MADURAI KAMARAJ UNIVERSITY
(University with Potential for Excellence)
DISTANCE EDUCATION

Recognised by DEC
www.mkudde.org

B.Sc.,

MATHEMATICS

Second Year

Paper - IV

STATISTICS

Volume 2
Units : 6 - 10

S 85

MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

DIRECTORATE OF DISTANCE EDUCATION

MATHEMATICS

B.Sc (Mathematics)
Second Year

PAPER - IV

STATISTICS

Volume - II : 6 to 10 Units

S 85

Unit VI

Probability

Objectives

In this unit, we shall discuss how to find a Probability of an event, addition theorem of Probability, multiplication theorem of Probability, Baye's theorem and the applications of Baye's theorem.

Introduction

The word *probability* or *chance* is commonly used in *day to day* life and it has a vague meaning. For example, "probably today the train may come late by an hour", "The chance of two teams X and Y winning cricket game are equal", "It is possible to me that complete the job in time". In these examples *probably, chance, possible* are convey the same meaning, (i.e) the event is not certain to take place or in other words, there is uncertainness about happening of the event in question. The theory of probability has its origin in the games of chance related to gambling such as throwing a die, drawing a cards from a pack of cards etc., Jerame Cardon, an Italian mathematician, was the firs man to write a book on "*Book of Games of chance*" which was published after his death in 1663. Galileo, an Italian mathematician, was the first man to attempt quantitative measure of probability while dealing with some problems related to the theory of dice in gambling. However, the systematic and scientific foundation of the mathematical theory of probability was laid by *B.Pascal* and *Pierre definition Fermat*. Later T*homas Bayes* introduced the concept of Inverse Probability.

Probability theory is being applied in the solution of social, economic, political and business problems. The insurance industry, which emerged in the 19^{th} century, required précised knowledge about the risk of loss in order to

Space for Hint

Space for Hint

calculate premium. Probability theory, in fact, is the foundation of statistical inference.

6. 1 Probability of an Event

(1) **Experiment** : An experiment is defined as an action which we conceive and do or intend to do.

(2) **Random Experiment :** In each trial of an experiment conducted under identical conditions, the outcome is not unique, but may be any of the possible outcomes, then such experiment is called a *random experiment.*

Example :

(a) Tossing a fair coin twice.

(b) Rolling a dice.

(c) Selecting a ball from an Urn.

(d) Selecting a card from a well shuffled pack of cards.

(3) **Outcome :** The result of a random experiment is called an outcome.

Example : When a coin is tossed the possible outcomes are *head* or *tail*.

(4) **Sample space :** The set of all possible outcomes is called a sample space. It is denoted by S.

Example :

(a) When a dice is thrown then the sample space is $S = \{1,2,3,4,5,6\}$.

(b) when two fair coins are tossed then the sample space is $S = \{HH, HT, TH, TT\}$.

Note : The elements of the sample space is called sample point.

(5) A subset of a sample space is called an event and it is denoted by the upper case of English alphabet.

Example :

(a) When two fair coins are tossed then $A = \{HH, TT\}$ is an event because $A \subseteq S$ where $S = \{HH, HT, TH, TT\}$.

Note :

(1) The event ϕ is called *impossible event* and S is called *sure event.*

(2) Non-occurrence of an event A is denoted by $\overline{A}$.

The following table shows the set theoretic expressions to solve probability problems (when there are three events in the problem).

S.No.	Event	Set theoretic expression
1.	Only A occurs	$A \cap \overline{B} \cap \overline{C}$
2.	All the three events A, B, C occurs	$A \cap B \cap C$
3.	None of the three events occurs	$\overline{A} \cap \overline{B} \cap \overline{C}$
4.	At least one of three events occurs	$A \cup B \cup C$

Definition :

Let S be a sample space of an experiment and A be an event. Suppose that experiment is conducted N times and suppose the event A occurs f times.

Then $\frac{f}{N}$ is called the relative frequency of the event A.

Note : $0 \le \frac{f}{N} \le 1$.

Definition :

Let S be a sample space for an experiment. Suppose a real number $P(A)$ is assigned to certain subsets of S and the set function P satisfies the following conditions.

(i) $P(A) \ge 0 \ \forall A \subseteq S$

(ii) $P(S) = 1$

(iii) If $\{A_n\}$ is any finite or infinite sequence of disjoint events then $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$

Then P is called the *probability set function* and the number $P(A)$ is called the probability of the event A.

Space for Hint

Definition :

Let $|S| = N$ and $P(\{a_i\}) = \frac{1}{N}$ for each $a_i \in S$.

Then P defines a probability set function on S and $P(A) = \frac{|A|}{N}$ $\forall A \subseteq S$.

Here P is known as the *uniform probability function.*

Note :

The *uniform probability function* coincide with the classical definition of probability, viz $P(A) = \frac{\text{Number of favourable cases to } A}{\text{Total number of cases}}$.

Theorem 6. 1 :

Let S be a sample space and let $A \subseteq S$. Then $P(\overline{A}) = 1 - P(A)$ and $P(\phi) = 0$.

Proof :

Let S be a sample space and let $A \subseteq S$.

Clearly $A \cup \overline{A} = S$ and $A \cap \overline{A} = \phi$.

$\therefore$ $P(A \cup \overline{A}) = P(S)$

(i.e.) $P(A \cup \overline{A}) = 1$ -- (6.1)

From the third condition of probability set function we have

$P(A \cup \overline{A}) = P(A) + P(\overline{A})$ -- (6.2)

From (6.1), (6.2) and the second condition of probability set function, we have, $P(A) + P(\overline{A}) = 1$

(i.e.) $P(\overline{A}) = 1 - P(A)$ -- (6.3)

This proves the first part of the theorem.

By considering $A = S$.

Therefore (6.3) becomes $P(\overline{S}) = 1 - P(S)$

(i.e.) $P(\phi) = 1 - 1$

(i.e.) $P(\phi) = 0$.

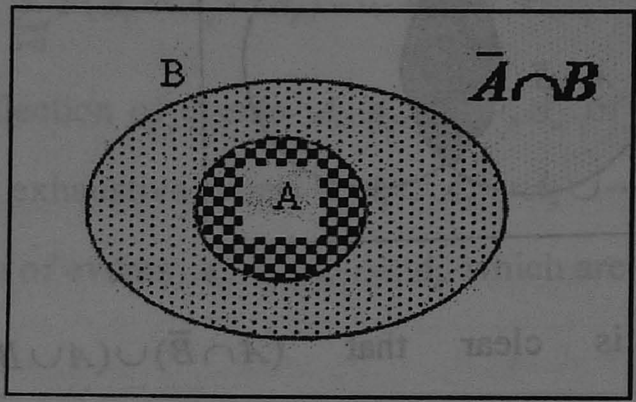
This proves the theorem.

Theorem 6. 2 :

Let A and B be two events of a sample space S such that $A \subseteq B$. Then $P(A) \leq P(B)$.

Proof :

Let A and B be two events of a sample space S such that $A \subseteq B$.



Clearly from the figure $A \cup (\bar{A} \cap B) = B$ and $A \cap (\bar{A} \cap B) = \phi$.

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

(i.e.) $P(A) + P\left(\bar{A} \cap B\right) = P(B)$ using the third condition of probability set function.

We know that $P\left(\bar{A} \cap B\right) \geq 0$

$\therefore\ P(B) - P(A) \geq 0$

(i.e.) $P(B) \geq P(A)$

(i.e.) $P(A) \leq P(B)$

This proves the theorem.

Corollary : If A is any event $0 \leq P(A) \leq 1$.

Proof of the corollary

We know that $\phi \subseteq A \subseteq S$

$\therefore\ P(\phi) \leq P(A) \leq P(S)$

(i.e.) $0 \leq P(A) \leq 1$.

The following theorem is called addition theorem of probability.

Theorem 6. 3

Let A and B be two events of a sample space S. Then

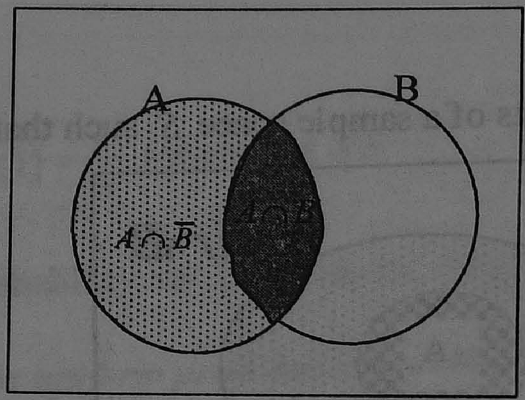
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof :

Let A and B be two events of a sample space S.

Space for Hint

Space for Hint



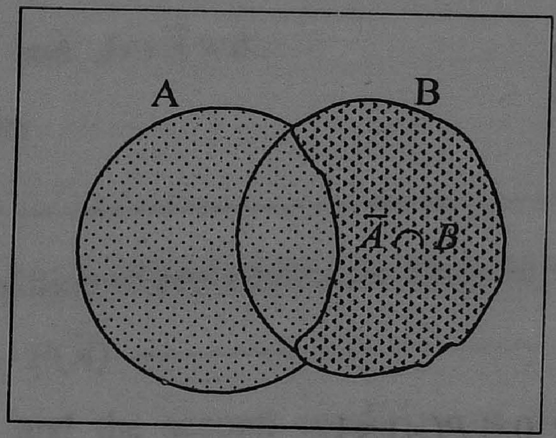
From the figure it is clear that $(A \cap \overline{B}) \cup (A \cup B) = A$ and $(A \cap \overline{B}) \cap (A \cap B) = \phi$.

Thus using third condition of probability set function, we have,

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

(i.e.) $P(A \cap \overline{B}) + P(A \cap B) = P(A)$

(i.e.) $P(A \cap \overline{B}) = P(A) - P(A \cap B)$ ------------------------------------ (6.4)



Again $B \cup (A \cap \overline{B}) = A \cup B$ and $B \cap (A \cap \overline{B}) = \phi$

$\therefore$ using third condition of probability set function, we have,

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

(i.e.) $P(A \cap \overline{B}) = P(A \cup B) - P(B)$ ------------------------------------ (6.5)

From (6.4) and (6.5), we have, $P(A) - P(A \cap B) = P(A \cup B) - P(B)$

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

Note : The addition theorem of probability can be extended to n events $A_1, A_2, A_3, \cdots, A_n$.

(i.e.) $P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{i=1}^{n}\sum_{j=1}^{n} P(A_i\cap A_j)+$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} P(A_i\cap A_j\cap A_k)-\cdots+(-1)^n P(A_1\cap A_2\cap A_3\cap\cdots\cap A_n)$$

Definition : A collection of events $A_1, A_2, A_3, \cdots, A_n$ of a sample space S is said to be mutually exhaustive events if $A_1\cup A_2\cup A_3\cup\cdots\cup A_n=S$

Note : A collection of events $A_1, A_2, A_3, \cdots, A_n$ which are mutually disjoint and exhaustive then

$$\boxed{P(A_1\cup A_2\cup A_3\cup\cdots\cup A_n)=P(A_1)+P(A_2)+P(A_3)+\cdots+P(A_n)}$$

Definition : Two events are said to be mutually exclusive events when both cannot occur simultaneously in a single trial, or, the occurrence of one event preventing the occurrence of the other and vice versa.

6. 2 Conditional Probability

Let A and B two events with $P(B)>0$. Then the conditional probability of A given B is denoted by $P(A/B)$ and it is defined as $P(A/B)=\dfrac{P(A\cap B)}{P(B)}$.

Note :

From the conditional probability of A given that of B is $P(A/B)=\dfrac{P(A\cap B)}{P(B)}$

(i.e.) $P(A\cap B)=P(B)\cdot P(A/B)$

The above relation is called *multiplication theorem for probabilities.*

Definition :

Let A and B be two events with $P(B)\neq 0$. The event A is said to be independent with the event B if $P(A/B)=P(A)$.

Space for Hint

Space for Hint

Note : If A and B are independent events then $P\left(A/B\right) = P(A)$

(i.e.) $\dfrac{P(A \cup B)}{P(B)} = P(A)$

(i.e.) $\boxed{P(A \cup B) = P(A).P(B)}$

Definition : A set of events $A_1, A_2, A_3, \cdots, A_n$ are said to be *pairwise independent* if $P(A_i \cup A_j) = P(A_i).P(A_j)$ for all $i \neq j$.

Definition : A set of events $A_1, A_2, A_3, \cdots, A_n$ are said to be *independent* if $P(A_1, A_2, A_3, \cdots, A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdots P(A_n)$.

<u>Properties of independent events</u>

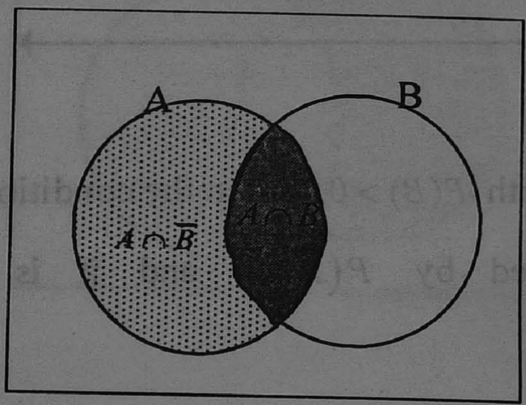
Theorem 6. 4

If A and B are independent events then A and $\overline{B}$ are independent events.

Proof

Given that A and B are independent events.

$\therefore \ P(A \cup B) = P(A).P(B)$ -- (6.6)



From the figure it is clear that $(A \cap \overline{B}) \cup (A \cup B) = A$ and $(A \cap \overline{B}) \cap (A \cap B) = \phi$.

Thus using third condition of probability set function, we have,

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

(i.e.) $P(A \cap \overline{B}) + P(A \cap B) = P(A)$

(i.e.) $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

(i.e.) $P(A \cap \overline{B}) = P(A) - P(A)P(B)$ (from (6.6))

(i.e.) $P(A \cap \overline{B}) = P(A)(1 - P(B))$

(i.e.) $P(A \cap \overline{B}) = P(A)P(\overline{B})$

Space for Hint

Hence A and $\overline{B}$ are independent events.

This proves the theorem.

Theorem 6. 5

If A and B are independent events then $\overline{A}$ and $\overline{B}$ are independent events.

Proof :

Given that A and B are independent events.

$\therefore\ P(A\cup B) = P(A).P(B)$ -- (6.7)

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

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

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

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

$= P(\overline{A}) - P(B) + P(A)P(B)$ {from (6.7)}

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

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

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

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

(i.e.) $P(\overline{A}\cap\overline{B}) = P(\overline{A})P(\overline{B})$

(i.e.) $\overline{A}$ and $\overline{B}$ are independent events

Thus if A and B are independent events then $\overline{A}$ and $\overline{B}$ are independent events.

This proves the theorem.

Theorem 6. 6

If A, B and C are mutually independent events, then $A\cup B$ and C are mutually independent events.

Proof : Given that A, B and C are mutually independent events.

