=y_j, X=x_i]/P[X=x_i]\ =P_{ij}/P_i^*$$

The set { $y_{ij}\ \ P/P_i^*$ }, j=1,2,3,…is called the conditional probability distribution of Y given $X=x_i$

(ii) Continuous case:

- The conditional probability density function of X is given by $Y=y_j$ is defined as

$f(x/y)=\dfrac{f(x,y)}{h(y)}$, where h(y) is a marginal probability density function of Y.

- The conditional probability density function of Y is given by $X = x_i$ is defined as

$f(y/x)=\dfrac{f(x,y)}{g(x)}$, where g(x) is a marginal probability density function of X.

Independent random variables:

(i) Discrete case:

Two random variable (x,y) are said to be independent if $P(X = x_i \cap Y = y_j) = P(X = x_i)(Y = y_j)$ (ie) $P_{ij} = P_i^{*}P_{*j}$ for all i ,j.

(ii) Continuous case:

Two random variables (x,y) are said to be independent if $f(x,y) = g(x)h(y)$, where $f(x,y) =$ joint probability density function of x and y,

$g(x) =$ Marginal density function of x,

$h(y) =$ Marginal density function of y.

Marginal Distribution Tables:

Table – I

To calculate marginal distribution when the random variables X takes horizontal values and Y takes vertical values

Y/X	x1	x2	x3	p (y) = p(Y=y)
y1	p11	p21	p31	p(Y=y1)
y2	p12	p22	p32	p(Y=y2)
y3	p13	p23	p33	p(Y=y3)
$P_X(X) = P(x = x)$	$P(x = x1)$	$p(x = x2)$	$p(x = x3)$	

Table – II

To calculate marginal distribution when the random variables X takes vertical values and Y takes horizontal values

Y\X	y1	y2	y3	P_X (x) = P(X=x)
x1	p11	p21	p31	p(X=x1)
x2	p12	p22	p32	p(X=x2)
x3	p13	p23	p33	p(X=x3)
$p(y) = p(y = y)$	$P(y = y1)$	$P(y = y2)$	$P(y = y3)$	

Solved Problems on Marginal Distribution:

Example :11

From the following joint distribution of X and Y find the marginal distribution

X/Y	0	1	2
0	3/28	9/28	3/28
1	3/14	3/14	0
2	1/28	0	0

Solution:

The marginal distribution are given in the table below

Y\X	0	1	2	$P_Y(y)$ = P(Y=y)
0	3/28	9/28	3/28	15/28

1	3/14	3/ 14	0	6/14
2	1/28	0	0	1/28
$P_X(x) = P(Y = y)$	$P_X(0) =$ 5/14	$P_X(1) =$ 15/28	$P_X(2) =$ 3/28	1

The marginal Distribution of X

$P_X(0) = P(X = 0) = p(0,0) + p(0,1) + p(0,2)$ = 3/28 + 3/14 + 1/28 = 5/14

$P_X(1) = P(X = 1) = p(1,0) + p(1,1) + p(1,2)$ = 9/28 + 3/14 + 0 = 15/28

$P_X(2) = P(X = 2) = p(2,0) + p(2,1) + p(2,2)$ = 3/28 + 0 + 0 = 3/28

Marginal probability function of X is $P_x(x) = \begin{cases} 5/14, & x = 0 \\ 15/28, & x = 1 \\ 3/28, & x = 2 \end{cases}$

The marginal distributions are

Y/X	1	2	3	$P_Y(y) = p(y = y)$
1	2/21	3/21	4/21	9/21
2	3/21	4/21	5/21	12/21
$P_x(x) = P(x = x)$	5/21	7/21	9/21	1

The marginal distribution of X

$P_x(1) = p(1,1) + (1,2)$ = 2/21 + 3/21

$P_x(1) = 5/21$

$P_x(2) = p(2,1) + (2,2)$ = 3/21 + 4/21

$P_x(2) = 7/21$

$P_x(3) = p(3,1) + p(3,2)$ = 4/21 + 5/21

$P_x(3) = 9/21$

Marginal probability function of X is, $P_x(x) = \begin{cases} 5/21, & x = 1 \\ 7/21, & x = 2 \\ 9/21, & x = 3 \end{cases}$

The marginal distribution of Y

$P_Y(1) = p(1,1) + p(2,1) + p(3,1)$ = 2/21 + 3/21 +4/21

$P_Y(1) = 9/21$

$P_Y(2) = p(1,2) + p(2,2) + p(3,2)$ = 3/21 + 4/21 +5/21

$P_Y(2) = 12/21$

Marginal probability function of Y is $P_Y(y) = \begin{cases} 3/21, & y = 1 \\ 4/21, & y = 2 \end{cases}$

Example :12

From the following table for joint distribution of (X, Y) find

i) $P(X \le 1)$ ii) $P(Y \le 3)$ iii) $P(X \le 1, Y \le 3)$ iv) $P(X \le 1 / Y \le 3)$

v) $P(Y \le 3 / X \le 1)$ vi) $P(X + Y \le 4)$.

X/Y	0	2	3	4	5	6
0	0	0	1/32	2/32	2/32	3/32

1	1/16	1/16	1/8	1/8	1/8	1/8
2	1/32	1/32	1/64	1/64	0	2/64

Solution:

The marginal distributions are

X / Y	1	2	3	4	5	6	$P_X(x) = P(X = x)$
0	0	0	1/32	2/32	2/32	3/32	8/32 P(x=0)
1	1/16	1/16	1/8	1/8	1/8	1/8	10/16 P(x=1)
2	1/32	1/32	1/64	1/64	0	2/64	8/64 P(x=2)
$P_Y(y) = P(Y = y)$	3/32	3/32	11/64	13/64	6/32	16/64	1
	P(Y=1)	P(Y=2)	P(Y=3)	P(Y=4)	P(Y=5)	P(Y=6)	

i) $P(X \le 1)$

$P(X \le 1) = P(X = 0) + P(X = 1)$

= 8/32 + 10/16

$P(X \le 1)$ = 28/32

ii) $P(Y \le 3)$

$P(Y \le 3) = P(Y = 1) + P(Y = 2) + P(Y = 3)$

= 3/32 + 3/32 + 11/64

$P(Y \le 3)$ = 23/64

iii) $P(X \le 1, Y \le 3)$

$P(X \le 1, Y \le 3)$ = P(0,1) + P(0,2) + P(0,3) + P(1,1) + P(1,2) + P(1,3)

= 0+0+1/32 + 1/16 + 1/16 + 1/8

$P(X \le 1, Y \le 3)$ = 9/32

iv) $P(X \le 1 / Y \le 3)$

By using definition of conditional probability

$$P[x = x_i / y = y_j] = \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

The marginal distribution of Y

$P_Y(0) = P(Y = 0) = p(0,0) + p(1,0) + p(2,0)$ = 3/28+9/28+3/28 = 15/28

$P_y(1) = P(y = 1) = p(0,1) + p(1,1) + p(2,1)$ = 3/14+3/14+0= 3/7

$P_y(2) = P(y = 2) = p(0,2) + p(1,2) + p(2,2)$ = 1/28+0+0 = 1/2

Marginal probability function of Y is $P_y(Y) = \begin{cases} 15/28, & y = 0 \\ 3/7, & y = 1 \\ 1/28, & y = 2 \end{cases}$

Example 13:

The joint distribution of X and Y is given by f(X, Y) =X+Y/21, x=1,2,3 y=1,2.Find the marginal distributions.

Solution:

Given f (X, Y) =X+Y/21, x=1, 2, 3 y=1,2

f (1,1) = 1+1/21 =2/21 =P(1,1)
f (1,2) = 1+2/21 =3/21 =P(1,2)
f (2,1) = 2+1/21 =3/21 =P(2,1)
f (2,2) = 2+2/21 =4/21 =P(2,2)
f (3,1) = 3+1/21 =4/21 =P(3,1)
f (3,2) = 3+2/21 =5/21 =P(3,2)

$$P[X \le 1/Y \le 3] = \frac{P[X \le 1, Y \le 3]}{P[Y \le 3]} = \frac{9/23}{23/64}$$

$P[X \le 1/Y \le 3] =$ 18/32

v) $P[Y \le 3/X \le 1]$

$$P[Y \le 3/X \le 1] = \frac{P[X \le 3, Y \le 1]}{P[Y \le 1]} = \frac{9/23}{7/8}$$

P[Y≤3/ X≤1]= 9/28

vi) $P(X+Y \le 4)$

P(X + Y≤4)) = P(0,1)+P(0,2)+P(0,3)+ P(0,4)+ P(1,1)+ P(1,2)+P(1,3)+P(2,1)+ P(2,2)
= 0+0+1/32 +2/32 + 1/16 + 1/16 + 1/8+1/32 +1/32
P(X + Y≤4) = 13/32

Example : 14

If the joint P.D.F of (X,Y) is given by p(X,Y)=K(2x+3y),x=0,1,2, y=1,2,3,. Find all the marginal probability distribution .Also find the probability of (X+Y) and P(X+Y >3).

Solution:

Given P(X,Y)= K(2x+3y)
P(0,1)= K(0+3) = 3K
P(0,2)= K(0+6) = 6K
P(0,3)= K(0+9) = 9K
P(1,1)= K(2+3) = 5K
P(1,2)= K(2+6) = 8K
P(1,3)= K(2+9) = 11K
P(2,1)= K(4+3) = 7K
P(2,2)= K(4+6) = 10K
P(2,3)= K(4+9) = 13K

To find K:

The marginal distribution is given in the table.

Y\X	0	1	2	$P_Y(y) = P(Y=y)$
1	3K	5K	7K	15K
2	6K	8K	10K	24K
3	9K	11K	13K	33K
PX(x)=P(X=x)	18K	24K	30K	72K

Total Probability =1
72K = 1
K = 1/72

Marginal probability of X & Y:

Substituting K = 1/72 in the above table, we get

Y\X	0	1	2	P_Y (y)=P(Y=y)
1	3/72	5/72	7/72	5/24
2	6/72	8/72	10/72	1/3
3	9/72	11/72	13/72	11/24
P_X (x)=P(X=x)	1/4	11/72	5/12	1

From table, $P_x(0) = 1/4$, $p_x(1) = 1/3$, $p_x(2) = 5/12$

Marginal probability function of x is , $P_x(X) = \begin{cases} 1/4, x = 0 \\ 1/3, x = 1 \\ 5/2, x = 2 \end{cases}$

From table, $p_y(1) = 5/24$, $P_y(2) = 1/3$, $P_Y(3) = 11/24$

Marginal Probability function of Y is , $P_Y(Y) = \begin{cases} 5/24, Y = 1 \\ 11/24, y = 2 \\ \end{cases}$

Example :15

From the following table for joint distribution of (X, Y) find
The marginal distributions are

Y/X	1	2	3	$P_Y(y) = P(Y = y)$
1	2/21	3/21	4/21	9/21
2	3/21	4/21	5/21	12/21
$P_X(x) = P(X = x)$	5/21	7/21	9/21	1

The marginal distribution of X

P $_X$ (1) = P(1,1)+P(1,2) = 2/21 + 3/21= P x (3)=9/21

P $_X$ (2) = P(2,1)+P(2,2) = 3/21 + 4/21= Py(2)=7/21

P $_X$ (3) = P(3,1)+P(3,2) = 4/21 + 5/21= P $_X$ (3)=9/21

Marginal probability function of X is P $_x$ (x) = $\begin{cases} 5/21, x = 1 \\ 7/21, x = 2 \\ 9/21, x = 3 \end{cases}$

The marginal distribution of Y

P $_y$ (1) = P (1, 1) + P (2, 1) +P (3, 1)

= 2/21 + 3/21 +4/21= 9/21

$P_y(2)$ = P (1, 2) + P (2, 2) +P (3, 2)

= 3/21 + 4/21 +5/21= 12/21

Marginal probability function of Y is $P_Y(y) = \begin{cases} 3/21, y = 1 \\ 4/21, y = 2 \end{cases}$

Exercises:

1. Given is the joint distribution of X and Y

Y/X	0	1	2

0	0.02	0.08	0.10
1	0.05	0.20	0.25
2	0.03	0.12	0.15

Obtain 1) Marginal Distribution.
2) The conditional distribution of X given Y =0.

2. The joint probability mass function of X & Y is

X/Y	0	1	2
0	0.10	0.04	0.02
1	0.08	0.20	0.06
2	0.06	0.14	0.30

Find the M.D.F of X and Y. Also $(X \leq 1, Y \leq 1)$ and check if X & Y are independent.

