Chapter 01

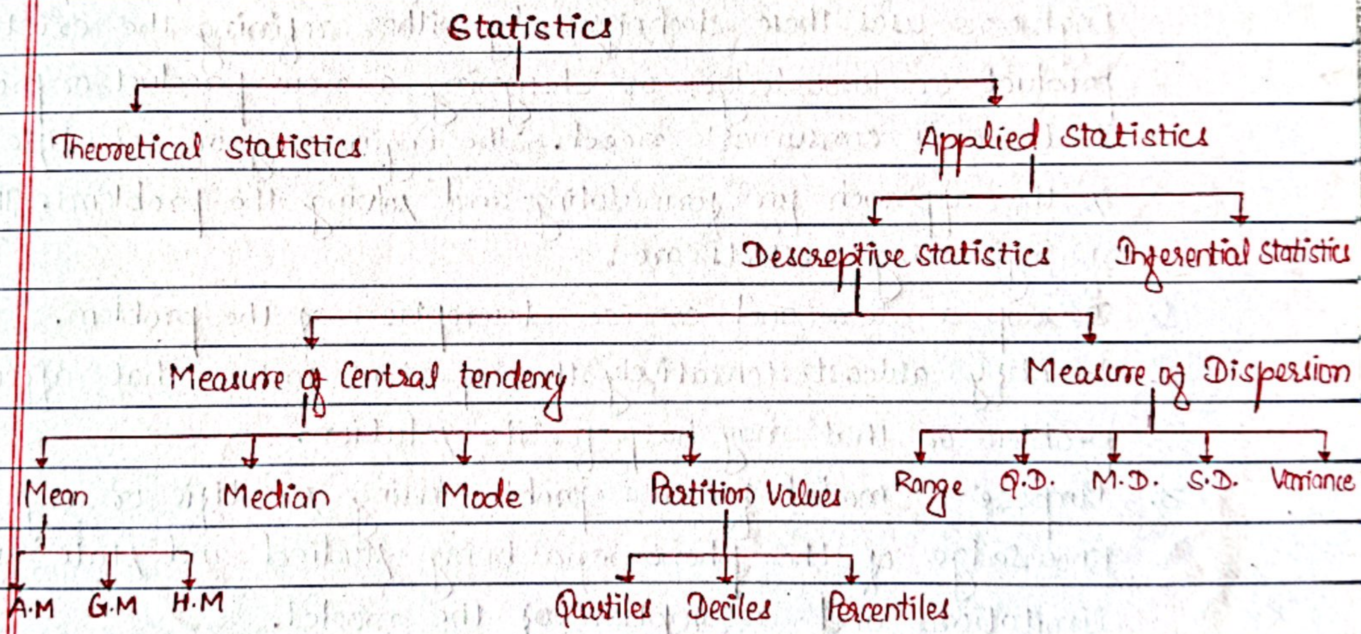
Role of Statistics in Engineering

DATE : 2020/05/12
PAGE : 01

Background

In the modern world of computer's and information technology, the importance of statistics and probability is very well organized by all the disciplines. Statistics has originated as science of data and statehood and slowly it has an important applications in agriculture, engineering, economics, commerce, biology, medicine, industry, planning, education, and so on.

Statistics is the science of learning from data. It is concerned with scientific methods for collecting, organizing, summarizing, presenting and analysing data as well as deriving valid conclusion and making resonable decision on the basis of analysis of sample data.



1. Descriptive Statistics

Descriptive Statistics is termed as the analysis of data that helps to describe, show, and summarize the data in meaningful way. It is the simple way to describe the data and moreover descriptive statistics is very useful to present raw data in effective and meaningful way using numerical calculations or graphs or tables.

DATE: 2080/05/18
PAGE:

2. **Inferential Statistics**

In the inferential statistics, predictions are made by taking any group of data in which you are interested. It can be defined as a random sample of data taken from population to describe and make inference about population. It basically allow you to make predictions by taking small sample instead of working on whole population.

The Engineering Methods and Statistical Thinking (Why Engineers need statistical techniques)

An Engineer is someone who solves problems of interest to society by using the efficient application of scientific principles. Engineers uses these techniques by either refining the existing product or knowledge, or designing a new product or process that needs consumer's need. The Engineering or scientific method is the approach for formulating and solving the problems. The steps of Engineering methods are:

1. Develop a clear and concise descreption of the problem.
2. Identify atleast tentatively the important factors that affect this problem or that may help for its solution.
3. Purpose a model for the problem using scientific or engineering knowledge of the phenomena being studied and state any limitations and assumption of the model.
4. Conduct appropriate experiment and collect data to test or validate the tentative model for conclusion made in state 2 or 3.
5. Refine the model on the basis of observed data.
6. Manipulate the model to assist in developing sol? to the problem.
7. Construct an appropriate experiment to confirm that the proposed sol? to the problem is both effective and efficient.
8. Draw a conclusion or make recommendation based on the problem's solution.

It is clear that theory of probability and statistics has an important role in the field of engineering. Statistics has various Engineering applications, for example in testing of materials, performance test of system, robotics and so on.

Specially the use of statistics in the different field of Engineering are as follows:

1. In Electrical Engineering signals and noise are analyzed by means of probability theory. Engineer study probability for ranging from quality control and quality assurance of communication theory.
2. Civil, Mechanical, and, Industrial Engineers are statistics and probability to test account for variation in materials and goods.
3. Civil Engineer uses it to determine the consistency of the all types of structures.

So, Engineering methods and statistical thinking are highly correlated to eachother.

Functions of Statistics

1. To represent facts from numerical figure into a definite form.
2. To summarize huge data using appropriate statistical measure.
3. To help in classification of data according in nature.
4. To help in formulating policies.
5. To determine relationship between different phenomena.
6. To help in predicting future trend.
7. To formulate test of hypothesis.
8. To draw a valid (i.e acceptable) appropriate conclusion.

chapter 02

Probability

DATE: 2020/05/15
PAGE:

Sample Space

When a random experiment is performed then the set of all possible outcomes of the random experiment is known as sample space.

e.g.

(i) If a coin is tossed then sample space $S = \{H, T\}$.

(ii) If two coins are tossed together then sample space $S = \{HH, HT, TH, TT\}$.

(iii) If three coins are tossed together then sample space $S = \{HHH, HHT, HTH, THT, HTT, THH, TTH, TTT\}$.

(iv) If a dice is rolled then sample space $S = \{1, 2, 3, 4, 5, 6\}$.

(v) If two dice rolled together then sample $S = \{(1,1) \ldots (1,6), (2,1) \ldots (2,6), (3,1) \ldots (3,6), (4,1) \ldots (4,6), (5,1) \ldots (5,6), (6,1) \ldots (6,6)\}$.

Events

An event is sub-set of sample space. Suppose if 'E' denotes even when a dice is rolled then,

$$E = \{2, 4, 6\}.$$

Clearly, $E = \{2, 4, 6\} \subseteq \{1, 2, 3, 4, 5, 6\}$

Note: If 'E' be an event then chance of occurrence of 'E' OR probability of E is,

$$P(E) = \frac{n(E)}{n(S)} = \frac{\text{Favourable case to event E}}{\text{All possible Case}}$$

And,

chance of none occurrence of 'E' is,

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

$$\Rightarrow P(E) + P(\bar{E}) = 1$$

Types of Event

1. Equally likely Event

Two or more than two events of a random experiment are called equally likely event if their chance of occurrence is same.

DATE : __/__/__
PAGE : ____

e.g., Let E_1 denotes red card, E_2 denotes black card, E_3 denotes face card, E_4 denotes card of heart and E_5 denotes card of diamond, when a card is drawn randomly from pack of 52 card, then

$$P(E_1) = \frac{n(E_1)}{n(S)} = \frac{26}{52} = \frac{1}{2}$$

$$P(E_2) = \frac{n(E_2)}{n(S)} = \frac{26}{52} = \frac{1}{2}$$

$$P(E_3) = \frac{n(E_3)}{n(S)} = \frac{12}{52} = \frac{3}{13}$$

$$P(E_4) = \frac{n(E_4)}{n(S)} = \frac{13}{52} = \frac{1}{4}$$

$$P(E_5) = \frac{n(E_5)}{n(S)} = \frac{13}{52} = \frac{1}{4}$$

Here events E_1 and E_2 are equally likely event, and, E_4 and E_5 are equally likely event but others are not equally likely.

2. Mutually Exclusive Event

Two or more than two events are said to be mutually exclusive events if their simultaneous occurrence is impossible i.e. they can't occur together otherwise they are not mutually exclusive.

e.g. Suppose a card is drawn from a bag of card numbered 1 to 20. Let E_1 denotes multiple of 4, $E_1 = \{x : x = 4K, K \in N\}$, $E_2 = \{y : y = 5K, K \in N\}$, E_3 denotes even card, E_4 denotes odd card.

Here, E_1 and E_2 are not mutually exclusive event, since $E_1 = \{4, 8, 12, 16, 20\}$ and $E_2 = \{5, 10, 15, 20\}$. E_1 and E_3 are also not mutually exclusive. However, E_3 and E_4 are mutually exclusive.

Note: If A and B are any two events then the chance of occurrence of atleast one of them is denoted by P(either A or B) OR P(A or B) OR $P(A \cup B)$ and given by,

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

DATE: 2080/05/19
PAGE:

3. **Independent Event**

Two or more than two events of a random experiment are to be independent if the chance of occurrence of one event isn't affected by chance of occurrence of other event.

e.g. Let a coin and a dice is thrown together, suppose E_1 denotes head on coin and E_2 denotes prime on dice. Clearly $P(E_1) = \frac{1}{2}$ and $P(E_2) = \frac{1}{2}$, both are not related to eachother.

Notes: 1. If A and B and C are three independent event, then
$$P(A \cap B \cap C) = P(A) * P(B) * P(C)$$

2. $P(A \cap \bar{B})$ = Occurrence of A but not B $= P(A) * P(\bar{B})$
3. $P(\bar{A} \cap B)$ = Occurrence of B but not A $= P(\bar{A}) * P(B)$
4. $P(\bar{A} \cap \bar{B})$ = Occurrence of none of them $= P(\bar{A}) * P(\bar{B})$
5. $P(A \text{ or } B)$ = Atleast one of them = 1 - none of them $= 1 - P(\bar{A}) * P(\bar{B})$

Conditional Probability

Additional law of probability

Let A and B are any two events of a random experiment, then the probability of occurrance of atleast one of them is denoted by $P(A \text{ or } B)$ or $P(A \cup B)$, where
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Note: 1. If A and B are mutually exclusive events then $P(A \cap B) = 0$, so
$$P(A \cup B) = P(A) + P(B)$$

2. If A, B, C, D are mutually exclusive events then $P(A \text{ or } B \text{ or } C \text{ or } D)$ or $P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D)$

Multiplication law of probability

Let A and B are any two independent event then the probability of their simultaneous occurance, is denoted by $P(A \text{ and } B)$ or $P(A \cap B)$ and

Scanned with CamScanner

DATE : __/__/__
PAGE :

given by, $P(A \cap B) = P(A) * P(B)$.

Note: 1. If A, B, C are three independent event then $P(A \cap B \cap C) = P(A) * P(B) * P(C)$

2. $P(\bar{A} \cap B)$ = chance of occurrance of B not A $= P(\bar{A}) * P(B)$

3. $P(A \cap \bar{B})$ = " " " A but not B $= P(A) * P(\bar{B})$

4. $P(\bar{A} \cap \bar{B})$ = " " " " none of them $= P(\bar{A}) * P(\bar{B})$

5. $P(A \text{ or } B)$ = " " " " atleast one of them $= 1 - P(\bar{A} \cap \bar{B})$

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

Q.no.1. A problem of statistics given to three students A, B and C whose chance of solving it $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ respectively. Find the probability that
(i) all of them can solve the problem
(ii) none of them can solve
(iii) problem will be solved (atleast one of them can solve)
(iv) only one of them can solve.

Sol^n →

Given, $P(A) = 1/2$, $P(\bar{A}) = 1/2$

$P(B) = 1/3$, $P(\bar{B}) = 2/3$

$P(C) = 1/4$, $P(\bar{C}) = 3/4$

(i) Probability of all of them can solve,

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

$$= \frac{1}{2} * \frac{1}{3} * \frac{1}{4}$$

$$= \frac{1}{24}$$

(ii) Probability of none of them can solve,

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

$$= \frac{1}{2} * \frac{2}{3} * \frac{3}{4}$$

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

DATE : __/__/__
PAGE : ____

(iii) Probability of problem will be solved, $= 1 - P(\bar{A}\cap\bar{B}\cap\bar{C})$

$= 1 - 1/4$

$= 3/4$

(iv) Probability of only one can solve

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

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

$= \frac{1}{2}*\frac{2}{3}*\frac{3}{4} + \frac{1}{2}*\frac{2}{3}*\frac{1}{4} + \frac{1}{2}*\frac{1}{3}*\frac{3}{4}$

Conditional Probability

When the chance of occurance of uncertain event is calculated on the basis of previously known information or on the basis of past record then such a probability is known as conditional probability.

Let A and B are any two events of a random experiment then the conditional probability of event A given that B is known or B is given is denoted by,

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{n(A\cap B)}{n(S)}}{\frac{n(B)}{n(S)}} = \frac{n(A\cap B)}{n(B)}$$

Similarly, conditional probability of event B when A is known is,

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

Note: 1. Since $P(A/B) = \frac{P(A\cap B)}{P(B)}$

$$\Rightarrow P(A\cap B) = P(B) * P(A/B) \quad \text{—— (1)}$$

Similarly,

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

$$\Rightarrow P(A\cap B) = P(A) * P(B/A) \quad \text{—— (11)}$$

If A and B are independent, $P(A/B) = P(A)$ and $P(B/A) = P(B)$

then, $P(A\cap B) = P(A) * P(B)$

DATE : 2080/05/22
PAGE :

Q.no. 2 Two dice are tossed together, what is the probability of
(i) Sum of turning faces is 7 given that one face is 5.
(ii) sum of two faces is 7.

Sol^n,

Here, when two dice tossed together, then

$$S = \{(1,1), --- (6,6)\} = 36$$

Let sum is $7 = A$

and one face is $5 = B$

(i) We need to find $P(A|B) = ?$

Here, $A = \{(1,6), (6,1), (2,5), (5,2), (3,4), (4,3)\} \subset S$

$B = \{(1,5), (5,1), (5,5), (2,5), (5,2), (3,5), (5,3), (4,5), (5,4), (6,5), (5,6)\}$

$$\therefore P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{n(A\cap B)}{n(B)} = \frac{2}{11}$$

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

Partition of Sample Space

Let S be the sample space of a random experiment consider that the set of events $\{E_1, E_2, E_3, ---, E_n\}$ of S then this set of events $\{E_1, E_2, --- E_n\}$ is said to be partition of S if

① $E_1 \cup E_2 \cup E_3 \cup --- \cup E_n = S$ i.e $\bigcup_{i=1}^{n} E_i = S$

(ii) $E_i \cap E_j = \phi$ for $i \neq j$

(iii) $P(E_i) \geq 0 \quad \forall \; i = 1 --- n$

eg; Let A dice is thrown once, suppose E_1 denotes prime and E_2 denotes even and E_3 denotes odd. Discuss whether the collection $\{E_1, E_2, E_3\}$ forms partition of S or not.

DATE : __/__/__
PAGE : ______

Sol^n → Here, $S = \{1, 2, 3, 4, 5, 6\}$

$E_1 = \{2, 3, 5\}$

$E_2 = \{2, 4, 6\}$

$E_3 = \{1, 3, 5\}$

(i) $P(E_1) \geq 0$, $P(E_2) \geq 0$, $P(E_3) \geq 0$

(ii) $E_1 \cup E_2 \cup E_3 = 0$

But $E_1 \cap E_2 \neq \phi$ so the collection of events $\{E_1, E_2, E_3\}$ doesn't form the partition of S.

If we take E_1 is odd and E_2 is even then the set $\{E_1, E_2\}$ forms partition of S.

Total Probability theorem

Statement : Let the collection $\{E_1, E_2, \ldots E_n\}$ be the partition of sample space S of a random experiment. Suppose, A be any arbitrary event of S, such that $P(A) \geq 0$ then

$$P(A) = P(E_1)\, P(A|E_1) + P(E_2)\, P(A|E_2) + \cdots + P(E_n)\, P(A|E_n)$$

i.e. $$P(A) = \sum_{i=1}^{n} P(E_i)\, P(A|E_i)$$

Proof:

Here, $\{E_1, E_2, E_3, \ldots, E_n\}$ forms partition of S so,

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

(ii) $E_i \cap E_j = \phi \;\; \forall \; i \neq j$

(iii) $P(E_i) \geq 0 \;\; \forall \; i = 1, \ldots n$;

Let $A \subseteq S$ be arbitory event then $A = A \cap S$

or, $A = A \cap (E_1 \cup E_2 \cup \cdots \cup E_n)$

$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots (A \cap E_n)$

$\therefore P(A) = P\{(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)\}$

$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n)$

$(\because (A \cap E_i) \cap (A \cap E_j) = \phi \;\; \forall \; i \neq j)$

(15) $P(A) = P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + \dots + P(E_n)P(A|E_n)$

proved

Q. In a particular campus 25% of boys and 10% of girls studies mathematics. The girl constitute 60% of the total student. If a student is selected at random what is the probability that the student is of studying mathematics.

Sol^n →

Here, Let B denotes boys, G denotes girl and M denotes student studying mathematics. Now,

$$P(B) = 0.4 \qquad P(G) = 0.6 \qquad P(M) = ?$$

$$P(M|B) = 0.25 \qquad P(M|G) = 0.10$$

Since, $G \cup B = S$, $G \cap B = \phi$ and $P(G) \geq 0$, $P(B) \geq 0$

so, $\{G, B\}$ forms the partition.

Now. $M = M \cap S$

$$M = M \cap (G \cup B)$$

$$M = (M \cup G) \cup (M \cap B)$$

$$\therefore P(M) = P(M \cap G) + P(M \cap B)$$

$$P(M) = P(G)\,P(M|G) + P(B)\,P(M|B)$$

$$P(M) = 0.6 \times 0.1 + 0.4 \times 0.25$$

$$= 0.16$$

Baye's Theorem (extension of conditional probability)

Conditional probability is a technique of determining the probability unknown event on the basis of past information or pre-given condition. This concept of conditional probability can be revised on the basis of the New information obtained on the period of experiment. So, the process of revising the conditional probability on the basis of new information is called Baye's theorem.

In Baye's theorem,

Initial or prior probability	New Information	Application of Baye's theorem	Calculation of posterior probability

DATE : __/__/__
PAGE : ______

Statement : Let the set of events $\{E_1, E_2, \ldots, E_n\}$ be the partition of sample space S and $A \subset S$, with $P(A) \neq 0$ be any event of S. Then, the conditional probability of E_r $(r = 1 \ldots n)$ given that A has been already occured is,

$$P(E_r|A) = \frac{P(E_r)\, P(A/E_r)}{P(A)} \quad \text{where } r = 1, 2, 3, \ldots n$$

OR

$$P(E_r|A) = \frac{P(E_r)\, P(A|E_r)}{P(E_1)\, P(A|E_1) + P(E_2)\, P(A|E_2) + \cdots + P(E_n)\, P(A|E_n)}$$

Proof :
Let, S be the ~~partition of~~ sample space of any random experiment. Suppose the set of event $\{E_1, E_2, E_3, \ldots, E_n\}$ forms the partition of S and A be any arbitrary event of S.
Then by total probability theorem,

$$P(A) = P(E_1)\, P(A|E_1) + P(E_2)\, P(A|E_2) + \cdots + P(E_n)\, P(A|E_n)$$

Now,

$$P(E_r|A) = \frac{P(E_r \cap A)}{P(A)}$$

$$= \frac{P(E_r)\, P(E_r|A)}{P(A)} \quad \text{Proved}$$

Q.no.1 Suppose that there are three machines A, B and C which produces 45%, 30%, and 25% of total production of a factory. The percentage of defective outputs produced by these machines respectively 6%, 4% and 3%. An item randomly selected from the total production of the factory and it is found to be defective. What is the probability that the item selected is produced by machine A?