$\therefore\ P(A\cap B) = P(A)P(B)$, $P(B\cap C) = P(B)P(C)$, $P(A\cap C) = P(A)P(C)$ and

$P(A\cap B\cap C) = P(A)P(B)P(C)$ -- (6.8)

Now $P((A\cup B)\cap C) = P((A\cap C)\cup(B\cap C))$

Space for Hint

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

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

$$= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \text{ \{from (6.8)\}}$$

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

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

$$= P(C) \cdot P(A \cup B)$$

Hence $P((A \cup B) \cap C) = P(A \cup B) \cdot P(C)$

(i.e.) $A \cup B$ and C are mutually independent events.

This proves the theorem.

Example 6. 1 :

What is the probability that a leap year contains 53 Fridays ?

Solution : We know that a leap year contains 52 weeks and 2 days.

$\therefore$ there must be 52 Fridays,

and the remaining two days may be the following :

(i) Sunday and Monday,

(ii) Monday and Tuesday

(iii) Tuesday and Wednesday

(iv) Wednesday and Thursday

(v) Thursday and Friday

(vi) Friday and Saturday

(vii) Saturday and Sunday

Thus the sample space contains the above seven possibilities and therefore $n(S) = 7$.

Again let A be the event of getting 53 Fridays.

Hence $n(A) = 2$.

Now P(A) = $\dfrac{n(A)}{n(S)} = \dfrac{2}{7}$.

(i.e) the probability of getting 53 Fridays is $\dfrac{2}{7}$.

Example 6. 2 :

Find the probability that a leap year contains 53 Fridays or 53 Saturdays ?

Solution : We know that a leap year contains 52 weeks and 2 days.

$\therefore$ there must be 52 Fridays,

and the remaining two days may be the following :

(i) Sunday and Monday,

(ii) Monday and Tuesday

(iii) Tuesday and Wednesday

(iv) Wednesday and Thursday

(v) Thursday and Friday

(vi) Friday and Saturday

(vii) Saturday and Sunday

Thus the sample space contains the above seven possibilities and therefore $n(S) = 7$.

Let A be the event of getting 53 Fridays and let B be the event of getting 53 Saturdays.

Thus $A \cap B = 1$

Hence $n(A) = 2$, $n(B) = 2$.

Now $\text{P(A)} = \frac{n(A)}{n(S)} = \frac{2}{7}$,

$\text{P(B)} = \frac{n(B)}{n(S)} = \frac{2}{7}$,

$\text{P}(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{7}$,

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

$$= \frac{2}{7} + \frac{2}{7} - \frac{1}{7}$$

$$= \frac{3}{7}$$

(i.e) the probability of getting 53 Fridays or 53 Saturdays is $\frac{3}{7}$.

Space for Hint

Space for Hint

Example 6. 3 :

A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is $\frac{1}{7}$ and that of wife's selection is $\frac{1}{5}$.

What is probability that

(a) both of them will be selected,

(b) only one of them selected and

(c) none of them will be selected?

Solution :

Let A, B be the events that denote husband, wife will be selected respectively.

Given that $P(A)=\frac{1}{7}$ and $P(B)=\frac{1}{5}$.

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

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

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

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

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

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

(i) The probability that both husband and wife will be selected

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

$$= P(A)P(B)$$

$$=\frac{1}{7}\cdot\frac{1}{5}$$

$$=\frac{1}{35}$$

(ii) The probability that only husband or wife will be selected

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

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

$$=\frac{1}{7}\cdot\frac{4}{5}+\frac{6}{7}\cdot\frac{1}{5}$$

Space for Hint

$$= \frac{4+6}{35}$$

$$= \frac{10}{35}$$

$$= \frac{2}{7}$$

(iii) The probability that both husband and wife will not be selected

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

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

$$= \frac{6}{7} \cdot \frac{4}{5}$$

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

Example 6. 4 :

A piece of electronic equipment has two identical parts A and B. In the past, part A has failed 40% of the time; part B 50% of the time. Parts A and B operate independently. Assume that both parts must operate to enable the equipment to function. What is the probability that the equipment will function?

Solution :

Let A and B be the events that part A and B fails to function respectively.

Given that $P(A) = 40\%$

$= 0.40$

and $P(B) = 50\%$

$= 0.50$

The probability that the equipment will function

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

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

$$= (0.6)(0.5)$$

$$= 0.30$$

Space for Hint

Example 6. 5 :

If $P(A)=\frac{1}{2}$, $P(B)=\frac{1}{2}$, $P(A\cup B)=\frac{2}{3}$ find (i) $P(A\cap B)$, (ii) $P(\overline{A}\cap B)$ and (iii) $P(\overline{A}\cup B)$

Solution :

Given that $P(A)=\frac{1}{2}$, $P(B)=\frac{1}{2}$, $P(A\cup B)=\frac{2}{3}$.

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

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

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

(ii) $P(\overline{A}\cap B)=P(B)-P(A\cap B)$

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

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

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

$$=\frac{1}{2}+\frac{1}{2}-\frac{1}{6}$$

$$=\frac{5}{6}.$$

Example 6. 6 :

Let A and B be two events in a sample space S. If $P(A\cup B)=\frac{5}{6}$, $P(A\cap B)=\frac{1}{3}$ and $P(\overline{B})=\frac{1}{2}$. Find (i) $P(B)$, (ii) $P(A)$ and (iii) $\overline{A}$ and $\overline{B}$ are independent.

Solution :

Given that $P(A\cup B)=\frac{5}{6}$, $P(A\cap B)=\frac{1}{3}$ and $P(\overline{B})=\frac{1}{2}$.

(i) $P(B)=1-P(\overline{B})$

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

(ii) We know that $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

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

(i.e.) $P(A)=\frac{5}{6}+\frac{1}{3}-\frac{1}{2}$

(i.e.) $P(A)=\frac{5+2-3}{6}$

(i.e.) $P(A)=\frac{2}{3}$

(iii) Now $P(A)\cdot P(B) = \frac{1}{2}\cdot\frac{1}{3}$

$= \frac{1}{6}$

$= P(A\cap B)$

$\therefore$ $\bar{A}$ and $\bar{B}$ are independent.

Example 6. 7:

A, B, C are any three independent events such that $P(A)=0.3$, $P(B)=0.2$ and $P(C)=0.1$. Find the probability of occurrence of at least one of three events.

Solution :

Given that $P(A)=0.3$, $P(B)=0.2$ and $P(C)=0.1$.

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

$= 1-0.3$

$= 0.7$,

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

$= 1-0.2$

$= 0.8$,

and $P(\bar{C})=1-P(AC)$

$= 1-0.1$

$= 0.9$

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

$= (0.7)(0.8)(0.9)$

$= 0.504$

Space for Hint

Space for Hint

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

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

$=1-0.504$

$=0.496$

Example 6. 8 :

If A and B are mutually exclusive events such that $P(B)=2P(A)$ and $A\cup B=S$, find $P(A)$.

Solution :

Given that $P(B)=2P(A)$ and $A\cup B=S$.

Now $A\cup B=S$

$\Rightarrow\ P(A\cup B)=P(S)$

$\Rightarrow\ P(A)+P(B)=1$ {since A, B are independent events}

$\Rightarrow\ P(A)+2P(A)=1$

$\Rightarrow 3P(A)=1$

$\Rightarrow P(A)=\dfrac{1}{3}$

Example 6. 9:

Let A and B are two events such that $P(A)=\dfrac{3}{4}$ and $P(B)=\dfrac{5}{8}$. Show that

(i) $P(A\cup B)\geq\dfrac{3}{8}$ and (ii) $\dfrac{3}{8}\leq P(A\cap B)\leq\dfrac{5}{8}$.

Solution :

Given that $P(A)=\dfrac{3}{4}$ and $P(B)=\dfrac{5}{8}$.

We know that $P(A\cap B)\leq 1$

(i.e.) $P(A)+P(B)-P(A\cup B)\leq 1$

(i.e.) $\dfrac{3}{4}+\dfrac{5}{8}-P(A\cup B)\leq 1$

(i.e.) $P(A\cup B)\geq\dfrac{3}{8}$

This proves (i).

Proof of (ii) :

We know that $P(A \cup B) \leq 1$

(i.e.) $P(A)+P(B)-P(A \cap B) \leq 1$

(i.e.) $\frac{3}{4}+\frac{5}{8}-P(A \cap B) \leq 1$

(i.e.) $P(A \cap B) \geq \frac{3}{8}$ -- (6.9)

Again $A \cap B \subseteq B$

$\Rightarrow P(A \cap B) \leq P(B)$

$\Rightarrow P(A \cap B) \leq \frac{5}{8}$ -- (6.10)

Form (6.9) and (6.10) we have $\frac{3}{8} \leq P(A \cap B) \leq \frac{5}{8}$.

Example 6. 10

A can hit a target four times in five shots; B three times in four shots and C twice in three shots. They each fire once at the same target. What is the probability that at least two shots hit the target?

Solution :

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

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

$= 1-\frac{4}{5}$

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

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

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

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

and $P(\bar{C})=1-P(AC)$

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

Space for Hint

$= \frac{1}{3}$

Thus the probability that at least two shots hit the target

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

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

$= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} + \frac{4}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} + \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3}$

$= \frac{12+8+6+24}{5 \times 4 \times 3}$

$= \frac{5}{6}$

Example 6. 11 :

In a shooting test the probability of hitting the target are $\frac{1}{2}$ for A, $\frac{2}{3}$ for B and $\frac{3}{4}$ for C. If all of them fore at the same target find the probabilities that

(i) only one of them hits the target, (ii) at least one of them hits the target.

Solution :

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

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

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

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

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

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

$= \frac{1}{3}$,

and $P(\bar{C}) = 1 - P(AC)$

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

Space for Hint

$= \frac{1}{4}$

(i) Thus the probability that only one hit the target

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

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

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

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

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

(ii) Now the probability that none of them hit the target

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

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

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

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

Thus the probability that at least one hit the target

$$= P(A \cup B \cup C)$$

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

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

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

Example 6. 12 :

Saravan can solve 75% of the problems given in Mathematics book, where as Sree can only 90%. What is the probability that the problem being solved ?

Solution :

Let A be the event that Saravan can solve the problem.

Let B be the event that Sree can solve the problem.

Given the P(A) = 75% = 0.75 and P(B) = 90% = 0.90.

Thus $P(\bar{A})$ = 1– P(A) = 1 – 0.75 = 0.25

Space for Hint

and $P(\overline{B}) = 1 - P(B) = 1 - 0.90 = 0.10$

Probability that both not solve the problem = $P(\overline{A} \cap \overline{B})$

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

$= 0.25 \times 0.10$

$= 0.025$

Thus the problem being solved by any one

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

$= 1 - 0.025$

$= 0.975$

Example 6. 13 :

In a survey of 100 readers, it was found that 40 read Hindu magazine, 15 read Indian Express magazine, 10 reads both. Find the probability that a person read at least one of the magazines.

Solution :

Let A be the event that the person reads Hindu magazine.

Let B be the event that the person reads Indian Express magazine.

Given the n(S) = 100, n(A) = 40, n(B) = 15 and $n(A \cap B) = 10$.

Thus $P(A) = \frac{40}{100} = 0.4$,

$P(B) = \frac{15}{100} = 0.15$

and $P(A \cap B) = \frac{10}{100} = 0.10$

$\therefore$ the probability that the person read at least one of magazines

$= P(A \cup B)$

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

$= 0.4 + 0.15 - 0.10$

$= 0.45.$

Space for Hint

Example 6. 14 :

State and prove **generalized Boole's inequality.**

Statement : If $A_1, A_2, A_3, \cdots, A_n$ are events in a sample space S then $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \geq 1-(P(\bar{A}_1)+P(\bar{A}_2)+P(\bar{A}_3)+\cdots+P(\bar{A}_n))$.

Proof :

Step 1 : First we shall prove the result for two events A_1 and A_2.

We know that $0 \leq P(A_1 \cup A_2) \leq 1$

$\therefore\ 1-P(A_1 \cup A_2) \geq 0$ -- (6.11)

Again $P(A_1 \cap A_2) = P(A_1)+P(A_2)-P(A_1 \cup A_2)$

$= 1-P(\bar{A}_1)+1-P(\bar{A}_2)-P(A_1 \cup A_2)$

$= [1-P(\bar{A}_1)-P(\bar{A}_2)]+[1-P(A_1 \cup A_2)]$

$\geq 1-P(\bar{A}_1)-P(\bar{A}_2)$ {form (6.11)}

Thus $P(A_1 \cap A_2) \geq 1-P(\bar{A}_1)-P(\bar{A}_2)$

Hence the result true for two events.

Step 2 : Now we shall prove the result for n events.

Now $A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n = A_1 \cap B$ where $B = A_2 \cap A_3 \cap \cdots \cap A_n$

Now from step 1, $P(A_1 \cap B) \geq 1-P(\bar{A}_1)-P(\bar{B})$

(i.e.) $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \geq 1-P(\bar{A}_1)-P(\overline{A_2 \cap A_3 \cap \cdots \cap A_n})$

$= 1-P(\bar{A}_1)-P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 \cup \cdots \cup \bar{A}_n)$

$\geq 1-P(\bar{A}_1)-\left[P(\bar{A}_1)+P(\bar{A}_2)+P(\bar{A}_3)+\cdots+P(\bar{A}_n)\right]$

$= 1-(P(\bar{A}_1)+P(\bar{A}_2)+P(\bar{A}_3)+\cdots+P(\bar{A}_n))$

Hence $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \geq 1-(P(\bar{A}_1)+P(\bar{A}_2)+P(\bar{A}_3)+\cdots+P(\bar{A}_n))$

This prove the generalized Boole's indequality

Check Your Progress

(1) If A, B, C are any three events in a sample space and if A, B, C are pair-wise independent and A is independent of $B \cup C$ then A, B, C are mutually independent.

Space for Hint

(2) If A and B are two events such that $P(A)=\frac{1}{3}$, $P(B)=\frac{3}{4}$ and $P(A\cup B)=\frac{11}{12}$, find $P(A/B)$ and $P(B/A)$.

(3) Among the workers in a factory only 30% receive a bonus. Among those receiving the bonus only 20% are skilled. What is the probability of a randomly selected workers who is skilled and receiving the bonus.

(4) A letter of the English alphabet is chosen at random. Calculate the probability that the letter so chosen (i) is a vowel, (ii) precedes m and is a vowel and (iii) follows m and is a vowel.

(5) The probability that a student Mr. Saravanan will passed Mathematics is $\frac{2}{3}$, the probability that he passes Statistics is $\frac{4}{9}$. If the probability of passing at least one subject is $\frac{4}{5}$, what is the probability that Mr. Saravanan will pass both the subjects?

(6) A class consists of 100 students, 25 of them are girls and 75 boys, 20 of them are rich and remaining poor, 40 of them are fair complexioned. What is the probability of selecting a fair complexioned rich girl?

6. 3 Baye's Theorem

State and prove Baye's theorem.

Statement:

Let $A_1, A_2, A_3, \cdots, A_n$ be a collection of mutually exclusive and exhaustive events in a sample space S such that $P(A_i)>0$ for all i. Let B be any event with $P(B)>0$. Then $P(A_i/B)=\dfrac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)}$

Proof :

Let $A_1, A_2, A_3, \cdots, A_n$ be a collection of mutually exclusive and exhaustive events in a sample space S.