3. Let X and Y have the following joint probability distribution

Y/X	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

Show that X and Y are independent.

4. The joint probability distribution of X and Y is given by the following table.

X/Y	1	3	9
2	1/8	1/24	1/12
4	¼	¼	0
6	1/8	1/24	1/12.

i) Find the probability distribution of Y.
ii) Find the conditional distribution of Y given X=2.
ii) Are X and Y are independent.

5. Given the following distribution of X and Y. Find
i) Marginal distribution of X and Y.
ii) The conditional distribution of X given Y=2.

X/Y	-1	0	1
0	1/15	2/15	1/15
1	3/15	2/15	1/15
2	2/15	1/15	2/15

Example : 16

If the joint probability density function of (X, Y) is given by $f(x,y)=2, \quad 0 \le x \le y \le 1.$ Find marginal density function of X.

Solution:

Given $f(x,y)=2,\ 0 \le x \le y \le 1$

To find marginal density function of x:

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy = \int_{x}^{1} 2dy = 2[1-x],\ 0 \le x \le y.$$

Example:17

If the joint probability density function of X and Y is given by

$$f(x,y)=\begin{cases} \frac{1}{8}(6-x-y), & 0<x<2,\ 2<y<4 \\ 0\ , & \text{otherwise} \end{cases}$$

Find (i) $P(X<1 \cap Y<3)$ (ii) $P\left(X<1 / Y<3\right)$ (iii) $f\left(\frac{y}{x}\right)$.

Solution:

Given $f(x,y)=\begin{cases} \frac{1}{8}(6-x-y), & 0<x<2,\ 2<y<4 \\ 0\ , & \text{otherwise} \end{cases}$

i) To find $P(X<1 \cap Y<3)$:

$$P(X<1 \cap Y<3) = \int_0^1\int_2^3 f(x,y)dydx$$

$$= \frac{1}{8}\int_0^1\int_2^3 (6-x-y)dydx$$

$$= \frac{3}{8}$$

ii) To find $P\left(X<1 / Y<3\right)$

$$P\left(X<1 / Y<3\right) = \frac{P(X<1 \cap Y<3)}{P(Y<3)} \qquad \text{..................(1)}$$

To find $P(Y<3)$:

$$P(Y<3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx$$

$$= \int_0^2\int_2^3 \frac{1}{8}(6-x-y)dydx$$

$$= \frac{5}{8}$$

Equation (1) becomes $P\left(X<1 / Y<3\right) = \frac{3}{5}$

iii) To find $f(y/x)$:

We know that $f(y/x) = \frac{f(x,y)}{f_x(x)}$

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{8}\int_{2}^{4}(6-x-y)dy$$

$$= \tfrac{1}{4}(3-x), 0 < x < 2.$$

$$f(y/x) = \frac{\frac{1}{8}(6-x-y)}{\frac{1}{4}(3-x)} = \frac{6-x-y}{2(3-x)}, \quad 0 < x < 2, \quad 2 < y < 4.$$

Example : 18

If the joint distribution of X and Y is given by

$F(x,y) = (1-e^{-x})(1-e^{-y})$, for $x > 0, y > 0$

$= \ 0$, otherwise

(i) Find the marginal densities of X and Y (ii) Are X and Y independent?

(iii) $P(1 < X < 3, 1 < Y < 2)$

Solution:

Given $F(x,y) = (1-e^{-x})(1-e^{-y})$

$= 1 - e^{-x} - e^{-y} + e^{-(x+y)}$

The joint pdf is given by $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}(1 - e^{-x} - e^{-y} + e^{-(x+y)})$$

$= e^{-(x+y)}$

$f(x,y) = e^{-(x+y)}, x \ge 0, y \ge 0$

i) The marginal density function of X is $f(x) = \int_{-\infty}^{\infty} f(x,y)dy$

$$f(x) = \int_{0}^{\infty} e^{-(x+y)}dy = e^{-x}, x \ge 0$$

The marginal density function of Y is $f(y) = \int_{-\infty}^{\infty} f(x,y)dx$

$$f(y) = \int_{0}^{\infty} e^{-(x+y)}dx = e^{-y}, y \ge 0$$

ii) Consider $f(x).f(y) = e^{-x}e^{-y} = e^{-(x+y)} = f(x,y)$

ie) X and Y are independent.

iii) $P(1 < X < 3, 1 < Y < 2) = P(1 < X < 3).P(1 < Y < 2)$

$$= \int_{1}^{3} f(x)dx.\int_{1}^{2} f(y)dy = \int_{1}^{3} e^{-x}dx \int_{1}^{2} e^{-y}dy$$

$$= \frac{(1-e^2)(1-e)}{e^5}$$

Exercises:

1. The joint p.d.f. of the two dimensional random variable is,

$$f(x,y)=\begin{cases}\frac{8xy}{9}, & 1<x<y<2\\ 0, & \text{otherwise}\end{cases}$$

(i) Find the marginal density functions of X and Y.
(ii) Find the conditional density function of Y given X=x.

2. If the joint Probability density function of two dimensional R.V (X,Y) is given by

$$f(x,y)=\begin{cases}x^2+\frac{xy}{3}, & 0\le x\le 1, 0\le y\le 2\\ 0\ , & \text{otherwise}\end{cases}.$$

Show that X and Y are not independent.

UNIT - III

BASIC STATISTICS

Discrete distributions:-

The important discrete distributions of a random variable 'X' are

1. Binomial distribution
2. Poisson distribution

Binomial distribution:-

Let us consider ' $'n'$ independent trails. If the successes (S) and failures (F) are recorded successively as the trails are repeated we get a result of the type

$$SSFFS...............FS$$

Let x be the number of success and hence we have(n-x) number of failures.

$$P(SSFFS...............FS) = P(S)P(S)P(F)P(F)P(S).......P(F)P(S)$$

$$= ppqqp...........q.p$$

$$= \underset{x-\text{factors}}{ppp...........p} \times \underset{(n-x)\text{factors}}{qqq...........q}$$

$$= p^x.q^{n-x}$$

But x successes in n trails can occur in nC_x ways

$\therefore$ The probability of x successes in n trails is given by $nC_x p^x.q^{n-x}$

$$P(X = x \quad \text{successes}) = nC_x p^x q^{n-x}$$

$$P(X = x) = p(x) = nC_x p^x q^{n-x}, \; x = 0,1,2,.........n, p + q = 1$$

Note:1

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

Here $nC_x p^x.q^{n-x}$ is the $(x+1)^{th}$ term in the expansion of $(q+p)^n$

[$\because (q+p)^n = q^n + nC_1 q^{n-1} p + ... + nC_x q^{n-x} p^x + ...$ Which is a Binomial series and hence the distribution is called binomial distribution]

Find the moment generating function (MGF) of a Binomial distribution about origin

Solution:

We know that the moment generating function of a random variable X about origin whose probability function $p(x)$ is given by

$$M_X(t) = \sum_{x=0}^{n} e^{tx} p(x) \qquad [\, p(x) \text{ is a pmf}]$$

Let X be a random variable which follows binomial distribution.

Then its MGF about origin is given by,

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} p(x)$$

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

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

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

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

$$= q^n + nC_1 q^{n-1}(pe^t)^1 + nC_2 q^{n-2}(pe^t)^2 ....$$

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

$\therefore$ MGF of binomial distribution is,

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

Find the mean and variance of binomial distribution.

Solution:

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

$\therefore M'_X(t) = n(q+pe^t)^{n-1}.pe^t$

Put $t=0$ we get

$M'_X(0) = n(q+p)^{n-1}.p$

Mean= $E(X) = np$ $[\because q+p=1]$ (Mean = $\therefore M'_X(0)$)

$M''_X(t) = np[(q+pe^t)^{n-1}.e^t + e^t(n-1)(q+pe^t)^{n-2}.pe^t]$

Put $t=0$ we get,

$M''_X(0) = np[(q+p)^{n-1} + (n-1)(q+p)^{n-2}.p]$

$= np[(1+(n-1)p] \quad = np + n^2p^2 - np^2$ $[q+p=1]$

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

$M''_X(0) = n^2p^2 + npq$ $[\because 1-p=q]$

$M''_X(0) = E(X^2) = n^2p^2 + npq$

$Var(X) = E(X^2) - [E(X)]^2$

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

$Var(X) = npq$

Binomial			
MGF	Mean	Variance	S.D
$(q+pe^t)$	np	npq	$\sqrt{npq}$

Example:1

The mean and S.D of a binomial distribution are 5 and 2. Determine the distribution.

Solution:

Given mean = $np = 5$ 1

S.D = $\sqrt{npq} = 2$

$npq = 4$ 2

$$\frac{2}{1} \Rightarrow \frac{npq}{np} = \frac{4}{5}$$

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

$$\therefore p = 1 - q = 1 - \frac{4}{5} = \frac{5-4}{5} = \frac{1}{5}$$

$p = \frac{1}{5}$ 3

Substituting 3 in 1, we get

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

$$n = 25$$

$\therefore$ The binomial distribution is,

$$P(X = x) = p(x) = nC_x p^x q^{n-x} = 25C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{25-x}, \; x = 0,1,.........,25$$

Example:2

The mean and variance of a binomial variate are 8 and 6, find $p(X \geq 2)$.

Solution:

Given mean = $np = 8$ 1

Variance = $npq = 6$ 2

$$\frac{2}{1} \Rightarrow \frac{npq}{np} = \frac{6}{8}$$

$$q = \frac{3}{4}$$

$$\therefore p = 1 - q = 1 - \frac{3}{4} = \frac{4-3}{4} = \frac{1}{4}$$

$p = \frac{1}{4}$ 3

Substituting 3 in 1, we get

$$n \times \frac{1}{4} = 8$$

$$n = 32$$

$\therefore$ The binomial distribution is,

$$P(X = x) = p(x) = nC_x p^x q^{n-x} \quad = 32C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{32-x}$$

Now $P(X \geq 2) = 1 - P(X < 2)$

$$= 1 - [P(X = 0) + P(X = 1)]$$

$$= 1 - [P(0) + P(1)]$$

$$= 1 - \left[32C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^{32-0} + 32C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{32-1}\right]$$

$$= 1 - \left[\left(\frac{3}{4}\right)^{32} + 32\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{31}\right]$$

$$= 1 - \left(\frac{3}{4}\right)^{31}\left[\frac{3}{4} + \frac{32}{4}\right]$$

$$= 1 - \frac{35}{4} \times \left(\frac{3}{4}\right)^{31} = 0.9988.$$

Example: 3

6 dice are thrown 729 times. How many times do you expect atleast three dice to show a five or six?

Solution:

Let X be the random variable denoting number of successes when 6 dice are thrown.

p = probability of getting 5 or 6 with one die

$$= \frac{2}{6} = \frac{1}{3}$$

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

$p = \frac{1}{3},\ q = \frac{2}{3},\ n = 6$

$$\therefore P(X = x) = 6C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}$$

$P(X=x) = 6cx\ (1/3)^x\ (2/3)^{6-x}$

P (at least three dice showing five or six) $= p(X \geq 3)$

$$= P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$$

$$= 6C_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + 6C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + 6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right) + 6C_6\left(\frac{1}{3}\right)^6$$

$$= 160 \times \frac{1}{3^6} + 60 \times \frac{1}{3^6} + 12 \times \frac{1}{3^6} + 1 \times \frac{1}{3^6}$$

$$= \frac{1}{3^6}\left[160 + 60 + 12 + 1\right] = \frac{233}{3^6}.$$

For 729 times, the expected number of times atleast 3 dice showing five or six

$$= N \times \frac{233}{3^6} = 729 \times \frac{233}{3^6} = 233 \text{ times}.$$

Example: 4

4 coins were tossed simultaneously. What is the probability of getting (i) 2 heads

(ii) atleast 2 heads (iii) atmost 2 heads

Solution:

Let X be the random variable denoting the number of heads obtained when 4 coin were tossed.

Here we are dealing with the coin problem.