Soln.

Here, Sample Space (S) = total production of factory
Clearly the production of these three machines $\{A, B, C\}$ forms partition of S.

DATE : 2080/06/04
PAGE :

Given, $P(A) = 45\% = 0.45$

$P(B) = 30\% = 0.3$

$P(C) = 25\% = 0.25$

Let D denotes the defective, then

$P(D/A) = 6\% = 0.06$

$P(D/B) = 4\% = 0.04$

$P(D/C) = 3\% = 0.03$

Since,

$$D = D \cap S$$
$$= D \cap (A \cup B \cup C)$$
$$= (D \cap A) \cup (D \cap B) \cup (D \cap C)$$
$$\Rightarrow P(D) = P(D \cap A) + P(D \cap B) + P(D \cap C)$$
$$= P(A)\,P(D/A) + P(B)\,P(D/B) + P(C)\,P(D/C)$$
$$= 0.45 * 0.06 + 0.3 * 0.04 + 0.25 \times 0.03$$
$$= 0.046 \quad \text{; total probability of defective}$$

Now, the probability that the obtained defective is production of machine A is

$$P(A/D) = \frac{P(A)\,P(D/A)}{P(D)}$$
$$= \frac{0.45 * 0.06}{0.046}$$
$$= 0.58$$

Random Variable and probability distribution

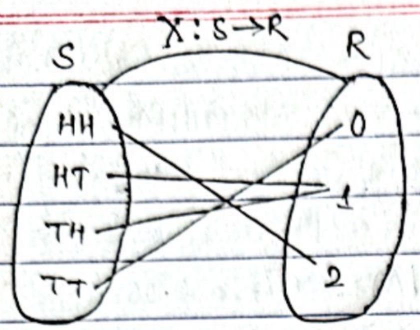
Random Variable

A random variable is a function from the set of sample space of a random experiment i.e. A function $X : S \rightarrow R$, which associates each sample point to a unique real number.

e.g., Suppose a coin is tossed twice and X denotes no. of heads. Then X associates each sample point to a unique real number as, $S = \{HH, HT, TH, TT\}$

$X(HH) = 2$, $X(HT) = 1$, $X(TH) = 1$, $X(TT) = 0$

Scanned with CamScanner



Note: 1. Since a random variable is a function $X : S \to R$. If the values assumed by random variable X are finite or countably infinite then X is called discrete random variable. Otherwise if X takes infinite values or any values lie between two real numbers a and b then it is called continuous random variable.

e.g. Suppose X denotes height of the plant between 80m to 100m in particular region then X can take any value between 80 to 100. So X is continuous random variable.

Suppose S be the set of students of particular class and X denotes their weights between 20 to 30 Kg. Then X can assume any values between 20 and 30. So X is continuous random variable.

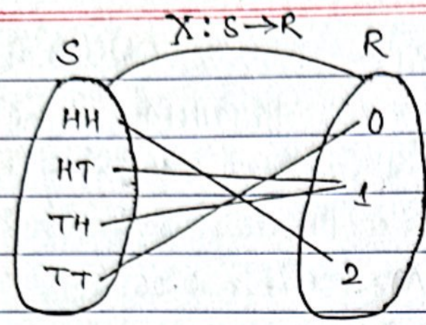
Suppose two dice are tossed together and X denotes sum of the turning faces. Here X can take any of the values $= \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. So X is discrete random variable.

Probability Distribution

A distribution that shows the values taken by the random variable X with their corresponding probabilities is called probability distribution. If X is a discrete random variable then the corresponding probability distribution is called discrete probability distribution and the probability function associated to it is called probability mass function which satisfy following properties:

1. Probability of $P(x=x_i)$ or $f(x=x_i) \geq 0$ for $i = 1, 2, \dots, n$
2. $\sum_{i=1}^{n} P(X=x_i) = 1$

DATE : __/__/__
PAGE : ______



Note: 1. Since a random variable is a function $X:S\to R$. If the values assumed by random variable X are finite or countably infinite then X is called discrete random variable. Otherwise if X takes infinite values or any values lie between two real numbers a and b then it is called continuous random variable.

e.g. Suppose X denotes height of the plant between 80m to 100m in particular region then X can take any value between 80 to 100. So X is continuous random variable.

Suppose S be the set of students of particular class and X denotes their weights between 20 to 30 kg. Then X can assume any values between 20 and 30. So X is continuous random variable.

Suppose two dice are tossed together and X denotes sum of the turning faces. Here X can take any of the values $= \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. So X is discrete random variable.

Probability Distribution

A distribution that shows the values taken by the random variable X with their corresponding probabilities is called probability distribution. If X is a discrete random variable then the corresponding probability distribution is called discrete probability distribution and the probability function associated to it is called probability mass function which satisfy following properties:

1. Probability of $P(x=x_i)$ or $f(x=x_i) \geq 0$ for $i = 1, 2, \ldots, n$
2. $\sum_{i=1}^{n} P(X=X_i) = 1$

Scanned with CamScanner

DATE : / /
PAGE :

Moreover, if X is continuous random variable then the corresponding probability distribution associated to it is called continuous probability distribution and the probability function is called probability density function which satisfy the following properties:

(1) $P(X=x)$ or $f(X=x) \geq 0$ where $a \leq x \leq b$

(2) $\int_a^b f(x)\,dx$ or $\int_a^b P(x)\,dx = 1$

Q. A coin is tossed three times. Find the probability distribution of no. of heads.

Sol^n:

If a coin is tossed thrice then sample space is,

$$S = \{HHH, HHT, HTH, THT, THH, TTH, HTT, TTT\}$$

Let X denotes no. of head, then X ~~denotes~~ can take values $X = 0, 1, 2, 3$.

Now,

$$P(X=0) = \frac{C(3,0)}{2^3} = \frac{1}{8}$$

$$P(X=1) = \frac{C(3,1)}{2^3} = \frac{3}{8}$$

$$P(X=2) = \frac{C(3,2)}{2^3} = \frac{3}{8}$$

$$P(X=3) = \frac{C(3,3)}{2^3} = \frac{1}{8}$$

Now probability distribution of X is,

X	0	1	2	3	Total
$P(X=x)$	1/8	3/8	3/8	1/8	$\sum_{i=1}^{4} P(x_i) = 1$

Q. Four coins are tossed together. If X denotes no. of head. Find probability distribution of X.

Q. Two dice are rolled. If X denotes sum of turning faces. Find probability distribution of X.

DATE : 2080/06/05
PAGE :

Mean and Variance of probability distribution

(i) Mean or Mathematical expectation

Let, X be a random variable then the mathematical expectation or mean of the random variable X is denoted by μ and defined by

$$\mu = \text{Expectation of } X = E(X)$$

where, $\mu = E(X) = \sum_{i=1}^{n} x_i P(x_i)$ (if X is discrete)

OR

$$\mu = E(X) = \int_a^b x P(x)\, dx \quad (a \le x \le b \text{ and } X \text{ is continuous})$$

(ii) Variance of a random variable

Let, X be random variable, then the variance of X is denoted by variance of X i.e. Var(X) and defined by

$$var(X) = E[(X-\mu)]^2$$

where if X is discrete random variable

$$var(X) = E(X-\mu)^2 = E(x_i-\mu)^2 f(x)$$
$$= \Sigma(x^2 - 2\mu x + \mu^2) f(x)$$
$$= \Sigma x^2 f(x) - \Sigma 2\mu x f(x) + \Sigma \mu^2 f(x)$$
$$= \Sigma x^2 f(x) - 2\mu \Sigma x f(x) + \mu^2 \Sigma f(x)$$
$$= E(X^2) - 2\mu E(X) + \mu^2 . 1$$

$$\therefore Var(X) = E(X^2) - 2\mu.\mu + \mu^2$$
$$= E(X^2) - \mu^2$$

Again if X is continuous then, $var(X) = E(X-\mu)^2$

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

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

DATE : ___/___/___
PAGE : ________

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

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

$$= E(X^2) - 2\mu E(X) + \mu^2 \cdot 1$$

$$= E(X^2) - 2\mu \cdot \mu + \mu^2$$

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

$$\therefore \quad Var(X) = E(X^2) - \mu^2$$

Q.1 A coin is tossed 4-times and X denotes number of head. Find probability distribution of X, also find mean and variance of X.

Sol^n:

Here a coin is tossed four times and X denotes no. of head. So X associates each sample point to one of the integers 0, 1, 2, 3, 4.

Now,

$$P(X=0) = \frac{C(4,0)}{16} = \frac{1}{16}$$

$$P(X=1) = \frac{C(4,1)}{16} = \frac{4}{16} = \frac{1}{4}$$

$$P(X=2) = \frac{C(4,2)}{16} = \frac{6}{16} = \frac{3}{8}$$

$$P(X=3) = \frac{C(4,3)}{16} = \frac{4}{16} = \frac{1}{4}$$

$$P(X=4) = \frac{C(4,4)}{16} = \frac{1}{16}$$

Now, probability distribution of X is

X	0	1	2	3	4	Total
$P(X=x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	$\sum_{n=1} P(x_i) = 1$

Again, the mean of random variable,

$$\mu = E(X)$$

or, $$\mu = \sum_{i=1}^{5} x_i P(x_i)$$

$$= x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3) + x_4 P(x_4) + x_5 P(x_4)$$

$$= 0 \times \frac{1}{16} + 1 \times \frac{1}{4} + 2 \times \frac{3}{8} + 3 \times \frac{1}{4} + 4 \times \frac{1}{16}$$

$$= 2$$

$$Var(X) = E(X^2) - \mu^2$$

$$= \sum_{i=1}^{5} x_i^2 P(x_i) - \mu^2$$

$$= x_1^2 P(x_1) + x_2^2 P(x_2) + x_3^2 P(x_3) + x_4^2 P(x_4) + x_5^2 P(x_5) - \mu^2$$

$$= 0^2 \times \frac{1}{16} + 1^2 \times \frac{1}{4} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{4} + 4^2 \times \frac{1}{16} - 2^2$$

$$= \frac{1}{4} + \frac{3}{2} + \frac{9}{4} + 1 - 2^2$$

$$= \frac{20}{4} - 4$$

$$= \frac{4}{4}$$

$$= 1$$

Q.2. Construct probability distribution corresponding to the probability mass function

(i) $f(x) = {}^{2}C_{x}\, {}^{4}C_{3-x}$ for $x = 0, 1, 2$

(ii) $f(x) = {}^{5}C_{x}\, (0.2)^{x}\, (0.4)^{5-x}$ for $x = 0, 1, 2, 3, 4, 5$

also find mean and variance of the distribution.

Q.3. A continuous random variable has probability function,

$$f(x) = \begin{cases} cx & \text{for } 0 \le x \le 1 \\ c(2-x) & \text{for } 1 \le x \le 2 \\ 0 & \text{else} \end{cases}$$

DATE : ___/___/___
PAGE : ______

Again, the mean of random variable,

$$\mu = E(X)$$

or, $$\mu = \sum_{i=1}^{5} x_i P(x_i)$$

$$= x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3) + x_4 P(x_4) + x_5 P(x_4)$$

$$= 0 \times \frac{1}{16} + 1 \times \frac{1}{4} + 2 \times \frac{3}{8} + 3 \times \frac{1}{4} + 4 \times \frac{1}{16}$$

$$= 2$$

$$Var(X) = E(X^2) - \mu^2$$

$$= \sum_{i=1}^{5} x_i^2 P(x_i) - \mu^2$$

$$= x_1^2 P(x_1) + x_2^2 P(x_2) + x_3^2 P(x_3) + x_4^2 P(x_4) + x_5^2 P(x_5) - \mu^2$$

$$= 0^2 \times \frac{1}{16} + 1^2 \times \frac{1}{4} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{4} + 4^2 \times \frac{1}{16} - 2^2$$

$$= \frac{1}{4} + \frac{3}{2} + \frac{9}{4} + 1 - 2^2$$

$$= \frac{20}{4} - 4$$

$$= \frac{4}{4}$$

$$= 1$$

Q.2. Construct probability distribution corresponding to the probability mass function

(i) $f(x) = {}^{2}C_x \, {}^{4}C_{3-x}$ for $x = 0, 1, 2$

(ii) $f(x) = {}^{5}C_x \, (0-2)^x \, (0-4)^{5-x}$ for $x = 0, 1, 2, 3, 4, 5$

also find mean and variance of the distribution.

Q.3. A continuous random variable has probability function,

$$f(x) = \begin{cases} cx & \text{for } 0 \le x \le 1 \\ c(2-x) & \text{for } 1 \le x \le 2 \\ 0 & \text{else} \end{cases}$$

DATE : __/__/__
PAGE : ____

find value of c and $E(X)$, $var(X)$.

Sol^n

Since the given function is probability density function, so $\int_{-\infty}^{\infty} f(x)\,dx$

$$\Rightarrow \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$\Rightarrow 0 + \int_{0}^{1} cx\,dx + \int_{1}^{2} c(2-x)\,dx + 0$$

$$\Rightarrow c\left(\frac{x^2}{2}\right)_0^1 + c\left(2x - \frac{x^2}{2}\right)_1^2 = 1$$

$$\Rightarrow \frac{c}{2} + c\left(4 - \frac{2^2}{2} - 2 + \frac{1}{2}\right) = 1$$

$$\Rightarrow \frac{c}{2} + \frac{c}{2} = 1$$

$$\Rightarrow 2c = 2$$

$$\Rightarrow c = 1$$

Now,

$$\mu = E(X) = \int_{0}^{2} x\,f(x)\,dx$$

$$= \int_{0}^{1} x f(x)\,dx + \int_{1}^{2} x f(x)\,dx$$

$$= \int_{0}^{1} x \cdot x\,dx + \int_{1}^{2} x(2-x)\,dx$$

$$= \left(\frac{x^3}{3}\right)_0^1 + \left(\frac{2x^2}{2} - \frac{x^3}{3}\right)_1^2$$

$$= \frac{1}{3} + \left\{\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right)\right\}$$

$$= \frac{1}{3} + \left\{\frac{4}{3} - \frac{2}{3}\right\} = 1$$

PAGE :

Also, $var(X) = E(X^2) - \mu^2$

$$= \int_0^2 x^2 f(x) - 1^2$$

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

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

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

$$= \left(\frac{x^4}{4}\right)_0^1 + \left(\frac{2x^3}{3} - \frac{x^4}{4}\right)_1^2 - 1$$

$$= \frac{1}{4} + \left(\frac{2\times8}{3} - \frac{16}{4} - \frac{2\times1}{3} + \frac{1}{4}\right) - 1$$

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

Q.4. The probability density function of random variable X, is given by

$$f(x) = \begin{cases} 1/5 & \text{for } 2 \le x \le 7 \\ 0 & \text{else} \end{cases}$$

verify that

(i) Total are under the given function is 1.

(ii) Find $P(3 < x < 5)$.

(iii) Find $P(x \geq 4)$.

Moments of a Random Variable

1. Moments about origin

Let X be a random variable then the r^{th} moment of X about origin is denoted by μ'_r and given by $\mu'_r = E[X^r]$

Note: 1. If X is discrete then $\mu'_r = E[X^r] = \sum x^r f(x)$, where $r = 0, 1, 2, \ldots$

2. If X is continuous random variable, then $\mu'_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x)\,dx$

DATE : ___/___/___

PAGE : ___

2. Moment about arbitrary numbers

The r^{th} moment of any random variable X about arbitrary number A is $\mu'_r = E(X-A)^r$, where, $\mu'_r = \sum (x-A)^r f(x)$ if X is discrete and $\mu'_r = \int_{-\infty}^{\infty} (x-A)^r f(x)\,dx$ if X is continuous.

Note : 1. Since $\mu'_r = E(X^r)$ or $\mu'_r = E(X-A)^r$

Putting $r=0$, $\mu'_0 = E(X^0) = E(1)$ $\begin{cases} \sum 1 * f(x) = 1 \\ \int_{-\infty}^{\infty} 1 \times f(x)\,dx = 1 \end{cases}$

$\therefore \mu'_0 = 1$

2. Putting $r=1$, $\mu'_1 = E(X) = \mu$

3. The moment about arbitrary number is called raw moment and the moment about mean is called central moment.

3. Moment about mean

The r^{th} moment of random variable X about mean is denoted by $\mu_r = E(X-\mu)^r$ where $\mu_r = \sum (X-\mu)^r f(x)$ or $\int_{-\infty}^{\infty} (X-\mu)^r f(x)\,dx$.

when $r=0$, $\mu_0 = E[X-\mu]^0 = E(1) = 1$

when $r=1$, $\mu_1 = E[X-\mu]^1 = \sum (X-\mu) f(x) = 0$

$= \int_{-\infty}^{\infty} (X-\mu) f(x)\,dx = 0$

when $r=2$, $\mu_2 = E(X-\mu)^2 = \sigma^2$ i.e. second central moment is variance.

Again,

$$\sigma^2 = E(X-\mu)^2 = E(X^2) - \mu^2$$

$$= \mu'_2 - \mu^2$$

$$= \mu'_2 - (\mu'_1)^2$$

Moment Generating Function

Let X be a random variable then the moment generating function of X is denoted by $M_x(t)$ and defined by,

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

DATE : ___/___/___
PAGE : ___

where, $M_x(t) = \sum e^{tx} f(x)$

or $M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$

Note: 1. With the help of moment generating function we can easily find mean and variance of the given random variable.

2. The r^{th} raw moment about origin μ_r' is the r^{th} derivative of moment generating function about at $t=0$.

i.e. $\mu_r' = \frac{d^r M_x(t)}{dt^r}$ at time $t=0$

For example: Let X be any random variable. We need to find $E(X)$ and $Var(X)$ then

(i) First we find the moment generating function of X say $M_x(t)$

(ii) $\mu_1' = \left[\frac{d\, M_x(t)}{dt}\right]_{t=0}$ Since, $\mu_1' = \mu$

$\mu_2' = \left[\frac{d^2 M_x(t)}{dt^2}\right]_{t=0}$

But $\sigma^2 = \mu_2' - \mu^2$

$\Rightarrow \mu_2 = \mu_2' - (\mu_1')^2$

Q. A coin is tossed twice, if X denotes head. Find probability distribution of X and also find mean and variance of X.

Soln →

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

Since X denotes number of head, $X(HH) = 2$, $X(TH) = 1$, $X(HT) = 1$, $X(TT) = 0$.

So X takes values 0, 1, 2.

Hence, Probability mass function (pmf) of X is,

X	0	1	2	
P(X)	1/4	2/4	1/4	$\sum P(X) = 1$

Now,

mean $(\mu) = E(X) = \sum_{x=1}^{3} x\, P(x) = x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3)$

DATE : __/__/__
PAGE : ______

where, $M_x(t) = \sum e^{tx} f(x)$

or $M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$

Note: 1. With the help of moment generating function we can easily find mean and variance of the given random variable.