$\therefore\ A_1\cup A_2\cup A_3\cup\cdots\cup A_n=S$ and $A_i\cap A_j=\phi\ \forall i,j$

Space for Hint

Now $P(B/A_i)=\dfrac{P(A_i\cap B)}{P(A_i)}$

$\Rightarrow P(A_i\cap B)=P(A_i)\cdot P(B/A_i)$ for $i=1,2,3,\cdots,n$. ----------- (6.12)

Now $B = B\cap S$

$= B\cap(A_1\cup A_2\cup A_3\cup\cdots\cup A_n)$

$= (B\cap A_1)\cup(B\cap A_2)\cup(B\cap A_3)\cup\cdots\cup(B\cap A_n)$

Again $A_i\cap A_j=\phi\ \forall i,j$

$\therefore\ B\cap(A_i\cap A_j)=\phi\ \forall i,j$

(i.e.) $(B\cap A_i)\cap(B\cap A_j)=\phi\ \forall i,j$

(i.e.) $B\cap A_i$ for $i=1,2,3,\cdots,n$ are mutually excusive.

$\therefore\ P(B)=P((B\cap A_1)\cup(B\cap A_2)\cup(B\cap A_3)\cup\cdots\cup(B\cap A_n))$

$\Rightarrow P(B)=P(B\cap A_1)+P(B\cap A_2)+P(B\cap A_3)+\cdots+P(B\cap A_n)$

$\Rightarrow P(B)=P(A_1)\cdot P(B/A_1)+P(A_2)\cdot P(B/A_2)+\cdots+P(A_n)\cdot P(B/A_n)$ {by (6.12)

}

$\Rightarrow P(B)=\sum_{i=1}^{n}P(A_i)P(B/A_i)$ --- (6.13)

Thus $P(A_i/B)=\dfrac{P(A_i\cap B)}{P(B)}$

$\Rightarrow P(A_i/B)=\dfrac{P(A_i)\cdot P(B/A_i)}{P(B)}$ {from (6.12)}

$\Rightarrow P(A_i/B)=\dfrac{P(A_i)\cdot P(B/A_i)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)}$ {from (6.13)}

This proves the Baye's theorem.

Example 6. 15

Bowl I contains 3 red chips and 7 blue chips. Bowl II contains t red chips and 4 blue chips. A bowl is selected at random and then 1 chip is drawn from this bowl.

(i) Compute the probability that this chip is red.

Space for Hint

(ii) Relative to the hypothesis that the chip is red, find the conditional probability that it is drawn from bowl II.

Solution :

Let A_1, A_2 be the event of selecting either Bowl I or Bowl II respectively.

$\therefore\ P(A_1)=\dfrac{1}{2}$ and $P(A_2)=\dfrac{1}{2}$.

(i) Let B be the event of selecting 1 red chip.

Now probability of selecting a red chip from bowl I

$$= P(B/A_1)$$

$$= \frac{{}^{3}C_1}{{}^{10}C_1}$$

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

and probability of selecting a red chip from bowl II

$$= P(B/A_2)$$

$$= \frac{{}^{6}C_1}{{}^{10}C_1}$$

$$= \frac{6}{10}$$

$\therefore$ Probability of selecting red chip

$$= P(B)$$

$$= P(A_1)\cdot P(B/A_1)+P(A_2)\cdot P(B/A_2)$$

$$= \frac{1}{2}\cdot\frac{3}{10}+\frac{1}{2}\cdot\frac{6}{10}$$

$$= \frac{9}{20}$$

(ii) The conditional probability that drawing a red chip from bowl II.

$$= P(A_2/B)$$

$$= \frac{P(A_2)P(B/A_2)}{P(B)}$$

$$= \frac{\frac{1}{2}.\frac{6}{10}}{\frac{9}{20}}$$

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

Example 6. 16

The probability of Saravanan, Uma and Sree Vathssa becoming manager is $\frac{4}{9}$, $\frac{2}{9}$, and $\frac{1}{3}$ respectively. The probabilities that the bonus scheme will be introduced by Saravanan, Uma and Sree Vathssa are $\frac{3}{10}$, $\frac{1}{2}$ and $\frac{4}{5}$ respectively. What is the probability that the bonus scheme will be introduced? If the bonus scheme has been introduced, what is the probability the manager appointed was Sree Vathssa?

Solution :

Let A_1 be the event that Saravanan becoming manager.

Let A_2 be the event that Uma becoming manager.

Let A_3 be the event that Sree Vathssa becoming manager.

Let B be the event that bonus scheme be introduced.

Event	$P(A_i)$	$P(B/A_i)$	$P(A_i)P(B/A_i)$
A_1	$\frac{4}{9}$	$\frac{3}{10}$	$\frac{4}{9}\times\frac{3}{10} = \frac{12}{10} = \frac{6}{45}$
A_2	$\frac{2}{9}$	$\frac{1}{2}$	$\frac{2}{9}\times\frac{1}{2} = \frac{1}{9}$
A_3	$\frac{1}{3}$	$\frac{4}{5}$	$\frac{1}{3}\times\frac{4}{5} = \frac{4}{15}$
Total			$\frac{23}{45}$

Space for Hint

Space for Hint

(i) The probability the bonus scheme be introduced

$$= P(B)$$

$$= \sum_{i=1}^{3} P(A_i)P(B/A_i)$$

$$= \frac{6}{45} + \frac{1}{9} + \frac{4}{15}$$

$$= \frac{23}{45}$$

(ii) The probability that the bonus scheme was introduced by the manager Sree Vathssa = $P(A_3/B)$

$$= \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

$$= \frac{\frac{4}{15}}{\frac{23}{45}}$$

$$= \frac{4}{\cancel{15}} \times \frac{\cancel{45}^{\,3}}{23}$$

$$= \frac{12}{23}.$$

Example 6. 17 :

A factory manufacturing televisions has four units A, B, C, D. The units A, B, C, D manufactures 15%, 20%, 30%, 35% of the total output respectively. it was found that out of their outputs 1%, 2%, 2% and 3% defectives. A television is chosen at random from the output and found to be defective. What is the probability that, it came from unit A ?

Solution : Let A_1, A_2, A_3, A_4 be the events that the television be produced from unit A, B, C, D respectively.

Thus we have the following table written according to the given data.

Event	$P(A_i)$	$P(B/A_i)$	$P(A_i)P(B/A_i)$
A_1	$15\% = 0.15$	$1\% = 0.01$	0.015
A_2	$20\% = 0.20$	$2\% = 0.02$	0.040
A_3	$30\% = 0.30$	$2\% = 0.02$	0.060
A_4	$35\% = 0.35$	$3\% = 0.03$	0.0105
Total			0.1255

Space for Hint

The probability that the chosen defective television produced from unit A

$$= P(A_1/B)$$

$$= \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

$$= \frac{0.015}{0.1255}$$

$$= 0.1195.$$

Example 6. 18 :

In the coming year there will be three candidates Dr. Saran, Dr. Karan and Dr. Maran for the position of Pricipalship in a well established codeduational college whose chances of getting appointment are in the ration $4:2:3$ respectively. The probability that Dr. Saran if appointed will introduce M.B.A. course in the college is 0.3. The probability of Dr. Karan and Dr. Maran doing the same are respectively 0.5 and 0.8. Find the probability that M.B.A is introduced in the college next year.

Solution :

Let A_1, A_2, A_3 be the events respectively that the Dr. Saran, Dr. Karan and Dr. Maran will selected as Principal.

Thus we have the following table written according to the given data.

Space for Hint

Event	$P(A_i)$	$P(B/A_i)$	$P(A_i)P(B/A_i)$
A_1	$\frac{4}{9}$	$\frac{3}{10}$	$\frac{4}{9}\times\frac{3}{10}=\frac{12}{10}=\frac{6}{45}$
A_2	$\frac{2}{9}$	$\frac{1}{2}$	$\frac{2}{9}\times\frac{1}{2}=\frac{1}{9}$
A_3	$\frac{1}{3}$	$\frac{4}{5}$	$\frac{1}{3}\times\frac{4}{5}=\frac{4}{15}$
Total			$\frac{23}{45}$

The probability that M.B.A course will be introduced next year in the college

$$=\sum_{i=1}^{n} P(A_i)P(B/A_i)$$

$$=\frac{23}{45}$$

Check Your Progress

(1) Suppose that a product is produced in three factories X, Y and Z. It is known that factory X produces thrice as many items as factory Y, and that factories Y and Z produce the same number of products. Assume that it is known 3 percent of the items produced by each of the factories X and Z are defective while 5 per cent of those manufactured by factory Y are defective. All the items produced in three factories are stocked, and an item of product is selected at random. What is the probability that this item is defective?

(2) Urn I contains one white, 2 black, 3 red balls; urn II contains 2 white, one black, one red balls and urn III contains 4 white, 5 black, 3 red balls. One urn is selected at random and two balls are drawn. They happen to be white and red. Find the probability that they came from urn III.

(3) First factory produces 1000 toys and 20 of them being defective, second factory produces 4000 toys 40 of them being defective and the third factory produces 5000 toys, 50 of them being defective. All these toys are put in one stock pile. One of them is chosen and is found to be defective. What is the probability that it is from the second factory.

(4) In a state election in 2005 there were 3 major political parties X, Y, Z fighting of Chief Ministership. The chances of winning the election of the three parties are in the ratio $1:2:3$ respectively. The probability that the party X if selected, will introduce total prohibition in that state is $\frac{1}{2}$. The probability that the party y if selected, will introduce total prohibition in that state is $\frac{1}{4}$ and the probability that the party Z if selected, will introduce total prohibition in that state is $\frac{3}{4}$. What is the probability that there will be a total prohibition in the state after the election in 2005?

(5) If A and B are two events and $P(B) \neq 1$ prove that $P(A/B) = \dfrac{P(A) - P(A \cap B)}{1 - P(B)}$ and hence prove that $P(A \cap B) = P(A) + P(B) - 1$.

(6) In an examination, 30% of the students have failed in Mathematics, 20% of the students have failed in Chemistry and 10% failed in both Mathematics and Chemistry. A student is selected at random.

(i) What is the probability that the student has failed in Mathematics if it is known that he has failed in chemistry?

(ii) What is the probability that the student has failed either in Mathematics or in Chemistry?

Space for Hint

Space for Hint

(7) A manufacturing firm produces pipes in two plants I and II with daily production 1500 and 2000 pipes respectively. The fraction of defective pipes produced by two plants I and II are 0.006 and 0.008 respectively. If a pipe is selected at random from the daily production is found to be defective, what is the probability that it has come from plant I, plant II ?

(8) A factory produces a certain type of outputs by three types of machine. The respective daily production figure are :

Machine I	:	3000 units
Machine II	:	2500 units
Machine III	:	4500 units

Past experience shows that 1 per cent of the output produced by Machine I is defective. The corresponding fraction of defectives for the other two machines are 1.2 per cent and 2 per cent respectively. An item is drawn at random from the day's production run and is found to be defective. What is the probability that it comes from the output of Machine III?

SUMMARY

In this unit, we came to know that the calculation of probability of an event, addition theorem of probability, multiplication theorem of probability, and Bayes' theorem. Also we have discussed how to apply Baye's theorem.

Unit VII

Mathematical Expectation

Objectives

In this unit, we are going to discuss the probability density function of random variable, distribution function of a random variable, mathematical expectation and it properties. Also we shall discuss the method finding mathematical expectation for discrete and continuous random variables.

7. 1 Mathematical Expectation

Definition :

A real valued function defined on a sample space is called a random variable. It is denoted by X.

Let $X : S \to \mathbf{R}$ be random variable. The space of the random variable is $\mathscr{A} = \{X(x) \,/\, s \in S\}$.

Let $A \subseteq \mathscr{A}$ and $B = \{x \in S \,/\, X(x) \in A\} = X^{-1}(A)$.

The probability of A is defined as $P(A) = P(B) = P(X^{-1}(A))$

Note : The function P is a probability set function on $\mathscr{A}$.

Notation :

A random variable is generally denoted by the capital letters $X, Y, Z, \cdots$ and its realization value by small letters $x, y, z, \cdots$ If x is a real number then $\{\omega \in S \,/\, X(\omega) = x\}$ is denoted by $X = x$.

Hence $P(X = x) = P(\{\omega / X(\omega) = x\})$

Similarly if $a, b \in \mathbf{R}$ then $P(X \leq a) = P(\{\omega / X(\omega) \in (-\infty, a]\})$ and

$P(a < X \leq b) = P(\{\omega / a < X(\omega) \leq b\})$

Space for Hint

Space for Hint

Definition :

Let X be a random variable. Then the function $F:\mathbf{R}\to\mathbf{R}$ defined by $F(x)=P(X\le x)$ where $-\infty<x<\infty$ is called a distribution function of the random variable X.

Problem :

If $F(x)$ is a distribution function of the random variable X and if $a<b$ then $P(a<X\le b)=F(b)-F(a)$.

Proof :

Let $F(x)$ is a distribution function of the random variable X and if $a<b$.

Clearly the events $X\le a$ and $a<X\le b$ are mutually exclusive events and the union is $X\le b$.

$\therefore\ P(X\le a)+P(a<X\le b)=P(X\le b)$

$\Rightarrow\ P(a<X\le b)=P(X\le b)-P(X\le a)$

$\Rightarrow\ P(a<X\le b)=F(b)-F(a)$.

This proves the problem.

7. 2 Discrete random variable

Definition :

If a random variable X takes at most countable number of values $x_1,x_2,x_3,\cdots$ is called a **discrete random variable.**

Note :

The distribution function of a random variable X having probability density function is given by $\boxed{F(x)=P(X\le x)=\sum_{x_i\le x}p_i}$

Example 7. 1 :

Let $p(x)=\begin{cases}\dfrac{x}{15} & ;\ x=1,2,3,4,5\\ 0 & ;\ otherwise\end{cases}$ be the probability density function of the discrete random variable X. Find (i) $P(X=1\ or\ 2)$, (ii) $P\left(\dfrac{1}{2}\le X\le\dfrac{5}{2}\right)$ and (iii) $P(1\le X\le 2)$.

Solution :

Given that $p(x)=\begin{cases}\dfrac{x}{15} & ; x=1,2,3,4,5\\ 0 & ; otherwise\end{cases}$ is a probability density function of a random variable X.

(i) Thus $P(X=1 \text{ or } 2) = p(1)+p(2)$

$$= \frac{1}{15}+\frac{2}{15}$$

$$= \frac{3}{15}$$

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

(ii) now $P\left(\frac{1}{2}\le X \le \frac{5}{2}\right) = p(1)+p(2)$

$$= \frac{1}{15}+\frac{2}{15}$$

$$= \frac{3}{15}$$

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

(iii) and $P(1\le X \le 2) = p(1)+p(2)$

$$= \frac{1}{15}+\frac{2}{15}$$

$$= \frac{3}{15}$$

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

Example 7. 2 :

A random variable X has the following probability distribution.