$\therefore p = \frac{1}{2}, q = \frac{1}{2}.$ Also given $n = 4$

p (getting x head in throwing 4 coins) $= p(X = x)$

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

(i) p (getting 2 heads) $= p(X = 2)$

$$= 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$$

$$= \frac{4 \times 3}{1 \times 2} \bullet \frac{1}{16} = \frac{3}{8}$$

(ii) p (getting atleast 2 heads) $= p(X \geq 2)$

$$= p(X = 2) + p(X = 3) + p(X = 4)$$

$$= 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 + 4C_3 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + 4C_4 \left(\frac{1}{2}\right)^4$$

$$= \left(\frac{1}{2}\right)^4 \left[4C_2 + 4C_3 + 4C_4\right] = \frac{1}{16}\left[6 + 4 + 1\right] = \frac{11}{16}.$$

(iii) p (getting atmost 2 heads) $= p(X \leq 2)$

$$= p(X = 0) + p(X = 1) + p(X = 2)$$

$$= 4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 + 4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 + 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$$

$$= \left(\frac{1}{2}\right)^4 \left[4C_0 + 4C_1 + 4C_2\right] = \left(\frac{1}{2}\right)^4 \left[6 + 4 + 1\right] = \frac{11}{16}.$$

Poisson distribution:

Poisson distribution is a limiting case of binomial distribution under the following assumptions:

(i) The number of trails n should be indefinitely large *i.e.,* $n \to \infty$

(ii) The probability of successes $'p'$ for each trail is indefinitely small.

(iii) $np = \lambda,$ should be finite where λ is a constant.

Find the moment generating function of the Poisson distribution

Solution:

We know that the MGF of a random variable X is given by,

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) \qquad \left[p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \right]$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\lambda} (\lambda e^t)^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda} \left\{ 1 + \lambda e^t + \frac{(\lambda e^t)^2}{2!} + \ldots\ldots\ldots \right\} \qquad \left[\because e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots\ldots \right]$$

$$= e^{-\lambda} e^{\lambda e^t} \quad = e^{\lambda(e^t - 1)}$$

Hence $M_X(t) = e^{\lambda(e^t - 1)}$.

Moment generating function of a poisson random variable X is $M_X(t) = e^{\lambda(e^t - 1)}$.

The probability mass function of a random variable X which follows poisson distribution is given by

$$P(X = x) = p(x) = \begin{cases} \dfrac{e^{-\lambda} \lambda^x}{x!}, & x = 0,1,2,\ldots\ldots\infty \\ 0, & otherwise \end{cases}$$

Find the mean and variance of the poisson distribution

Solution:

We know that, for discrete probability distribution mean is given by,

$$\mu_1' = E(X) = \sum_{x=0}^{\infty} x.p(x) = \sum_{x=0}^{\infty} x.\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda\lambda^{x-1}}{x!}$$

$$= 0 + e^{-\lambda}.\lambda\sum_{x=1}^{\infty} \frac{x\lambda^{x-1}}{x!}$$

$$= e^{-\lambda}.\lambda\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \qquad \left[\frac{x}{x!} = \frac{1}{(x-1)!}\right]$$

$$= e^{-\lambda}.\lambda\left[1 + \lambda + \frac{\lambda^2}{2!} + ...\right] = \lambda e^{-\lambda}.e^{\lambda} = \lambda$$

Hence the mean of the poisson distribution is λ

Now $\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2.p(x) = \sum_{x=0}^{\infty} x^2.\frac{e^{-\lambda}\lambda^x}{x!}$

$$= \sum_{x=0}^{\infty}\{x(x-1) + x\}\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^{x-2}\lambda^2}{x(x-1)(x-2).......1} + \lambda$$

$$= e^{-\lambda}\lambda^2 \sum_{x=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)(x-3).......1} + \lambda \qquad \left[\because \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda^x}{x!} = \lambda\right]$$

$$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \qquad \left[using \ \ \frac{1}{n!} = 0 \ when \ \ n \ is \ negative\right]$$

$$= e^{-\lambda}\lambda^2 \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + ...\right] + \lambda = e^{-\lambda}e^{-\lambda}\lambda^2 = \lambda^2 + \lambda$$

Variance $\mu_2 = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Variance $= \lambda$

Hence mean = variance $= \lambda$

Poisson			
MGF	Mean	Variance	S.D
$e^{\lambda(e^t - 1)}$	λ	λ	$\sqrt{\lambda}$

Example:5

If X and Y are independent poisson variate such that $P(X = 1) = P(X = 2)$ and $P(Y = 2) = P(Y = 3)$ find the variance of $X - 2Y$.

Solution:

we know that $P(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

Given,

$P(X=1) = P(X=2)$

$$e^{-\lambda}\lambda = \frac{e^{-\lambda}\lambda^2}{2!} \qquad 1$$

Also given,

$P(Y=2) = P(Y=3)$

$$\frac{e^{-\mu}\mu^2}{2!} = \frac{e^{-\mu}\mu^3}{3!} \qquad 2$$

From 1 we get, $2e^{-\lambda}\lambda = e^{-\lambda}\lambda^2$

$$\lambda^2 - 2\lambda = 0$$

$$\lambda(\lambda - 2) = 0$$

Since $\lambda > 0$, $\lambda - 2 = 0 \Rightarrow \lambda = 2$

From 2 we get, $6e^{-\mu}\mu^2 = 2e^{-\mu}\mu^3$

$$3\mu^2 = \mu^3 \Rightarrow \mu^2(\mu - 3) = 0$$

Since $\mu > 0$, $\mu - 3 = 0 \Rightarrow \mu = 3$

$\text{var}(X) = \lambda = 2$ 3

$\text{var}(Y) = \mu = 3$ 4

$\therefore \text{var}(X - 2Y) = 1^2\ \text{var}(X) + (-2)^2\ \text{var}(Y)$ 5

$\left[\because \text{var}(a_1X_1 + a_2X_2) = a_1^2\ \text{var}(X_1) + a_2^2\ \text{var}(X_2)\right] = 2 + 4 \times 3 = 14 \quad [\text{using } 3 \text{ and } 4 \text{ in } 5]$

Covariance

It is useful to measure of the relationship between two random variables is called covariance. To define the covariance we need to describe the expected value of a function of two random variables C(x,y).

Covariance:

If X and Y are random variables, than covariance between X and Y is defined as

$$Cov(X,Y) = E\{[X - E(x)][Y - E(y)]\}$$

$$= E\{XY - XE(Y) - E(x)y + E(X)E(Y)\}$$

$$= E(XY) - E(X) - E(Y) - E(X)E(Y) + E(X)E(Y)$$

Covariance (X, Y) = $E(XY) - E(X)E(Y)$ ……………. (A)

If X and y are independent, then $E(XY) = E(X)E(Y)$ ……………… (B)

Substituting (B) in (A), we get Covariance (x, y) = 0

If X and Y are independent, then $Cov(X_, Y) = 0$

Correlation:

If the change in are variable affects a change in the other variable, the variable are said to be correlated In a invariable distribution we may be interested to find out if there is any correlation or co-variance between the two variables under study.

Types of correlation:

1) Positive correlation
2) Negative Correlation

Positive Correlation:

If the two variables deviate in the same direction i.e. If the increase (or decrease) in one results in a corresponding increase (or decrease) in the other, correlation is said to be direct or positive.

Example: The Correlation between

a) The height, and weight of a group of person and
b) Income and expenditure

Negative Correlation:

If the two variable constancy deviate in opposite directions i.e. if (increase 9or decrease) in one result in corresponding decrease (or increase) in the other correlation , is said to be negative.

Example: The Correlation between

a) Price and demand of a commodity and
b) The correlation between volume and pressure of a perfect gas.

Measurement of Correlation:

We can measure the correlation between the two variables by using Karl-Pearson's co -efficient of correction.

Karl-Pearson's Co-Efficient of Correlation:

Correlation co-efficient between two random variable X and Y usually denotes by (X,Y) is a numerical measure of linear.

Karl Pearson's co-efficient of correlation between x & y is

r = $1 - 6 \sum_{i=1}^{n} d_i^{\,2} / n(n^2 - 1)$, , where $d_i = x_i - y_i$

Relationship between them and detained as

$$r(X,Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y} \quad \text{Where COV}\,(X,Y) = \frac{1}{n}\sum XY - \overline{XY}$$

$$\sigma_X = \sqrt{\frac{1}{n}\sum X^2 - \overline{X_2}},\ \overline{X} = \frac{\sum X}{n}$$

$$\sigma_Y = \sqrt{\frac{1}{n}\sum Y^2 - \overline{Y_2}}$$ (n is the number of items in the given data)

Note:

1. Correlation coefficient may also be denoted by r(x,y)

2. If r(x,y) = 0, we say that x & y are uncorrelated.

3. When r = 1 , the correlation is perfect.

Example :6

Calculate the Correlation co-efficient for the following heights (in inches) of father x and their sons y.

Solution:

Method : 1

X	**Y**	**XY**	$\mathbf{X^2}$	$\mathbf{Y^2}$
67 66 67 67 68 69 70 72	67 68 65 68 72 72 69 71	4355 4488 4355 4556 4896 4968 4836 5112	4225 4356 4489 4489 4624 4761 4900 5184	4489 4624 4225 4624 5184 5184 4761 5041
∑(x) =544	∑ (y) = 552	∑XY = 37560	$\sum x^2 = 37028$	$\sum y^2 = 38132$

Now

$\overline{X} = 544/8$ =68

$\overline{Y} = 544/8$ =69

$\overline{X}\ \overline{Y} = 68*69 = 4692$

$\sigma_X = \sqrt{1/n \sum x^2 - \overline{x^2}}$

$= \sqrt{37028/8 - 4624}$ =2.121

$= \sqrt{38132/8 - 4761}$ =2.345

r(X,Y)= $\frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

=1/n $\sum xy - \overline{x\ y} / \sigma_{x} . \sigma_y$

=1/8*37560-4692/2.121*2.345

= 3/4.973

=0.6032

It is positive correlation.

Example:7 Find the co-efficient of Correlation between industrial productions and expose using the following data

Production (x)	55	56	58	59	60	60	62
Export (y)	35	38	37	39	44	43	44

Solution :

X	Y	U =X-58	V =Y-40	UV	U2	V2
55	35	-3	-5	15	9	25
56	38	-2	-2	4	4	4
58	37	0	-3	0	0	9
59	39	1	-1	-1	1	1
60	44	2	4	8	4	16
60	43	2	3	6	4	9
62	44	4	4	16	16	16
		$\sum U=4$	$\sum U=0$	$\sum UV=48$	$\sum U^2 = 38$	$\sum V^2 = 8$

Now $\overline{U} = \sum U/n = 4/7$ =0.5714

$\overline{V} = \sum V/n = 0$……… (1)

$\sigma_U = \sqrt{\sum U^2/n - \overline{U}^2} = \sqrt{38/7 - (0.5714)^2}$ =2.2588……………….. (2)

$\sigma_V = \sqrt{\sum V^2 - \overline{V} = \sqrt{80/7 - 0} =}$ 3.38………… (3)

$\therefore r = (X,Y) = r(U,V) = COV(U,V)/\sigma_U * \sigma_V = 6.857/2.258*3.38 =$ 0.898[using (1), (2) & (3)]

r = 0.79

The value between 0 to 1. So it is positive correlation.

Example :8

Find the Correlation co-efficient for the following data.

X	10	14	18	22	26	30
Y	18	12	24	6	30	36

Solution:

X	**Y**	**U =X-22/4**	**V =Y-24/6**	**UV**	**U^2**	**V^2**
10	18	-3	-1	3	9	1
14	12	-2	-2	4	4	4
18	24	-1	0	0	1	0
22	6	0	-3	0	0	9
26	30	1	1	1	1	1
30	36	2	2	4	4	4
		$\sum U=-3$	$\sum V=-3$	$\sum UV=12$	$\sum U^2 = 19$	$\sum V^2 = 19$

Now $\overline{U} = \sum U/n = -3/6 = -0.5$ ………………(1)

$\overline{V} = \sum V/n = -3/6 = 0.5$ …………………(2)

$$COV(U,V) = \frac{\sum UV}{n\overline{UV}} = 1.75$$

$$\sigma_U = \sqrt{\sum U^2 - \overline{U}^2}$$

$= \sqrt{19/6 - (0.5)^2} = 1.708$ ……… (3)

$\therefore r(x,y) = 0.6$

The value between 0 to 1. So it is positive correlation

Rank Correlation:

Let us suppose that a group of n individuals are arranged in order of merit or proficiently in possession of two characteristics A & B.

$$r = 1 - 6\sum_{i=1}^{n} d_i^2 / n(n^2-1), \quad \text{, where } d_i = x_i - y_i$$

Note:

This formula is called a Spearman's formula.