2. The r^{th} raw moment about origin μ_r' is the r^{th} derivative of moment generating function about at $t=0$.

i.e. $\mu_r' = \frac{d^r M_x(t)}{dt^r}$ at time $t=0$

For example: Let X be any random variable. We need to find $E(X)$ and $Var(X)$ then

(i) First we find the moment generating function of X say $M_x(t)$

(ii) $\mu_1' = \left[\frac{d\, M_x(t)}{dt}\right]_{t=0}$ Since, $\mu_1' = \mu$

$\mu_2' = \left[\frac{d^2\, M_x(t)}{dt^2}\right]_{t=0}$

But $\sigma^2 = \mu_2' - \mu^2$

$\Rightarrow \mu_2 = \mu_2' - (\mu_1')^2$

Q. A coin is tossed twice, if X denotes head. Find probability distribution of X and also find mean and variance of X.

Sol^n →

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

Since X denotes number of head, $X(HH) = 2$, $X(TH) = 1$, $X(HT) = 1$, $X(TT) = 0$.

So X takes values $0, 1, 2$.

Hence, Probability mass function (pmf) of X is,

X	0	1	2	
$P(X)$	$1/4$	$2/4$	$1/4$	$\sum P(X) = 1$

Now,

mean $(\mu) = E(X) = \sum_{x=1}^{3} x\, P(x) = x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3)$

DATE : __/__/__
PAGE : ______

$= 0 * \frac{1}{4} + 1 * \frac{2}{4} + 2 * \frac{1}{4}$

$= 1$

And,

$Var(X) = E(X-\mu)^2$ or $E X^2 - \mu^2 = \sum_{i=1}^{3} x_i^2 P(x_i) - \mu^2$

$= \sum_{i=1}^{3} (x_i - \mu)^2 P(x_i)$

$= (x_1-\mu)^2 P(x_1) + (x_2-\mu)^2 P(x_2) + (x_3-\mu)^2 P(x_3)$

$= (0-1)^2 * \frac{1}{4} + (1-1)^2 * \frac{2}{4} + (2-1)^2 * \frac{1}{4}$

$= \frac{1}{4} + \frac{1}{4}$

$= \frac{1}{2}$

Another way.

The moment generating function of X is, $M_x(t) = E(e^{tx})$

$= \sum_{i=1}^{3} e^{tx_i} P(x_i)$

$\therefore M_x(t) = e^{tx_1} P(x_1) + e^{tx_2} P(x_2) + e^{tx_3} P(x_3)$

$= e^{t*0} * \frac{1}{4} + e^{t} * \frac{2}{4} + e^{2t} * \frac{1}{4}$

$= \frac{1}{4} + \frac{1}{2} e^t + \frac{1}{4} e^{2t}$

$= \frac{1}{4} (1 + 2e^t + e^{2t})$

Now,

$\mu_1' = \left[\frac{d}{dt} M_x(t)\right]_{t=0} = \left[\frac{1}{4}(2e^t + 2e^{2t})\right]_{t=0} = 1$

Since,

$\mu_1' = \mu = 1$

$\Rightarrow$ mean $= 1$

Scanned with CamScanner

DATE : 2080/ 06 / 11
PAGE :

Again, $\mu_2' = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0}$

$$= \left[\frac{1}{4}\left(2e^t + 4e^{2t}\right)\right]_{t=0}$$

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

$\therefore$ Variance $(\sigma^2) = \mu_2 = \mu_2' - (\mu_1')^2$

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

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

Binomial Distribution

- Binomial trial or Binomial experiment

A trial is said to be an binomial trial, if it satisfies following properties:

(i) The trial is repeated n-times

(ii) Each trial has only two outcomes one is success (whose probability denoted by p) and another is failure (whose probability is denoted by $q = 1-p$).

(iii) Each trial is independent i.e the probability of success and failure in each trial is constant.

Note: 1. A random variable X satisfying above condition is called binomial random variable.

2. If X is a binomial random variable with p is the probability of success and q is the probability of failure, then probability mass function of X is denoted by

$$b(x;n,p) = c(n,x)\, p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

i.e out of 'n' trial the probability of $X = x$ success is $P(X=x) = b(x;n,p)$.
$= c(n,x)\, p^x q^{n-x}$, $x = 0, 1, 2, \ldots, n$.

Then the corresponding probability distribution of X is,

X	0	1	2	...	n
$P(X=x)$	$c(n,0)p^0q^{n-0}$	$c(n,1)p^1q^{n-1}$	$c(n,2)p^2q^{n-2}$	...	$c(n,n)p^nq^{n-n}$

DATE :__/__/__
PAGE :____

Question Five cards are drawn successively with replacement from pack of 52 cards. Find probability distribution of X where X denotes heart. Also find

(a) probability of exactly 3 hearts

(b) " " atleast " "

(c) " " atmost 4 hearts

Sol.

Here, total trial $= 5 = n$

In each trial probability $p = \frac{13}{52} = \frac{1}{4}$

Probability of failure $q = 1 - \frac{1}{4} = \frac{3}{4}$

Since out of n-trials the probability of x success in binomial distribution (experiment) is,

$$P(X=x) = b(x;n,p) = c(n,x)\, p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

$$\therefore P(X=x) = c(5,x)\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5$$

Now, $P(X=0) = c(5,0)\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{5-0} = 0.2373$

$P(X=1) = c(5,1)\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.3955$

$P(X=2) = c(5,2)\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = 0.2636$

$P(X=3) = c(5,3)\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 = 0.087891$

$P(X=4) = c(5,4)\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1 = 0.014648$

$P(X=5) = c(5,5)\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0 = 0.000977$

Now probability distribution of X is,

X	0	1	2	3	4	5	
$P(X=x)$	0.2373	0.3955	0.2636	0.087891	0.014648	0.000977	$\Sigma P(x) = 1$

DATE : ___/___/___

PAGE : ________

(a) $P(X=3) = 0.007891$

(b) $P(X \geq 3) = P(3) + P(4) + P(5) = 0.103516$

(c) $P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$

- **Moment generating function of binomial distribution**

If X be the binomial ~~distribution~~ random variable then its probability mass function is,

$$P(X=x) = C(n,x)\, p^x q^{n-x}, \quad x = 0, 1, \dots n$$

Now moment generating function of X is

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} P(X=x)$$

$$\text{or, } M_x(t) = \sum_{x=0}^{n} e^{tx}\, C(n,x)\, p^x q^{n-x}$$

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

$$= C(n,0)q^n + C(n,1)(pe^t)^1 q^{n-1} + C(n,2)(pe^t)^2 q^{n-2} + \dots + C(n,n)(pe^t)^n$$

$$= (pe^t + q)^n \quad (\because \text{binomial distribution } (a-x)^n)$$

$$= (pe^t + 1 - p)^n$$

$$= [1 + p(e^t - 1)]^n$$

$$\therefore M_x(t) = [1 + p(e^t - 1)]^n$$

- **Mean and Variance of binomial distribution**

Since, $M_x(t) = [1 + p(e^t - 1)]^n$

Now, differentiating both side w.r.t 't',

$$\frac{d}{dt}(M_x(t)) = n[1 + p(e^t - 1)]^{n-1} * pe^t$$

$$\text{So, } \mu_1' = \left[\frac{d}{dt}(M_x(t))\right]_{at\ t=0}$$

$$\Rightarrow \mu_1' = n[1 + p(e^0 - 1)]^{n-1} * pe^0$$

$$= n[1]^{n-1} p * 1$$

$$= np$$

DATE : __/__/__
PAGE : ______

(a) $P(X=3) = 0.087891$

(b) $P(X \geq 3) = P(3) + P(4) + P(5) = 0.103516$

(c) $P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$

- **Moment generating function of binomial distribution**

If X be the binomial ~~distribution~~ random variable then its probability mass function is,

$$P(X=x) = C(n,x)\, p^x q^{n-x}, \quad x = 0, 1, \dots n$$

Now moment generating function of X is

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} P(X=x)$$

$$\text{or,} \quad M_x(t) = \sum_{x=0}^{n} e^{tx}\, C(n,x)\, p^x q^{n-x}$$

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

$$= C(n,0)\, q^n + C(n,1)(pe^t)^1 q^{n-1} + C(n,2)(pe^t)^2 q^{n-2} + \dots + C(n,n)(pe^t)^n$$

$$= (pe^t + q)^n \quad (\because \text{binomial distribution } (a-x)^n)$$

$$= (pe^t + 1 - p)^n$$

$$= [1 + p(e^t - 1)]^n$$

$$\therefore \quad M_x(t) = [1 + p(e^t - 1)]^n$$

- **Mean and Variance of binomial distribution**

Since, $M_x(t) = [1 + p(e^t - 1)]^n$

Now, differentiating both side w.r.t 't',

$$\frac{d}{dt}(M_x(t)) = n[1 + p(e^t - 1)]^{n-1} * pe^t$$

So, $$\mu_1' = \left[\frac{d}{dt}(M_x(t))\right]_{at\ t=0}$$

$$\Rightarrow \mu_1' = n[1 + p(e^0 - 1)]^{n-1} * pe^0$$

$$= n[1]^{n-1} p * 1$$

$$= np$$

Since, $\mu_1' = \mu = E(X)$

$\therefore$ Mean $= np$

Variance $(\sigma^2) = \mu_2 = \mu_2' - (\mu_1')^2 = npq$

Poisson Distribution

Poisson distribution is the limiting case of binomial distribution. In binomial distribution when no. of trials $n \to \infty$ (i.e 'n' is large) and probability of success $p < 0.5$ i.e $p \to 0$ or $q \to 1$ then poisson distribution provides more accurate approximation than binomial distribution or binomial probability mass of function.

More particularly, the binomial distribution becomes poisson distribution in the following cases:

1. If no. of trials 'n' is indefinitely large i.e. $n \to \infty$.
2. The probability of success in each trial constant but $p \to 0$ i.e. $q \to 1$ or $p < 0.5$.
3. The mean of binomial distribution $np = \lambda$ (say) finite.

Probability Mass Function of Poisson Distribution

Since probability mass function of binomial distribution is denoted by,

$$b(x; n; p) = {}^nC_x p^x q^{n-x} \quad ; \quad x = 0, 1, \dots n$$

Since Poisson distribution is limiting case of binomial distribution, so

$$\lim_{n\to\infty} b(x; n; p) = \lim_{x\to\infty} \frac{n!}{(n-x)!\,x!} p^x q^{n-x} \quad , \quad x = 0, 1, 2, \dots$$

or, $$P(X) = \lim_{n\to\infty} b(x; n; p) = \lim_{n\to\infty} \frac{n(n-1)(n-2)\dots(n-x+1)(n-x)!}{(n-x)!\; x!} p^x q^{n-x}$$

Since, $np = \lambda$ and $q = 1 - p$

$\Rightarrow p = \frac{\lambda}{n}$ $\qquad = 1 - \frac{\lambda}{n}$

$$\therefore \quad P(X) = \lim_{n\to\infty} \frac{n(n-1)(n-2)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

DATE : __/__/__
PAGE : ____

$$\Rightarrow P(X) = \lim_{n\to\infty} \frac{n^x\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!} \frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$\Rightarrow P(X) = \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!} e^{-\lambda} * 1 \qquad \left[\because \left(1-\frac{\lambda}{\infty}\right)^{-x} = (1-0)^{-x} = (1)^{-x} = 1\right]$$

$$\therefore P(X) = \frac{e^{-\lambda}\lambda^x}{x!} \quad ; x = 0, 1, 2, \ldots$$

Hence a discrete random variable x is said to have a Poisson distribution with parameter λ (i.e. mean) >0 if its probability mass function (Pmf) is given by,

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{if } x = 0, 1, 2, \ldots$$

$$= 0 \quad \text{Otherwise} \quad \text{eg. at } x = 1.5$$

Note: 1. Total Probability = 1

Here, $$\sum_{X=0}^{\infty} P(X) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}$$

$$= e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right)$$

$$= e^{-\lambda} e^{\lambda}$$

$$= e^{\lambda-\lambda}$$

$$= e^0$$

$$= 1$$

2. Mean $(\mu) = E(X) = \sum_{x=0}^{\infty} x P(x)$

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

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

$(-*)! = \infty$ $(-1)! = \infty$

$(0)! = 1$

DATE : / /
PAGE :

$$= e^{-\lambda}\left\{\frac{\lambda^1}{0!} + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \frac{\lambda^4}{3!} + \cdots\right\}$$

$$= \lambda e^{-\lambda}\left\{1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right\}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

$$= \lambda$$

$$\therefore \boxed{\mu = \lambda}$$

3. Variance, $Var(X) = E(X^2) - \mu^2$

Now,

$$E(X^2) = \sum_{x=0}^{\infty} x^2 P(x)$$

$$= \sum_{x=0}^{\infty} x(x-1)\,P(x) + \sum_{x=0}^{\infty} x\,P(x)$$

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

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

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

$$= e^{-\lambda}\left\{\frac{\lambda^2}{0!} + \frac{\lambda^3}{1!} + \frac{\lambda^4}{2!} + \cdots\right\} + \lambda$$

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

$$= e^{-\lambda}\cdot\lambda^2\cdot e^{\lambda} + \lambda$$

$$= \lambda^2 + \lambda$$

Now,

$$Var(X) = E(X^2) - \mu^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$= \lambda$, so mean and variance of Poisson distribution is same.

DATE : / /
PAGE :

Moment Generating Function of Poisson Distribution

Since moment generating function is $M_x(t)$, here x is poisson variable

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

$$= \sum_{x=0}^{\infty} e^{tx} P(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!}$$

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

$$= e^{-\lambda}\left\{\frac{(e^t\lambda)^0}{0!} + \frac{(e^t\lambda)^1}{1!} + \frac{(e^t\lambda)^2}{2!} + \cdots\right\}$$

$$= e^{-\lambda}\left\{1 + \frac{e^t\lambda}{1!} + \frac{(e^t\lambda)^2}{2!} + \cdots\right\}$$

$$= e^{-\lambda} e^{e^t\lambda}$$

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

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

Note : 1.

$$\frac{d}{dt}\left[M_x(t)\right] = \frac{d}{dt}\left[e^{\lambda(e^t-1)}\right]$$

$$= e^{\lambda(e^t-1)} \cdot \frac{d}{dt}\left[\lambda(e^t-1)\right]$$

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

$$\therefore \left[\frac{d\, M_x(t)}{dt}\right]_{t=0} = \mu_1' = \mu = \lambda$$

The following are the cases where Poisson distribution is applied:

1. Numbers of phone calls arriving per unit time.
2. The no. of accidents per unit time.
3. The no. of suicides reported at particular time.
4. The no. of death due to heart attack per unit time, etc.

DATE : 2020 / 06 / 17
PAGE :

Q. In a particular location between 2 Pm and 4 Pm, the average no. of phone calls per minute is 2.35. Find the probability that during a particular minute there will be

(i) at most 2 phone calls
(ii) at least 2 phone calls
(iii) exactly 3 phone calls
(iv) at least 3 phone calls

Sol:

Here, if X denotes no. of phone calls at a particular minute, then X follows Poisson distribution where average no. of phone calls per minute $\lambda = 2.35$.

Since in Poisson distribution $P(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$

(iii) $P(X = 3) = \dfrac{e^{-\lambda}\lambda^3}{3!} = \dfrac{e^{-2.35}(2.35)^3}{3!} = 0.20$

(i) $P(X \le 2) = P(0) + P(1) + P(2)$

$$= \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!}$$

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

$$= e^{-2.35}\left\{1 + \frac{2.35}{1!} + \frac{(2.35)^2}{2!}\right\}$$

$= 0.58$

(ii) $P(X > 2) = 1 - P(X \le 2)$

$= 1 - 0.58$

$= 0.41$

(iv)

DATE : __/__/__
PAGE :

Fitting a Binomial Distribution

If random variable consisting of n-trials satisfying condition of binomial distribution repeated N times, then expected frequency of "r" success is given by,

$$f(r) = N * C(n,r)\, p^r q^{n-r}, \quad \text{where } r = 0, 1, 2, --- n$$

e.g. Q. Calculate expected frequency table from following table using binomial law

No. of success (X)	5	4	3	2	1	0
observed frequency (f)	190	500	900	960	500	150

$N = \sum f = 3200$

where probability of success $p = 0.5$.

$n = 5$

$P(X=r) = C(n,r)\, p^r q^{n-r}$

Sol^n →

Expected frequency table :

Success (X)	Observed Frequency (f)	$P(X=x)$	Expected frequency $f(x) = NP(x)$
0	150	0.0312	3200 * 0.0312 = 100
1	500	0.1562	3200 * 0.1562 = 500
2	960	0.3125	3200 * 0.3125 = 1000
3	900	0.3125	3200 * 0.3125 = 1000
4	500	0.1562	3200 * 0.1562 = 500
5	190	0.0312	3200 * 0.0312 = 100

Q. Out of 9000 families 4 children each, how many families do you expect

(i) 2 boys and 2 girls

(ii) at least one boy

(iii) No girls

(iv) 3 boys and 1 girls

(v) at most 2 girls

Sol^n →

Let X denotes no. of boy then probability of success (P) = 0.5

Total families i.e Population size (N) = 9000

Sample size (n) = 4

DATE : __/__/__
PAGE : ______

Since, by binomial distribution, $P(X=x) = C(n,x)\, p^x q^{n-x}$
and expected frequency of x sucess, $f(x) = N\, P(X=x)$
$= N * C(n,x)\, p^x q^{n-x}$

So,

$$P(X=2) = C(4,2)\,(0.5)^2\,(0.5)^2$$
$$= 0.375$$

(i) expected no. of families having 2 boys (or having 2 boys and 2 girls) $= N * P(X=2)$
$= 9000 * 0.375$
$= 3375$

Again,

$$P(X\geq 1) = 1 - P(X=0) = 1 - C(4,0)(0.5)^0(0.5)^4$$
$$= 1 - 0.0625$$
$$= 0.9375$$

(ii) expected frequency having atleast 1 boy $= N * P(X\geq 1)$
$= 9000 * 0.9375$
$= 8437.5$
≈ 8437

(iii) $P(X=4) = C(4,4)\,(0.5)^4\,(0.5)^{4-4}$
$= 0.0625$

$\therefore$ No. of families having all boys or no girl $= 9000 * 0.0625$
$= 562.5$
≈ 562

(iv) $P(X=3) = C(4,3)\,(0.5)^3\,(0.5)^{4-3}$
$= 0.25$

$\therefore$ No. of families having 3 boys and 1 girls $= 9000 * 0.25 = 2250$

(v) $P(X\leq 2) = P(0) + P(1) + P(2) = 0.0625 + 0.25 + 0.375 = 0.6875$

$\therefore$ No. of families having at most 2 girls $= 9000 * 0.6875 = 6187.5$
≈ 6187

DATE : 2020/08 /05
PAGE :

Hypergeometric Distribution

Negative Binomial distribution

A binomial and -ve binomial experiment have exactly same properties except one thing, with binomial experiment we are concerned with finding the probability of 'r' success out of 'x' trials where ($r \leq x$) where x is fixed. However with -ve binomial distribution we are concerned with finding the probability that r^{th} success occurs in x^{th} trial i.e. Here 'r' is fixed but no.of trials are not fixed. The conditions of -ve binomial distribution are as follows:

(i) The experiment consist of 'x' repeated trials.
(ii) The experiment consist of a sequence of repeated independent trials.
(iii) Each trial have exactly two outcomes (s and f)
(iv) The probability of success (p) in each trial is constant.
(v) The experiment continues until required no.of 'r' success have been observed.
(vi) We count the no.of ~~trials~~ success until r^{th} success occurs and 'x' denotes no.of trials until the r^{th} success occurs.