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

(i) Determine the value of a,

(ii) Find $P(X<3)$; $P(X\ge 3)$,

(iii) $P(0<X<5)$ and

Space for Hint

Space for Hint

(iv) Find the distribution of X

Solution :

x	$p(x)$	$F(x) = P(X \le x)$
0	a	$\frac{1}{81}$
1	$3a$	$\frac{4}{81}$
2	$5a$	$\frac{9}{81}$
3	$7a$	$\frac{16}{81}$
4	$9a$	$\frac{25}{81}$
5	$11a$	$\frac{36}{81}$
6	$13a$	$\frac{49}{81}$
7	$15a$	$\frac{64}{81}$
8	$17a$	$\frac{81}{81}$
Total	$81a$	

(i) Now $\sum_{i=0}^{8} p(x_i) = 1$

(i.e.) $81a = 1$

$\therefore \; a = \frac{1}{81}$,

) $P(X<3) = P(X \le 2)$

$$= F(2)$$

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

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

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

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

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

(iii) $P(0 < X < 5) = P(0 < X \leq 4)$

$$= F(4) - F(0)$$

$$= \frac{25}{81} - \frac{1}{81}$$

$$= \frac{24}{81}$$

$$= \frac{8}{27},$$

(iv) the last column of the table shows the distribution function of X.

Example 7. 3 :

The probability density function of random variable X is $p(x) = \begin{cases} \frac{1}{3} & ; \ x = -1, 0, 1 \\ 0 & ; \ otherwise \end{cases}$. Find the distribution function of X.

Solution :

Given that $p(x) = \begin{cases} \frac{1}{3} & ; \ x = -1, 0, 1 \\ 0 & ; \ otherwise \end{cases}$ is a probability density function of the random variable X.

x	$p(x)$	$F(x) = P(X \leq x)$
-1	$\frac{1}{3}$	$\frac{1}{3}$
0	$\frac{1}{3}$	$\frac{2}{3}$
1	$\frac{1}{3}$	$\frac{3}{3} = 1$

The last column of the table shows the distribution function of X

Space for Hint

Example 7. 4 :

The probability density function of random variable X is $p(x) = \begin{cases} \frac{x}{15} & ; \; x = 1,2,3,4,5 \\ 0 & ; \; otherwise \end{cases}$ Find the distribution function of X.

Solution :

Given that $p(x) = \begin{cases} \frac{1}{3} & ; \; x = -1,0,1 \\ 0 & ; \; otherwise \end{cases}$ is a probability density function of the random variable X.

x	$p(x)$	$F(x) = P(X \le x)$
1	$\frac{1}{15}$	$\frac{1}{15}$
2	$\frac{2}{15}$	$\frac{3}{15}$
3	$\frac{3}{15}$	$\frac{6}{15}$
4	$\frac{4}{15}$	$\frac{10}{15}$
5	$\frac{5}{15}$	$\frac{15}{15} = 1$

The last column of the table shows the distribution function of X

Check Your Progress

(1) Let X be discrete random variable with the following probability density

x	-2	1	2	4
$p(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

(2) Find the constant c so that $p(x)=\begin{cases} c\left(\frac{2}{3}\right)^x & ; x=1,2,3,\cdots \\ 0 & ; otherwise \end{cases}$ is a probability density function of a random variable X.

7. 3 Continuous random variable

Definition :

A random variable X is said to be a continuous random variable if its range is uncountable or infinite.

Definition :

Let X be a continuous random variable taking the values in the interval $(-\infty,\infty)$. Let $f(x)$ be a function such that

(i) $f(x)$ is integrable in $(-\infty,\infty)$.

(ii) $f(x)\geq 0$ for all $-\infty<x<\infty$.

(iii) $\int\limits_{-\infty}^{\infty} f(x)\,dx=1$.

Then $f(x)$ is called the **probability density function** of the continuous random variable X.

Note :

(i) If $f(x)$ is called the **probability density function** of the continuous random variable X and $A=\{x/a<x<b\}$ then $P(A)$ is defined as

$$P(A)=P(a<X<b)=\int\limits_a^b f(x)\,dx$$

(ii) If $A=\{a\}$ then $P(A)=P(a<X<a)$

$$=\int\limits_a^a f(x)\,dx$$

$$=0$$

(iii) $P(a<X\leq b)\ =\ P(a\leq X\leq b)$

Space for Hint

Space for Hint

Definition :

Let X be a continuous random variable with probability density function $f(x)$. Define $F:\mathbf{R}\to\mathbf{R}$ given by $F(x)=\int_{-\infty}^{x} f(t)\,dt$. Then $F(x)$ is called the distribution function of the continuous random variable X.

Properties of distribution function $F(x)$

(i) $F(\infty)=1$,

(ii) $F(-\infty)=0$,

(iii) $P(a\le X\le b)=F(b)-F(a)$,

(iv) $F(x)$ is an increasing function of x.

Proof :

(i) $F(\infty) = \lim_{x\to\infty} F(x)$

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

$$=1$$

Thus $F(\infty)=1$.

(ii) $F(-\infty) = \lim_{x\to-\infty} F(x)$

$$=0$$

Thus $F(-\infty)=0$.

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

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

$$= F(b)-F(a)$$

Hence $P(a\le X\le b)=F(b)-F(a)$.

(iv) Let $a<b$.

We know that $P(a\le X\le b)\ge 0$

(i.e.) $F(b)-F(a)\ge 0$

$\therefore\ F(b)\ge F(a)$

Thus $a < b \Rightarrow F(b) \geq F(a)$

(i.e.) $F(x)$ is an increasing function of x.

Example 7. 5 :

If $f(x) = \begin{cases} k(2x+3) & ; 0 < x < 2 \\ 0 & ; otherwise \end{cases}$ is the probability density function of the continuous random variable X. Find k and also find the distribution function of X.

Solution :

Given that $f(x) = \begin{cases} k(2x+3) & ; 0 < x < 2 \\ 0 & ; otherwise \end{cases}$ is the probability density function of the continuous random variable X.

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

(i.e.) $\int_{-\infty}^{0} f(x)\, dx + \int_{0}^{2} f(x)\, dx + \int_{2}^{\infty} f(x)\, dx = 1$

(i.e.) $0 + \int_{0}^{2} k(x+2)\, dx + 0 = 1$

(i.e.) $k\left[x^2 + 3x\right]_0^2 = 1$

(i.e.) $k[(4+6)-0] = 1$

(i.e.) $k = \frac{1}{10}$

Thus $f(x) = \begin{cases} \dfrac{2x+3}{10} & ; 0 < x < 2 \\ 0 & ; otherwise \end{cases}$

Let $F(x)$ be the distribution function of X.

If $x \leq 0$ then $F(x) = P(X \leq x) = 0$.

If $0 < x < 2$ then $F(x) = P(X \leq x)$

$$= \int_{-\infty}^{x} f(t)\, dt$$

$$= \int_{-\infty}^{0} f(t)\, dt + \int_{0}^{x} f(t)\, dt$$

Space for Hint

Space for Hint

$$= 0 + \int_0^x \frac{2t+3}{10}\,dt$$

$$= \frac{1}{10}\left[t^2 + 3t\right]_0^x$$

$$= \frac{1}{10}\left[x^2 + 3x\right]$$

If $x \geq 2$ then $F(x) = P(X \leq x)$

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

$$= 1$$

$$\text{Thus } F(x) = \begin{cases} 0 & ; x \leq 0 \\ \dfrac{x^2+3x}{10} & ; 0 < x < 2 \\ 1 & ; x \geq 2 \end{cases}$$

Example 7. 6 :

Is the function defined as follows a probability density function?

$$f(x) = \begin{cases} \dfrac{3+2x}{18} & ; 2 < x < 4 \\ 0 & ; otherwise \end{cases}$$ If so find the probability that a random variable X having the probability density function will fall in the interval $2 \leq x \leq 3$.

Solution :

$$\text{Given that } f(x) = \begin{cases} \dfrac{3+2x}{18} & ; 2 < x < 4 \\ 0 & ; otherwise \end{cases}$$

$$\text{Now } \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{2} f(x)\,dx + \int_{2}^{4} f(x)\,dx + \int_{4}^{\infty} f(x)\,dx$$

$$= 0 + \int_2^4 \frac{3+2x}{18}\,dx + 0$$

$$= \frac{1}{18}\left[3x + x^2\right]_2^4$$

$$= \frac{1}{18}[(12+16) - (6+4)]$$

$$= \frac{1}{18}[18]$$

$$= 1$$

Thus $f(x)$ is a probability density function of X.

Now $P(2 \le X \le 3) = \int_{2}^{3} \frac{3+2x}{18} dx$

$$= \int_{2}^{3} \frac{3+2x}{18} dx$$

$$= \frac{1}{18}\left[3x + x^2\right]_2^3$$

$$= \frac{1}{18}[(9+9)-(6+4)]$$

$$= \frac{1}{18}[8]$$

$$= \frac{4}{9}$$

Example 7. 7 :

The distribution function of a random variable X is given by $F(x) = \begin{cases} 1-(1+x)e^{-x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$. Find the corresponding density function of the random variable X.

Solution :

Given the distribution function of X is $F(x) = \begin{cases} 1-(1+x)e^{-x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$

Let $f(x)$ be the probability density function of X.

Thus $f(x) = \frac{d}{dx}\left(F(x)\right)$

$$= \begin{cases} (1+x)e^{-x} - e^{-x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$$

$$= \begin{cases} xe^{-x} & ; x \ge 0 \\ 0 & ; x < 0 \end{cases}$$

Space for Hint

Space for Hint

Thus the probability density function of X is $f(x)=\begin{cases} xe^{-x} & ; x \geq 0 \\ 0 & ; x<0 \end{cases}$.

Check Your Progress

(1) Given the distribution function $F(x)=\begin{cases} 0 & ; x<-1 \\ \dfrac{x+2}{4} & ; -1 \leq x<1 \\ 1 & ; x \geq 1 \end{cases}$. Find

(i) $P\left(-\frac{1}{2}<X \leq \frac{1}{2}\right)$,

(ii) $P(X=0)$,

(iii) $P(X=1)$ and

(iv) $P(2<X \leq 3)$.

(Answer : (i) $\frac{1}{4}$, (ii) 0, (iii) $\frac{1}{4}$ and (iv) 0)

(2) Let $f(x)=\begin{cases} kx^2 & ; 0<x<3 \\ 0 & ; \textit{otherwise} \end{cases}$ be a probability density function of a random variable X. Find

(i) the constant k

(ii) $P(1<X<2)$

(iii) the distribution function of X.

Answer : (i) $\frac{1}{9}$, (ii) $\frac{7}{27}$ and (iii) $F(x)=\begin{cases} 0 & ; x \leq 1 \\ \dfrac{x^3}{27} & ; 1<x \leq 3 \\ 1 & ; x>3 \end{cases}$

(3) Let $f(x)=\begin{cases} \dfrac{x^2}{18} & ; -3<x<3 \\ 0 & ; \textit{otherwise} \end{cases}$ be a probability density function of a random variable X. Find

(i) $P(|X|<1)$

(ii) $P(X^2<9)$

(Answer : (i) $\frac{1}{27}$, (ii) 1)

Space for Hint

(4) Let $f(x)=\begin{cases}\frac{x+2}{18} & ;-2<x<4\\ 0 & ;\ otherwise\end{cases}$ be a probability density function of a random variable X. Find

(i) $P(|X|<1)$

(ii) $P(X^2<9)$

(Answer : (i) $\frac{2}{9}$, (ii) $\frac{25}{36}$)

(5) Find the distribution function of X whose probability density function is given by $f(x)=\begin{cases} x & ;\ 0<x<1\\ 2-x & ;\ 1\le x<2\\ 0 & ;\ x\ge 2\end{cases}$.

Answer : $F(x)=\begin{cases} 0 & ;\ x\le 0\\ \frac{x^3}{27} & ;\ 0<x\le 1\\ 2x-\frac{x^2}{2}-1 & ;\ 1<x\le 2\\ 1 & ;\ x>2\end{cases}$

7. 4 Mathematical Expectation

In this section we shall discuss the mathematical expectation of a random variable X.

Definition :

Let X be a random variable. Then the mathematical expectation of X is denoted by $E(X)$ and defined as

$$\boxed{E(X)=\begin{cases}\sum_i p_i x_i & \text{if } X \text{ is a discrete random variable}\\ \int_{-\infty}^{\infty} f(x)\,dx & \text{if } X \text{ is a continuous random variable}\end{cases}}$$

provided the summation or integration are absolutely convergent.

Space for Hint

Properties of mathematical expectation :

(i) $E(c) = c$ where c is a constant.

(ii) $E(cX) = cE(X)$.

(iii) $E(aX + b) = aE(X) + b$ where a and b are constants.

Proof of the properties :

For our convenient we prove the properties for continuous random variables. (i.e.) we assume both X and Y are independent random variables and a, b and c are constants.

(i) LHS = $E(c)$

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

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

$= c(1)$ $\{\because\ f(x)$ is a probability density function of $X\}$

$= c$

= RHS

Thus $E(c) = c$.

(ii) LHS = $E(cX)$

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

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

$= c\, E(X)$

= RHS

Thus $E(cX) = cE(X)$.

(iii) LHS = $E(aX + b)$

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

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

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

$$= aE(X) + b(1)$$

$$= aE(X) + b$$

$$= \text{RHS}$$

Thus $E(aX+b) = aE(X)+b$.

Definition :

Expectation $E(X)$ of random variable X is called mean of the random variable X and it is denoted by μ.

$$\boxed{(i.e.)\ \overline{x} = \mu = E(X)}$$

Definition :

Let X be random variable X. r^{th} moment of X about the origin is defined as $E(X^r)$ and it is denoted by μ_r'.

$$\boxed{(i.e.)\ \mu_r' = E(X^r)}$$

Note : $\boxed{\mu_1' = E(X) = \mu}$

Definition :

Let X be random variable X. r^{th} moment of X about the mean or r^{th} central moment of X is defined as $E\left[(X-\mu)^r\right]$ and it is denoted by μ_r.

$$\boxed{(i.e.)\ \mu_r = E\left[(X-\mu)^r\right]}$$

Note : $(i.e.)\ \mu_2 = E\left[(X-\mu)^2\right] = \sigma^2$ variance of the random variable X

and $(i.e.)\ \mu_1 = E\left[(X-\mu)\right] = 0$.

Problem :

Prove that $\sigma^2 = \mu_2' - (\mu_1')^2$

Proof :

LHS $= \sigma^2$

Space for Hint

Space for Hint

$$= E\left[(X-\mu)^2\right]$$
$$= E\left[X^2 - 2X\mu + \mu^2\right]$$
$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$
$$= E(X^2) - 2\mu E(X) + \mu^2$$
$$= \mu_2' - 2\mu_1'\mu_1' + (\mu_1')^2$$
$$= \mu_2' - (\mu_1')^2$$
$$= \text{RHS}$$

Hence $\sigma^2 = \mu_2' - (\mu_1')^2$.