Solved Problems on Rank Correlation:

Example :9

Find the rank correlation co-efficient from the following data:

Rank in X	1	2	3	4	5	6	7
Rank in Y	4	3	1	2	6	5	7

Solution

X	Y	$d = x_i - y_i$	di^2
1	4	-3	9
2	3	-1	1
3	1	2	4
4	2	2	4
5	6	-1	1
6	5	1	1
7	7	0	0
		$\sum d_i = 0$	$\sum d_i^2 = 20$

Rank Correlation co-efficient

$$r = 1 - 6 \sum_{i=1}^{n} d_i^2 / n(n^2 - 1), \quad \text{where } d_i = x_i - y_i$$

$= 1 – 6 \times 20/ 7(49-1) = 0.6429$

Example : 10 The ranks of some 16 students in mathematics & physics are as follows. Calculate rank correlation co-efficient for proficiency in mathematics & physics.

Rank in Mathematics	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Rank in Physics	1	10	3	4	5	7	2	9	8	11	15	9	14	12	16	13

Solution:

Rank in Mathematics(X)	Rank in Physics(Y)	$d_i = X_i - Y_i$	d_i^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	9	-1	1
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9
		$\sum d_i = 0$	$\sum d_i^2 = 136$

Rank correlation co-efficient

$$r = 1 - 6 \sum_{i=1}^{n} d_i^2 / n(n^2 - 1), \qquad \text{where } d_i = x_i - y_i$$

r = 0.8

Example : 11

10 competitors in a musical test were ranked by the 3 judges X, Y, Z in the following order

	A	B	C	D	E	F	G	H	I	J
Rank in X	1	6	5	10	3	2	4	9	7	8
Y	3	5	8	4	7	10	2	1	6	9
Z	6	4	9	8	1	2	3	10	5	7

Using Rank correlation method, discuss which panel of Judges has the nearest approach to common likings of music.

X	Y	Z	$D_1 = x_i - y_i$	$D_2 = y_i - z_i$	$D_3 = x_i - z_i$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	-5	4	9	25
6	5	4	1	1	2	1	1	4
5	8	9	-3	-1	-4	9	1	16
10	4	8	6	-4	2	36	16	4
3	7	1	-4	6	2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	-1	2	1	1	4	1
						$\sum d_1^2 = 200$	$\sum d_2^2 = 214$	$\sum d_3^2 = 60$

The rank correlation between X & Y is

$$r_1 = 1 - 6 \sum_{i=1}^{n} d_i^2 / n(n^2 - 1) = -0.212$$

The rank correlation between Y& Z is

$r_2 = 1 - 6 \sum_{i=1}^{n} d_i^2 / n(n^2 - 1) = -0.296$

The rank correlation between X & Z is

$r_3 = 1 - 6 \sum_{i=1}^{n} d_i^2 / n(n^2 - 1) = 0.636$

Since the rank correlation between X & Z is maximum and also positive, We conclude that the pair of Judges X & Z has the nearest approach to common likings of music.

Exercises:

1) Calculate the Karl Pearson's co-efficient of correlation from the following data

X	25	26	27	30	32	35
Y	20	22	24	25	26	27

2) Find the co-efficient of correlation of the advertisement cost & sales from the following data

Cost:	39	65	62	90	82	75	98	36	78
Sales:	47	53	58	86	62	68	91	51	84

REGRESSION

Definition:

Regression is a mathematical measure of the average relationship between two or more variables in terms of the original limits of the data.

Lines of regression:

If the variables in a bivariate distribution are related we will cluster around some curve called of regression. If the curve is a straight line, it is and called the line of regression and there is said to be linear regression is said to be curve linear.

The line of regression of y on x is given by $y - \overline{y} = r.\frac{\partial y}{\partial x}(x - \overline{x})$

where r is the correlation coefficient, ∂_y and ∂_X are standard deviation.

The line of regression of X on Y is given by $x - \overline{x} = r.\frac{\partial y}{\partial x}(y - \overline{y})$

Angle between two line of regression:

If the equation of lines of regression of Y on X and X on Y are

$y-\overline{y}=r.\frac{\partial y}{\partial x}(x-\overline{x})$ and $x-\overline{x}=r.\frac{\partial x}{\partial y}(y-\overline{y})$

The angle 'θ' between the two line of regression is given by

$$\tan\theta=\frac{l-r^2}{r}\left(\frac{\partial y\partial x}{\partial x^2+\partial y^2}\right)$$

Regression coefficients:

Regression coefficient of Y on X , $r\frac{\partial Y}{\partial X}=b_{YX}$ …………. (1)

Regression coefficient of X on Y , $r\frac{\partial X}{\partial Y}=b_{XY}$ ………….. (2)

From (1) and (2) we get

$$r\frac{\partial Y}{\partial X}r\frac{\partial X}{\partial Y}=b_{YX}*b_{YX}$$

Correlation coefficient $r=\pm\sqrt{b_{XY}*b_{YX}}$

The regression coefficients b_{YX} and b_{yx} can be easily obtained by using the following formula.

$$b_{YX}=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$$

$$b_{XY}=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(y-\overline{y})^2}$$

Solved Problems on Regression:

Example:12

The equations of two regression lines are 3x+12y=19, 3y+9x=46. Obtain the mean value of X and Y.

Solution:

Given the lines are 3x+12y=19,

3y+9x=46

Since both are passing through $(\overline{x},\overline{y})$, we get

$3\bar{x}+12\bar{y}=19$ …………(1)

$9\bar{x}+3\bar{y}=46$ …………..(2)

Solving equation (1) & (2) we get $33\bar{y}=11$

$\bar{y}=\frac{11}{33}=0.33$, $\bar{y}$ value sub in equation (1) we get $\bar{x}=5$

$(\bar{x},\bar{y})$ = (5, 0.33)

Example:13

From the following data, find

i)The two regression equations.

ii) The co-efficient of correlation between the marks in economics and statistics.

iii)The most likely marks in statistics when marks in economics are 30.

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution:

X	Y	$X-\bar{X}=X-32$	$Y-\bar{Y}=Y-38$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24

29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
$\sum X =$ 320	$\sum Y =$ 380	$\sum(X-\overline{X})=0$	$\sum(Y-\overline{Y})=0$	$\sum(X-\overline{X})^2 =$ 140	$\sum(Y-\overline{Y})^2 =$ 398	$\sum(X-\overline{X})(Y-\overline{Y}) =$ -93

Here $\overline{X} = \dfrac{\sum X}{n}$ and $\overline{Y} = \dfrac{\sum y}{n}$

$= \dfrac{320}{10} = 32$ $\quad = \dfrac{380}{10} = 38$

Coefficient of regression of Y on X is

$$b_{YX} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (X-\overline{X})^2} = \frac{-93}{140}$$

Coefficient of regression of X on Y is

$$b_{XY} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (Y-\overline{Y})} = \frac{-93}{398} = -0.2337$$

Equation of the line of regression of X on Y is

$$x-\overline{x} = b_{XY}(y-\overline{y})$$

X – 32 =0.2337(Y-38)

X= -0.2337 y+0.2337 *38 +32

X=-0.23374 +40.8806

Equation of the line of regression of Y on X is

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

Y-38 = -0.6643(x-32)

Y = - 0.6643 x+38+0.6643*32 =-0.6642x+59.2576

Now we have to find the most likely marks in statistics (Y) when marks in economics (X) are 30.we use the line of regression of Y on X.

Y = -0.6643x+59.2575

Put x=30, we get

Y = -0.6643*30+59.2536 =39.3286 =39

Example :14

Height of father and sons are given in centimeters

X:Height of father	150	152	155	157	160	161	164	166
Y:Height of son	154	156	158	159	160	162	161	164

Find the two lines of regression and calculate the expected average height of the son when the height of the father is 154 cm.

Solution:

Let 160 and 159 be assured means of x and y.

x	y	U=X-160	V=Y-159	u^2	v^2	uv
150	154	-10	-5	100	25	50
152	156	-8	-3	64	9	24
155	158	-5	-1	25	1	5
157	159	-3	0	9	0	0
160	160	0	1	0	1	0
161	162	1	3	1	9	3
164	161	4	2	16	4	8
166	164	6	5	36	25	30
		$\sum U = -15$	$\sum V = 2$	$\sum U^2 = -15$	$\sum V^2 = 74$	$\sum UV = 120$

Now $\overline{X} = 158.13$ and $\overline{Y} = 159.25$

Since regression coefficient are independent of change and of origin we have regression coefficient of Y on X

Coefficient of regression of Y on X is

$$b_{YX} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (X-\overline{X})^2} = \frac{990}{1783} = 0.555$$

Coefficient of regression of X on Y is

$$b_{XY} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (Y-\overline{Y})} = 1.68.$$

Exercise:

1. The two lines of regression are $8x - 10y = 66$ *and* $40x - 18y - 214 = 0$.The variance of X is 9. Find i) the mean values of X and Y ii) Correlation between X and Y.

UNIT - IV

APPLIED STATISTICS

Introduction:

Many problems in engineering require that we decide whether to accept or reject a statement about some parameter. The statement is called a hypothesis and the decision making procedure about the hypothesis is called hypothesis testing.

Population:

A population in statistics means a set of objects or mainly the set of numbers which are measurements or observations pertaining to the objects.

Sampling:

A part selected from the population is called a sample. The process of selection of a sample is called sampling.

Sampling Distribution:

The sample mean, the sample median and the sample standard deviation are examples of random variables whose values will vary from sample to sample. Their distributions, which reflect such chance variations, play an important role in statistics and they are referred to as sampling distributions.

If we draw a sample of size n from a given finite population of size N then the total number of possible samples is NC_n

$$NC_n = \frac{N!}{n!(N-n)!} = k$$

For each of these K samples we can compute some statistics say $t = t(x_1, x_2, x_3 ...., x_n)$ in particular the mean $\bar{x}$, variance (s^2) etc. The set of the values of the statistic so obtained, one for each sample constitutes the sampling distribution of the statistic.

Standard Error:

The standard deviation of sampling distribution of a statistic is known as standard error and it is denoted by (S.E)

Testing a hypothesis:

On the basis of sample information, we make certain decisions about the population . In taking such decisions we make certain assumptions. These assumptions are known as statistical hypothesis are tested.

Assuming the hypothesis correct we calculate the probability of getting the observed sample. If this probability is less than a certain assigned value, the hypothesis is to be rejected.

Null hypothesis:

Null hypothesis is based for analyzing the problem . Null hypothesis is the hypothesis of np difference. It is denoted by H_0. It is defined as a definite statement about the population parameter.

Alternative hypothesis:

Any hypothesis which is complementary to the null hypothesis is called alternative hypothesis . It is denoted by H_1.

Rule:

1.If we want to test the significance of the difference between a statistic and the parameter or between two sample statistics then we setup the null hypothesis that the difference is not significant.

2.If we want to test any statement about the population , we set up the null hypothesis that it is true.

Types of errors:

Type 1 Error: Reject H_0 when it is true.

Type 2 error: Accept H_0 when it s wrong.

P (Reject H_0 when it is true) = P (Type I error) = α

P (Accept H_0 when it is wrong) = P (Type II error) = β

The sizes of the type I and type II errors are also known as producer's risk and consumer's risk respectively.

α = P (Rejecting a good lot)

β = P (Accepting a bad lot)

Level of significance:

The probability that the value of static lies in the critical region is called as level of significance.

Test of significance for single mean (Normal):

Large samples:

If the size of the sample n>30, then that sample is called large sample. There are 4 important test to test the significance of large samples.

1. Test of significance for single proportion
2. Test of significance for test of difference of proportions

3. Test of significance for single mean
4. Test of significance for difference of means.

Test of significance for single mean:

Suppose we want to test whether the given sample of size n has been drawn form a population with mean μ. We set up a null hypothesis that there is no difference between $\bar{x}$ and μ where $\bar{x}$ is the sample mean.

The test statistic is $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ where σ is the standard deviation of the population. If the population standard deviation is not known use the sample standard deviation $z = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

Note:

The values $\bar{x} \pm 1.96\sigma/\sqrt{n}$ is called 95% fiducial limits or confidence limits and similarly $\bar{x} \pm 2.58\sigma/\sqrt{n}$ is called 99% confidence limits.