A random variable having above properties is called -ve binomial random variable.

e.g. Let us suppose that a coin is tossed repeatedly until 6 ~~months~~ heads occurs then this is a -ve binomial experiment because,

a) The experiment consist of repeated trials.
b) We toss the coin until 6 heads occurs.
c) Each trial is independent.
d) The process of tossing the coin we continue until the last trial is required no.of success.

Note : If X denotes no.of trials until r^{th} success occurs then the discrete random variable x is said to have -ve binomial distribution or pascal distribution with parameter p and r and the probability mass f^n of

x is,

$$P(X=x) = nb(x; r; p)$$
$$= C(x-1, r-1)\, p^r q^{x-r}$$

where,

$$x = r, r+1, r+2, \ldots$$
$$= 0 \quad \text{otherwise}$$

'OR'

if we count no. of failure before r^{th} success and random variable x denotes no. of failure trials before the last success is obtained then total no. of trials are $x+r$ and the probability of $r-1$ success in $x+r-1$ trials is,

$$C(x+r-1, r-1)\, p^{r-1} q^{(x+r-1)-(r-1)}$$
$$= C(x+r-1, r-1)\, p^{r-1} q^{x}$$

Since the last probability is always success and it is so by law of compound probability 'r' success out of "$x+r$" trials is

$$P(X=x) = C(x+r-1, r-1)\, p^{r-1} q^{x} p$$
$$= C(x+r-1, r-1)\, p^{r} q^{x}$$
$$= C(x+r-1, x)\, p^{r} q^{x}$$

where $x = 0, 1, 2, 3, \ldots$

So if X denotes no. of failure before r^{th} success. Then the discrete random variable 'x' has nb distribution with parameter p and r and its probability mass f^n is,

$$P(X=x) = nb(x; r; p) = \binom{x+r-1}{r-1} p^r q^x \quad \text{where } x = 0, 1, 2, \ldots$$
$$= 0 \quad \text{otherwise}$$

Note: probability mass f^n $\binom{x-1}{r-1} p^r q^{x-r}$ is equivalent with $\binom{x+r-1}{r-1} p^r q^x$

i.e both provide the same result.

DATE : 2024 08 /07
PAGE :

Note: 1. The mathematical expectation i.e. Mean of negative binomial distribution is

$$\mu = E(X) = \frac{r(1-p)}{p}$$

and Variance $(\sigma^2) = \frac{r(1-p)}{p^2}$

Q. A man is throwing a stone in a target, what is the probability that his 10^{th} throw is 5^{th} hit if his probability to strike the target is 0.5. Also find $E(X)$ and $V(X)$.

Sol^n →

Let X be a negative binomial random variable. Suppose X denotes total no. of trials before r^{th} success.

Then, $P(X=x) = \binom{x-1}{r-1} p^r q^{x-r}$; $x = r, r+1, r+2, ---$

Here,

$$x = 10, \; p = 0.5, \; q = 0.5 \; ; \; r = 5$$

$$P(X=x) = \binom{9}{4} (0.5)^5 (0.5)^5$$

$$= 0.123$$

OR

If X denotes failure trials before r^{th} success. Then,

$$x + r = 10$$
$$\Rightarrow x + 5 = 10$$
$$\Rightarrow x = 5$$

So,

$$P(X=x) = \binom{x+r-1}{r-1} p^r q^x = \binom{9}{4} (0.5)^5 (0.5)^5$$

$$= 0.123.$$

DATE : __/__/__
PAGE : ______

Continuous Probability Distribution

Continuous Random Variable

The random variable which can assume not only integral values and as well as decimal values also, then it is called continuous random variable.

OR,

If it is impossible to establish one to one correspondence between Natural number and the values assumed by X then it is called continuous random variable.

Moreover, the continuous random variable can take any value between two limits a and b.

e.g. If X denotes height, weight, length, temperature, pressure, electrical current, etc then X can take any value.

Moreover, a random variable X defined from, R to (0 to ∞) i.e. $f(x): R \to (0,\infty)$ is said to be continuous if its probability is given,

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$

Note: The probability f^n $f(x)$ of continuous random variable x is said to be probability density f^n or simply density f^n if

(i) $f(x) \geq 0 \quad \forall \; -\infty < x < \infty$

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

Normal Distribution (Gaussian distribution or Bell-shaped distribution or Symmetrical distribution)

Out of the different continuous probability distribution, the normal distribution has important role throughout the theory of probability and statistics. The mathematical eq^n of normal distribution depends upon two parameters mean μ and standard deviation (σ) of continuous random variable (X). So, a continuous random variable X is said to

DATE : __/__/__
PAGE : ______

have normal distribution if its probability density f^n is given by.

$$f(x) = c e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{where } -\infty < x < \infty.$$

where μ and σ are arbitrary parameter of random variable 'x' and constant c is determined by using the property of density function i.e. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\text{or } \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} c e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1 \quad \text{—— (I)}$$

Put $\frac{x-\mu}{\sigma} = z$

or $\sigma z = x - \mu$

or $\sigma dz = dx$

Now from (I),

$$c \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \sigma dz$$

$$\Rightarrow c\sigma \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = 1$$

$$\Rightarrow c = \frac{1}{\sigma \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz} \quad \text{—— (II)}$$

Since the integral $\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$ is improper integral called poisson integral

where,

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$$

So, $C = \dfrac{1}{\sigma\sqrt{2\pi}}$

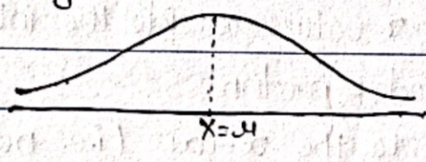
Hence,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

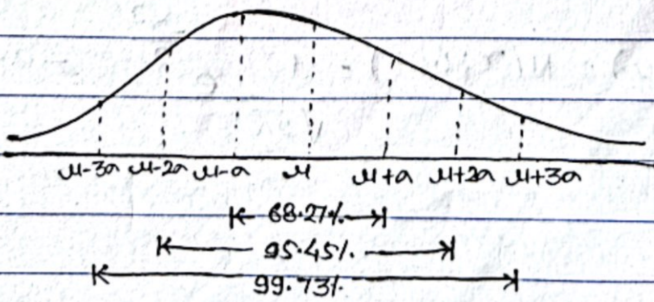
Thus, a continuous random variable X is said to have normal distribution if its pdf is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty<x<\infty$$
$$= 0 \quad \text{otherwise.}$$

Note: 1. The shape of normal distribution is Bell shaped

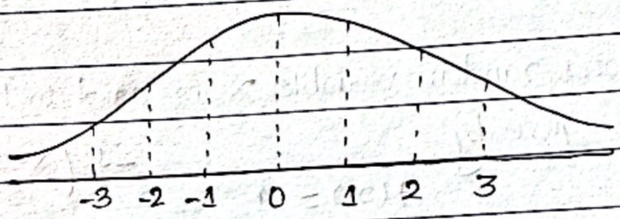


2. $f(x) \geq 0$ for all $x \in R$
3. The area bounded by the curve above x-axis is always 1.
4. The curve is symmetrical (mean = median = mode) about $X=\mu$ i.e. $P(X<\mu) = \frac{1}{2}$ and $P(X>\mu) = \frac{1}{2}$.
5. The normal distribution with mean μ and standard deviation σ is denoted by $N(x;\mu;\sigma)$ i.e. $P(X) = N(x;\mu;\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty<x<\infty$
6. The area under the normal curve is as follows:



Scanned with CamScanner

7. If we take $\mu=0$ and $\sigma=1$ then the normal distribution is called standard normal distribution and denoted by $N(x;0;1)$ and pdf becomes $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty<x<\infty$



8. The standard normal distribution is symmetrical about $z=0$ and is bell shaped and the curve is maximum at $x=\mu$.
9. The curve extends infinitely on either side of mean (μ).
10. The curve is concave downward between $z=-1$ to $z=1$ and concave upward for other values outside the interval $[-1, 1]$.
11. The curve has mean = mode = median.
12. The curve is asymptotic to the x-axis (i.e never intersect x-axis).

Suppose X be a normal variable with mean μ and S.D σ, then pdf of X is

$$P(X=x) = N(x;\mu;\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty<x<\infty$$

There are infinite normal variable among them, then distribution having mean $\mu=0$ and $\sigma=1$ is called standard normal distribution and its pdf is

$$P(X=x) = N(x;0;1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad -\infty<x<\infty$$

DATE : __/__/__
PAGE : ____

If X is a normal variable with mean $\mu = 0$, $\sigma = 1$, then standard normal variable $z = \frac{X-\mu}{\sigma}$ and the pdf of standard normal variable z is

$$P(Z=z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Let X be normal variable with mean μ and S.D σ, then to find $P(X_1 \le X \le X_2)$ is calculated as follows:
Here, the standard normal variable corresponding to x_1 is $Z_1 = \frac{X_1-\mu}{\sigma}$

the standard normal variable corresponding to x_2 is $Z_2 = \frac{X_2-\mu}{\sigma}$

Now $P(X_1 \le X \le X_2) = P(Z_1 \le Z \le Z_2)$

Q. Suppose x be a normal variable with mean 50 and S.D. is 10. Find the probability that, X assumes value between 45 and 62.

Sol:
Here, X be normal variable where $\mu = 50$, $\sigma = 10$ and $x_1 = 45$, $x_2 = 62$.
We need to find $P(x_1 \le x \le x_2)$ or $P(45 \le x \le 62) = ?$
Now, the standard normal variable corresponding to $x_1 = 45$ is $Z_1 = \frac{x_1-\mu}{\sigma}$

$$\Rightarrow Z_1 = \frac{45-50}{10} = -0.5$$

Again, the standard normal variable corresponding to $x_2 = 62$ is,

$$Z_2 = \frac{x_2-\mu}{\sigma}$$

$$= \frac{62-50}{10}$$

$$= 1.2$$

DATE : __/__/__
PAGE : ______

Now, $P(45 \le X \le 62) = P(-0.5 \le Z \le 1.2)$

$= P(0 \le Z \le 0.5) + P(0 \le Z \le 1.2)$

$= 0.1915 + 0.3849$

$= 0.5464$

Q. Find the area between,

(i) $Z=0$ to $Z=0.65$

(ii) $Z=-0.53$ to $Z=0$

(iii) $Z=1.36$ and $Z=1.28$

Soln →

(i) $P(0 \le Z \le 0.65)$

0.6 — 0.05

$= 0.2422$

$P(Z<0.65)$

$= P(Z<0) + P(0 \le Z \le 0.65)$

$= P(Z>0) + P(0 \le Z \le 0.65)$

$= 0.5 + 0.2422$

(ii) $P(-0.53 \le Z \le 0)$

0.5 — 0.03

$= P(0 \le Z \le 0.53)$

$= 0.2019$

(iii) $P(1.28 \le Z \le 1.36)$

1.3 — 0.06

$= P(0 \le Z \le 1.36) - P(0 \le Z \le 1.28)$

1.2 — 0.08

$= 0.4131 - 0.3997$

$= 0.0134$

Q. Suppose X is a normal variable with $\mu = 100$ and $\sigma = 20$. Find,

(i) $P(X<120)$

(ii) $P(X>70)$

(iii) $P(75 < X < 110)$

Soln,

Here, $\mu = 100$, $\sigma = 20$

Suppose Z_1 be the standard normal variable corresponding to $x_1 = 120$ is

$$Z_1 = \frac{120-100}{20} = 1$$

Scanned with CamScanner

DATE : __/__/__

PAGE : ______

Then, $P(X < 120) = P(Z < 1)$

$= P(Z < 0) + P(0 < Z < 1)$

$= 0.5 + 0.3413$

$= 0.8413$

(ii) $P(X > 70) = P\left(Z > \frac{70-100}{20}\right)$

$= P(Z > -1.5)$ OR $P(-1.5 < Z < 0) + P(0 < Z < \infty)$

$= P(Z < 1.5)$ $= P(0 < Z < 1.5) + 0.5$

$= P(Z < 0) + P(0 < Z < 1.5)$

$= 0.5 + 0.4332$

$= 0.9332$

(iii) $P(75 < X < 110) = P\left(\frac{75-100}{20} < X < \frac{110-100}{20}\right)$

$= P(-1.25 < X < 0.5)$

$= P(-1.25 < X < 0) + P(0 < X < 0.5)$

$= P(0 < X < 1.25) + P(0 < X < 0.5)$

$= 0.3944 + 0.1915 = 0.5859$

Note: $P(-\infty < Z < \infty) = P(-\infty < Z < 0) + P(0 < Z < \infty)$

$= P(0 < Z < \infty) + P(0 < Z < \infty)$ (Due to symmetry)

$= 2P(0 < Z < \infty)$

$= 2 \times 0.5$

$= 1$

Moment Generating Function of Normal Distribution

Since the pdf of Normal distribution is,

$$P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

DATE : 2020/08/12
PAGE :

We know, moment generating f^n of any random variable is

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

So,
$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

or,
$$M_x(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{--- (1)}$$

Put $z = \frac{x-\mu}{\sigma}$ $\Rightarrow \sigma z = x-\mu$

$\Rightarrow x = \sigma z + \mu$

Differentiating, $\sigma dz = dx$

Now from (1),
$$M_x(t) = \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)} e^{-\frac{z^2}{2}} \cancel{\sigma} dz$$

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

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

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

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

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

DATE : / /
PAGE :

$$M_x(t) = \frac{e^{t\mu} \cdot e^{\frac{t^2a^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-ta)^2} dz$$

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

Put $z - ta = y$

$\Rightarrow dz = dy$

$$\therefore \quad M_x(t) = \frac{e^{t\mu + \frac{t^2a^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

$$= \frac{e^{t\mu + \frac{t^2a^2}{2}}}{\sqrt{2\pi}} \cdot \sqrt{2\pi}$$

Thus, $\boxed{M_x(t) = e^{t\mu + \frac{t^2a^2}{2}}}$

If $\mu = 0$, $a = 1$, then $M_x(t) = e^{\frac{t^2}{2}}$, which is moment generating function of standard normal distribution.

Mean and Variance of Normal Distribution

Q. Prove that mean and variance of normal distribution are μ and a^2.

Sol^n

Since the moment generating f^n of normal distribution is,

$$M_x(t) = e^{t\mu + \frac{t^2a^2}{2}}$$

Differentiating w.r.t 't',

$$\frac{d}{dx} M_x(t) \text{ or } M_x'(t) = e^{t\mu + \frac{t^2a^2}{2}} \left(\mu + \frac{2ta^2}{2}\right)$$

$$= e^{t\mu + \frac{t^2a^2}{2}} (\mu + ta^2)$$

Now, at $t = 0$, $M_x'(t) = \mu$

$\Rightarrow \mu = \text{mean} \quad (\because \mu_1' = \text{mean})$

Scanned with CamScanner

DATE : __/__/__
PAGE : ______

Again differentiating,

$$M''_x(t) = \frac{d}{dt}\left\{\left(e^{t\mu+\frac{t^2\sigma^2}{2}}\right).(\mu+t\sigma^2)\right\}$$

$$M''_x(t) = e^{t\mu+\frac{t^2\sigma^2}{2}}\frac{d}{dt}(\mu+t\sigma^2) + (\mu+t\sigma^2)\frac{d}{dt}\left(e^{t\mu+\frac{t^2\sigma^2}{2}}\right)$$

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

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

At $t=0$,

$$\left[M''_x(t)\right]_{t=0} = \mu'_2 = \mu^2+\sigma^2$$

Thus, $Var(X) = \mu_2$

$$= \mu'_2 - (\mu'_1)^2$$
$$= \mu^2+\sigma^2-\mu^2$$
$$= \sigma^2$$

Q. Prove that mean and variance of standard normal variable are 0 and 1.

Sol^n →

Since moment generating function of standard normal variable is,

$$M_x(t) = e^{t^2/2}$$

So, mean $(\mu) = \left[M'_x(t)\right]_{t=0} = \left[\frac{d}{dt}e^{\frac{t^2}{2}}\right]_{t=0} = \left[e^{\frac{t^2}{2}}.t\right]_{t=0}$

$\therefore \quad \mu = 0$

Again,

$$M''_x(t) = e^{\frac{t^2}{2}}.1 + t.te^{\frac{t^2}{2}}$$

$$= e^{\frac{t^2}{2}} + t^2e^{\frac{t^2}{2}}$$

$$\Rightarrow \left[M''_x(t)\right]_{t=0} = \mu'_2 = 1$$

Since, $Var(X) = \mu_2 = \mu'_2 - (\mu'_1)^2 = 1-0^2 = 1$

Q. The achievement scores of a college are normally distributed with mean 75 and S.D. 10. Find the probability that the student will achieve the scores between 80 and 90.

Sol?

Let X denotes the scores where mean of X is 75 and S.D. of X is 10.

We need to find $P(80 < X < 90) = P\left(\frac{x_1-\mu}{\sigma} < z < \frac{x_2-\mu}{\sigma}\right)$

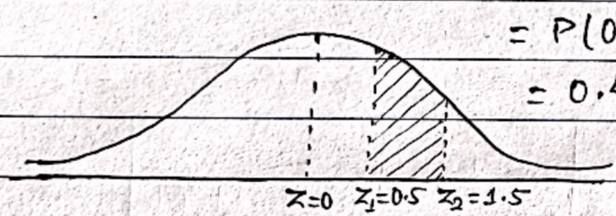
$$= P\left(\frac{80-75}{10} < z < \frac{90-75}{10}\right)$$

$$= P(0.5 < z < 1.5)$$

$$= P(0<z<1.5) - P(0<z<0.5)$$

$$= 0.4332 - 0.1915$$

$$= 0.2417$$



Normal Approximation to Binomial Distribution

DATE : 2081_08/27

PAGE :

Q. The achievement scores of a college are normally distributed with mean 75 and S.D. 10. Find the probability that the student will achieve the scores between 80 and 90.

Sol?

Let X denotes the scores where mean of X is 75 and S.D. of X is 10.

We need to find $P(80 < X < 90) = P\left(\frac{x_1-\mu}{\sigma} < z < \frac{x_2-\mu}{\sigma}\right)$

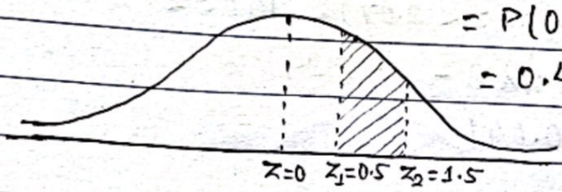
$$= P\left(\frac{80-75}{10} < z < \frac{90-75}{10}\right)$$

$$= P(0.5 < z < 1.5)$$

$$= P(0<z<1.5) - P(0<z<0.5)$$

$$= 0.4332 - 0.1915$$

$$= 0.2417$$



Normal Approximation to Binomial Distribution

The formula for calculating binomial probability of,

$$P(X=x) = C(n,x)\, p^x q^{n-x}, \quad x=0,1,2,\dots n$$

becomes complicated when the no. of trials are so large. So in that case we can approximate the binomial probability by normal distribution.

Let us suppose that a binomial experiment is repeated 15 times with probability of success is $\frac{1}{2}$.