Problem :

Prove that $\mu_r = \mu_r' - {}^rC_1\mu\mu_{r-1}' + {}^rC_2\mu^2\mu_{r-2}' - {}^rC_3\mu^3\mu_{r-3}' + \cdots$

Proof :

LHS $= \mu_r$

$$= E\left[(X-\mu)^r\right]$$
$$= E(X^r) - {}^rC_1\mu E(X^{r-1}) + {}^rC_2\mu^2 E(X^{r-2}) - {}^rC_3\mu^3 E(X^{r-3}) + \cdots$$
$$= \mu_r' - {}^rC_1\mu\mu_{r-1}' + {}^rC_2\mu^2\mu_{r-2}' - {}^rC_3\mu^3\mu_{r-3}' + \cdots$$
$$= \text{RHS}$$

Hence $\mu_r = \mu_r' - {}^rC_1\mu\mu_{r-1}' + {}^rC_2\mu^2\mu_{r-2}' - {}^rC_3\mu^3\mu_{r-3}' + \cdots$

Note :

Substituting $r = 1, 2, 3, 4$ successively in the above problem, we get,

$\mu_1 = \mu_1' - \mu = 0$,

$$\mu_2 = \mu_2' - 2\mu\mu_1' + \mu^2\mu_0'$$
$$= \mu_2' - 2\mu_1'\mu_1' + (\mu_1')^2$$
$$= \mu_2' - (\mu_1')^2$$

Thus $\mu_2 = \mu_2' - (\mu_1')^2$,

$$\mu_3 = \mu_3' - 3\mu\mu_2' + 3\mu^2\mu_1' - \mu^3\mu_0'$$
$$= \mu_3' - 3\mu_1'\mu_2' + 3(\mu_1')^2\mu_1' - (\mu_1')^3$$
$$= \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

Thus $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$,

and $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 + -3(\mu_1')^4$.

Example 7.8 :

A random variable X has the following probability density function. Find the mathematical expectation of X.

x	-1	0	1	2
$p(x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

Solution :

We know that $E(X) = \sum_i p_i x_i$

Thus

x	$p(x)$	$x\,p(x)$
-1	$\frac{1}{3}$	$-\frac{1}{3}$
0	$\frac{1}{6}$	0
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{3}$	$\frac{2}{3}$
Total		$\frac{1}{2}$

Thus $E(X) = \frac{1}{2}$.

Space for Hint

Space for Hint

Example 7. 9 :

Find the mean and standard deviation

x	-3	-1	0	4
$p(x)$	0.2	0.4	0.3	0.1

Solution :

Now

x	$p(x)$	$x\,p(x)$	
-3	0.2	-0.6	1.8
-1	0.4	-0.4	0.4
0	0.3	0	0
4	0.1	0.4	1.6
Total		-0.6	3.8

We know that mean = $E(X)$

$$= \sum_i x_i p_i$$

$$= -0.6$$

Now $\mu_2' = 3.8$ and $\mu_1' = -0.6$

Thus variance = σ^2

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

$$= 3.8 - (-0.6)^2$$

$$= 3.8 - 0.36$$

$$= 3.44$$

Hence standard deviation = σ

$$= \sqrt{3.44}$$

$$= 1.855$$

Example 7. 10 :

A random variable X has the following probability density function.

x	1	2	3	4	5	6	7
$p(x)$	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(i) find k, (ii) $P(X \geq 6)$, (iii) $P(X < 6)$, (iv) $P(1 \leq X < 5)$ and (v) $E(X)$.

Solution :

We know that $E(X) = \sum_i p_i x_i$

Thus

x	$p(x)$	$F(x) = P(X \leq x)$	$x\,p(x)$
1	k	k	$k = \frac{1}{10}$
2	$2k$	$3k$	$4k = \frac{4}{10}$
3	$2k$	$5k$	$6k = \frac{6}{10}$
4	$3k$	$8k$	$12k = \frac{12}{10}$
5	k^2	k^2+8k	$5k^2 = \frac{5}{100}$
6	$2k^2$	$3k^2+8k$	$12k^2 = \frac{12}{100}$
7	$7k^2+k$	$10k^2+9k$	$49k^2+7k = \frac{49}{100}+\frac{7}{10}$
Total			$\frac{366}{100}$

Space for Hint

Space for Hint

(i) To find the value of k.

Given that X is a probability density function and therefore

$$\sum_i p_i = 1$$

(i.e.) $10k^2 + 9k = 1$

(i.e.) $10k^2 + 9k - 1 = 0$

(i.e.) $(k+1)\left(k - \frac{1}{10}\right) = 0$

(i.e.) $k = -1$ or $k = \frac{1}{10}$

Since $k > 0$, we have, $k = \frac{1}{10}$,

(ii) To find $P(X \geq 6)$

Now $P(X \geq 6) = P(X = 6) + P(X = 7)$

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

$$= \frac{2}{100} + \frac{7}{100} + \frac{1}{10}$$

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

(iii) To find $P(X < 6)$

Now $P(X < 6) = P(X \leq 5)$

$$= F(5)$$

$$= 8k + k^2$$

$$= \frac{8}{10} + \frac{1}{100}$$

$$= \frac{81}{100},$$

(iv) To find $P(1 \leq X < 5)$

Now $P(1 \leq X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

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

$$= 8k$$

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

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

and (v) To find $E(X)$.

Now $E(X) = \frac{366}{100} = 3.66$.

Example 7. 11 :

A random variable X has the following probability density function.

x	-3	6	9
$p(x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Find (i) $E(X)$, (ii) $E(X^2)$, (iii) $E\left[(2X+1)^2\right]$

Solution :

We know that $E(X) = \sum_i p_i x_i$

Thus

x	$p(x)$	$xp(x)$	$x^2p(x)$
-3	$\frac{1}{6}$	$-\frac{1}{2}$	$\frac{3}{2}$
6	$\frac{1}{2}$	3	18
9	$\frac{1}{3}$	3	27
Total		5.5	46.5

Space for Hint

Space for Hint

(i) from the table, $E(X)=5.5$

(ii) from the table, $E(X^2)=46.5$

(iii) $E\left[(2X+1)^2\right] = 4E(X^2)+4E(X)+1$

$= 4(46.5)+4(5.5)+1$

$= 209$.

Example 7. 12 :

For the continuous random variable X whose probability density function is given by $f(x)=\begin{cases} cx(2-x) & ; 0\le x\le 2 \\ 0 & ; otherwise \end{cases}$. Find c, mean, median, mode and variance.

Solution :

Given that $f(x)=\begin{cases} cx(2-x) & ; 0\le x\le 2 \\ 0 & ; otherwise \end{cases}$ is the probability density function of random variable X.

Step 1 : To find c.

Given that $f(x)$ is a probability density function of X.

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

(i.e.) $\int_{-\infty}^{0} f(x)\,dx+\int_{0}^{2} f(x)\,dx+\int_{2}^{\infty} f(x)\,dx=1$

(i.e.) $0+\int_{0}^{2} cx(2-x)\,dx+0=1$

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

(i.e.) $c\left[\left(4-\frac{8}{3}\right)-0\right]=1$

(i.e.) $c=\frac{3}{4}$

Step 2 : To find $E(X)$.

Now $f(x) = \begin{cases} \dfrac{3x(2-x)}{4} & ; 0 \le x \le 2 \\ 0 & ; \textit{otherwise} \end{cases}$

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

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

$$= \frac{3}{4}\left[\frac{2}{3}x^3 - \frac{x^4}{4}\right]_0^2$$

$$= \frac{3}{4}\left[\frac{16}{3} - \frac{16}{4}\right]$$

$$= 1$$

Step 3 : To find the median

Let m be the median of the distribution.

$\therefore \int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = \frac{1}{2}$

Now $\int_{-\infty}^{m} f(x)\,dx = \frac{1}{2}$

(i.e.) $\frac{3}{4}\int_0^m x(2-x)\,dx = \frac{1}{2}$

(i.e.) $\left[x^2 - \frac{x^3}{3}\right]_0^m = \frac{2}{3}$

(i.e.) $3m^2 - m^3 = 2$

(i.e.) $m^3 - 3m^2 + 2 = 0$

(i.e.) $(m-1)(m^2 - 2m - 2) = 0$

(i.e.) $m = 1$ or $m^2 - 2m - 2 = 0$

(i.e.) $m = 1$ or $m = 1 \pm \sqrt{3}$

Since $1 \pm \sqrt{3} \notin (0,2)$ and therefore required median is 1.

Space for Hint

Space for Hint

Step 4 : To find the mode of the distribution.

Here $f(x)=\frac{3}{4}x(2-x)$

Differentiate $f(x)$ twice with respect to x we get,

$f'(x)=\frac{3}{4}(2-2x)$ and $f''(x)=-2$.

For maximum or minimum $f(x)$, put $f'(x)=0$

If $f'(x)=0$ then we have $x=1$

When $x=1$, then $f''(1)=-2<0$

(i.e.) $f(x)$ attains its maximum at $x=1$.

Thus mode of the distribution is 1.

Step 5 : To find the variance of the distribution.

Now $\mu_2' = E(X^2)$

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

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

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

$$=\frac{3}{4}\left[2\frac{x^4}{4}-\frac{x^5}{5}\right]_0^2$$

$$=\frac{3}{4}\left[\left(2\frac{16}{4}-\frac{32}{5}\right)-(0)\right]$$

$$=\frac{3}{4}\left[8-\frac{32}{5}\right]$$

$$=\frac{3}{4}\cdot\frac{8}{5}$$

$$=\frac{6}{5}$$

Thus the variance $= \mu_2' - (\mu_1')^2$

$$= \frac{6}{5} - (1)^2$$

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

Hence, mean = median = mode = 1 and variance $= \frac{1}{5}$.

Example 7.13 :

Let X be a probability density function is given by $f(x) = \begin{cases} \frac{1}{4} & ; -2 < x < 2 \\ 0 & ; \textit{otherwise} \end{cases}$

Find (i) $P(X < 1)$ and (ii) $P(|x| > 1)$.

Solution :

Given that $f(x) = \begin{cases} \frac{1}{4} & ; -2 < x < 2 \\ 0 & ; \textit{otherwise} \end{cases}$ is the probability density function.

(i) Now $P(X < 1)$

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

$$= 0 + \int_{-2}^{1} \frac{1}{4}\,dx$$

$$= \frac{1}{4}[x]_{-2}^{1}$$

$$= \frac{1}{4}[1+2]$$

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

Thus $P(X < 1) = \frac{3}{4}$

and (ii) $P(|x| > 1)$

$$= 1 - P(X \le 1)$$

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

Space for Hint

$$= 1 - \int_{-1}^{1} \frac{1}{4} dx$$

$$= 1 - \frac{1}{4}[x]_{-1}^{1}$$

$$= 1 - \frac{1}{4}[2]$$

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

Thus $P(|x| > 1) = \frac{1}{2}$.

Example 7. 14 :

Let X have the probability density function $f(x) = \begin{cases} \frac{x+2}{18}; & -2 < x < 4 \\ 0 \quad ; & elsewhere \end{cases}$

Find E(X), E $[(X + 2)^3]$ and E $[6X - 2 (X + 2)^3]$

Solution :

Given that $f(x) = \begin{cases} \frac{x+2}{18}; & -2 < x < 4 \\ 0 \quad ; & elsewhere \end{cases}$ is a probability density function of X.

Now $E(X) = \int_{-2}^{4} x \frac{x+2}{18} dx$

$$= \frac{1}{18}\left[\frac{x^3}{3} + x^2\right]_{-2}^{4}$$

$$= \frac{1}{18}\left[\left(\frac{64}{3} + 16\right) - \left(\frac{-8}{3} + 4\right)\right]$$

$$= \frac{1}{18} \quad [36]$$

$$= 2$$

and $E\left[(X+2)^2\right] = \int_{-2}^{4} (x+2)^2 \frac{x+2}{18} dx$

Space for Hint

$$= \frac{1}{18} \int_{-2}^{4} (x+2)^4 dx$$

$$= \frac{1}{18} \left[\frac{(x+2)^5}{5} \right]_{-2}^{4}$$

$$= \frac{1}{18 \times 5} \left[6^5 - 0 \right]$$

$$= \frac{432}{5}$$

and $E(6X - 2\,(X+2)^3)$

$$= 6\,E(X) - 2\,E((X+2)^3)$$

$$= 6\,(2) - 2\left(\frac{432}{5}\right)$$

$$= -\frac{805}{5}$$

Check Your Progress

(1) A card is drawn from a pack of 52 cards. If aces are counted as 1 and face cards – Jack, Queen, King – as 10 and other according to their denominations denoting the random variable X as the score on the card, find the expectation of the score on the card.

(Answer : (i) $\frac{85}{13}$)

(2) Obtain the probability distribution of the number of heads in three tosses of a coin. Hence find the mean and variance from the probability distribution.

(Answer : mean = $\frac{3}{2}$, variance = $\frac{3}{4}$)

(3) Suppose that X is the discrete random variable with probability density function $f(x) = \begin{cases} \frac{1}{5} ; & 1,2,3,4,5 \\ 0 ; & otherwise \end{cases}$. Compute

(i) $E(X)$

(ii) $E(X^2)$

(iii) $E\left[(X+2)^2\right]$

Space for Hint

(Answer : (i) $E(X)=3$, (ii) $E(X^2)=11$, (iii) $E\left[(X+2)^2\right]=27$)

(4) A random variable X has the probability density function $f(x)=\dfrac{k}{1+x^2}$; $-\infty<x<\infty$. Find k

(Answer : $k=\dfrac{1}{\pi}$)

(5) A continuous distribution of X is defined as follows

$$F(x)=\begin{cases} 0 & if\ \ x\le 1 \\ \dfrac{(x-4)^4}{16} & if\ \ 1<x\le 3 \\ 1 & if\ \ x>2 \end{cases}$$

. Find the probability density function $f(x)$. Also find the mean of X.

(Answer : mean = 2.65)

(6) Suppose that X is the discrete random variable with probability density function $f(x)=\begin{cases} c\,x(2-x)\ ;\ 0\le x\le 2 \\ 0 \quad ;\ otherwise \end{cases}$. Compute

(i) c

(ii) mean

(iii) μ_2', μ_3'

(iv) μ_2, μ_3

(Answer : (i) $c=\dfrac{3}{4}$, (ii) mean =1, (iii) $\mu_2'=\dfrac{6}{5}$, $\mu_3'==8$,

(iv) $\mu_2=\dfrac{1}{5}$, $\mu_3=-\dfrac{7}{5}$)

SUMMARY

In this unit, we have discussed probability density function, distribution function and mathematical expectations and also we learned how find the distribution function, expectation of a random variable.

Space for Hint

Unit VIII

Moment Generating Function
and
Theoretical Distributions

Objectives

In this unit, we are going to discuss moment generating function, cumulants, Binomial distribution, Poisson distribution and Normal distribution with examples and how to fit these distribution functions.

8. 1 Moment generating function

Definition :

Let X be a random variable. The moment generating function of X is defined as $E\left(e^{tX}\right)$ where $-h<t<h$, (h is a positive real number).

Note :

(1) The moment generating function is abbreviated as m.g.f.

(2) The moment generating function of X is exists provided either

$$\int_{-\infty}^{\infty} e^{tX} f(x)\,dx \text{ or } \sum_{-\infty}^{\infty} e^{tX} f(x) \text{ exists.}$$

(3) The moment generating function of X is denoted by $M_X(t)$.