Example :1

A sample of 900 members has a mean 3.4 cm. and SD 2.61 cms. Is the same from a large population of means 3.25 cms and SD 2.61 cms. If the population is normal and its mean is unknown find the 95% and 98% fiducial limits of the true mean.

Solution:

Given n=900, $\mu = 3.25cms$, $\sigma = 2.61$, $\bar{x} = 3.4cms$

1. The parameter of interest is μ

2. H_0: The sample has been drawn from the population with mean $\mu = 3.2cms$ and standard deviation $\sigma = 2.61cms$

3. $H_1 : \mu \neq 3.25$

4. $\alpha = 0.05$

5. Test Statistic : $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\because$ n is large

6. Reject H_0 $|z| > 1.96$ $\therefore z = 1.73$

Since $|z| < 1.96$, we accept H_0 at 5% level of significance. Therefore the sample has been drawn from the large population with mean $\mu = 3.25 cms$

95% fiducial limits for the population mean μ are $\bar{x} \pm 1.96\sigma/\sqrt{n} = 3.5705 \,\&\, 0.1705$

98% fiducial limits for the population mean μ are $\bar{x} \pm 2.33\sigma/\sqrt{n} = 3.40 \pm 2.33(2.61/\sqrt{900})$

Exercise:

The average marks in Mathematics of a sample of 100 students was 51 with a S.D of 6 marks. Could this have been a random sample from a population with average marks 50.

Test of significance for difference of mean:

Let $\overline{x_1}$ be the mean of a sample of size n_1 from a population with mean μ_1 and variance $\sigma_1^{\ 2}$. Let $\overline{x_2}$ be the mean of a sample of size n_2 from a population with mean μ_2 and variance $\sigma_2^{\ 2}$.

$H_0 : \mu_1 = \mu_2$

Test Statistic : $z = \dfrac{\bar{x}_1 - \overline{x_2}}{\sqrt{\dfrac{\sigma_1^{\ 2}}{n_1} + \dfrac{\sigma_2^{\ 2}}{n_2}}} \sim N(0,1)$

Note :

If $\sigma_1^{\ 2} = \sigma_2^{\ 2} = \sigma^2$, then under $H_0 : \mu_1 = \mu_2$,

$$z = \frac{\bar{x}_1 - \overline{x_2}}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim N(0,1)$$

Example :2

The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches.

Solution:

Given $n_1 = 1000$ $n_2 = 2000$ $\bar{x}_1 = 67.5$ $\bar{x}_2 = 68$

$H_0 : \mu_1 = \mu_2$ and $\sigma = 2.5$ inches.

$H_1 : \mu_1 \neq \mu_2$

Test Statistic : $z = \frac{\overline{x}_1 - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

= -5.1

$\because |z| > 3$ H_0 is rejected.

Example:3

In a survey of buying habits 400 women shoppers are chosen at random in supermarket 'A' located in a certain section of the city. Their average weekly food expenditure is Rs.250 with a standards deviation of Rs.40. For 400 women shoppers chosen at random in supermarket 'B', the average weekly food expenditure is Rs.220 with a SD of Rs.55. Test at 1% level of significance whether the average weekly food expenditures of the two populations of shoppers are equal.

Sol: Given $n_1 = 400$, $n_2 = 400$, $\overline{x}_1 = 250$, $\overline{x}_2 = 220$, $\sigma_1 = 40$, $\sigma_2 = 55$

1. The parameter of interest is μ_1 and μ_2

2. $H_0 : \mu_1 = \mu_2$

3. $H_1 : \mu_1 \neq \mu_2$

4. $\alpha = 0.01$

5. Test Statistic : $z = \frac{\overline{x}_1 - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

= $\frac{30}{\sqrt{11.5625}}$ = $\frac{30}{3.4}$ =8.8235

$\because |z| = 8.8235 > 2.58$. H_0 is rejected.

Exercise:

A simple sample of heights of 6400 Englishmen has a mass of 6785 inches and a S.D of 2.56 inches, while a simple sample of heights of 1600 Australians has a mean of 68.55 inches and a S.D of 2.52 inches. Do the data indicate that Australians are on the average taller than Englishmen.

Test of significance for single proportion:

If X is the number of success in n independent trials with constant probability P of successes for each trial,

$E(x) = nP \& V(x) = nPQ$ where $Q = 1 - P$ is the probability of failure. It has been proved that for large n the Binomial distribution tends to normal distribution. Hence for large n $X \sim N(nP, nPQ)$ i.e.,

$$z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - nP}{X - nPQ} \sim N(0,1)$$

and we can apply the normal test.

Example:4

In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

Solution:

Given n =1000, X=540

p = sample proportion of rice eaters $= \frac{540}{1000} = 0.54 = x$

P = Population proportion of rice eaters = ½ = 0.5

Q = 0.5

1. The parameter of interest is P

2. $H_0 : P = 0.5$, Both rice and wheat eater are equally popular in the state

3. $H_1 : P \neq 0.5$

4. $\alpha = 0.01$

5. Test Statistic: $z = \frac{x - P}{\sqrt{PQ/n}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.532$

6. Conclusion: Since $z_{0.01} = 2.58$ $|z| < z_{0.01}$ $\therefore H_0$ is accepted at 1% level of significance. We conclude that rice and wheat eaters are equally popular in Karnataka.

Example:5

In a study designed to investigate whether certain detonators used with explosives in a coal mining meet the requirement that at least 90% will ignite the explosives when charged it is found that 174 of f 200 detonators function properly. Test the null hypothesis P = 0.90 against the alternative hypothesis P< 0.90 at the 0.05 level of significance

Solution:

$H_0 : P = 0.90$

$H_1 : P < 0.90$

X = 174 n = 200 P = 0.90 Q = 0.10

$$z = \frac{X - nP}{\sqrt{nPQ}} = \frac{174 - 200(0.90)}{\sqrt{200 \times 0.90 \times 0.10}} = -1.41$$

Tabulated value of Z at 5% level of significance for right tail test is 1.645

$|z| < 1.645 \qquad \therefore H_0$ is accepted.

Exercise:

Twenty people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease is 85% in favour of the hypothesis that its more at 5% level.

Test of significance for difference of proportions:

Suppose we want to compare two distinct population with respect to the preval of a certain attribute say A, among their members. Let X1, X2 be number of persons possessing the given attribute A in random samples of sizes n_1 and n_2 from the two population respectively. Then the sample proportions are given by

$$P_1 = \frac{X_1}{n_1} \;\&\; P_2 = \frac{X_2}{n_2}$$

If $P_1 \& P_2$ are populations, then $E(P_1) = P_1 \& E(P_2) = P_2$

$$V(P_1) = \frac{P_1 Q_1}{n_1} \;\&\; V(P_2) = \frac{P_2 Q_2}{n_2}$$

Under $H_0 : P_1 = P_2$ the test statistic for the difference of proportions is

$$z = \frac{\overline{P_1} - \overline{P_2}}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

Example:6

Random samples of 400 men and 600 women were asked whether they would like to have a fly over near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportion of men and women in favour of the proposal are same against that they are not at 5% level.

Solution:

Given $n_1 = 400$, X_1 = No of men favouring the proposal = 200

$n_2 = 600$, $X_2 = 325$

$P_1 = \frac{X_1}{n_1} = \frac{200}{400} = 0.5$ Similarly, $P_2 = \frac{X_2}{n_2} = \frac{325}{600} = 0.54$

1. The parameter of interest is P_1 & P_2 the difference

2. $H_0 : P_1 = P_2 = P$

3. $H_1 : P_1 \neq P_2$

4. $\alpha = 0.05$

5. The test statistic is $\therefore z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{200 + 325}{400 + 600} = 0.525$$

$$\hat{Q} = 1 - \hat{P} = 1 - 0.525 = 0.475$$

$$\therefore z = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.475 \times \left(\frac{1}{400} + \frac{1}{600}\right)}} = -1.24$$

$|z| = 1.24$ 6. Reject H_0 if $|z| > 1.96$

7. Conclusion: $|z| < 1.96 \therefore H_0$ is accepted.

Exercise:

Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons found to be tea drinkers. After an increase in duty 800 people were tea drinkers in sample of 1200 people. Using standard error of proportion state whether there is a significant decrease in the consumption in tea after the increase in excise duty.

UNIT - V

SMALL SAMPLES

Test of significance of small samples:

When the size of the sample (n) is less than 30, then that sample is called a small sample. The following are some important test for small samples.

(i) Student's 't' test

(ii) F- Test

Test for single mean (Student's 't' test):

Suppose we want to test

(a) If a random sample x_i of size n has been drawn from a normal population with a specified mean μ_0

(b) If the sample mean differs significantly from the hypothetical value μ_0 of the population mean.

In this case, the statistic is given by,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ follows students t- distribution with (n-1) degrees of freedom. We now compare the calculated value of t with tabulated value at a certain level of significance. If calculated $|t|$ > tabulated t, null hypothesis is rejected and if calculated $|t|$ < tabulated t, null hypothesis may be accepted.

Assumption for students t – test:

1. The parent population from which the sample is drawn is normal.
2. The sample observations are independent

The population standard deviation is unknown.

Example : 1

A machinist is making engine parts with axle diameters of 0.700 inch. A random sample of ten parts shows a mean diameter of 0.742 inch with a standard deviation of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specifications

Solution:

Given, $\mu = 0.700$ inches

$\bar{x} = 0.742$ inches

$s = 0.040$ inches and n = 10.

$H_0 : \mu = 0.700$ i.e. the product is conforming to specifications.

$H_1 : \mu \neq 0.700$

Under H_0, the test statistic is

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2}/\sqrt{n}} = \frac{\bar{x} - \mu}{\sqrt{s^2}/\sqrt{n-1}} \sim t_{(n-1)}$$

$$t = \frac{0.742 - 0.700}{\sqrt{(0.040)^2} / \sqrt{10-1}} = 3.15$$

t follows student t distribution.

The table value of t at 5% level of significance for 9 degrees of freedom is $t_{0.05} = 1.833$

$\because$ Calculated t > tabulated t, H_0 rejected.

Example :2

A random sample of 10 boys had the following IQs. 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean IQ of 100? Find a reasonable range in which most of the mean IQ values of samples of 10 boys lie.

Solution

H_0: The data support the assumption of a population mean IQ of 100 in the population

$H_0 : \mu = 100$

$H_1 : \mu \neq 100$

Here n = 10, $\bar{x} = \frac{972}{10} = 97.2$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 203.73$$

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{197.2 - 1001}{14.27/\sqrt{10}} = 0.62$$

Tabulated value of t_0 for (10-1) degrees of freedom is 2.26

$\because t < t_0$ H_0 may be accepted and we conclude that the data are consistent.

The 95% confidence limits are given by $\overline{x} \pm t_{0.05} \frac{s}{\sqrt{n}} = 97.2 \pm 2.26(4.514)$

$= 107.40$ and 87.00

Exercise

1. A sample of 26 bulbs gives a mean life of 990 hours with a SD of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not up to the standard?
2. The following data gives the length of 12 samples of Egyptian cotton taken from a large consignment 48,46,49,46,52,45,43,47,47,46,47,50. Test if the mean length of the consignment be taken as 46.

Students 't' test for difference of Means: (Corrected t test or Paired t test)

To test the significant difference between two means $\overline{x_1}$ and $\overline{x_2}$ of samples of sizes n_1 and n_2 the statistic is

$$t = \frac{(\overline{x} - \overline{y}) - (\mu_x - \mu_y)}{\sqrt{s^2 \frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\overline{x} = \frac{1}{n_1} \sum_{i=1}^{n} x_i , \quad \overline{y} = \frac{1}{n_2} \sum_{j=1}^{n} y_j ,$$

$$s^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{j=1}^{n} (y_i - \overline{y})^2 \right]$$

Degrees of freedom = $n_1 + n_2 - 2$

Example: 3

Below are the gain in weight in lbs of pigs fed on the 2 diets A and B.

Gain in weight

Diet A: 25,63,30,34,24,14,32,24,30,31,35,25

Diet B: 44,34,22,10,47,31,40,30,32,35,80,21,35,29,22

Test if the two diets differ significantly as regards to their effect on increasing the weight.

Solution:

$H_0 : \mu_x = \mu_y$ i.e. There is no significant difference between the mean increase in the weights due to diet A and B.