Now out of 15 ~~trials~~ the probability of 4 success by binomial pmf is,

$$P(X=4) = C(15,4)\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{11}$$

$$= 0.041$$

We now calculate same propability by normal distribution,

Since, $\mu = np = 15 * \frac{1}{2} = 7.5$

Scanned with CamScanner

DATE : __/__/__
PAGE : ______

$$\sigma = \sqrt{npq} = \sqrt{15 \times \frac{1}{2} \times \frac{1}{2}} = 1.93$$

Here mean of random variable is 7.5 and S.D. is 1.93. We now convert the binomial random variable into standard normal variable as

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{3.5 - 7.5}{1.93}$$

$$= -2.07$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{4.5 - 7.5}{1.93} = -1.55$$

Now,

$$P(X=4) = P(z_1 \le X \le z_2)$$
$$= P(-2.07 \le z \le -1.55)$$
$$= P(0<z<2.07) - P(0<z<1.55)$$
$$= 0.4808 - 0.4394$$
$$= 0.041$$

So,

$$P(X=4) \approx P(-2.07 < z < -1.55)$$

Note: Continuity correction to improve the approximation:

The addition of 0.5 and or subtraction of 0.5 from the values from the random variable x, when the continuous ~~approximation~~ distribution used as approximate to discrete distribution is called continuity correction and the factor 0.5 is called continuity correction factor.

	Binomial Variable	Approximation by Normal Variable
1.	$P(X=10)$	$P(9.5 < X < 10.5)$
2.	$P(X<10)$	$P(X \le 9.5)$
3.	$P(X \le 10)$	$P(X < 10.5)$
4.	$P(X>10)$	$P(X \geq 10.5)$
5.	$P(X \geq 10)$	$P(X > 9.5)$
6.	$P(10 \le X \le 20)$	$P(9.5 \le X \le 20.5)$
7.	$P(10 < X < 20)$	$P(10.5 < X < 19.5)$

DATE : / /
PAGE :

Q. A dice is rolled 120 times, find the probability that 4 will turn 18 times or less.

Sol→

Here, no. of trials $(n) = 120$

Probability of success $(p) = \frac{1}{6}$

Probability of failure $(q) = \frac{5}{6}$

We need to find $P(X \leq 18)$.

Since by Normal Approximation to binomial distribution,

$$P(X \leq 18) \approx P(X \leq 18.5)$$

i.e. $P(X \leq 18) = P(X < 18.5)$ — ①

Since, mean $(\mu) = 120 \times \frac{1}{6} = 20$

$$\sigma = \sqrt{120 \times \frac{1}{6} \times \frac{5}{6}} = 4.082$$

$\therefore$ $P(X \leq 18) = P\left(Z < \frac{18.5 - 20}{4.08}\right)$

$= P(Z < -0.3604)$

$= P(Z < -0.37)$

$= 0.5 - P(0 \leq Z \leq 0.37)$

$= 0.5 - 0.1443$

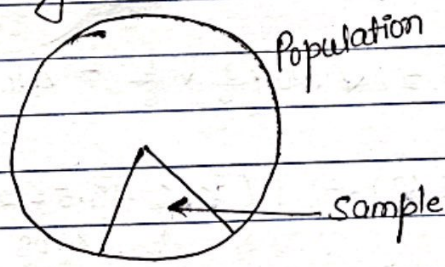
$= 0.3557$

Sampling Distribution

Population

The totality of study or the universe of the study is called the population. The statistical values of the population are called parameters such as mean is denoted by μ, population S.D. is denoted by σ, population proportion is denoted by P, etc. population size denoted by N. In any research it is generally impossible to investigate on whole population. So we need to take a small representative set of data from the population (called sample) and with the help of statistical technique we analyse the sample data and the information obtained from it is tried to generalize on the whole population.

So, sample is a small representitive set of data drawn from the population for investigation.



The statistical values of sample are called statistic where the sample mean is denoted by $\bar{x}$, sample S.D. is denoted by S, and sample proportion is denoted by $\hat{P}$, etc. Sample size dendted by 'n'.

Sampling Distribution

Let population size is 'N'. Suppose from the population samples of 'n' size are drawn (either with replacement or without replacement).
Suppose there are 'K' samples of size 'n' from the population.
Let us calculate, sample statistics $\hat{\theta}$ of each K samples. Then the value of $\hat{\theta}$ with its probability is called sampling distribution of sample statistics $\hat{\theta}$.

DATE : __/__/__
PAGE :

Thus, if $\hat{\theta}$ is sample statistic, then the probability distribution of $\hat{\theta}$ is called sampling distribution of $\hat{\theta}$.

Note : 1. If samples of size 'n' are drawn from the population of size 'N', with replacement then total no. of samples are N^n.

2. If samples of size 'n' are drawn from the population of size 'N' without replacement then total no. of samples are $C(N, n)$.

Question If population is 1, 2, 3, 4 then
(i) Find all samples of size 2 without replacement
(ii) Find sampling distribution of sample mean
(iii) Prove that sample mean is equal to population mean

Sol^n →
Here, Population is 1, 2, 3, 4,
Population size, $N = 4$
We need to draw samples of size $n = 2$
So total samples are $C(4, 2) = 6$
which are, $\{1,2\}$ $\{1,3\}$ $\{1,4\}$ $\{2,3\}$ $\{2,4\}$ $\{3,4\}$
Now,
Sampling Distribution of sample mean :

Samples	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$
$\bar{x}$	1.5	2	2.5	2.5	3	3.5

Mean ($\bar{x}$)	Frequency (f)	$P(\bar{x})$
1.5	1	1/6
2	1	1/6
2.5	2	2/6
3	1	1/6
3.5	1	1/6

DATE : __/__/__
PAGE : ______

Now, mean of sampling distribution, sample Mean = $E(\bar{X})$ or $\bar{\bar{X}}$

$$\Rightarrow E(\bar{X}) = 1.5 * \frac{1}{6} + 2 * \frac{1}{6} + 2.5 * \frac{2}{6} + 3 * \frac{1}{6} + 3.5 * \frac{1}{6}$$

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

$$= 2.5$$

Again,

Population Mean $(\mu) = \frac{1+2+3+4}{4} = 2.5$

Population S.D. $(\sigma) = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$

$$= \sqrt{\frac{(1-2.5)^2+(2-3.5)^2+(3-2.5)^2+(4-2.5)^2}{4}}$$

$$= 1.118$$

~~Sample S.D.~~ Variance $(\sigma_{\bar{x}}^2) = E(\bar{X}-\mu)^2$

$$= \Sigma(\bar{x}-\mu)^2 * P(\bar{x})$$

$$= (1.5-2.5)^2 * \frac{1}{6} + (2-2.5)^2 * \frac{1}{6} + (2.5-2.5)^2 * \frac{2}{6} + (3-2.5)^2 * \frac{1}{6} + (3.5-2.5)^2 * \frac{1}{6}$$

$= 0.417$, S.D $(\sigma_{\bar{x}}) = \sqrt{0.417} = 0.645$

Here, $\sigma \neq \sigma_{\bar{x}}$

But $\cancel{\frac{\sigma}{\sqrt{n}} = \frac{1.11}{\sqrt{2}} = 0.78}$ $\quad \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.118}{\sqrt{2}}\sqrt{\frac{4-2}{4-1}} = 0.645$

$\Rightarrow$ standard deviation of sample S.D (called standard error) $= \cancel{\frac{\sigma}{\sqrt{n}}}$

$$= \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

DATE: 2080/09/03
PAGE:

Sampling Distribution of sample mean with replacement

Q. If population is 1, 2, 5, draw all possible samples of size 2 with replacement from the population and also,
(i) Find the sampling Distribution of sample mean
(ii) standard error of sampling distribution sample
(iii) Show that Sample Mean is equal to Population Mean.

Sol^n→

Here, given population 1, 2, 5.
$N = 3$
Sample size $(n) = 2$

So, all possible samples of size 2 are $N^n = 3^2 = 9$ which are
{1,1}, {1,2} {1,5} {2,1} {2,2} {2,5} {5,1} {5,2} {5,5}.

Now, sampling Distribution of sample mean

Samples	$\bar{x}$	$P(\bar{x})$
{1,1}	1	1/9
{1,2}	1.5	1/9
{1,5}	3	1/9
{2,1}	1.5	1/9
{2,2}	2	1/9
{2,5}	3.5	1/9
{5,1}	3	1/9
{5,2}	3.5	1/9
{5,5}	5	1/9

i.e.

$\bar{x}$	f	$P(\bar{x})$
1	1	1/9
1.5	2	2/9
2	1	1/9
3	2	2/9
3.5	2	2/9
5	1	1/9

Note: Standard error of any sampling distribution of sample statistics $\hat{\theta}$ is its standard Deviation i.e. $SE(\hat{\theta})$ = S.D. of sampling distribution of $\hat{\theta}$

Now,

Mean of Sampling Distribution of Sample Mean.

$$\mu_{\bar{x}} = E(\bar{x})$$

$$= \sum \bar{x} P(\bar{x})$$

DATE : __/__/__
PAGE : ____

$= 1 * \frac{1}{9} + 1.5 * \frac{2}{9} + 2 * \frac{1}{9} + 3 * \frac{2}{9} + 3.5 * \frac{2}{9} + 5 * \frac{1}{9}$

$= 2.67$

Population Mean $(\mu) = \frac{2+1+5}{3} = 2.67$

$\therefore \boxed{\mu_{\bar{x}} = \mu}$

Universal truth :- Mean of Sampling Distribution of Sample Mean is equal to population mean (whatever be the process of drawing samples and whatever be the size of sample).

~~Standard Deviation~~ Variance $(\sigma^2_{\bar{x}}) = E(\bar{x} - \mu)^2$

$= \Sigma (\bar{x} - \mu)^2 P(\bar{x})$

$= (1-2.67)^2 * \frac{1}{9} + (1.5-2.67)^2 * \frac{2}{9} + (2-2.67)^2 * \frac{1}{9} + (3-2.67)^2 * \frac{2}{9} +$

$(3.5-2.67)^2 * \frac{2}{9} + (5-2.67)^2 * \frac{1}{9}$

$= 1.42$

$\therefore S.D. (\sigma_{\bar{x}}) = \sqrt{1.42} = 1.19$

Population S.D. $(\sigma) = \sqrt{\frac{\Sigma (x-\mu)^2}{N}}$

$= \sqrt{\frac{(1-2.67)^2 + (2-2.67)^2 + (5-2.67)^2}{3}}$

$= 1.69$

Here, $\boxed{\sigma_{\bar{x}} \neq \sigma}$

Scanned with CamScanner

Now, $\frac{\sigma}{\sqrt{n}} = \frac{1.69}{\sqrt{2}} = 1.19$

∴ If μ be the mean of population and σ be S.D. population then Sampling distribution of sample mean with replacement is,

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

Note: 1. If Samples are drawn without replacement from finite population, then

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

2. When $N \xrightarrow{\text{tends to}} \infty$, $\sqrt{\frac{N-n}{N-1}} \rightarrow 1$ then $S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$

If the samples are drawn from infinite population, in that case

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$ (whether the samples are drawn with replacement or without replacement)

Sampling Distribution of Sample Proportion

DATE : 2020/09/05
PAGE :

Now, $\frac{\sigma}{\sqrt{n}} = \frac{1.69}{\sqrt{2}} = 1.19$

$\therefore$ If μ be the mean of population and σ be S.D. population then Sampling distribution of sample mean with replacement is,

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

Note: 1. If Samples are drawn without replacement from finite population, then

$$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

2. When $N \xrightarrow{\text{tends to}} \infty$, $\sqrt{\frac{N-n}{N-1}} \rightarrow 1$ then $S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$

If the samples are drawn from infinite population, in that case $S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ (whether the samples are drawn with replacement or without replacement)

Sampling Distribution of Sample Proportion

Let size of the population be N and suppose K samples of size n are drawn from the population (either with replacement or without replacement) then population proportion is denoted by

$P = \frac{x}{N}$, where x is the no. of success or no. of interests.

Let us calculate the sample proportion of each sample $\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_K$ where $\hat{P}_i = \frac{x_i}{n}$, where x_i is no. of success in i^{th} sample then the probability distribution of $\hat{P}$ is called sampling distribution of sample proportion.

DATE: 2080/09/05

PAGE:

Now, $\frac{\sigma}{\sqrt{n}} = \frac{1.69}{\sqrt{2}} = 1.19$

$\therefore$ If μ be the mean of population and σ be S.D. population then sampling distribution of sample mean with replacement is,

$$S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Note: 1. If samples are drawn without replacement from finite population, then

$$S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

2. When $N \xrightarrow{\text{tends to}} \infty$, $\sqrt{\frac{N-n}{N-1}} \to 1$ then $S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

If the samples are drawn from infinite population, in that case

$$S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$ (whether the samples are drawn with replacement or without replacement)

Sampling Distribution of Sample Proportion

Let size of the population be N and suppose K samples of size n are drawn from the population (either with replacement or without replacement) then population proportion is denoted by

$$P = \frac{x}{N}$$, where x is the no. of success or no. of interests.

Let us calculate the sample proportion of each sample $\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_K$ where $\hat{P}_i = \frac{x_i}{n}$, where x_i is no. of success in i^{th} sample then the probability distribution of $\hat{P}$ is called sampling distribution of sample proportion.

Prime : 2, 3, 5, ...

DATE : __/__/__

PAGE : ______

Q. If the population is {1, 2, 5}. Find the sampling distribution of sample proportion of prime number with replacement of sample size 2. Also find mean of sampling distribution of sample proportion and show that $\mu_{\hat{p}} = P$. Find standard error of sampling distribution of sample proportion.

Sol?

Here, Population is 1, 2, 5.

No. of success $x = 2$

So, population proportion $(P) = \frac{2}{3}$.

Now, all possible samples of size 2 with replacement are

{1, 1}, {1, 2}, {1, 5}, {2, 1}, {2, 2}, {2, 5}, {5, 1}, {5, 2}, {5, 5}

Now sampling distribution of sample proportion of prime is

Samples	Sample proportion ($\hat{p}$)	$P(\hat{p})$
{1, 1}	1	1/9
{1, 2}	1/2	1/9
{1, 5}	1	1/9
{2, 1}	1/2	1/9
{2, 2}	0	1/9
{2, 5}	1/2	1/9
{5, 1}	1	1/9
{5, 2}	1/2	1/9
{5, 5}	1	1/9

Now, mean of sampling distribution of sample proportion is

$$\mu_{\hat{p}} = E(\hat{p})$$

$$= \Sigma\, \hat{p}.P(\hat{p})$$

$$= 0 * \frac{1}{9} + 0.5 * \frac{4}{9} + 1 * \frac{4}{9}$$

$$= \frac{2}{3} = P$$

DATE : __/__/__
PAGE : ______

$\therefore$ $\boxed{\mu_{\hat{p}} = P \text{ or } E(\hat{p}) = P}$

Now,

Standard deviation of sampling distribution of sample proportion,

$$Var(\hat{p}) \text{ or } \sigma^2_{\hat{p}} = E(\hat{p} - \mu_{\hat{p}})^2$$

$$= \sum (\hat{p} - \mu_{\hat{p}})^2 \cdot P(\hat{p})$$

$$= \left(0 - \frac{2}{3}\right)^2 * \frac{1}{9} + \left(0.5 - \frac{2}{3}\right)^2 * \frac{4}{9} + \left(1 - \frac{2}{3}\right)^2 * \frac{4}{9}$$

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

So, S.E of $\hat{p}$ i.e. $S.E(\hat{p})$ or $\sigma_{\hat{p}} = \frac{1}{3}$ (i.e $\sqrt{\frac{1}{9}}$)

Again, $\frac{pq}{n} = \frac{\frac{2}{3} * \frac{1}{3}}{2} = \frac{1}{9}$

$$\therefore \sqrt{\frac{pq}{n}} = \frac{1}{3} = \sigma_{\hat{p}}$$

So,

If samples are drawn with replacement or from infinite population, then

$$\mu_{\hat{p}} = P$$

$\Rightarrow$ Mean of sampling distribution of sample proportion = Population proportion

And

Standard Deviation or Standard Error of sampling distribution of sample proportion with replacement is

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

$$\Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

DATE : __/__/__

PAGE : ____

Q. Suppose population consist of 1, 2, 3, 4, 5. Find the Sampling distribution of sample proportion of odd when samples are drawn without replacement of size 2.

Sol^n:

Here, Population is 1, 2, 3, 4, 5.

$N = 5$

No. of success, $x = 3$

$\therefore$ Population proportion $(P) = \frac{3}{5}$

Total samples of size without replacement $= C(N, n) = C(5, 2)$
$= 10$

Now,

Sampling Distribution of ~~sampling distribution~~ sample proportion of odd without replacement of size 2 are,

{1, 2} {1, 3} {1, 4} {1, 5} {2, 3} {2, 4} {2, 5} {3, 4} {3, 5} {4, 5}

Samples	Sample proportion ($\hat{P}$)	$P(\hat{P})$
{1, 2}	1/2	1/10
{1, 3}	1	1/10
{1, 4}	1/2	1/10
{1, 5}	1	1/10
{2, 3}	1/2	1/10
{2, 4}	0	1/10
{2, 5}	1/2	1/10
{3, 4}	1/2	1/10
{3, 5}	1	1/10
{4, 5}	1/2	1/10

Now $\mu_{\hat{p}} = E(\hat{P})$

$= \Sigma \hat{P} . P(\hat{P})$

$= 0 * \frac{1}{10} + 0.5 * \frac{6}{10} + 1 * \frac{3}{10}$

Note:

Scanned with CamScanner

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

$$\therefore \boxed{\mu_{\hat{p}} = P}$$

Now,

$$\text{Variance } (\sigma^2_{\hat{p}}) = E(\hat{p} - \mu_{\hat{p}})^2$$

$$= \sum (\hat{p} - \mu_{\hat{p}})^2 * P(\hat{p})$$

$$= \left(0 - \frac{3}{5}\right)^2 * \frac{1}{10} + \left(0.5 - \frac{3}{5}\right)^2 * \frac{6}{10} + \left(1 - \frac{3}{5}\right)^2 * \frac{3}{10}$$

$$= \frac{9}{100}$$

$$\therefore \sigma_{\hat{p}} = \sqrt{\frac{9}{100}} = \frac{3}{10}$$

Now,

$$\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{\frac{3}{5} * \frac{2}{5}}{2}}\sqrt{\frac{5-2}{5-1}}.$$

$$= \sqrt{\frac{3}{25} * \frac{3}{4}}$$

$$= \frac{3}{\cancel{\chi}\,10}$$

$$\therefore \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)}$$ is the standard error when samples are drawn without replacement.

Note: In the case of infinite population i.e. when $N \to \infty$, both standard error of sampling distribution of sample proportion (with replacement or without replacement) approximately same to $\sqrt{\frac{pq}{n}}$.

DATE : __/__/__
PAGE : ______

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

$$\therefore \boxed{\mu_{\hat{p}} = P}$$

Now,

$$\text{Variance } (\sigma^2_{\hat{p}}) = E(\hat{p} - \mu_{\hat{p}})^2$$

$$= \sum (\hat{p} - \mu_{\hat{p}})^2 * P(\hat{p})$$

$$= \left(0 - \frac{3}{5}\right)^2 * \frac{1}{10} + \left(0.5 - \frac{3}{5}\right)^2 * \frac{6}{10} + \left(1 - \frac{3}{5}\right)^2 * \frac{3}{10}$$

$$= \frac{9}{100}$$

$$\therefore \sigma_{\hat{p}} = \sqrt{\frac{9}{100}} = \frac{3}{10}$$

Now,

$$\sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{\frac{3}{5} * \frac{2}{5}}{2}} \sqrt{\frac{5-2}{5-1}}$$

$$= \sqrt{\frac{3}{25} * \frac{3}{4}}$$

$$= \frac{3}{10}$$

$$\therefore \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)}$$ is the standard error when samples are drawn without replacement.