(i.e.) $\boxed{M_X(t)=E\left(e^{tX}\right)}$

(4) If two random variables have the same moment generating function, then they have the same distribution.

(5) $\left[\frac{d^m}{dt^m}M(t)\right]_{t=0} = E(X^m)$ and it is called m^{th} moments of X.

(6) $\mu=M'(0)$ and $\sigma^2=M''(0)-(M'(0))^2$

Space for Hint

That is the moment generating function of random variable X is defined as

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x) & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{if } X \text{ is a continuour random variable} \end{cases}$$

Note :

$$M_X(t) = E(e^{tX})$$

$$= E\left(1 + \frac{tX}{1!} + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots + \frac{t^rX^2}{r!} + \cdots\right)$$

$$= 1 + \frac{t}{1!}E(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots + \frac{t^r}{r!}E(X^r) + \cdots$$

$$\therefore\ M_X(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots$$

Thus the coefficient of $\frac{t^r}{r!}$ in $M_X(t)$ is μ_r'.

Definition :

The moment generating function of random variable X about a is defined as $E(e^{t(X-a)})$ and is denoted by $M_{X=a}(t)$.

Note (1) :

Thus $M_{X=a}(t) = E(e^{t(X-a)})$

$$= E\left(1 + \frac{t(X-a)}{1!} + \frac{t^2(X-a)^2}{2!} + \frac{t^3(X-a)^3}{3!} + \cdots + \frac{t^r(X-a)^2}{r!} + \cdots\right)$$

$$= 1 + \frac{t}{1!}E(X-a) + \frac{t^2}{2!}E(X-a)^2 + \frac{t^3}{3!}E(X-a)^3 + \cdots + \frac{t^r}{r!}E(X-a)^r + \cdots$$

$\therefore\ M_{X=a}(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots$ where $\mu_r' = E\{(X-a)^r]$ is r^{th} moment of X about a.

Note (2) :

In the above note, if we consider $a = m$ mean of the distribution then

$$M_{X=m}(t) = E(e^{t(X-m)})$$

$$= E\left(1 + \frac{t(X-m)}{1!} + \frac{t^2(X-m)^2}{2!} + \frac{t^3(X-m)^3}{3!} + \cdots + \frac{t^r(X-m)^2}{r!} + \cdots\right)$$

Space for Hint

$$=1+\frac{t}{1!}E(X-m)+\frac{t^2}{2!}E(X-m)^2+\frac{t^3}{3!}E(X-m)^3+\cdots+\frac{t^r}{r!}E(X-m)^r+\cdots$$

$$\therefore\ M_{X=a}(t)=1+t\mu_1+\frac{t^2}{2!}\mu_2+\frac{t^3}{3!}\mu_3+\cdots+\frac{t^r}{r!}\mu_r+\cdots \quad \text{where } \mu_r=E\{(X-m)^r\}$$

is r^{th} central moment of X.

Properties of moment generating function

Property 1 : $M_{X=a}(t)=e^{-at}M_X(t)$

Proof :

LHS = $M_{X=a}(t)$

$= E(e^{t(X-a)})$

$= E(e^{-at}e^{tX})$

$= e^{-at}E(e^{tX})$

$= e^{-at}M_X(t)$

Hence $M_{X=a}(t)=e^{-at}M_X(t)$.

Property 2 : r^{th} derivative of moment generating function of a random variable X at $t=0$ is μ_r'.

Proof :

If $M_X(t)$ is the moment generating function of a random variable X then

$$M_X(t)=1+t\mu_1'+\frac{t^2}{2!}\mu_2'+\frac{t^3}{3!}\mu_3'+\cdots+\frac{t^r}{r!}\mu_r'+\cdots$$

$$\therefore\ \frac{d^r}{dt^r}(M_X(t))=\mu_r'+t\mu_r'+\cdots$$

Thus $\left[\frac{d^r}{dt^r}(M_X(t))\right]_{t=0}=\mu_r'$

Hence r^{th} derivative of moment generating function of a random variable X at $t=0$ is μ_r'.

This proves the property 2.

Space for Hint

Property 3 : If $M_X(t)$ is the moment generating function of X and $y=\alpha X+\beta$ then the moment generating function of Y is $M_Y(t)=e^{\beta t}M_X(\alpha t)$

Proof :

If $M_X(t)$ is the moment generating function of a random variable X then the moment generating function of $y=\alpha X+\beta$ is $M_Y(t)=E(e^{tY})$

(i.e.) $M_Y(t) = E(e^{\alpha tX}e^{\beta t})$

(i.e.) $M_Y(t) = e^{\beta t}E(e^{\alpha tX})$

This proves property 3.

Property 4 :

If X and Y are two independent random variables and $Z=X+Y$ then $M_Z(t)=M_X(t)M_Y(t)$.

Proof :

Let $M_X(t)$ be the moment generating function of X and $M_Y(t)$ be the moment generating function of Y.

If $M_Z(t)$ is the moment generating function of Z then

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) \\ &= E(e^{t(X+Y)}) \\ &= E(e^{tX}e^{tY}) \\ &= E(e^{tX})E(e^{tY}) \text{ \{since } X \text{ and } Y \text{ are independent random variables)} \\ &= M_X(t)M_Y(t) \end{aligned}$$

(i.e.) $M_Z(t)=M_X(t)M_Y(t)$

This proves the property 4.

Example 8. 1 :

Let the random variable X assume the value r with the probability law : $P(X=x)=p\cdot q^{r-1}; r=1,2,3,\cdots$ and $p+q=1$. Find the moment generating function and hence find the mean and variance.

Solution :

Given that $P(X=x)=p\cdot q^{r-1}; r=1,2,3,\cdots$ and $p+q=1$.

Space for Hint

Step 1 :

$\therefore$ the moment generating function of $X = M_X(t)$

$$= E(e^{tX})$$

$$= \sum_{r=1}^{\infty} e^{tr} p(r)$$

$$= \sum_{r=1}^{\infty} e^{tr} p q^{r-1}$$

$$= pe^t \sum_{r=1}^{\infty} (e^t q)^{r-1}$$

$$= pe^t \left(1 + (qe^t) + (qe^t)^2 + (qe^t)^3 + \cdots\right)$$

$$= pe^t \frac{1}{1-qe^t}$$

$$= \frac{pe^t}{1-qe^t}$$

Thus the moment generating function of $X = \dfrac{pe^t}{1-qe^t}$.

Step 2 :

Now $M_X(t) = \dfrac{pe^t}{1-qe^t}$

(i.e.) $M_X(t) = pe^t(1-qe^t)^{-1}$

$\therefore\ M'_X(t) = pe^t(1-qe^t)^{-1} + pe^t(-1)(1-qe^t)^{-2}(-qe^t)$

(i.e.) $M'_X(t) = pe^t(1-qe^t)^{-1} + pqe^{2t}(1-qe^t)^{-2}$

$$= \frac{pe^t}{1-qe^t} + \frac{pqe^{2t}}{(1-qe^t)^2}$$

$$= \frac{pe^t}{(1-qe^t)^2}$$

and $M''_X(t) = \dfrac{(1-qe^t)^2 pe^t - pe^t(2)(1-qe^t)(-qe^t)}{(1-qe^t)^4}$

Now $M'_X(0) = \dfrac{p}{p^2}$

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

Space for Hint

and $M''_X(0) = \dfrac{p^2 p + 2ppq}{p^4}$

$= \dfrac{p^2(p+2q)}{p^4}$

$= \dfrac{p+2q}{p^2}$

$= \dfrac{p+q+q}{p^2}$

$= \dfrac{1+q}{p^2}$

Thus mean $= \mu$

$= M'_X(0)$

$= \dfrac{1}{p}$

and variance $= \sigma^2$

$= \dfrac{1+q}{p^2} - \dfrac{1}{p^2}$

$= \dfrac{q}{p^2}$

Example 8. 2 :

The random variable X take the value n with probability $\dfrac{1}{2^n}$; $n = 1, 2, 3, \cdots$

Find the moment generating function of X and hence fine the mean and variance.

Solution :

Given that $P(X = x) = \begin{cases} \dfrac{1}{2^n} ; & n = 1, 2, 3, \cdots \\ 0 ; & otherwise \end{cases}$.

Step 1 :

$\therefore$ The moment generating function of $X = M_X(t)$

$= E(e^{tX})$

$= \sum_{n=1}^{\infty} e^{tn} p(n)$

Space for Hint

$$= \sum_{n=1}^{\infty} e^{tn} \frac{1}{2^n}$$

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

$$= \sum_{n=1}^{\infty} x^n \text{ where } x = \frac{e^t}{2}$$

$$= x + x^2 + x^3 + \cdots$$

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

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

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

Thus the moment generating function of $X = \dfrac{e^t}{2-e^t}$.

Step 2 : To find the mean and the variance

Now $M_X(t) = \dfrac{e^t}{2-e^t}$

$$\therefore\ M'_X(t) = \frac{(2-e^t)e^t - e^t(-e^t)}{(2-e^t)^2}$$

(i.e.) $M'_X(t) = \dfrac{2e^t}{(2-e^t)^2}$

and $M''_X(t) = \dfrac{(2-e^t)^2 2e^t - 2e^t(2)(2-e^t)(-e^t)}{(2-e^t)^4}$

Now $M'_X(0) = \dfrac{2}{1}$

$= 2$

and $M''_X(0) = \dfrac{2+4}{1}$

$= 6$

Thus mean $= \mu$

$= 2$

Space for Hint

and variance = σ^2

$= 6-4$

$= 2$

Example 8. 3 :

Obtain the moment generating function of the random variable X having probability density function $f(x)=\begin{cases} x & ; 0 \le x<1 \\ 2-x & ; 1\le x<2 \\ 0 & ; otherwise \end{cases}$

Solution :

Given that $f(x)=\begin{cases} x & ; 0 \le x<1 \\ 2-x & ; 1\le x<2 \\ 0 & ; otherwise \end{cases}$ is a probability density function of a random variable X.

Let $M_X(t)$ be the moment generating function of X.

Thus $M_X(t) = E(e^{tX})$

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

$$= \int_0^1 x e^{tx}\,dx + \int_1^2 e^{tx}(2-x)\,dx$$

$$= \left[x\frac{e^{tx}}{t}-\frac{e^{tx}}{t^2}\right]_0^1 + \left[(2-x)\frac{e^{tx}}{t}+\frac{e^{tx}}{t^2}\right]_1^2$$

$$= \left[\frac{e^t}{t}-\frac{e^t}{t^2}\right]-\left[1-\frac{1}{t^2}\right]+\left[0+\frac{e^{2t}}{t^2}\right]-\left[\frac{e^t}{t}+\frac{e^t}{t^2}\right]$$

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

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

Thus the moment generating function of $X = \left(\frac{e^t-1}{t}\right)^2$

Example 8. 4 :

If $\overline{X}$ is the mean of n independent random variables having the moment generating function $M_X(t)$ then prove that $M_{\overline{X}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$.

Solution :

Let $X_1, X_2, X_3, \cdots, X_n$ be n random variables.

Then $\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$.

Now moment generating function of $\overline{X} = M_{\overline{X}}(t)$

$$= E(e^{t\overline{X}})$$

$$= E\left(e^{t\left(\frac{X_1+X_2+X_3+\cdots+X_n}{n}\right)}\right)$$

$$= E\left(e^{\frac{t}{n}X_1} e^{\frac{t}{n}X_2} e^{\frac{t}{n}X_3} \cdots e^{\frac{t}{n}X_n}\right)$$

$$= E\left(e^{\frac{t}{n}X_1}\right) E\left(e^{\frac{t}{n}X_2}\right) E\left(e^{\frac{t}{n}X_3}\right) \cdots E\left(e^{\frac{t}{n}X_n}\right)$$ **{since $X_1, X_2, X_3, \cdots, X_n$ are independent random variables}**

$$= M_{X_1}\left(\frac{t}{n}\right) M_{X_2}\left(\frac{t}{n}\right) M_{X_3}\left(\frac{t}{n}\right) \cdots M_{X_n}\left(\frac{t}{n}\right)$$

$$= M_X\left(\frac{t}{n}\right) M_X\left(\frac{t}{n}\right) M_X\left(\frac{t}{n}\right) \cdots M_X\left(\frac{t}{n}\right)$$

$$= \left[M_X\left(\frac{t}{n}\right)\right]^n$$

Thus $M_{\overline{X}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$

This proves the problem.

Space for Hint

Space for Hint

8. 2 Cumulant generating function

Definition :

The cumulant generating function $K_X(t)$ of random variable X is defined as $K_X(t) = \log_e(M_X(t))$ provided the right hand side can be expanded as a convergent series in powers of t.

Note :

$K_X(t) = \log_e(M_X(t))$

(i.e.) $K_X(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \cdots + \kappa_r \frac{t^r}{r!} + \cdots$

Thus κ_r = coefficient of $\frac{t^r}{r!}$ in $K_X(t)$ is called r^{th} cumulant of X.

Properties of cumulant generating function

Property 1 :

r^{th} derivative of cumulant generating function of a random variable X at $t = 0$ is κ_r.

Proof :

If $K_X(t)$ is the moment generating function of a random variable X then

$K_X(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \cdots + \kappa_r \frac{t^r}{r!} + \cdots$

$\therefore \frac{d^r}{dt^r}(K_X(t)) = \kappa_r + t\kappa_{r+1} + \cdots$

Thus $\left[\frac{d^r}{dt^r}(K_X(t))\right]_{t=0} = \kappa_r$

r^{th} derivative of cumulant generating function of a random variable X at $t = 0$ is κ_r.

This proves the property 1.

Property 2 : $\kappa_1 = \mu_r'$, $\kappa_2 = \mu_2$, $\kappa_3 = \mu_3$, $\mu_4 - 3\mu_2^2$ and so on.

Proof :

Let $K_X(t)$ be cumulant generating function of X.

Then $K_X(t) = \log_e(M_X(t))$

(i.e.) $\kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \cdots + \kappa_r \frac{t^r}{r!} + \cdots =$

$$\log_e\left(1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots\right)$$

$$= \left(t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots\right)$$

$$-\frac{1}{2}\left(t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots\right)^2$$

$$+\frac{1}{3}\left(t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots\right)^3$$

$$-\frac{1}{4}\left(t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots\right)^4 + \cdots$$

Comparing coefficients of like powers of t on both sides, we get,

$\kappa_1 = \mu_r'$,

$\kappa_2 = \mu_2' - (\mu_1')^2 = \mu_2$,

$\kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = \mu_3$,

$\kappa_4 = \mu_4 - 3\mu_2^2 = \mu 4$

This proves property 2.

Property 3 :

If $X_1, X_2, X_3, \cdots, X_n$ are independent random variables then

$K_{X_1+X_2+X_3+\cdots+X_n}(t) = K_{X_1}(t) + K_{X_2}(t) + K_{X_3}(t) + \cdots + K_{X_n}(t)$

Proof :

Let $X_1, X_2, X_3, \cdots, X_n$ are independent random variables.