$H_1 : \mu_x \neq \mu_y$

Diet A $\sum x = 336$ $\sum (x - \bar{x})^2 = 380$ $n_1 = 12$ $n_2 = 15$

Diet B $\sum y = 450$ $\sum (y - \bar{y})^2 = 1410$ $\bar{x} = 28$ $\bar{y} = 30$ $S^2 = 171.6$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

$$= \frac{30 - 28}{\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}} = 0.609$$

t_0 at 5% level of significance for (12+15-2)25 degrees of freedom is 2.06

$t < t_0$ $\therefore$ H_0 may be accepted and we may conclude that the two diets do not differ significantly.

Example : 4

The student of 6 randomly chosen sailors are in inches: 63,65,68,69,71,72

These of 10 randomly chosen sailors are (in inches): 61,62,65,66,69,69,70,71,72,73

Discuss the height that these data throw on the suggestions that the sailors are on the average taller than soldiers.

Solution:

Given $n_1 = 6$, $n_2 = 10$

$\bar{x} = 68$ $\bar{y} = 67.8$

<table>
<tr><td>x</td><td>$(x-\overline{x})$</td><td>$(x-\overline{x})^2$</td><td>y</td><td>$(y-\overline{y})$</td><td>$(y-\overline{y})^2$</td></tr>
<tr><td>63</td><td>-5</td><td>25</td><td>61</td><td>-6.8</td><td>46.24</td></tr>
<tr><td>65</td><td>-3</td><td>9</td><td>62</td><td>-5.8</td><td>33.64</td></tr>
<tr><td>68</td><td>0</td><td>0</td><td>65</td><td>-2.8</td><td>7.84</td></tr>
<tr><td>69</td><td>1</td><td>1</td><td>66</td><td>-1.8</td><td>3.24</td></tr>
<tr><td>71</td><td>3</td><td>9</td><td>69</td><td>1.2</td><td>1.44</td></tr>
<tr><td>72</td><td>4</td><td>16</td><td>69</td><td>1.2</td><td>1.44</td></tr>
<tr><td colspan="2" rowspan="5">Total</td><td rowspan="5">60</td><td>70</td><td>2.2</td><td>4.84</td></tr>
<tr><td>71</td><td>3.2</td><td>10.24</td></tr>
<tr><td>72</td><td>4.2</td><td>17.64</td></tr>
<tr><td>73</td><td>5.2</td><td>27.04</td></tr>
<tr><td colspan="2">Total</td><td>153.6</td></tr>
</table>

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 + \sum_{j=1}^{n}\left(y_i - \overline{y}\right)^2\right]$$

$= 15.2571.$

$H_0 : \mu_1 = \mu_2$

Test statistic is

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{68 - 67.8}{\sqrt{15.271\left(\frac{1}{6} + \frac{1}{10}\right)}} = 0.099$$

Reject H_0 if |t| <1.76, We accept H_0 at 5% level of significance.

$\therefore$ t = 0.099

Exercise:

1. Average number of articles produced by two machines per day is 200 and 250 with its standards deviations 20 and 25 respectively on the basis of records of 25 days production. Can you regard both the machine equally efficient at 1% level of significance?

2. Two horses A and B were tested according to the time (in seconds) to run a particular race with following results:

Horse A: 28 30 32 33 33 29 34

Horse B: 29 30 30 24 27 29

F -distribution (Test for ratio of variance) – Snedecor's F – distribution:

To test whether if there is any significant difference between two estimates of population variance. To test if the two samples have come from the same population, we use F test.

$H_0 : \sigma_1^2 = \sigma_2^2$ (i.e.) Population variations are same.

Under H_0 the test statistic is $F = \frac{S_x^2}{S_y^2}$,

Where $S_x^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(x_i - \bar{x}\right)^2$ & $S_y^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(y_j - \bar{y}\right)^2$ and it follows F distribution with (v_1, v_2) degrees of freedom where $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$.

Note:

1. We will take greater of the variance S_1^2 or S_2^2 in the numerator and adjust for the degrees of freedom accordingly

$$F = \frac{Greater\ \mathrm{var}iance}{Smaller\ \mathrm{var}iance}$$

2. If sample variance s^2 is given we can obtain population variance S^2 by using the relation $ns^2 = (n-1)S^2$

Example :5

In one sample of 8 observations the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in another sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.

Solution:

Given $n_1 = 8$ $n_2 = 10$

$\sum\left(x - \bar{x}\right)^2 = 84.4$ $\sum\left(y - \bar{y}\right)^2 = 102.6$

$$S_x^{\ 2} = \frac{1}{n_1 - 1}\sum\left(x - \bar{x}\right)^2 = 12.057$$

$$S_y^{\ 2} = \frac{1}{n_2 - 1}\sum\left(y - \bar{y}\right)^2 = 11.4$$

Steps:

1. The parameter of interest is $\sigma_x^{\ 2}\ \&\ \sigma_y^{\ 2}$

2. $H_0 : \sigma_x^{\ 2} = \sigma_y^{\ 2} = \sigma^2$

3. $H_1 : \sigma_x^{\ 2} =\neq \sigma_y^{\ 2}$

4. $\alpha = 0.05$, d.f (V_1) = n_1-1=7

d.f (V_2) = n_2-1=9

5. $F = \frac{S_x^{\ 2}}{S_y^{\ 2}} = 1.057$

6. Reject H_0 if F>3.29 (from table F)

7. Computations: $F = \frac{12.057}{11.42} = 1.057$

8. Conclusion: Since Tabulated $F_{0.05}$ for (7,9) degrees of freedom is 3.29 $F < F_0$

$\therefore$ H_0 is accepted.

Example :6

Two random samples gave the following results.

Sample	Size	Sample mean	Sum of the squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population.

Solution:

To test if two independent samples have been drawn from the same normal population, we have to test

1. The equality of population means and
2. The equality of population variances

$H_0 : \mu_1 = \mu_2 \; \& H_0 : \sigma_1^{\;2} = \sigma_2^{\;2}$

$n_1 = 10 \qquad n_2 = 12 \qquad \overline{x_1} = 15 \qquad \overline{x_2} = 14$

$\sum \left(x_1 - \overline{x_1}\right)^2 = 90 \qquad \sum \left(x_2 - \overline{x_2}\right)^2 = 108$

F Test

$$S_{x_1}^{\;2} = \frac{1}{n_1 - 1} \sum \left(x_1 - \overline{x_1}\right)^2 = \frac{90}{9} = 10$$

$$S_{x_2}^{\;2} = \frac{1}{n_2 - 1} \sum \left(x_2 - \overline{x_2}\right)^2 = \frac{108}{11} = 9.827$$

$$F = \frac{S_1^{\;2}}{S_2^{\;2}} = 1.078$$

Tabulated $F_{0.05}$ for (9,11) degrees of freedom is 2.90 $F < F_0$

$\therefore$ H_0 is accepted.

t Test:

$H_0 : \mu_1 = \mu_2$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2)$$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum \left(x_1 - \overline{x_1}\right)^2 + \sum \left(x_2 - \overline{x_2}\right)^2\right] = 9.9$$

$\therefore$ t = 0.742

$t_{0.05}$ for 20 degrees of freedom is 2.086

$|t| < t_0 \therefore\ H_0$ is accepted.

Both the hypothesis are accepted.

$\therefore$ We may consider that the given samples have been drawn from the same population.

Example:7

A group of 10 rats fed on diet A and another group of 8 rats fed on diet B, recorded the following increase in weight.

Diet A	5	6	8	1	12	4	3	9	6	10
Diet B	2	3	6	8	10	1	2	8		

Find if the variances are significantly different.

Solution:

Given $n_1 = 10$ $n_2 = 8$

Diet-A			Diet-B		
x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
5	-1.4	1.96	2	-3	9
6	-0.4	0.16	3	-2	4
8	1.6	2.56	6	1	1
1	-5.4	29.16	8	3	9
12	5.6	31.36	1	-4	16
4	-2.4	5.76	10	5	25
3	-3.4	11.56	2	-3	9
9	2.6	6.76	8	Total $= 82$	
6	-0.4	0.16	$\sum y = 40$		
10	3.6	12.96			
$\sum x = 64$	Total $= 102.4$				

$\bar{x} = 6.4$ and $\bar{y} = 5$

$$\sum\left(x-\bar{x}\right)^2 = 102.4 \qquad \sum\left(y-\bar{y}\right)^2 = 82$$

$$S_x^{\ 2} = \frac{1}{n_1 - 1}\sum\left(x-\bar{x}\right)^2 = 11.378$$

$$S_y^{\ 2} = \frac{1}{n_2 - 1}\sum\left(y-\bar{y}\right)^2 = 11.71$$

Steps:

1. The parameter of interest is $\sigma_x^{\ 2}\ \&\ \sigma_y^{\ 2}$

2. $H_0 : \sigma_x{}^2 = \sigma_y{}^2 = \sigma^2$

3. $H_1 : \sigma_x{}^2 =\neq \sigma_y{}^2$

4. $\alpha = 0.05$, $d.f(V_1) = n_1 - 1 = 9$, $d.f(V_2) = n_2 - 1 = 7$

5. $F = \dfrac{S_x{}^2}{S_y{}^2} = 1.02$

6. Reject H_0 if F > 5.19 (from table F)

7. Computations: 1.02 < 5.19

8. Conclusion: Since Tabulated $F_{0.05}$ for (9,7) degrees of freedom is 5.19 $F < F_0$

$\therefore$ H_0 is accepted. We conclude that the difference is not significant.

Exercise:

The nicotine content in milligram of two samples of tobacco were found to be as follows.

Sample A: 24, 27, 26, 21, 25

Sample B: 27, 30, 28, 31, 22, 36

Can it be said that two samples come from the same normal population?

Chi –square distribution:

(i) Chi – Square Test of Goodness of Fit:

A very powerful test for testing the significance of the difference between theory and experiment was given by Karl pearson is 1990 and is known as " Chi square test of goodness of fit".

If $O_i(i = 1,2,...,n)$ is a set of observed or experimental frequencies and $E_i(i = 1,2,...,n)$ is the corresponding set of expected frequencies, are significant or not.

Then Karl Pearson's χ^2 is given by,

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

χ^2 is used to test whether the differences between observed and expected frequencies are significant. Degrees of freedom for B.D= n-1.

Applications of χ^2 distribution:

1. To test the goodness of fit.

2.To test the independence of attributes.

3. To test if the hypothetical value of the population vaiance is σ^2

4. To test the homogeneity of independent estimates of the population variance.

5. . To test the homogeneity of independent estimates of the population correction coefficient.

Conditions for the application of χ^2-test:

1. The sample observations should be independent

2.Constraints on the cell frequencies if any must be linear

3. No.of the total frequency should be atleast 50

4. No theoretical cell frequency should be less than 5

Example:8

The number of automobile accidents per week in a certain community are as follows. 12,8,20,2,14,10,15,6,9,4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

Solution:

Expected frequency of accidents each week =100/10 = 10

Null hypothesis Ho: The accident conditions were the same during the 10 week period.

Observed Frequency	Expected Frequency	(0- E)	$(0- E)^2 /E$
12	10	2	0.4
8	10	-2	0.4
10	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0
15	10	5	3.5

6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] = 26.6$$

Tabulated value of χ^2 at 9 degrees of freedom is 16.9.

Calculated χ^2 >Tabulated χ^2

$\therefore H_0$ is rejected.

Exercise :

1. The no.of automobile accidents per week in a certain community are as follws: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident condition, were the same during this 10 week period.

2. The following figures show the distribution of digits in numbers chosen at random from a telephone directory .

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently?

Chi squire test for independence of attributes:

An attribute means a quality or characteristic. Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as 2 X 2 contingency table.

A	a	B
B	c	D

a	b	a+b
c	d	c+d
a+c	b+d	N

$E(a) = \frac{(a+c)(a+b)}{N}$	$E(a) = \frac{(b+d)(a+b)}{N}$	a+b

The expected frequencies are given by,

$E(a)=\frac{(a+c)(c+d)}{N}$	$E(a)=\frac{(b+d)(c+d)}{N}$	c+d
a+c	b+d	N

H_0 : Attributes are independent

Degrees of freedom = (r-1)(c-1)

r = no of rows

c = no of columns

Example:9

The following table gives the classification of 100 workers according to the sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	Unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	100

Solution:

1. The parameter of interest is χ^2

2. H_0 : Nature of work is independent of the sex of the workers.

3. H_1 : Nature of work is not independent of the sex of the workers.

4. $\alpha = 0.05$ d.f = (r-1)(c-1)=1

5. Reject H_0 if $\chi^2 > 3.841$ at 5%

6. Computation:

Expected frequencies are given in the table.