Note: In the case of infinite population i.e. when $N \to \infty$, both standard error of sampling distribution of sample proportion (with replacement or without replacement) approximately same to $\sqrt{\frac{pq}{n}}$.

DATE : 2080/09/06
PAGE :

Sampling Distribution of difference of two Sample Mean

Let there are two populations with sample size N_1 and N_2 with means and S.D μ_1, μ_2 and σ_1, σ_2 respectively. Let $\bar{x}_1$ be the mean of random samples of size n_1 drawn from the first population and $\bar{x}_2$ be the mean of random samples of size n_2 drawn from the second population then the probability distridution of $(\bar{x}_1 - \bar{x}_2)$ between two set of independent sample means is called sampling distribution of difference of two sample means.

Q. Let population 1st consist $\{3, 5, 7\}$ and population 2nd consist $\{0, 3\}$. Suppose from the population 1st samples of size 2 and from the population 2nd samples of size 3 are drawn with replacement. Find,

(i) Sampling Distribution of difference of two sample means.

(ii) Find expectation of Sampling Distribution of difference of sample means

(iii) Standard error of sampling distribution of difference of sample mean

Soln→

Here, 1st population is $\{3, 5, 7\}$

$$\Rightarrow \mu_1 = 5$$

and $$\sigma_1^2 = \frac{(3-5)^2 + (5-5)^2 + (7-5)^2}{3} = \frac{8}{3}$$

$$\Rightarrow \sigma_1 = \sqrt{\frac{8}{3}} = 1.63$$

Again,

Second population is $\{0, 3\}$

$$\Rightarrow \mu_2 = \frac{3}{2}$$

and $$\sigma_2^2 = \frac{\left(0-\frac{3}{2}\right)^2 + \left(3-\frac{3}{2}\right)^2}{2} = \cancel{\frac{15}{4}}\ \frac{9}{4}$$

$$\Rightarrow \sigma_2 = \sqrt{\frac{9}{4}} = \frac{3}{2} = 1.5$$

DATE : __/__/__
PAGE :

Now, mean of random samples drawn from the population,

Population 1

No.	Samples	$\bar{x}_1$
1	{3,3}	3
2	{3,5}	4
3	{3,7}	5
4	{5,3}	4
5	{5,5}	5
6	{5,7}	6
7	{7,3}	5
8	{7,5}	6
9	{7,7}	7

Population 2

No.	Samples	$\bar{x}_2$
1	{0,0,0}	0
2	{0,0,3}	1
3	{0,3,0}	1
4	{3,0,0}	1
5	{0,3,3}	2
6	{3,0,3}	2
7	{3,3,0}	2
8	{3,3,3}	3

Difference of Independent samples,

$\bar{x}_2$ \ $\bar{x}_1$	3	4	5	4	5	6	5	6	7
0	3	4	5	4	5	6	5	6	7
1	2	3	4	3	4	5	4	5	6
1	2	3	4	3	4	5	4	5	6
1	2	3	4	3	4	5	4	5	6
2	1	2	3	2	3	4	3	4	5
2	1	2	3	2	3	4	3	4	5
2	1	2	3	2	3	4	3	4	5
3	0	1	2	1	2	3	2	3	4

Sampling distribution of difference of two samples.

$\bar{x}_1 - \bar{x}_2$	f	$P(\bar{x}_1 - \bar{x}_2)$
0	1	1/72
1	5	5/72
2	12	12/72
3	18	18/72
4	18	18/72
5	12	12/72
6	5	5/72
7	1	1/72

DATE : __/__/__
PAGE : ____

Now, $E(\bar{x}_1 - \bar{x}_2)$ or $\mu_{\bar{x}_1 - \bar{x}_2} = \sum (\bar{x}_1 - \bar{x}_2)\, P(\bar{x}_1 - \bar{x}_2)$

$\Rightarrow 0 * \frac{1}{72} + 1 * \frac{5}{72} + 2 * \frac{12}{72} + 3 * \frac{18}{72} + 4 * \frac{18}{72} + 5 * \frac{12}{72} + 6 * \frac{5}{72} + 7 * \frac{1}{72}$

$= 3.5$

Now, $\mu_1 - \mu_2 = 5 - \frac{3}{2} = 3.5$

$\therefore \boxed{\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2}$

Thus,

Mean of Sampling Distribution of difference of two sample means is equal to the difference of their corresponding population mean.

Now,

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \sum \{(\bar{x}_1 - \bar{x}_2) - (\mu_{\bar{x}_1 - \bar{x}_2})\}^2\, P(\bar{x}_1 - \bar{x}_2)$$

$$= (0 - 3.5)^2 * \frac{1}{72} + (1 - 3.5)^2 * \frac{5}{72} + (2 - 3.5)^2 * \frac{12}{72} + (3 - 3.5)^2 * \frac{18}{72} + (4 - 3.5)^2 * \frac{18}{72} + (5 - 3.5)^2 * \frac{12}{72} + (6 - 3.5)^2 * \frac{5}{72} + (7 - 3.5)^2 * \frac{1}{72}$$

$= \cancel{\frac{7}{3}}\ 2.083$

$\therefore \sigma_{x_1 - x_2} = \sqrt{\cancel{\frac{7}{3}}} = \cancel{0.81}\ \sqrt{2.083} = 1.443$

Now,

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{1.63^2}{2} + \frac{1.5^2}{3}} = 1.442$$

So, S.E. of $(\bar{x}_1 - \bar{x}_2)$ i.e. $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$

Note : If Samples are drawn from finite population or without replacement, then

(i) $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

(ii) $\sigma_{\bar{x}_1-\bar{x}_2} \approx \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cancel{\sqrt{\frac{N-n}{N-1}}}$

Sampling Distribution of difference of two sample proportion

Q. Suppose the first population consist of $\{4, 6, 9, 11\}$ and second population consist of $\{2, 3, 5, 7\}$. If samples of size 2 from both population are drawn without replacement find the sampling distribution of difference of two sample proportions of even. Also find

(i) Expectation of Sampling Distribution of $(\hat{P}_1 - \hat{P}_2)$ and show that
$$\mu_{\hat{P}_1-\hat{P}_2} = P_1 - P_2 .$$

(ii) Standard error of $\hat{P}_1 - \hat{P}_2$.

Scanned with CamScanner

DATE: 2080/09/17
PAGE:

Central limit theorem

Suppose, random samples of size n are drawn from, infinite or large population (the population may or maynot be normal). Then the sampling distribution of sample statistics $\hat{\theta}$ follows, approximately the standard normal distribution, where

$$\text{Standard Normal Variable, } Z = \frac{\hat{\theta} - E(\hat{\theta})}{S.E.(\hat{\theta})}$$

The approximation is better when sample size $n \geq 30$.

In particular.

(i) If random sample of size 'n' are drawn from an infinite population with mean μ and variance σ^2, then the sampling distribution of sample mean i.e $\bar{X}$ has normal distribution with mean

$$\mu_{\bar{X}} = \mu \text{ and } S.D. = \frac{\sigma}{\sqrt{n}}, \text{ Variance} = \frac{\sigma^2}{n}$$

Then Sampling distribution of sample mean approximately follows the standard normal distribution, where

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{S.E(\hat{\theta})}$$

$$\text{i.e } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

and moreover the approximation is better and better when $n \to \infty$.

Q. An electrical firm manufactures light bulbs that have length of light is approximately normally distributed with mean 800 hrs and S.D. 40 hrs. Find the probability that the random sample of 16 bulbs will have average life time less than 775 hrs.

Soln →

Here, The population is normally distributed with mean $(\mu) = 800$ hrs and S.D. $(\sigma) = 40$ hrs

Scanned with CamScanner

DATE : __/__/__
PAGE : ____

Sample size, $n = 16$

We know that, $\mu_{\bar{x}} = \mu = 800$

$$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{16}} = 10$$

We need to find $P(\bar{x} \leq 775)$.

We now change the variable $\bar{x}$ into corresponding standard normal variable z.

So by Central Limit theorem,

$$z = \frac{\bar{x} - \mu}{S.E(\bar{x})} = \frac{775 - 800}{10} = -2.5$$

$$\therefore P(\bar{x} \leq 775) \approx P(z < -2.5)$$
$$\approx 0.5 - P(0 < z < 2.5)$$
$$\approx 0.5 - 0.4938$$
$$\approx 0.062$$

Note: 1. The meaning of central limit theorem can be used sampling distribution of sample proportion, sampling distribution of difference of two sample mean, sampling distribution of difference of two sample proportion.

2. Suppose samples of size 'n' are drawn from a large population with population proportion 'P' then the sampling distribution of sample proportion follows the normal distribution with mean $\mu_{\hat{p}} = P$ where corresponding standard normal variable,

$$z = \frac{\hat{P} - P}{\sqrt{\frac{pq}{n}}}$$

Estimation (Guessing or Forcasting)

When we are not sure about something we need to estimate or guess on the basis of some previously obtained information. In Statistics estimation is the part of inferential statistics, where

DATE : __/__/__
PAGE :

we make the inferences about the unknown population on the basis of sample information, so estimation means a process in which we obtain the value of unknown population parameter with the help of sample data or sample information.

There are two kinds of estimation :

1. Point Estimation

When an estimate for unknown population parameter is expressed by a single value then it is called point estimation. For example, when we find estimate for population mean (μ), we use sample mean ($\bar{x}$), let $\bar{x} = 10$ is the numeric value of estimator which is called the estimate and here estimate is single numeric value so its called point estimation.

2. Interval Estimation

When an estimate for unknown population parameter is expressed by the range of values with in which the unknown population parameter is expected to lie at certain level of confident or at certain level of risk is called interval estimation.

A random interval (L, U) or range of values in which unknown parameter expected to lie with certain probability or confident $(1-\alpha) * 100\%$ where 'α' is risk level

'$1-\alpha$' = confident level

is called confident interval.

i.e probability of, $P(L \leq \theta \leq U) = (1-\alpha) * 100\%$

For example : if risk level or significance level 5% then confident level = 95%.

Scanned with CamScanner

DATE : __/__/__
PAGE : ______

Interval Estimation for Population Mean

Suppose '$\bar{x}$' be sample mean, with population S.D. 'σ' or sample S.D. 'S', then the corresponding confident interval or risk interval when risk level is α or confident level is $(1-\alpha) * 100\%$ is

$$P\left(\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

ie lower limit of confident interval, $L = \bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ (If sample size $n \geq 30$)

Upper limit of confident interval, $U = \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

And

$$P\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}\right) \quad \text{(If sample size } n<30)$$

More particularly to estimate population μ on the basis of sample mean $\bar{x}$, we use either Z-test or t-test.

	Use of Z-test	Use of t-test
1.	If population is normal and σ^2 (known or unknown) and $n \geq 30$.	If σ^2 unknown and $n \le 30$ we t-test

Q. A random sample of size n=100 is taken from population with $\sigma = 5.1$. If sample mean is $\bar{x} = 21.6$, construct 95% confident interval for population mean μ.

Sol^n →

Here, $n = 100$ (large sample)

Population S.D., $\sigma = 5.1$

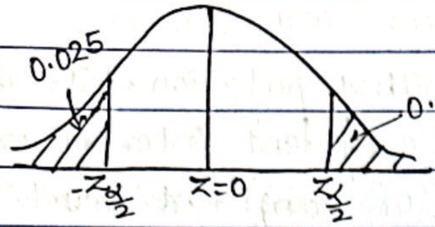
Sample Mean, $\bar{x} = 21.6$

Confident level $(1-\alpha) = 95\%$

$\Rightarrow \alpha = 5\% = 0.05$

$Z_{\alpha/2} = Z_{0.025}$

DATE : __/__/__
PAGE :



$P(0 \le z \le z_{\alpha/2})$
$= 0.5 - 0.025$
$= 0.475$

0.06
1.9 — 0.475

from z-table, the critical value $z_{0.025} = 1.96$.

So,

Lower limit of interval, $L = \overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$\Rightarrow L = 21.6 - 1.96 * \frac{5.1}{\sqrt{100}}$

$\Rightarrow L = 20.6$

Upper limit $U = \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$= 21.6 + 1.96 * \frac{5.1}{\sqrt{100}}$

$= 22.6$

Hence required 95% confident interval or 5% risk interval about population mean on the basis of sample mean is
$(L, U) = (20.6, 22.6)$
OR $P(20.6 \le \mu \le 22.6) = 95\%$

Maximum error of estimate or Margin Error

When we collect a set of samples we can calculate the sample mean $\overline{x}$ and that sample mean is typically different from the population mean μ. The difference between sample mean and population mean is assumed as error of estimation. Since we have $|\overline{x} - \mu| \le z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, if we intended to estimate

DATE : 2020/09/18
PAGE :

population mean (μ) when $n \geq 30$, With certain level of confidence $(1-\alpha) * 100\%$.

Since, Error (E) = $|\overline{x}-\mu|$, so error of estimation $E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is maximum error of estimation when $n \geq 30$.

Note: 1. Since $(1-\alpha)100\%$ confident interval for population mean, on the basis of sample mean $\left(\overline{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

$$\Rightarrow (\overline{x}-E, \overline{x}+E)$$

i.e. confident limits are $\overline{x} \pm E$.

2. If we use t-test for estimating population mean on the basis of sample mean, then $(1-\alpha) * 100\%$ confident limits

$$L = \overline{x} - t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}, \quad U = \overline{x} + t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}$$

3. If population is normal and σ^2 is known or unknown, if $n \geq 30 \Rightarrow$ use z-test.

But if σ^2 unknown, population normal and $n < 30 \Rightarrow$ use t-test

Q. A machine is producing metal pieces that are cylindrical on shape. A sample of 8 pieces is taken and diameters are (in cm) 1.01, 0.97, 1.04, 0.99, 0.98, 0.99, 1.01 and 1.03. Find 98% confident interval for mean diameter of all pieces produced by the machine.

Soln:

Here, Sample size $n = 8 < 30$ so we use t-test.

$$\text{Sample Mean } (\overline{x}) = \frac{1.01+0.97+1.04+0.99+0.98+0.99+1.01+1.03}{8}$$

$$= 1.0025$$

$$\text{Sample Variance}(s^2) = \frac{\sum(x-\overline{x})^2}{n-1}$$

Scanned with CamScanner

DATE : ___/___/___
PAGE : ________

$$= \frac{1}{7}\left\{ \begin{array}{l} (1.01-1.0025)^2 + (0.97-1.0025)^2 + (1.04-1.0025)^2 + (0.99-1.0025)^2 + \\ (0.98-1.0025)^2 + (0.99-1.0025)^2 + (1.01-1.0025)^2 + (1.03-1.0025)^2 \end{array} \right\}$$

$$= 5.91 * 10^{-4}$$

$$\therefore \ S.D.\ (S) = \sqrt{5.91 \times 10^{-4}}$$
$$= 0.022$$

Since, $(1-\alpha) * 100\%$ confident interval,

$$P\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \right) = (1-\alpha) * 100\%$$

Here,

$$(1-\alpha) * 100\% = 98\%$$
$$\Rightarrow \alpha = 2\% = 0.02$$
$$\Rightarrow \frac{\alpha}{2} = 0.01$$

degree of freedom $= n-1 = 8-1 = 7$

From t-table, at $\alpha = 2\%$ critical value for two tail test i.e.

$$t_{\frac{\alpha}{2}, n-1} \text{ or } t_{0.01, 7} = 2.998$$

So, Lower limit of confident interval,

$$L = \bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

$$L = 1.0025 - 2.998 * \frac{0.022}{\sqrt{8}}$$

$$= \cancel{0.023}\ 0.97$$

$$\text{Upper Limit } (U) = 1.0025 + 2.998 * \frac{0.022}{\sqrt{8}}$$

$$= 1.025$$

DATE : __/__/__
PAGE : ______

So, 98% confident interval or 2% risk interval for population mean on the basis of sample mean is (0.97, 1.025).

Detemining Sample size and Degree of Confident

If $\bar{x}$ be the sample mean of sample of size n drawn from population having mean μ and S.D σ.

Then $\bar{x}$ follows normal distribution, where, standard normal variable, $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ (By central limit theorem)

or, $z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$

or $\sqrt{n} = \frac{z\sigma}{\bar{x}-\mu}$

Squaring,

$$n = \left(\frac{z\sigma}{\bar{x}-\mu}\right)^2$$

$$\Rightarrow n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{\bar{x}-\mu}\right)^2$$, where $E = |\bar{x}-\mu|$

$z_{\frac{\alpha}{2}}$ = significance value or critical value of z at $\frac{\alpha}{2}$.

Note: If population S.D. is not known but $n \geq 30$ then

$$n = \left(\frac{z_{\frac{\alpha}{2}} S}{E}\right)^2$$

OR

If t-test used then, $n = \left[\frac{t_{\frac{\alpha}{2}, n-1} S}{E}\right]^2$

Q. A research worker wants to determine the average time it takes a machine to rotate the tires of car and he want to be able to assert with 95% confident that mean of his sample is off by the time 0.5 minute. If ~~he can~~ his past experience says that

DATE : __/__/__
PAGE : ____

standard deviation (σ) is 1.6 minute, how large sample should be take ?

Sol^n →

Here, Error of estimation $E = 0.5$

$\sigma = 1.6$

$(1-\alpha) * 100\% = 95\%$

$\Rightarrow \alpha = 5\% = 0.05$

$\Rightarrow z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

We know,

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{E} \right)^2$$

$$= \left(\frac{1.96 * 1.6}{0.5} \right)^2$$

$$= 39.33$$

$$\approx 39$$

Confidence Interval for difference of two normal population mean

Case I : For large sample ($n > 30$)

(i) When variance of both population is known

Let $x_1, x_2, \dots x_{n_1}$ and $y_1, y_2, \dots y_{n_2}$ be two independent samples of size n_1 and n_2 (where $n_1 > 30$, $n_2 > 30$) are drawn from two population having variance σ_1^2 and σ_2^2 and means μ_1 and μ_2.

Suppose $\bar{x}$ and $\bar{y}$ be the means of both samples then the sampling distribution of $\bar{x} - \bar{y}$ follows the normal distribution with mean $\mu_1 - \mu_2$ and S.D. $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Then $(1-\alpha) * 100\%$ confident interval for difference of two population mean $(\mu_1 - \mu_2)$ is

DATE : / /
PAGE :

$$P\left((\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 \le (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

$$= (1-\alpha) \times 100\%.$$

i.e. $$L = (\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$U = (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Note : If both different samples are drawn from same population i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then $(1-\alpha)\times 100\%$ confidence interval is

$$L = (\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\,\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$U = (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\,\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

(ii) If population variance is unknown but unequal, Then $(1-\alpha)\times 100\%$ confident interval for difference of two population mean is

$$P\left((\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1-\mu_2 \le (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right) = (1-\alpha)\times 100\%$$

i.e.

$$L = (\bar{X}-\bar{Y}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$U = (\bar{X}-\bar{Y}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

(iii) When population variance is unknown but equal i.e σ_1^2, σ_2^2 are unknown but assumed $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (say) then combined sample variance or pooled variance

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \approx \frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2} \quad \text{for large sample}$$

Then, $(1-\alpha) * 100\%$ confidence interval for difference of two population mean $\mu_1 - \mu_2$ is,

$$P\left((\bar{x}-\bar{y}) - Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}-\bar{y}) + Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

i.e.

$$L = (\bar{x}-\bar{y}) - Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$U = (\bar{x}-\bar{y}) + Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Case II : For small samples $(n<30)$

Assumptions :

(i) $X_1 \dots X_{n_1}$ and $Y_1 \dots Y_{n_2}$ are two independent samples drawn from two populations where $n_1, n_2 \le 30$.