Now $K_{X_1+X_2+X_3+\cdots+X_n}(t) = \log_e\left(M_{X_1+X_2+X_3+\cdots+X_n}(t)\right)$

Space for Hint

Space for Hint

$$= \log_e\left(M_{X_1}(t)M_{X_2}(t)M_{X_3}(t)\cdots M_{X_n}(t)\right)$$

$$= \log_e\left(M_{X_1}(t)\right)+\log_e\left(M_{X_2}(t)\right)+\log_e\left(M_{X_3}(t)\right)+\cdots+\log_e\left(M_{X_n}(t)\right)$$

$$= K_{X_1}(t)+K_{X_2}(t)+K_{X_3}(t)+\cdots+K_{X_n}(t)$$

Thus $K_{X_1+X_2+X_3+\cdots+X_n}(t)=K_{X_1}(t)+K_{X_2}(t)+K_{X_3}(t)+\cdots+K_{X_n}(t)$.

This proves the property 3.

8. 3 Binomial distribution

Definition :

A random variable X is said to have a binomial distribution if its probability density function is given by $p(x)=\begin{cases} {}^nC_x p^x q^{n-x} & ; x=0,1,2,3,\cdots,n \\ 0 & ; elsewhere \end{cases}$

where $p+q=1$

Note :

(1) In Binomial distribution n, p are called parameters

(2) A binomial distribution is denoted by $B(n,p)$ or $b(n,p)$.

Example 8. 5 :

Find the moment generating function of binomial distribution and also find its mean and variance.

Solution:

Step 1 :

Let X be a random variable having binomial distribution b(n,p).

(i.e) $p(x)=\begin{cases} {}^nC_x p^x q^{n-x} & ; x=0,1,2,3,\cdots,n \\ 0 & ; elsewhere \end{cases}$ where $p+q=1$.

The moment generating function of x is $M_X(t)$ is given by

$$M_X(t) = E(e^{tX})$$

$$= \sum_{x=-\infty}^{\infty} e^{tx}p(x)$$

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

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

$$= \left(q + pe^t\right)^n$$

Therefore the moment generating function of the Binomial distribution is

$M_X(t) = \left(q + pe^t\right)^n$

Step 2 : To find the mean and variance.

Now $M_X(t) = \left(q + pe^t\right)^n$

Differentiate $M_X(t)$ twice with respect to t, we get,

$$M'(t) = n\cdot\left(q + pe^t\right)^{n-1} \cdot pe^t$$

and $M''(t) = n\cdot\left(q + pe^t\right)^{n-1} \cdot pe^t + n(n-1)\cdot\left(q + pe^t\right)^{n-2} \cdot \left(pe^t\right)^2.$

$\therefore M(0) = 0,$

$M'(0) = np,$

and $M''(0) = np + n(n-1)p^2$

Now Mean $= M'(0) = np$

and variance $= M''(0) - (M'(0))^2$

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

$$= npq$$

Example 8. 6 :

Find the first three moments of Binomial distribution.

Proof :

We know that moment generating function of Binomial distribution is

$M_X(t) = \left(q + pe^t\right)^n.$

$$= \left[q + p\left(1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots\right)\right]^n$$

Space for Hint

Space for Hint

$$= \left[q + p + p\left(\frac{t}{1!} + \frac{t^2}{2!} + \cdots\right)\right]^n$$

$$= \left[1 + p\left(\frac{t}{1!} + \frac{t^2}{2!} + \cdots\right)\right]^n$$

$$= \left[1 + {}^nC_1\, p\left(\frac{t}{1!} + \frac{t^2}{2!} + \cdots\right) + {}^nC_2\, p^2\left(\frac{t}{1!} + \frac{t^2}{2!} + \cdots\right)^2 + {}^nC_3\, p^3\left(\frac{t}{1!} + \frac{t^2}{2!} + \cdots\right)^3 + \cdots\right]^n$$

$$= 1 + np\frac{t}{1!} + \left(np + n(n-1)p^2\right)\frac{t^2}{2!} + \left(np + 3n(n-1)p^2 + n(n-1)(n-2)p^3\right)\frac{t^2}{2!} + \cdots$$

Comparing the coefficient of $\frac{\mu_r'}{r!}$ the above moment generating function with

$M_X(t) = 1 + \mu_1'\frac{t}{1!} + \mu_2'\frac{t^2}{2!} + \mu_3'\frac{t^3}{3!} + \cdots$, we get,

$\mu_1' = np$,

$\mu_2' = np + n(n-1)p^2$

$\mu_3' = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3$

Example 8. 7 :

Find the first four central moments of Binomial distribution.

Proof :

Step 1 :

We know that moment generating function of Binomial distribution about mean

Now $M_{X=\mu}(t)$

$$= E(e^{t(X-\mu)})$$

$$= E(e^{t(X-np)})$$

$$= E(e^{tX}e^{-\mu t})$$

$$= e^{-\mu t}E(e^{tX})$$

$$= e^{-\mu t}M_X(t)$$

$$= e^{-\mu t}(q + pe^t)^n$$

$= e^{-npt}(q + pe^t)^n$ {since mean = $\mu = np$ }

$$= (qe^{-pt} + pe^{-pt}e^{t})^n$$

$$= (qe^{-pt} + pe^{qt})^n$$

which is the moment generating function of Binomial distribution about its mean np.

Step 2 : To find the first four central moments.

From step 1, the moment generating function of Binomial distribution about its mean is $(qe^{-pt} + pe^{qt})^n$.

(i.e.) $M_{X=\mu}(t) = (qe^{-pt} + pe^{qt})^n$

$$= \left[q\left(1 - p\frac{t}{1!} + p^2\frac{t^2}{2!} - + p^3\frac{t^3}{3!} + \cdots\right) + p\left(1 + p\frac{t}{1!} + p^2\frac{t^2}{2!} + p^3\frac{t^3}{3!} + \cdots\right)\right]^n$$

$$= \left[(p+q) + (p+q)pq\frac{t^2}{2!} + pq(q^2 - p^2)\frac{t^3}{3!} + pq(p^3 + q^3)\frac{t^4}{4!} + \cdots\right]^n$$

$$= \left[1 + pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + pq(p^2 - pq + q^2)\frac{t^4}{4!} + \cdots\right]^n$$

$$= 1 + n\left[pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + pq(p^2 - pq + q^2)\frac{t^4}{4!} + \cdots\right]$$

$$+ \frac{n(n-1)}{2!}\left[pq\frac{t^2}{2!} + pq(q-p)\frac{t^3}{3!} + pq(p^2 - pq + q^2)\frac{t^4}{4!} + \cdots\right]^2 + \cdots$$

$$= 1 + npq\frac{t^2}{2!} + npq(q-p)\frac{t^3}{3!} + \left[npq(1 - 3pq) + \frac{n(n-1)}{2}p^2q^2\right]\frac{t^4}{4!} + \cdots$$

$$= 1 + npq\frac{t^2}{2!} + npq(q-p)\frac{t^3}{3!} + \left[npq(1 - 3pq) + 3n(n-1)p^2q^2\right]\frac{t^4}{4!} + \cdots$$

Comparing the coefficient of $\frac{\mu_r'}{r!}$ the above moment generating function with

$M_{X=\mu}(t) = 1 + \mu_1\frac{t}{1!} + \mu_2\frac{t^2}{2!} + \mu_3\frac{t^3}{3!} + \mu_4\frac{t^4}{4!} + \cdots$, we get,

$\mu_2 = npq$,

$\mu_3 = npq(q-p)$,

$\mu_4 = npq(1 - 3pq) + 3n(n-1)p^2q^2$

$$= npq - 3np^2q^2 + 3n^2p^2q^2 - 3np^2q^2$$

Space for Hint

Space for Hint

$= npq(1-6pq)+3n^2p^2q^2$

Thus $\mu_4 = npq-3np^2q^2+3n^2p^2q^2-3np^2q^2$.

Example 8. 8 :

State and prove addition property of Binomial distribution.

Proof :

Statement : If $X \sim B(n_1, p)$ and $Y \sim B(n_2, p)$ are two independent random variables then $X+Y \sim B(n_1+n_2, p)$

Proof :

Let $X \sim B(n_1, p)$ and $Y \sim B(n_2, p)$ are two independent random variables.

Let $M_X(t)$ and $M_Y(t)$ be the moment generating functions of X and Y respectively.

Thus $M_X(t)=\left(q+pe^t\right)^{n_1}$ and $M_Y(t)=\left(q+pe^t\right)^{n_2}$

Hence $M_{X+Y}(t)=M_X(t)M_Y(t)$ {since X and Y are independent random variables}

(i.e.) $M_{X+Y}(t)=\left(q+pe^t\right)^{n_1}\left(q+pe^t\right)^{n_2}$

(i.e.) $M_{X+Y}(t)=\left(q+pe^t\right)^{n_1+n_2}$

which is the moment generating function of Binomial distribution with parameters n_1+n_2 and p.

(i.e.) $X+Y \sim B(n_1+n_2, p)$.

This proves the problem.

Example 8. 9 :

State and prove the recurrence relation for $p(x)$ in Binomial distribution.

Proof :

Statement : If $X \sim B(n, p)$ then $p(x+1)=\left(\frac{n-x}{x+1}\right)\left(\frac{p}{q}\right)p(x)$

Proof :

Let $X \sim B(n, p)$ then $p(x)={}^nC_x p^x q^{n-x}$

(i.e.) $p(x)=\dfrac{n!}{x!(n-x)!}p^x q^{n-x}$.

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

(i.e.) $p(x+1)=\dfrac{n!}{(x+1)!(n-x-1)!}p^{x+1}q^{n-x-1}$

$$\text{Now } \frac{p(x+1)}{p(x)}=\frac{\dfrac{n!}{(x+1)!(n-x-1)!}p^{x+1}q^{n-x-1}}{\dfrac{n!}{x!(n-x)!}p^x q^{n-x}}$$

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

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

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

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

Hence $p(x+1)=\left(\dfrac{n-x}{x+1}\right)\left(\dfrac{p}{q}\right)p(x)$.

This proves the recurrence relation for $p(x)$ in Binomial distribution.

Example 8. 10 :

State and prove the recurrence relation for the moments of the Binomial distribution.

Proof :

Statement : If $X\sim B(n,p)$ then $\mu_{r+1}=pq\left[nr\mu_{r-1}+\dfrac{d\mu_r}{dp}\right]$.

Let $X\sim B(n,p)$.

We know that the mean of Binomial distribution is np.

Then $\mu_r=E(X-\mu)^r$

$$=E(X-np)^r$$

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

Space for Hint

Space for Hint

Now $\frac{d\mu_r}{dp}$

$$= \sum_{x=0}^{n} {}^nC_x\left[r(-n)(x-np)^{r-1}p^xq^{n-x} + (x-np)^r xp^{x-1}q^{n-x} + (x-np)^r p^x(n-x)q^{n-x-1}\right]$$

$$= -nr\sum_{x=0}^{n} {}^nC_x(x-np)^{r-1}p^xq^{n-x} + \sum_{x=0}^{n} {}^nC_x(x-np)^r\left[p^xq^{n-x}\left(\frac{x}{p}-\frac{n-x}{q}\right)\right]$$

$$= -nr\sum_{x=0}^{n}(x-np)^{r-1}p(x) + \sum_{x=0}^{n}(x-np)^r p(x)\left(\frac{x-np}{pq}\right)$$

$$= -nr\sum_{x=0}^{n}(x-np)^{r-1}p(x) + \frac{1}{pq}\sum_{x=0}^{n}(x-np)^{r+1}p(x)$$

$$= -nr\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$$

$$\therefore\ pq\frac{d\mu_r}{dp} = -nrpq\mu_{r-1} + \mu_{r+1}$$

Thus $\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]$

This proves the recurrence relation for the moments of the Binomial distribution.

Example 8. 11 :

Find the mode of the Binomial distribution.

Proof :

Let $X \sim B(n,p)$ and x be the mode of the distribution.

$\therefore\ p(x-1) \le p(x) \ge p(x+1)$.

Now $p(x) = {}^nC_xp^xq^{n-x}$

(i.e.) $p(x) = \frac{n!}{x!(n-x)!}p^xq^{n-x}$

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

(i.e.) $p(x+1) = \frac{n!}{(x+1)!(n-x-1)!}p^{x+1}q^{n-x-1}$

Space for Hint

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

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

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

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

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

Now $p(x) \geq p(x+1)$

$\Rightarrow \frac{p(x+1)}{p(x)} \leq 1$

$\Rightarrow \left(\frac{n-x}{x+1}\right)\left(\frac{p}{q}\right) \leq 1$

$\Rightarrow (n-x)p \leq (x+1)q$

$\Rightarrow np - xp \leq xq + q$

$\Rightarrow np - q \leq xq + xp$

$\Rightarrow np - 1 + p \leq x$

$\Rightarrow (n+1)p - 1 \leq x$ -- (8.1)

Again $p(x-1) \leq p(x) \Rightarrow (n+1)p \geq x$ ---------------------------------- (8.2)

From (8.1) and (8.2), we have, $(n+1)p - 1 \leq x \leq (n+1)p$ ------------ (8.3)

If $(n+1)p$ in not an integer then the integral part of $(n+1)p$ is the mode of the distribution.

If $(n+1)p$ is an integer then $(n+1)p$ and $(n+1)p-1$ are the modes of the distribution.

Example 8. 12 :

If the moment generating function of a random variable X is $\left(\frac{1}{3}+\frac{2}{3}e^{t}\right)^{5}$ then find $\Pr(X = 2 or 3)$.

Space for Hint

Solution:

Given that moment generating function of X is $\left(\frac{1}{3}+\frac{2}{3}e^{t}\right)^{5}$, comparing this moment generating function with moment generating function $\left(q + pe^{t}\right)^{n}$ binomial distribution $B(n, p)$, we get,

$$n=5, \quad p=\frac{2}{3} \text{ and } q=\frac{1}{3}$$

$\therefore$ The probability density function of X is

$$p(x)=\begin{cases} {}^{n}C_{x}p^{x}q^{n-x} & ; x=0,1,2,3,\cdots,n \\ 0 & ; elsewhere \end{cases}$$

(i.e.) $$f(x)={}^{5}C_{x}\left(\frac{2}{3}\right)^{x}\left(\frac{1}{3}\right)^{5-x} \quad ; x=0,1,2,3,4,5$$

$$\therefore \quad \Pr(X=2 or 3) = p(2)+p(3)$$

$$= {}^{5}C_{2}\left(\frac{2}{3}\right)^{2}\left(\frac{1}{3}\right)^{3}+{}^{5}C_{3}\left(\frac{2}{3}\right)^{3}\left(\frac{1}{3}\right)^{2}$$

$$= 10.\frac{4}{9}.\frac{1}{29}+10.\frac{4}{27}.\frac{1}{9}$$

$$= \frac{80}{243}$$

Example 8. 13 :

If $X \sim P(n,p)$, show that $E\left(\frac{X}{n}\right)=p$ and $E\left[\left(\frac{X}{n}-p\right)^{2}\right] = \frac{pq}{n}$

Proof :

Given that $X \sim B(n,p)$.

$$\therefore \quad E(X) = \left(q+pe^{t}\right)^{n} \quad \text{where } p+q=1$$

$$\text{Now } E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)$$

$$=\frac{1}{n}\cdot n\cdot p$$

$$= p$$

Space for Hint

$$\therefore\ E\left(\frac{X}{n}\right) = \text{p}$$

and $E\left[\left(\frac{X}{n} - p\right)^2\right]$

$$= E\left[\frac{(X-np)^2}{n^2}\right]$$

$$= \frac{1}{n^2}E(X-\mu)^2 \text{ where } \mu = np$$

$$= \frac{1}{n^2}\sigma^2$$

$$= \frac{1}{n^2}npq$$

$$= \frac{pq}{n}$$

$$\therefore\ E\left(\frac{X}{n}\right) = \text{p, and } E\left[\left(\frac{X}{n} - p\right)^2\right] = \frac{pq}{n}$$

This proves the problem.