$\frac{50\times 60}{100}=30$	$\frac{50\times 60}{100}=30$	60

$\frac{50\times 40}{100}=20$	$\frac{50\times 40}{100}=20$	40
50	50	100

Calculation of χ^2:

Observed Frequency	Expected Frequency	(0- E)	(0- E)^2 /E
40	30	100	3.33
20	30	100	3.33
10	20	100	5
30	20	100	5

$$\chi^2=\sum_{i=1}^{n}\left[\frac{(O_i-E_i)^2}{E_i}\right]=16.66$$

7. Conclusion: Tabulated value of χ^2 for 1 degrees of freedom at 5% level of significance is 3.84

Calculated χ^2 >Tabulated χ^2

$\therefore H_0$ is rejected. We conclude that the nature of the workers are not independent.

Example:10

Out of 8000 graduates in a town 800 are females, out of 1600 graduate employees 120 are females, Use χ^2 to determine if any distinction is made in appointment on the basis of sex. Value of χ^2 at 5% level for one degree of freedom is 3.84.

Solution:

	Female	Male	Total
Graduates	800	7200	8000
Employees	120	1480	1600
Total	920	8680	9600

1. The parameter of interest is χ^2

2. H_0 : No difference between 2 treatments

3. H_1 : Difference between 2 treatments

4. $\alpha = 0.05$ d.f = (r-1)(c-1)=1

5. The test statistic $\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)}$

6. Reject H_0 if $\chi^2 > 3.841$ at 5%

7. Computation: $\chi^2 = \frac{(800X1480 - 7200X120)^2(9600)}{920X8680X8000X1600} = 9.617$

8. Conclusion: Tabulated value of χ^2 for 1 degrees of freedom at 5% level of significance is 3.84

Calculated χ^2 >Tabulated χ^2

$\therefore H_0$ is rejected. We conclude that the treatment are not independent

Exercise:

On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the treatment is comparatively superior to the conventional treatment.

	Favourable	Not favourable	Total
New	60	30	90
Conventional	40	70	110

Questions
A numerical measure of uncertainty is practiced by the important branch of statistics called _______
If P(A) is 1, the event A is called a _____
p + q = _____, here p is success and q is failure events
In rolling of single die, the chance of getting 2,4,6 (even numbers) are ______
The set of all possible outomes of an activity is the ___________
Events that cannot happen together are called________________
If one event is unaffected by the outcome of another event, the two events are said to be _____
If P(A or B)=P(A), then______________
If P(X≤ x)=______
If the outcome of one event does not influence another event, then the two events are_____
If P(A)=0.9, P(B/A)=0.8, find P(A∩B) = ______
If P(X>x)=______
For a discrete random variable, the probability density function represents the _____
Why are the events of a coin toss mutually exclusive________
What is the probabiity that a ball drawn at random from the urn is blue_____________
What is the probability of getting an even number when a die is tossed___
What is the probability of getting more than 2 when a die is tossed___
The probability of drawing a spade from a pack of cards is_____
If A and B are independent event P(A)=0.4 and P(B)=0.5 then P(AUB)=
What is the probability of getting a sum 9 from two throws of a dice?
Three unbiased coins are tossed. What is the probability of getting at most two heads?
A bag contains 6 black and 8 white balls. One ball is drawn at random. What is the probability that the ball drawn is white?
Total probability is _____
Probability of a single real value in a continuous random variable is _____
A random variable X is_____ if it assumes only discrete values.
If P(A)=0.35, P(B)=0.73, P(A∩B) =0.14 find P(A'UB')= ______
The classical school of thought on probability assumes that all possible outcomes of an experiment are____
Rolling of die is a
If P(A) = 0, the event A is called a
The simple probability of an occurrence of an event is called the _____________
E(ax+b)=
The impossible event is
A density function may correspond to different ______
If A and B are independent and P(A)=0.2, P(B)=0.6 find P(A∩B) =

opt1
Theory of mathematics
Cases
7
simple
sample space
mutually exculsive
dependent
A and B are mutually exclusive
1-P(X>x)
mutually exculsive
0.72
1-P(X≤x)
probability mass function
the out come of any toss is not affected by the out come of those preceding
0.1
1/3
1/3
1/52
0.7
1/6
3/4
3/4
0
two
spectrum
0.86
Equally likely
Trial
Trial
bayesian probability
ax+b
0
probability mass function
0.12

opt2
Theory of physics
Trial
9
Compound event
event
event
independent
venn diagram
1
dependent
0.17
P(X≤x)
probability distribution function
both a head and a tail cannot turn up on any one toss
0.4
1/2
1/2
1/13
0.1
1/9
1/4
4/7
1
three
complex
0.115
Mutually exclusive
Cases
Impossible event
joint probability
aE(x)+b
1
probability distribution function
0.8

opt3
Theory of statistics
Certain Event
1
Certain event
independent
exclusive
mutually exclusive
P(A)+P(B)
0
independent
0.1
1
probability density function
the probabiity of getting a head and the probability of getting a tail
0.6
1/6
2/3
4/13
0.3
8/9
3/8
1/8
-1
four
continuous
1.08
Mutually exclusive and equally likely
Event
Cases
mariginal probabiity
E(x)
-1
probability density function
0.2

opt4
Theory of probability
3
impossible event
Theory of probability
mode
11
deviation
conditional probability
P(X>x)
random variable
0.86
0
zero
1
mode
1/9
1/4
1/4
0.5
7/8
3/7
none of these
0.5
zero
Random experiment
0.66
1/9
Random experiment
Event
all of these
None of these
0.5
random variable
0.4

opt5	opt6

Answer
Theory of probability
Certain Event
1
Compound event
sample space
mutually exculsive
independent
A and B are mutually exclusive
1-P(X>x)
independent
0.72
1-P(X≤x)
probability mass function
both a head and tail cannot turn up on any one toss
0.6
1/2
2/3
1/4
0.7
1/9
3/8
4/7
1
zero
Random experiment
0.86
Mutually exclusive and equally likely
Trial
Impossible event
marginal probability
aE(x)+b
0
random variable
0.12

Questions
The correlation coefficient is used to determine______
If X and Y are independent , then
The relationship between three or more variables is studied with the help of _________ correlation.
The coefficient of correlation is under-root of two _____
The coefficient of correlation___
which of the following is the highest range of r__________
The coefficient of correlation is independent of ________
The coefficient of correlation___
COV(X,Y)= __________
Two random variables with non zero correlation are said to be___
Correlation means relationship between ______ variables
Two random variables X and Y with joint pdf f(x,y) is said to independent if ____
The covariance of two independent random variable is ________
Two random variables are said to be orthogonal if _______
Two random variables are said to be uncorrelated if correlation coefficient is____
Regression analysis is a mathematical measures of the average relationship between ________variable
The regresision analysis confined to ther study of only two variable at a time is called__________regression
If r=0, then the regression coefficient are________
The equation of the fitted stright line is _____
If X=Y , then correlation cofficient between them is _______
The greater the value of r__________obtained through regression analysis
Where r is zero the regression lines cut each other making an angle of______
The father the two regression lines cut each other_______
The regression lines cut each other at the point of :______
When the two regression lines coincide, then r is :_________
The variable , we are trying to predict is called the _________
Both the regression coefficients cannot ______one
The regression analysis measures ______between variables
If the possible values of (X,Y) are finite, then (X,Y) is called a______
If X & Y are continuous random variable , then f(x,y) is _____
Joint probability is the probability of the _________occurrence of two or more events.
The order of arrangement is important in _______
If X & Y are____random variable , then f(x,y) is called joint probability function.
If the value of y decreases as the value of x increases then there is ______correlation between two variables.
The correlation between the income and expenditure is _____
correlation between price and demand of commodity is ____
If X and Y are independent , then _____
correlation coefficient does not exceed _____
Two independent variables are _____
In Rank correlation the correction factor is added for each _____value.
When r = 1 or -1 the the line of regression are ____ to each other.
If the curve is a straight line, then it is called the ____
If the curve is not a straight line, then it is called the ____
when r is ____ the correlation is perfect and positive.
The coefficient of correlation is independent of change of_____ and ________

opt1
A specific value of the y-variable given a specific value of the x-variable
E(XY)=0
multiple
regression coefficients
has no limits
0 and 1
change of scale only
cannot be positive
E(XY)-E(X)E(Y)
correlation
two
f(x,y) = f(x) +f(y)
Zero
correlation is zero
zero
two or more
Simple
zero
y=ax+b
1
the better are estimates
30 degree
Greater will be degree of correlation
Average of X and Y
0
depentent variable
exceed
dependence
two dimensional random variable
joint probability function
Simultaneous (or) joint
permutation
discrete
negative
positive
positive
E(XY) = E(X) + E(Y)
unity
correlated
repeated
parallel
the line of correlation
covariance
1
scale,origin

opt2
A specific value of the x-variable given a specific value of the y-variable
E(X) E(Y)=0
rank
rank coefficient
can be less than 1
minus one and 0
change of origin only
cannot be negative
E(XY)+E(X)E(Y)
regression
one
f(x,y) = f(x) / f(y)
two
rank is zero
one
one
Multiple
one
y=a+bx
zero
the worst are the estimates
60 degree
The less will be the degree of correlation
Average of X only
-1
indepent variable
exact
independence
onedimensional random variable
joint probability density function
Conditional
Gambling
continuous
perfect positive
negative
finite
E(XY) = E(X) - E(Y)
5
uncorrelated
Non-repeated
perpendicular
the line of regression
the line of correlation
2
vector,origin

opt3
The strength of the relationship between the x and y variables
Cov(X,Y) =0
perferct
Regression equation
can be more than 1
minus one and one
both change of scale and origin
can be either positive or negative
E(XY)
rank
two or more
f(x,y) = f(x) * f(y)
three
covariance is zero
two or more
Two variables
Linear
threee
y=mx+c
less than one
really makes no difference
90 degree
does not matter
Average of Y only
1
constant
plus or minus
constant
both a and b
both a and b
Mariginal probability
joint
both a and b
both a and b
finite
negative
E(XY) = E(X) E(Y)
0
both a and b
indefinite
straight line
covariance
the curvilinear
3
variable, constant

opt4
none of these
E(XY)=1
spearman's rank
regression line
varies between + or - one
zero
change of variables
zero
Var(X,Y)
variables
three
f(x,y) = f(x) - f(y)
two or more
one
orthogonal
three
two
constant
y=mx
gerater than one
good estimates
neither of the above
the worst are the estimates
average of both(a) and (b)
0.5
normal
negative
normal
infinte
infinte
density function
density
infinte
infinte
both a and b
both a and b
E(XY) = E(X)/E(Y)
2
positive
both a and b
circular
both a and b
the line of regression
0
interer, origin

opt5	opt6

Answer
The strength of the relationship between the x and y variables
Cov(X,Y) =0
multiple
regression coefficient
varies between + or - one
minus one and one
both change of scale and origin
can be either positive or negative
E(XY)-E(X)E(Y)
regression
two or more
f(x,y) = f(x) * f(y)
Zero
correlation is zero
zero
two or more
Simple
zero
y=ax+b
1
the better are estimates
neither of the above
the less will be the degree of correlation
average of X and Y
1
dependent variable
exceed
dependence
two dimensional random variable
both a and b
Simultaneous (or) joint
permutation
continuous
negative
positive
negative
E(XY) = E(X) E(Y)
unity
uncorrelated
repeated
parallel
the line of regression
the curvilinear
1
scale,origin

When r = 0 the line of regression are ______to each other.
A Mathematical measure of the average relationship between two variables is called______
Cov(X,Y)=_____
The coefficient of correlation_____
The correlation between two variables is of order_____

parallel
correlation
E[{ X- E(X) }* { Y – E(Y) }]
is the square of the coefficient of determination
2

perpendicular
regression
E[{ X- E(X) }+ { Y – E(Y) }]
is the square root of the coefficient of determination
1

straight line
rank
E[{ X- E(X) }- { Y – E(Y) }]
is the same as r-square
0

circular
variables
E[{ X- E(X) } { Y – E(Y) }]
can never be negative
3

perpendicular
correlation
E[{ X- E(X) } { Y – E(Y) }]
is the square root of the coefficient of determination
0