(ii) Two populations are normally distributed.

(iii) The mean of both population are respectively μ_1, μ_2 and variance σ_1^2 and σ_2^2 but not given.

Then

1. If σ_1^2 and σ_2^2 are unknown and unequal we estimate sample variance

$$S_1^2 = \frac{1}{n_1-1} \sum (X_1-\bar{x})^2$$

$$S_2^2 = \frac{1}{n_2-1} \sum (Y_1-\bar{y})^2$$

Then $(1-\alpha) * 100\%$ confident interval is

$$P\left((\bar{x}-\bar{y}) - t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}-\bar{y}) + t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$$

$$= (1-\alpha) * 100\%$$

DATE : __/__/__
PAGE : ______

Then, $(1-\alpha)*100\%$ confidence interval for difference of two population mean $\mu_1-\mu_2$ is,

$$P\left((\bar{x}-\bar{y}) - Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1-\mu_2 \le (\bar{x}-\bar{y}) + Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

i.e.

$$L = (\bar{x}-\bar{y}) - Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$U = (\bar{x}-\bar{y}) + Z_{\frac{\alpha}{2}}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Case II : For small samples ($n<30$)

Assumptions :

(i) $X_1 \dots X_{n_1}$ and $Y_1 \dots Y_{n_2}$ are two independent samples drawn from two populations where $n_1, n_2 \le 30$.

(ii) Two populations are normally distributed.

(iii) The mean of both population are respectively μ_1, μ_2 and variance σ_1^2 and σ_2^2 but not given.

Then

1. If σ_1^2 and σ_2^2 are unknown and unequal we estimate sample variance

$$S_1^2 = \frac{1}{n_1-1}\sum (X_1-\bar{X})^2$$

$$S_2^2 = \frac{1}{n_2-1}\sum (Y_1-\bar{Y})^2$$

Then $(1-\alpha)*100\%$ confident interval is

$$P\left((\bar{X}-\bar{Y}) - t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \le \mu_1-\mu_2 \le (\bar{X}-\bar{Y}) + t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$$

$$= (1-\alpha)*100\%$$

DATE : __/__/__

PAGE : ______

2. If variances are unknown but equal,

Then $(1-\alpha) * 100\%$ confident interval,

$$L = (\bar{X} - \bar{Y}) - t_{\frac{\alpha}{2}, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$U = (\bar{X} - \bar{Y}) + t_{\frac{\alpha}{2}, n_1+n_2-2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where,

$$S_p^2 = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1+n_2-2}$$

Q. Two independent samples of size 30 and 40 are drawn from the normal population $N(\mu_1, 25^2)$ and $N(\mu_2, 30^2)$ with mean $\mu_1 = 350$ and $\mu_2 = 325$. Find 90% confidence interval for $(\mu_1 - \mu_2)$.

Sol^n,

Here, Population S.D. $\sigma_1 = 25$, $\sigma_2 = 30$

Sample size, $n_1 = 30$, $n_2 = 40$

Sample Mean, $\bar{X} = 350$, $\bar{Y} = 325$

$(1-\alpha) * 100\% = 90\%$

$\Rightarrow \alpha = 10\% = 0.10$

$\Rightarrow \frac{\alpha}{2} = 0.05$

$\Rightarrow Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.64$ (from z-table)

Now,

Lower limit $(L) = (\bar{X} - \bar{Y}) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$= (350-325) - 1.64 \sqrt{\frac{25^2}{30} + \frac{30^2}{40}} = 14.2$$

Upper Limit $(U) = (\bar{X} - \bar{Y}) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$= (350-325) + 1.64 \sqrt{\frac{25^2}{30} + \frac{30^2}{40}} = 35.79$$

DATE : 2080/09/19
PAGE :

Hypothesis

A statistical hypothesis is a quantitative statement (assumption or claim) made about population on the basis of sample.

Examples:

1. A consumer claims that there is less than 200 ml in cold drink bottles produced by factory n (we have to test the claim is true or false).

2. A doctor claims that a particular drug cures 60% of patient of certain disease.

Hypothesis Testing

A hypothesis testing is a standard process of testing the claim made about population whether it is true or false under certain level of significance or risk.

Types of Hypothesis

1. Null Hypothesis or Hypothesis of no any difference

The null hypothesis says that there is no any difference between sample statistic and its corresponding population parameter i.e. it says that the result obtained from the sample and the corresponding population parameter are same.

The null hypothesis is denoted by H_0 i.e. $H_0: \theta = \theta_0$ (There is no any significant difference between sample and population).

2. Alternative Hypothesis

The complement of null hypothesis is called alternative hypothesis. It is also called hypothesis of difference. The alternative hypothesis is denoted by H_1. The alternative hypothesis may be one of the following:

(i) $H_1: \theta \neq \theta_0$ Two Tail test (there is significance difference between sample & population).

(ii) $H_1 : \theta < \theta_0$ Left tail test (Population parameter is less than sample statistics)

(iii) $H_1 : \theta > \theta_0$ Right tail test (Population parameter is greater than sample statistics)

Note: To test, above claim, which one is true, we use test statistics : z-test, t-test, F-test, χ^2- test (chi-square test).

z-test

- Generally, z-test is used when sample size $n \geq 30$.
- Under z-test, we test,

(i) Test of significance of single mean
(ii) Test of significance of difference between two mean
(iii) Test of significance of single sample proportion
(iv) Test of significance of difference of two sample proportion.

Steps of Hypothesis Testing

1. Null Hypothesis
   $H_0 : \theta = \theta_0$, (there is no any significant difference between sample and population)

2. Alternative Hypothesis
   $H_1 : \theta \neq \theta_0$, (there is significance difference between sample and population)
   $H_1 : \theta < \theta_0$, (Population Parameter < Sample Statistics)
   $H_1 : \theta > \theta_0$ (" " > " ")

3. Level of Significance or Level of Risk
   $\alpha = 1\%$, or 5%, 10%, 2%, etc.

4. Test Statistics

$$t_{cal} \text{ or } Z_{cal} = \frac{\theta_0 - E(\theta_0)}{S.E.(\theta_0)}$$

5. Critical Value or Tabulated Value
Find tabulated value from z-table or t-table at 'α' level of risk. for one tail or two tail. Then we find z_{tab} or t_{tab}.

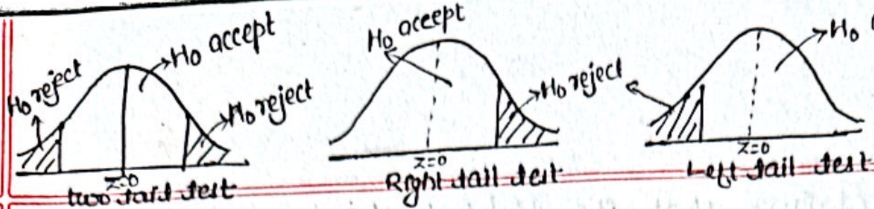
6. Decision
If $|z_{cal}| < |z_{tab}|$ OR $|t_{cal}| < |t_{tab}|$, then null hypothesis (H_0) is accepted.
But
If $|z_{cal}| > |z_{tab}|$ OR $|t_{cal}| > |t_{tab}|$, then H_0 is rejected and H_1 is accepted.

Note: If the question contains no any comparative words like smaller, taller, superior, inferior, greater, smaller, etc, then test is two tail otherwise the test is one tail.

Q. A random sample of 100 recorded ~~bags~~ ~~deaths~~ deaths in a certain hospital during past years shows an average life time is 71.8 yrs with S.D. 8.9 yrs. Does this seem to indicate that the average life ~~spend~~ span today is greater than 75 yrs? Test under, $\alpha = 5\%$.

Sol^n →
Here, sample size, $n = 100$
Sample mean $(\bar{x}) = 71.8$
Sample S.D. $(s) = 8.9$
Level of Risk $(\alpha) = 5\% = 0.05$

(i) Null Hypothesis
$H_0: \mu = 75$ (there is no any difference between sample mean and population mean).



(ii) Alternative Hypothesis:

$H_1 : \mu > 75$ (population mean is greater than sample mean).

(iii) Level of risk, $\alpha = 0.05$

(iv) Test statistics,

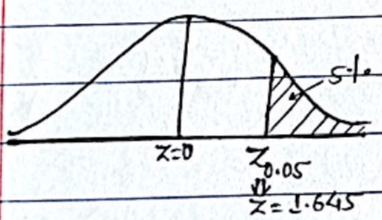
$$Z_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$= \frac{71.8 - 75}{\frac{8.9}{\sqrt{100}}}$$

$$= -3.59$$

(v) Critical Value

Z_{tab} from z-table at $\alpha = 5\%$ for one tail test.



$1 - 0.05 = 0.95$

$$P(z < 1.645) = 0.5 + P(0 < z < 1.645)$$

$$= 0.5 + \frac{P(0 < z < 1.64) + P(0 < z < 1.65)}{2}$$

$$= 0.5 + \frac{0.4495 + 0.4505}{2}$$

$$= 0.5 + 0.45$$

$$= 0.95$$

(vi) Compare And Decision

Here, $|Z_{cal}| = 3.59$

$|Z_{tab}| = 1.645$

Here $|Z_{cal}| > |Z_{tab}|$ so H_0 rejected i.e. H_1 accepted.

So population mean $(\mu) > 75$ is true.

DATE : __/__/__
PAGE : ______

Q. A company claims that its light bulbs has mean life time 1500 hrs of continuous use with S.D. 1200 hrs. However there is a complain against this claim from the regular customer that a sample of 60 bulbs has mean life time 1300 hrs of continuous use. Justify the claim at 0.02 (i.e $\alpha = 2\%$).

Sol^n →

Population mean (μ) = 1500
Sample mean ($\bar{x}$) = 1300
Population S.D (σ) = 1200
$\alpha = 0.02$
Sample size (n) = 60

(i) $H_0 : \mu = 1500$

(ii) $H_1 : \mu > 1500$

(iii) $\alpha = 0.02$

(iv) $$Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{1300 - 1500}{\frac{1200}{\sqrt{60}}}$$

=

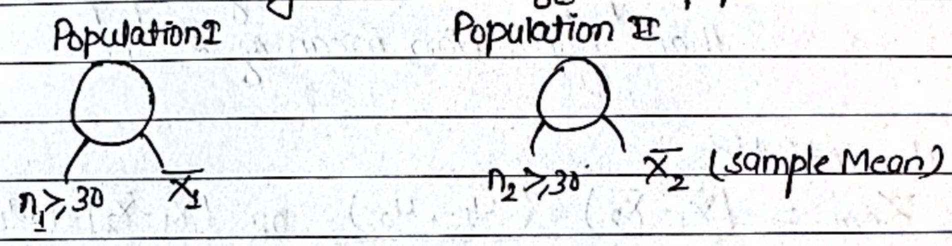
Test of Significance of difference of two Sample Mean

To test the significance of difference of two sample means, we use z-test if sample size $n \geq 30$ and we use t-test if sample size $n < 30$.

For the use of z-test:

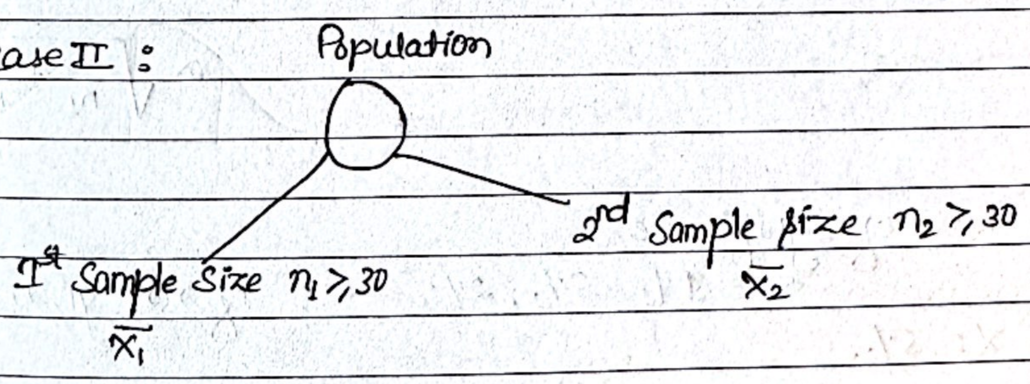
(i) Sample Size ≥ 30

Case I: There may be two different population



Now the variance of both population σ_1^2 and σ_2^2 may be given or may not be given.

Case II:



Now sample variance σ^2 may be given or may not be given.

Process:-

(1) Null Hypothesis

$H_0: \mu_1 = \mu_2$ (both population mean are ~~sample~~ same ~~and~~ or there is no any significant different betⁿ two population mean).

DATE : __/__/__
PAGE : ______

2. Alternative

$H_1 : \mu_1 \neq \mu_2$ (There is significant difference betn ~~both~~ both population mean)

$H_1 : \mu_1 < \mu_2$ (~~there is~~ the population mean of first population is less than population mean of 2nd)

$H_1 : \mu_1 > \mu_2$ (the population mean of 1st population is greater than population mean of 2nd).

3. Test statistic

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{OR} \quad \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{OR}$$

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

4. Level of Significance

$\alpha = 1\%, 5\%, 2\%$, etc, If not given by default $\alpha = 5\%$.

5. ~~Compare~~ ~~and~~ Find Critical value from z-table for α level of risk for one tail or two tail, which is called Z_{tab}.

6. Compare and Decision

If $|Z_{cal}| < |Z_{tab}| \Rightarrow H_0$ accepted

If $|Z_{cal}| > |Z_{tab}| \Rightarrow H_0$ rejected and H_1 accepted.

DATE : __/__/__
PAGE : ____

Q. A Mathematics test was given to two groups consisting of 40 and 50 students respectively. In the first group mean marks was 74 and S.D. is 8 and in the second group mean marks 78 with S.D. 7. Is there any significant difference betn the performance of both groups?

Soln:

Here, Sample size from first group, $n_1 = 40$

$\bar{x}_1 = 74$

$s_1 = 8$

For second group, $n_2 = 50$

$\bar{x}_2 = 78$

$s_2 = 7$

(i) Null Hypothesis

$H_0 : \mu_1 = \mu_2$ (there is no any significant difference betn mean or performance of both groups).

(ii) Alternative

$H_1 : \mu_1 \neq \mu_2$ (there is significant difference betn performance of both groups)

(iii) Level of risk

$\alpha = 5\% = 0.05$

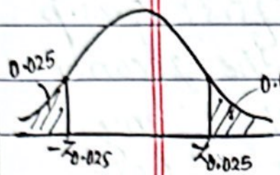
(iv) Test Statistics

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(74-78) - 0}{\sqrt{\frac{8^2}{40} + \frac{7^2}{50}}} = -2.49$$

DATE : __/__/__
PAGE : ______

(v) Critical value from z-table at $\alpha = 5\%$ for two tail test

$$z_{tab} \text{ or } z_{\frac{\alpha}{2}} \text{ or } z_{0.025} = 1.96$$



(vi) Compare and Decision

Here, $|z_{cal}| = |-2.49| = 2.49$

$|z_{tab}| = 1.96$

$\Rightarrow |z_{cal}| > |z_{tab}|$

$\Rightarrow H_0$ Rejected i.e. H_1 accepted.

i.e. Significance difference between performance of both group exist.

Note: 1. The above process of testing the significance of difference of two mean is based on the assumption that there is zero difference on the population mean.

But if there is non-zero difference betn the population mean i.e. the difference between both population mean is d_0 is significant or not, then

(i) Null Hypothesis

$$H_0: \mu_1 - \mu_2 = d_0$$

(ii) Alternative

$$H_1: \mu_1 - \mu_2 \neq d_0$$
$$H_1: \mu_1 - \mu_2 < d_0$$
$$H_1: \mu_1 - \mu_2 > d_0$$

Then remaining process is same.

2. When two different samples are drawn from same population and the population variance is unknown then test statistics is

$$z_{cal} \text{ or } t_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is pooled or combined variance.

Q. To test the claim that the resistance of electric wire can be replaced by more than 0.05 by allowing 32 values obtained from the wires provides $\bar{x}_1 = 0.136$, $S_1 = 0.004$ and next 32 values of wires provides $\bar{x}_2 = 0.083$, $S_2 = 0.005$ at 5% level of significance. Does the claim is supported?.

Sol^n →

For 1st population, $n_1 = 32$

$\bar{x}_1 = 0.136$

$S_1 = 0.004$

For 2nd population,

$n_2 = 32$

$\bar{x}_2 = 0.083$

$S_2 = 0.005$

(i) Null Hypothesis

$H_0 : \mu_1 - \mu_2 = 0.05$ (the difference bet^n two population mean is 0.05)

(ii) Alternative

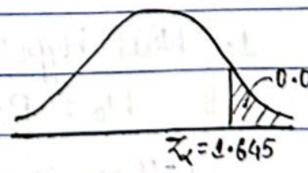
$H_1 : \mu_1 - \mu_2 > 0.05$ (the difference bet^n two population mean is greater than 0.05).

(iii) Level of risk, $\alpha = 5\% = 0.05$

(iv) Critical value

For one tail test at 5% risk level,

$Z_{tab} = 1.645$



(v) Test statistics

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(0.136 - 0.083) - 0.05}{\sqrt{\frac{0.004^2}{32} + \frac{0.005^2}{32}}} = 2.65$$

DATE : __/__/__
PAGE : ____

(vi) $|Z_{cal}| > |Z_{tab}|$

$\Rightarrow$ H_0 rejected and H_1 accepted.

i.e. the difference betn two population mean is greater than 0.05.

Thus the claim is supported.

Q. Suppose that we want to investigate whether the average earning of men is more than \$20 per week then women. If sample shows that 60 men earn in average $\bar{x}_1 = \$585$, and S.D $(S_1) = \$31.20$ while 60 women earn in average $\bar{x}_2 = \$532.20$ with S.D $(S_2) = \$36.4$. What do you conclude at 0.01 level of significance?

$\mu_1 - \mu_2 = \$20$
$\mu_1 - \mu_2 > \$20$

Q. If mean height of 60 science students of KU is 68.6 inch and mean height of 50 Eco. students is 69.51 inch. Would you conclude that Eco students are taller than science students assuming that S.D of all students of KU is 2.48.

$\mu_1 - \mu_2 = 0$
$\mu_1 \neq \mu_2$
$\mu_1 = \mu_2$

Test of Significance of Single Sample Proportion

For large sample (use z-test):

Suppose sample proportion of success is $\hat{p}$ and population proportion of success is P and sample proportion of failure is $\hat{q}$ and population proportion of failure is q.

1. Null Hypothesis

$H_0: P = P_0$, where P_0 is specific value.

(There is no any significant difference between sample proportion and population proportion).

2. Alternative

$H_1: P \neq P_0$ (two tail test)

$H_1 : P > P_0$ (Right tail test)
$H_2 : P < P_0$ (Left tail test)

(iii) Level of risk $\alpha = 1\%, 5\%, \ldots$ etc

(iv) Test statistics

$$Z_{cal} = \frac{\hat{P} - P}{\sqrt{\frac{pq}{n}}} \quad \text{(when p and q are known)}$$

OR

$$Z_{cal} = \frac{\hat{P} - P}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \quad \text{(when p and q are not known)}$$

(v) Critical Value
Find Z_{tab} for Z-table at α-level of risk for one tail or two tail.

(vi) Compare and Decision
If $|Z_{cal}| < |Z_{tab}| \Rightarrow H_0$ accepted
If $|Z_{cal}| > |Z_{tab}| \Rightarrow H_0$ Rejected i.e H_1 accepted.

Q. A salesman claims that only 4% of apples are defective. A sample of 1600 apples contain 36 defective apples. Test the claim of salesman.

Soln.

Population Proportion $(P) = \frac{4}{100} = 0.04$

Sample size $(n) = 600$
Defective $= 36$

DATE : __/__/__
PAGE : ______

Sample proportion ($\hat{p}$) = $\frac{36}{600}$ = 0.06

Now,

(I) Null Hypothesis

$H_0: P = 0.04$ (there is no any significance difference between sample proportion and population proportion).

(II) Alternative

$H_1: P > 0.04$ (population proportion is greater than 4%)

(III) Level of Risk

$\alpha = 5\% = 0.05$

(IV) At 5% level of risk,

Critical Value, $Z_{tab} = 1.645$

(V) Test statistics

$$Z_{cal} = \frac{\hat{P} - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.06 - 0.04}{\sqrt{\frac{0.04 \times 0.96}{600}}} = 2.5$$

(VI) Compare and Decision

$|Z_{cal}| > |Z_{tab}| \Rightarrow H_0$ rejected and H_1 accepted.

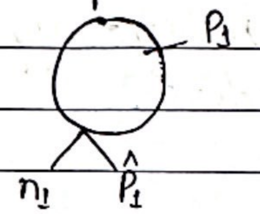
i.e

Defective Apples are more than 4%.

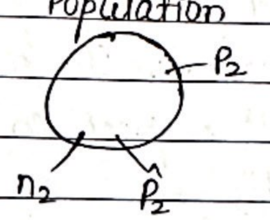
Scanned with CamScanner

Test of significance of difference of two sample proportion

1st Population



2nd Population



(I) Null Hypothesis

$H_0 : P_1 = P_2$ (there is no significance difference betn two ~~sample~~ population proportion).

(II) Alternative

$H_1 : P_1 \neq P_2$ (two tail test)

$H_1 : P_1 > P_2$ (Right tail test)

$H_1 : P_1 < P_2$ (Left tail test)

(III) Level of Risk $\alpha = 1\%, 5\%, \ldots$

(IV) Test statistics

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$ (Both population proportion known).

OR

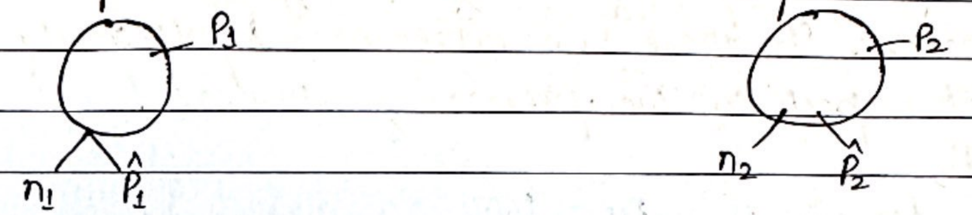
$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}}$$ (Both population proportion not known)

OR

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ (Samples are drawn from same population and population proportion is known)

Test of significance of difference of two sample proportion

I^{st} Population — 2^{nd} Population



(I) Null Hypothesis

$H_0 : P_1 = P_2$ (there is no significance difference betn two ~~sample~~ population proportion).

(II) Alternative

$H_1 : P_1 \neq P_2$ (two tail test)

$H_1 : P_1 > P_2$ (Right tail test)

$H_1 : P_1 < P_2$ (Left tail test)

(III) Level of Risk $\alpha = 1\%, 5\%, ---$

(IV) Test statistics

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$ (Both population proportion known).

OR

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}}$$ (Both population proportion not known)

OR

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ (Samples are drawn from same population and population proportion is known)

DATE :__/__/__
PAGE :____

Q. In a sample of 1600 from city A, 450 are found smokers and in a sample of 900 from city B, 450 are found smokers. Do these two cities are significantly different with respect to the presence of smokers?

Sol:→

For city A, $n_1 = 1600$

smokers $= 450$

Sample proportion of success $(\hat{p}_1) = \frac{450}{1600} = 0.28$

Sample proportion of failure $(\hat{q}_1) = 1 - 0.28 = 0.72$

For city B,

$n_2 = 900$

smokers $= 450$

$\hat{p}_2 = \frac{450}{900} = 0.5$

$\hat{q}_2 = 1 - 0.5 = 0.5$

Let P_1 and P_2 be the population proportion of success.

(i) Null Hypothesis

$H_0 : P_1 = P_2$ (there is no any significance difference betn the population proportion)

(ii) Alternative

$H_1 : P_1 \neq P_2$ (there is significance difference between the population proportion)

(iii) $\alpha = 5\% = 0.05$

(iv) From z-table, for two tail test,

$z_{tab} = 1.96$

DATE : __/__/__
PAGE : ______

(v) Test statistics

$$Z_{cal} = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}}$$

$$= \frac{(0.28 - 0.5) - 0}{\sqrt{\frac{0.28 * 0.72}{1600} + \frac{0.5 * 0.5}{900}}}$$

$$= -10.94$$

(vi) $|Z_{cal}| > |Z_{tab}| \Rightarrow H_1$ accepted.

Scanned with CamScanner

DATE : 2080.09.22
PAGE :

Joint Probability Distribution or Bivariate Probability Distribution or Two Dimensional Probability Distribution

Bi-variate Random Variable OR Two Dimensional Random Variable

Let S be a sample space associated with a random experiment. Suppose $X: S \to R$ and $Y: S \to R$ are two independent fn defined on S such that each assigning unique real number corresponding to for all $w \in S$. Then a new fn defined by $X \times Y: S \to R^2$ such that $(x, y)w = (X(w), y(w)) \in R^2 \ \forall\ w \in S$ is called two dimensional random variable and bivariate random variable.

For example:

Suppose A coin tossed three times let X denotes no. of tail in first trial and Y denotes no. of heads in all trials, then the two dimensional random variable (x, y) assumes the following values.

Soln.

When a coin is tossed three times the sample space is

$S = \{HHH, HHT, HTH, THT, HTT, THH, TTH, TTT\}$

Here,

X = tail in 1st trial = 0, 1

Y = no. of head in all three trial = 0, 1, 2, 3

Now,

$X \times Y: S \to R$ such that $(X, Y)(HHH) = \{X(HHH), Y(HHH)\}$
$= (0, 3)$

$(X, Y)(HHT) = \{X(HHT), Y(HHT)\}$
$= (0, 2)$

$(X, Y)(HTH) = \{X(HTH), Y(HTH)\}$
$= (0, 2)$

$(X, Y)(THT) = \{X(THT), Y(THT)\} = (1, 1)$

$(X, Y)(THH) = \{X(THH), Y(THH)\} = (1, 2)$

$(X, Y)(TTH) = \{X(TTH), Y(TTH)\} = (1, 1)$

$(X, Y)(TTT) = \{X(TTT), Y(TTT)\} = (1, 0)$

$(X, Y)(HTT) = \{X(HTT), Y(HTT)\} = (0, 1)$

Joint Probability Distribution or Bivariate Probability Distribution or Two Dimensional Probability Distribution

Bi-variate Random Variable OR Two Dimensional Random Variable

Let S be a sample space associated with a random experiment. Suppose $X : S \to R$ and $Y : S \to R$ are two independent fn defined on S such that each assigning unique real number corresponding to for all $w \in S$. Then a new fn defined by $X \times Y : S \to R^2$ such that $(x, y)w = (X(w), y(w)) \in R^2 \; \forall \; w \in S$ is called two dimensional random variable and bivariate random variable.

For example:

Suppose A coin tossed three times let X denotes no. of tail in first trial and Y denotes no. of heads in all trials, then the two dimensional random variable (x, y) assumes the following values.

Soln

When a coin is tossed three times the sample space is

$S = \{HHH, HHT, HTH, THT, HTT, THH, TTH, TTT\}$

Here,

X = tail in 1^{st} trial = 0, 1

Y = no. of head in all three trial = 0, 1, 2, 3

Now,

$X \times Y : S \to R$ such that $(X, Y)(HHH) = \{X(HHH), Y(HHH)\}$
$= (0, 3)$

$(X, Y)(HHT) = \{X(HHT), Y(HHT)\}$
$= (0, 2)$

$(X, Y)(HTH) = \{X(HTH), Y(HTH)\}$
$= (0, 2)$

$(X, Y)(THT) = \{X(THT), Y(THT)\} = (1, 1)$

$(X, Y)(THH) = \{X(THH), Y(THH)\} = (1, 2)$

$(X, Y)(TTH) = \{X(TTH), Y(TTH)\} = (1, 1)$

$(X, Y)(TTT) = \{X(TTT), Y(TTT)\} = (1, 0)$

$(X, Y)(HTT) = \{X(HTT), Y(HTT)\} = (0, 1)$

DATE : ___/___/___
PAGE : ________

Hence, two dimensional random variable (X,Y) associates each sample point $\omega \in S$ to unique point $(X(\omega), Y(\omega)) \in R^2$.

Note : 1) If $(X(\omega), Y(\omega)) \in R^2 \ \forall \ \omega \in S$ are finite then two dimensional random variable (X,Y) is called two dimensional discrete random variable.

2) If (X,Y) can assume any point on the region $a \le x \le b$, $c \le Y \le d$ of R^2 for the sample space then (X,Y) is called two dimensional continuous random variable.

Joint Discrete Probability Distribution OR Joint Discrete Probability Mass function

Suppose S be a sample space of any random experiment. Let X and Y are two discrete random variable defined on S such that X takes the values $x_1, x_2, \dots, x_n$ and Y takes values $y_1, y_2, \dots, y_n$ then joint pmf of (X,Y) is denoted by

$$P(X=x, Y=y) \text{ or } P(X=x \cap Y=y) \text{ and given by}$$

$$f_{x,y}(x,y) = P(X=x \cap Y=y) = \begin{cases} P_{ij} & \text{if } (X,Y) = (x_i, y_j) \text{ for } i=1,\dots m,\ j=1\dots n \\ 0 & \text{Otherwise} \end{cases}$$

Note: 1) $f_{XY}(x_i, y_j) \geq 0$

2) $\sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_i) = 1$

Q. Three balls are drawn from the bag containing 2 white balls, 3 red and 4 black balls without replacement. If X denotes no of white and Y denotes no. of Red. Find joint probability distribution of (X,Y).

Sol^n →

$n(W) = 2$, $n(R) = 3$, $n(B) = 4$

Total $n = 9$

X takes 0, 1, 2 and Y takes 0, 1, 2, 3.

Total possible outcomes $n(S) = C(9,3)$
$= 84$

Now the random variable (X,Y) can assume values $\{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (1,3), (2,0), (2,1), (2,2), (2,3)\}$

(i) $P(X=0, Y=0) = P_{00} = \frac{C(4,3)}{84} = 0.047$
all black

(ii) $P(X=0, Y=1) = P_{01} = \frac{C(3,1)\, C(4,2)}{84} = 0.21$
zero white, one red, two black

(iii) $P(X=0, Y=2) = P_{02} = \frac{C(3,2)\, C(4,1)}{84} = 0.1428$

(IV) $P(X=0, Y=3) = P_{03} = \frac{C(3,3)\, \cancel{C(4,0)}}{84} = 0.0119$

(V) $P(X=1, Y=0) = P_{10} = \frac{C(2,1)\, C(4,2)}{84} = 0.1428$

(VI) $P(X=1, Y=1) = P_{11} = \frac{C(2,1)\; C(3,1)\; C(4,1)}{84} = 0.2857$

(VII) $P(X=1, Y=2) = P_{12} = \frac{C(2,1)\, C(3,2)}{84} = 0.071$

(VII) $P(X=1, Y=3) = P_{13} = \frac{0}{84} = 0$

(VIII) $P(X=2, Y=0) = P_{20} = \frac{C(2,2) \times C(4,1)}{84} = 0.047$

(IX) $P(X=2; Y=1) = P_{21} = \frac{C(2,2) \times C(3,1)}{84} = 0.035$

(X) $P(X=2, Y=2) = 0$

(XI) $P(X=2, Y=3) = 0$

Now, Joint pmf is

Y \ X	0	1	2
0	0.047	0.1428	0.047
1	0.21	0.2857	0.035
2	0.1428	0.071	0
3	0.0119	0	0

Here, $f_{XY}(x_i, y_j) \geq 0 \quad \forall\ i = 0, 1, 2$ and $j = 0, 1, 2, 3$

$$\sum_{i=0}^{2} \sum_{j=0}^{3} P(x_i, y_j) = 1$$

$$L.H.S = \sum_{i=0}^{2} \{P(x_i, 0) + P(x_i, 1) + P(x_i, 2) + P(x_i, 3)\}$$

$$= P(0,0) + P(1,0) + P(2,0) + P(0,1) + P(1,1) + P(2,1) + P(0,2) + P(1,2) + P(2,2) + P(0,3) + P(1,3) + P(2,3)$$

Marginal Pmf and Marginal Distribution

If $P(x_i, y_j) = P(X = x_i, Y = y_j) = P_{ij}$ be the joint pmf, then the marginal probability of $X = x_i$ is

$$f_X(x_i) = P_i = P(x_i, y_1) + P(x_i, y_2) + \cdots + P(x_i, y_n)$$

where $Y = y_1, y_2, \ldots y_n$

$$= \sum_{j=1}^{n} P_{ij}$$

Then we calculate the marginal probability of all x_i for $i = 1, \ldots m$ and the set $\{x_i, P_i\}$ for $i = 1, 2, \ldots m$ is called marginal probability distribution of X.

Similarly,

The marginal probability of $Y = y_j$ is,

$$f_Y(y_j) = P_j = P(x_1, y_j) + P(x_2, y_j) + \cdots + P(x_m, y_j)$$

where X takes $x_1, \ldots x_m$

Now, Joint pmf is

Y \ X	0	1	2
0	0.047	0.1428	0.047
1	0.21	0.2857	0.035
2	0.1428	0.071	0
3	0.0119	0	0

Here, $f_{XY}(x_i, y_j) \geq 0 \quad \forall\ i = 0, 1, 2$ and $j = 0, 1, 2, 3$

$$\sum_{i=0}^{2} \sum_{j=0}^{3} P(x_i, y_j) = 1$$

$$L.H.S = \sum_{i=0}^{2} \left\{ P(x_i, 0) + P(x_i, 1) + P(x_i, 2) + P(x_i, 3) \right\}$$

$$= P(0,0) + P(1,0) + P(2,0) + P(0,1) + P(1,1) + P(2,1) + P(0,2) + P(1,2) + P(2,2) + P(0,3) + P(1,3) + P(2,3)$$

Marginal Pmf and Marginal Distribution

If $P(x_i, y_j) = P(X = x_i, Y = y_j) = P_{ij}$ be the joint pmf, then the marginal probability of $X = x_i$ is

$$f_X(x_i) = P_i = P(x_i, y_1) + P(x_i, y_2) + \dots + P(x_i, y_n)$$

where $Y = y_1, y_2, \dots y_n$

$$= \sum_{j=1}^{n} P_{ij}$$

Then we calculate the marginal probability of all x_i for $i = 1, \dots m$ and the set $\{x_i, P_i\}$ for $i = 1, 2, \dots m$ is called marginal probability distribution of x.

Similarly,

The marginal probability of $Y = y_j$ is,

$$f_Y(y_j) = P_j = P(x_1, y_j) + P(x_2, y_j) + \dots + P(x_m, y_j)$$

where X takes $x_1, \dots x_m$

DATE : __/__/__
PAGE :

We now calculate the marginal probability of all y_j for $j = 1 \ldots n$, then the set $\{y_j, P_j\}$ for $J = 1, \ldots n$ is called marginal probability distribution of Y.

Q. Given the joint pmf

P_{ij}		X		
Y		0	1	2
	0	0.1	0.4	0.1
	1	0.2	0.2	0

Find

(i) $P(X+Y>1)$

(ii) Find marginal pmf of X

(iii) Find marginal " " Y

(iv) Are X and Y are independent

sol^n ⟹

(i) $P((X+Y)>1) = P(1,1) + P(2,1) + P(2,0)$

$= 0.2 + 0 + 0.1$

$= 0.3$

(ii) Now marginal probabilities of $X = x_i$ is

$P(X=0) = P(0,0) + P(0,1) = 0.1 + 0.2 = 0.3$

$P(X=1) = P(1,0) + P(1,1) = 0.4 + 0.2 = 0.6$

$P(X=2) = P(2,0) + P(2,1) = 0.1 + 0 = 0.1$

Hence marginal pmf of X is

$X = x_i$	0	1	2
P_i	0.3	0.6	0.1

$\Sigma P_i = 1$

DATE : __/__/__
PAGE : ____

(iii) Again, Marginal probabilities $Y = y_j$

$$P(Y=0) = P(0,0) + P(1,0) + P(2,0) = 0.1 + 0.4 + 0.1 = 0.6$$

$$P(Y=1) = P(0,1) + P(1,1) + P(2,1) = 0.2 + 0.2 + 0 = 0.4$$

So, marginal pmf of Y is

$Y = y_j$	0	1
P_j	0.6	0.4

$\Sigma P_j = 1$

Note: Two random variable X and Y are said to be independent if.

$$P(X = x_i, Y = y_j) = P_{ij} = P(X = x_i) \times P(Y = y_j)$$
$$= P_i P_j$$

i.e. Joint Probability = Product of their marginal probability $\forall\ i = 1, \dots m,\ j = 1 \dots n$

(iv) Here, Joint probability

$$P(X=0, Y=0) = P_{00} = 0.1$$

But $P(X=0) = P_i = 0.3$

$P(Y=0) = P_j = 0.6$

Here,

$$P(X=0, Y=0) \neq P(X=0) * P(Y=0)$$

So, these variables are not independent.

Joint Continuous Probability Distribution OR Joint probability density function (pdf)

A function $f(X,Y) : S \to R^2$ is said to be two dimensional pdf if

$$f(x,y) \geq 0 \quad \forall\ (x,y) \in S$$

and

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

DATE : __/__/__
PAGE :

Q. If the joint pdf probability f^n of X and Y is

$$f(x,y) = \begin{cases} cxy & \text{for } 0<x<1,\ 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

(i) find value c
(ii) $P(x \le 0.5,\ y \le 0.8)$
(iii) Find marginal probability distribution of X and Y
(iv) Find $E(X)$, $E(Y)$, $E(XY)$

Note: If $f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$ is pdf then marginal

pdf of X is
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

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

Soln →

Here, the probability density is

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

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

$$\Rightarrow \int_0^1\int_0^1 cxy\,dy\,dx = 1$$

DATE : __/__/__
PAGE : ______

$$\Rightarrow \int_0^1 cx\left(\frac{y^2}{2}\right)_0^1 dx = 1$$

$$\Rightarrow \int_0^1 cx \cdot \frac{1}{2}\, dx = 1$$

$$\Rightarrow \frac{1}{2} \cdot \frac{1}{2} \cdot c = 1$$

$$\Rightarrow c = 4$$

(ii) $P(X<0.5, Y<0.8) = \int_0^{0.8}\int_0^{0.5} 4xy\, dx\, dy$

(iii) Marginal probability of X,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$= \int_0^1 4xy\, dy$$

$$= 2x$$

$\therefore$ Marginal pdf of X

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

Similarly marginal pdf of Y,

$$f_Y(y) = \int_0^1 4xy\, dx = 2y$$

$$\therefore f_Y(y) = \begin{cases} 2y & 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

Since $f(x,y) = f_X(x) * f_Y(y)$

$\Rightarrow 4xy = 2x \cdot 2y = 4xy$

So,

X and Y are independent variables.

$E(X) = \int_0^1 x P(x) \, dx$ $E(Y) = \int_0^1 y P(y) \cdot dy$