Questions
The population consisting of all real numbers in an example of ____
The propability distribution of a statistic is called ________
A part selected from the population is called a ________
________is the standard deviation of the sampling distribution
The chi square test was devised by ________
Null hypothesis is the hypothesis of ________
Alternative hypothesis complementary to________
Type I error is committed when the hypothesis is true but our test___it
A Type II error is made when__________
The best critical region consists of _______
The standard deviation of sampling distribution is called_____
Standard error provides an idea about the _________of sample
Normal distribution is a limiting form of________distribution
A hypothesis may be classified as______
The standard normal distribution is also known as _________distribution
if v tends to infinity , the chi-square distrbution tends to _______ distribution
The mean of sampling distribution of means is equal to the_______
If a test of hypothesis has a Type I error probability (α) of 0.01, we mean____
Students t- test is applicable only when____
A contingencies table should have frequencies in ____
Student's t- test is applicable in case of_______
Which distribution is used to test the equality of population means ____
The shape of t-distribution is similar to that of ________
The number of degrees of freedom for contingency table are on the basis of ________
Student's t- test was invented by_________
The degrees of freedom for contingency table are on the basis of _________
The calculated value of chi-square is :_________
Degrees of freedom for statistic chi-square in case of contingency table of order (2x2) is________
Normal distribution is applicable in case of_______
Degrees of freedom for chi-square in case of contingency table of order (4x3) are _______
E(ax+b)=
The impossible event is
A density function may correspond to different ______
If A and B are independent and P(A)=0.2, P(B)=0.6 find P(A∩B) =

opt1
An infinite population
normal distribution
sample
standard error
Fisher
difference
hypothesis
rejects
the null hypothesis is accepted when it is false.
extreme positive values
standard error
unreliability
Binomial
Simple
unit normal
normal distribution
Mean
if the null hypothesis is true, we don\\'t reject it 1% of the time.
the variate values are independent
percentages
Small samples
chi-square distribution
chi-square distribution
8
R.A.Fisher
n-1
always positive
3
Small samples
12
ax+b
0
probability mass function
0.12

opt2
An finite population
Sampling distribution
Population mean
Population mean
gauss
mean
testing of hypothesis
accept
the null hypothesis is rejected when it is true.
extreme negative values
mean error
normality
normal
Composite
normal
Sampling distribution
Population mean
if the null hypothesis is true, we reject it 1% of the time.
the variate is distributed normally
proporation
for samples of size between 5 and 30
F-Distribution
F-Distribution
4
G.W.Snedector
r-1
always negative
4
for samples of size between 5 and 30
9
aE(x)+b
1
probability distribution function
0.8

opt3
sample
binomial distribution
error
sample
laplace
no difference
null hypothesis
null hypthesis
the alternate hypothesis is accepted when it is false.
both (a) and (b)
error
reliability
uniform
null
uniform
binomial distribution
variance
if the null hypothesis is false, we don\'t reject it 1% of the time.
the sample is not large
frequencies
large samples
Normal distribution
Normal distribution
3
W.S.Gosset
c-1
either positive or negative
2
large samples
8
E(x)
-1
probability density function
0.2

opt4
normal
Sample
mean square
sampling
karl pearson
variance
Type-I
alternative hypothesis
the null hypothesis is accepted when it is true.
neither (a) nor (b)
variance
simple
sample
all the above
sample
Sample
Sample mean
if the null hypothesis is false, we reject it 1% of the time.
all the above
ratio
all the above
t- distribution
uniform distribution
2
W.G.Cochran
r-2
none of these
1
all the above
6
None of these
0.5
random variable
0.4

opt5	opt6

Answer
An infinite population
Sampling distribution
sample
standard error
karl person
no difference
null hypothesis
rejects
the null hypothesis is accepted when it is false.
both (a) and (b)
standard error
unreliability
Binomial
all the above
unit normal
normal distribution
Population mean
if the null hypothesis is true, we reject it 1% of the time.
all the above
frequencies
Small samples
F-Distribution
chi-square distribution
4
W.S.Gosset
r-1
always positive
1
large samples
6
aE(x)+b
0
random variable
0.12

Questions
The most widely used of all experimental design is ________
The experimental area should be in the form of _________
The word _________ in analysis of variance is used to refer to any factor in the experiment.
In the case of one-way classification the total variation can be split into ___________________
Analysis of variance can be used when there are samples of _____ sizes
Mean square of error =____________ for one way classification
Total variation SST = ________ for one way classification
Mean square between column mean ___________
_______ stands for mean square between samples
The stimulus to the development of theory and practice of experimental design came from__________
The analysis of variance originated in ___________
The Latin square model assumes that interactions between treatment and row and column groupings are _________
The science of experimental designs is associated with the name ________
In 4×4 Latin square, the total of such possibilities are___________
The latin sqares are most widely used in the field of ___________
The total number of possibilities in which arrangements can be made in 3×3 Latin square are_________
The one way classification is exhibited ____________ wise
The shape of the experimental material should be _________
The number of treatments should be ________ number of rows and number of columns
Latin square design controls variability in ________ directions of the experimental material
________ Latin square is not possible
In the case of two-way classification, the total variation (TSS) equals______
The assumptions in analysis of variance __________
In one way classification the dates are classified according to ______ factor
Equality of several normal population means can be tested by ___________
Analysis of variance technique was developed by__________
Analysis of variance technique originated in the field of____________________
One of the assumption of analysis of variance is that the population from which the samples are drawn is__________
In a two way classification the datas are classified to _____ factor
In the case of one-way classification with N observations and t treatments, the error degrees of freedom is________
In the case of one-way classification with t treatments, the mean sum of squares for treatment is________________
In the case of two-way classification with r rows and c columns, the degrees of freedom for error is______________
Latin square design controls variability in ________ directions of the experimental material
With 90, 35, 25 as TSS, SSR and SSC respectively in case of two way classification, SSE is____
One of the assumptions of Analysis of variance is observations are ________________
Total variation in two – way classification can be split into ___________ components.
In the case of one way classification with 30 observations and 5 treatment, the degrees freedom for SSE is ________
In the case of two-way classification with 120, 54, 45 respectively as TSS, SSC, SSE, the SSR is ______________
The origin of statistics can be traced to _________________
'Statistics may be called the science of counting' is the definition given by ___________
______ is one of the statistical tool plays prominent role in agricultural experiments.
The Latin square model assumes that interactions between treatment and row and column groupings are _________
In 4×4 Latin square, the total of such possibilities are___________
The sum of the squares between samples are denoted by

opt1
Randomised block design
Circle
Error
Two components
Equal
SSE
SSC+SSR
SSE
MSC
Agrarian research
Agrarian research
Existent
Randomised block design
8
Agriculture
6
Row
Circle
Equal
One
2×2
SSR + SSC + SSE
Normality
One
Bartlet’ s test
S. D. Poisson
Agriculture
Binomial
One
N-1
SST/N-1
(rc) – 1
One
50
independent
two
20
19
State
Croxton
Analysis of variance
Existent
8
SSR

opt2
latin square
parabola
normality
Three components
unequal
SSE\n-c
SSE+TSS
SSE\n-c
SSE
industry research
industry research
non-existent
latin square
10
industry
9
column
parabola
unequal
two
3×3
SSR -SSC + SSE
Homogencity
two
F - test
Karl – Pearson
industry
Poisson
two
t -1
SST/ t-1
(r-1).c
two
40
dependent
three
19
21
Commerce
A.L.Bowley
Normality
non-existent
10
SSE

opt3
Mean square of error
square
treatment
Four components
greater than
SSE\1-c
SSC+SSE
SSE\c-1
SSR
astronomy research
astronomy research
experimental error
Mean square of error
200
astronomy
12
both a & b
rectangular
greater than
three
4×4
SSR + SSC – SSE
independence of error
three
chi square-test
R.A. Fisher
astronomy
Chi-square
three
N-t
SST/N-t
(r-1) (c-1)
three
30
Industry
four
24
20
Economics
Boddington
Homogenecity
experimental error
200
TSS

opt4
experimental error
ellipse
Homogencity
Only one component
less than
SSE\r-1
TSS
SSE\r-1
SST
medicine research
medicine research
Mean square of error
None of these
576
medicine
120
None of these
ellipse
less than
four
5×5
SSR + SSC
both a,b& c
four
t- test
W. S. Gosset
medicine
Normal
four
Nt
SST/t
(c-1).r
four
20
Genetics
five
25
27
Industry
Webster
independence of error
Mean square of error
576
SSC

opt5	opt6

Answer
Randomised block design
square
treatment
Two components
unequal
SSE\n-c
SSC+SSE
SSE\c-1
MSC
Agrarian research
Agrarian research
non-existent
latin square
576
Agriculture
12
both a & b
rectangular
Equal
two
2×2
SSR + SSC + SSE
both a,b& c
One
F - test
R.A. Fisher
Agriculture
Normal
two
N- t
SST/ t-1
(r-1) (c-1)
two
30
independent
three
25
21
State
A.L.Bowley
Analysis of variance
non-existent
576
SSC

Questions
A control chart contains ________ horizontal lines.
Attributes are characteristics of products which are _______
The theoretical basis for c chart is ________ distribution
When the quality of a product is measurable quantitatively, we use control charts are _____
The theoretical basis for X chart is ________ distribution
Standard error of means ________________
Variable are those quality characteristics of a product or item which are _______
The theoretical basis for the np-chart is _____ distribution.
Control chart for number of defects is called _____
The total number of defects in 15 pieces of cloth of equal length is 90. Then the UCL for c-char
The variation of a quality characteristics can be divided under ________ heads.
The total number of defects in 20 pieces of cloth is 220. The UCL is________________
Whenever LCL is ≤0, it is taken as _____________
The total number of defects in 15 pieces of cloth of equal length is 90. Then the UCL
Parallel series configuration is also known as ________
In R- chart, if σ is known, UCL = D2 σ and LCL = _______
The variation of a quality characteristics can be divided under two heads, chance variartion and ________
Control chart for fraction defective is also called _______
Control chart for number of defectives is called _____
The theoretical basis for R- chart is ________ distribution
The total number of defects in 20 pieces of cloth is 220. The LCL is________________
The total number of defects in 15 pieces of cloth of equal length is 90. Then the LCL for c-chart ________
The theoretical basis for c- chart mean is ________
For n=2 to 6, the value of D1 is________________
In the preparation of R-chart, if D3=0 then LCL is
A ________________is the partial and total loss of a device
________________is only for non repairable items.
________________is only for repairable items.
The reliability R(t) is________________function of t
If the repair time is negligible then MTBF_______________
Series is ________ in which the components of the system are connected in series
Parallel is ________ in which the components of the system are connected in parallel
The control limits of R-chart are UCLand LCL_______
The theoretical basis for np- chart mean is ________

opt1
UCL
Normal
Uniform
np chart
Binomial
A1
Measurable
Uniform
np chart
13.35
one
19.95
0
13.35
low level redundancy
D1σ
Measurable
np chart
np chart
Binomial
1.05
0
λ
2.2
0
failure
MTTF
MTTF
normal
MTTF
one
one
D3R
np

opt2
Central line CL
Measurablev
Binomial
R chart
Poisson
A2
Not Measurable
Binomial
R chart
14.35
two
20.95
1
14.35
High level redundancy
D2 σ
Normal
R chart
R chart
Poisson
0
-0.003
np
0.1
1
reliability
MTBF
MTBF
increasing
MTBF
two
two
D4R
n/p

opt3
LCL
Not Measurable
Poisson
X chart
normal
A4
Normal
Poisson
p chart
0
three
0.05
2
0
Redundant configuration
σ
Not Measurable
p chart
p chart
normal
0.05
0.3
npq
0.2
2
unreliability
Hazard rate
Hazard rate
decreasing
Hazard rate
three
three
A2R
p/q

opt4
All
Poission
Geomentric
both np &X chart
uniform
A5
Poission
Geomentric
c chart
25.7
four
1.05
3
1
Down time
0
Assignable variation
c chart
c chart
uniform
1.04
0.001
λp
0
3
availability
Failure
Failure
failure
MMTF
zero
zero
both D3R&D4R
npq

opt5	opt6

Answer
All
Not Measurable
Poisson
both np &X chart
normal
A2
Measurable
Binomial
c chart
13.35
two
20.95
0
13.35
low level redundancy
D1σ
Assignable variation
p chart
np chart
normal
1.05
0
λ
0
0
failure
MTTF
MTBF
decreasing
MTTF
one
one
both D3R&D4R
np