Example 8. 14 :

Check the validity of the statement :

"The mean of a Binomial distribution is 3 and the variance is 4"

Solution :

Given that mean = 3 and variance = 4.

(i.e.) $\bar{x} = 3$ and $\sigma^2 = 4$.

(i.e.) $np = 3$ and $npq = 4$.

Now $\frac{npq}{np} = \frac{4}{3}$

(i.e.) $q = \frac{4}{3} > 1$ which is not possible.

Hence the given statement is not a valid statement.

Space for Hint

Example 8. 15 :

The mean and the variance of binomial variate X with parameters n and p are 16 and 8 respectively. Find (i) $P(X=0)$, (ii) $P(X=1)$, (iii) $P(X \geq 2)$.

Solution :

Given that mean = 16 and variance = 8.

(i.e.) $\bar{x}=16$ and $\sigma^2=8$.

(i.e.) $np=16$ and $npq=8$.

Now $\dfrac{npq}{np}=\dfrac{8}{16}$

(i.e.) $q=\dfrac{1}{2}$

$\therefore\ p=1-q$

$=1-\dfrac{1}{2}$

$=\dfrac{1}{2}$

and $np=16$

(i.e.) $n\left(\dfrac{1}{2}\right)=16$

(i.e.) $n=32$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X=x)={}^{n}C_x p^x q^{n-x},\ x=0,1,2,3,\cdots,n$

(i.e.) $P(X=x)={}^{32}C_x\left(\dfrac{1}{2}\right)^x\left(\dfrac{1}{2}\right)^{32-x},\ x=0,1,2,3,\cdots,32$

(i.e.) $P(X=x)={}^{32}C_x\left(\dfrac{1}{2}\right)^{32},\ x=1,2,3,\cdots,32$

(i) Now $P(X=0)={}^{32}C_0\left(\dfrac{1}{2}\right)^{32}$

$=\left(\dfrac{1}{2}\right)^{32}$

$=\dfrac{1}{2^{32}}$

Space for Hint

Thus $P(X=0)=\frac{1}{2^{32}}$

(ii) Now $P(X=1)={}^{32}C_1\left(\frac{1}{2}\right)^{32}$

$$=32\left(\frac{1}{2^{32}}\right)$$

$$=\frac{1}{2^{27}}$$

(ii) Now $P(X=1)={}^{32}C_1\left(\frac{1}{2}\right)^{32}$

$$=32\left(\frac{1}{2^{32}}\right)$$

$$=\frac{1}{2^{27}}$$

Thus $P(X=1)=\frac{1}{2^{27}}$

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

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

$$=1-\left[{}^{32}C_0\left(\frac{1}{2^{32}}\right)+{}^{32}C_1\left(\frac{1}{2^{32}}\right)\right]$$

$$=1-\left(\frac{1}{2^{32}}\right)[1+32]$$

$$=1-\frac{33}{2^{32}}$$

Thus $P(X\geq 2)=1-\frac{33}{2^{32}}$.

Example 8. 16 :

Assume that on an average one telephone number out of 15 is busy. What is the probability that if six randomly selected telephone numbers are called

(a) not more than 3 will be busy?

(b) at least 3 of them will be busy?

Solution :

Let $p=$ probability that a telephone number is busy $=\frac{1}{15}$

Space for Hint

$\therefore\ q = 1-p$

$= 1-\dfrac{1}{15}$

$= \dfrac{14}{15}$

Given that $n = 6$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X=x) = {}^nC_x p^x q^{n-x},\ x = 0,1,2,3,\cdots,n$

(i.e.) $P(X=x) = {}^6C_x\left(\dfrac{1}{15}\right)^x\left(\dfrac{14}{15}\right)^{6-x},\ x = 0,1,2,3,\cdots,6$

(i) The probability that out of six randomly selected telephone numbers are called not more than 3 will be busy

$$= P(X \le 3)$$

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

$$= {}^6C_0\left(\frac{14}{15}\right)^6 + {}^6C_1\left(\frac{1}{15}\right)\left(\frac{14}{15}\right)^5 + {}^6C_2\left(\frac{1}{15}\right)^2\left(\frac{14}{15}\right)^4 + {}^6C_3\left(\frac{1}{15}\right)^3\left(\frac{14}{15}\right)^3$$

$$= \left(\frac{1}{15}\right)^6 (14)^3\left\{14^3+6(14)^2+14(14)+20\right\}$$

$$= 0.9997$$

The probability that out of six randomly selected telephone numbers are called not more than 3 will be busy is 0.9997

(ii) The probability that out of six randomly selected telephone numbers are called in which at least 3 will be busy

$$= P(X \ge 3)$$

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

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

$$= 1-\left\{{}^6C_0\left(\frac{14}{15}\right)^6 + {}^6C_1\left(\frac{1}{15}\right)\left(\frac{14}{15}\right)^5 + {}^6C_2\left(\frac{1}{15}\right)^2\left(\frac{14}{15}\right)^4\right\}$$

$$= 1-\left\{\left(\frac{1}{15}\right)^6 (14)^3\left\{14^3+6(14)^2+14(14)\right\}\right\}$$

$$= 1-0.7588\times 0.0044\times 295$$

$$= 1-0.9849$$

= 0.0151

The probability that out of six randomly selected telephone numbers are called in which at least 3 will be busy = 0.0151 .

Space for Hint

Example 8. 17 :

if 20% of a bolts produced by a machine are defective, determine the probability that out of 4 bolts (i) 0, (ii)1, (iii) at the most 2 bolts will be defective

Solution :

Let p = probability that selected bolt will be defective

Given that $p = 20\%$

$$= \frac{20}{100}$$

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

$\therefore\ q = 1 - p$

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

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

Given that $n = 4$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X = x) = {}^nC_x p^x q^{n-x},\ x = 0, 1, 2, 3, \cdots, n$

(i.e.) $P(X = x) = {}^4C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{4-x},\ x = 1, 2, 3, 4$

(i) The probability that 0 bolt will be defective

$$= P(X = 0)$$

$$= {}^4C_0 \left(\frac{4}{5}\right)^4$$

$$= 0.4096$$

(ii) The probability that 1 bolt will be defective

$$= P(X = 1)$$

$$= {}^4C_1 \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^3$$

Space for Hint

$= 0.4096$

(iii) The probability that at the most 2 bolts will be defective

$= P(X \leq 2)$

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

$$= {}^4C_0\left(\frac{4}{5}\right)^4 + {}^4C_1\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^3 + {}^4C_2\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$$

$= 0.4096+0.4096+0.1536$

$= 0.9728$

Example 8. 18 :

The probability that a will make a profit on any business deal is 0.8, what is the probability that he will make a profit exactly 8 times in 10 successive deals?

Solution :

Let $p=$ probability that A will make a profit on any business deal

Given that $p=0.8$

$\therefore\ q=1-p$

$=1-0.8$

$=0.2$

Given that $n=10$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X=x) = {}^nC_x p^x q^{n-x},\ x=1,2,3,\cdots,n$

(i.e.) $P(X=x) = {}^{10}C_x\,(0.8)^x\,(0.2)^{10-x},\ x=0,1,2,3,\cdots,10$

Thus the probability that a will make a profit exactly 8 times in 10 successive deals $= P(X=0)$

$= {}^{10}C_8\,(0.8)^8\,(0.2)^2$

$= 45\times0.1678\times0.04$

$= 0.3020$

Example 8. 19 :

A box contains 100 transistors, 20 of which are defective, 10 are selected for inspection. Compute the probability that (i) all 10 are defectives, (ii) all 10 are good, (iii) at least one is defective and (iv) at most 3 are defective.

Solution :

Let p = probability that a selected transistor be defective

Given that $p = 20\%$

$$= \frac{20}{100}$$

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

$\therefore\ q = 1 - p$

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

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

Given that $n = 10$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X = x) = {}^{n}C_{x} p^{x} q^{n-x},\ x = 0, 1, 2, 3, \cdots, n$

(i.e.) $P(X = x) = {}^{10}C_{x}\left(\frac{1}{5}\right)^{x}\left(\frac{4}{5}\right)^{10-x},\ x = 0, 1, 2, 3, 4, \cdots, 100$

(i) The probability that all 10 transistors will be defective

$$= P(X = 10)$$

$$= {}^{10}C_{10}\left(\frac{1}{5}\right)^{10}\left(\frac{4}{5}\right)^{0}$$

$$= \frac{1}{5^{10}}$$

(ii) The probability that all 10 transistors will be good

= The probability that 0 transistor will be defective

$$= P(X = 0)$$

$$= {}^{10}C_{0}\left(\frac{1}{5}\right)^{0}\left(\frac{4}{5}\right)^{10}$$

$$= 0.1074$$

Space for Hint

Space for Hint

(iii) The probability that at least 1 transistor will be defective

$$= P(X \geq 1)$$

$$= 1 - P(X < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - {}^{10}C_{10}\left(\frac{1}{5}\right)^{10}\left(\frac{4}{5}\right)^{0}$$

$$= 1 - \frac{1}{5^{10}}$$

(iv) The probability that at most 3 transistors will be defective

$$= P(X \leq 3)$$

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

$$= {}^{10}C_0\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10} + {}^{10}C_1\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^9 + {}^{10}C_2\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^8 + {}^{10}C_3\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7$$

$$= 0.1074 + 0.2684 + 0.3020 + 0.2013$$

$$= 0.8791$$

Example 8. 20 :

Out of 1000 families with 5 children each, how many would you expect to have (i) 3 boys, (ii) 5 girls, (iii) either 2 or 3 boys.(Assume equal probabilities for boys and girls)

Solution :

Let $p =$ probability that a child is a boy

Given that $p = \frac{1}{2}$

$\therefore\ q = 1 - p$

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

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

Given that $n = 5$ and $N = 1000$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X = x) = {}^{n}C_x p^x q^{n-x}$, $x = 0, 1, 2, 3, \cdots, n$

(i.e.) $P(X=x)={}^5C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x}$, $x=0,1,2,3,4,5$

(i.e.) $P(X=x)={}^5C_x\left(\frac{1}{2}\right)^5$, $x=0,1,2,3,4,5$

(i.e.) $P(X=x)=\frac{1}{32}\,{}^5C_x$, $x=0,1,2,3,4,5$

(i) The probability that a family has 3 boys

$= P(X=3)$

$= \frac{1}{32}\,{}^5C_3$

$= \frac{1}{32}(10)$

$= 0.3125$

Thus number of families having 3 boys $= N\cdot P(X=3)$

$= 1000\times 0.3125$

$= 312.5$

≈ 313

(ii) The probability that a family has 5 girls

= The probability that a family has 0 boys

$= P(X=0)$

$= \frac{1}{32}\,{}^5C_0$

$= \frac{1}{32}$

$= 0.0313$

Thus number of families having 3 boys $= N\cdot P(X=3)$

$= 1000\times 0.0313$

$= 31.3$

≈ 31

(iii) The probability that a family has 2 or 3 boys

$= P(X=2)+P(X=3)$

$= \frac{1}{32}\,{}^5C_2+\frac{1}{32}\,{}^5C_3$

Space for Hint

Space for Hint

$$= \frac{1}{32}(10) + \frac{1}{32}(10)$$

$$= \frac{20}{32}$$

$$= 0.6250$$

Thus number of families having 3 boys $= N \cdot P(X=3)$

$$= 1000 \times 0.6250$$

$$= 625$$

Example 8. 21 :

Seven unbiased coins are tossed and number of heads noted. The experiment is repeated 128 times and the following distribution is obtained. Fit a Binomial distribution and find the expected frequencies.

No. of heads	0	1	2	3	4	5	6	7
Frequencies	7	6	19	35	30	23	7	1

Solution :

Let $p =$ probability of getting a head when a coin is tossed

Given that $p = \frac{1}{2}$

$\therefore\ q = 1 - p$

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

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

Given that $n = 7$ and $N = 128$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X = x) = {}^nC_x p^x q^{n-x},\ x = 0, 1, 2, 3, \cdots, n$

(i.e.) $P(X = x) = {}^7C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{7-x},\ x = 0, 1, 2, 3, \cdots, 7$

(i.e.) $P(X = x) = {}^7C_x \left(\frac{1}{2}\right)^7,\ x = 0, 1, 2, 3, \cdots, 7$

(i.e.) $P(X=x)=\frac{1}{128}\,{}^{7}C_x,\ x=0,1,2,3,\cdots,7$

The recurrence formula of Binomial distribution is

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

The following table shows the probability of the distribution and the expected frequencies.

x	$P(X=x)=p(x)$	$N\cdot p(x)$
0	$\frac{1}{128}\,{}^{7}C_0=\frac{1}{128}$	$128\cdot\frac{1}{128}=1$
1	$\left(\frac{7-0}{0+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{1}{128}\right)=\frac{7}{128}$	$128\cdot\frac{7}{128}=7$
2	$\left(\frac{7-1}{1+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{7}{128}\right)=\frac{21}{128}$	$128\cdot\frac{21}{128}=21$
3	$\left(\frac{7-2}{2+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{21}{128}\right)=\frac{35}{128}$	$128\cdot\frac{35}{128}=35$
4	$\left(\frac{7-3}{3+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{35}{128}\right)=\frac{35}{128}$	$128\cdot\frac{35}{128}=35$
5	$\left(\frac{7-4}{4+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{21}{128}\right)=\frac{35}{128}$	$128\cdot\frac{21}{128}=21$
6	$\left(\frac{7-5}{5+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{21}{128}\right)=\frac{7}{128}$	$128\cdot\frac{7}{128}=7$
7	$\left(\frac{7-6}{6+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{7}{128}\right)=\frac{1}{128}$	$128\cdot\frac{1}{128}=1$

Space for Hint

Space for Hint

The second column of the above table shows the probability of getting $0,1,2,\cdots,7$ heads and the last column of the table shows the expected frequencies of the experiment.

Example 8. 22 :

Five dice were thrown together 96 times. The number of times 4, 5 or 6 was actually thrown in experiment is given below. Calculate the expected frequencies.

No. of dice	0	1	2	3	4	5
Frequencies	1	10	24	35	18	8

Solution :

Let $p =$ probability of getting a 4, 5 or 6 when a dice is thrown

Given that $p=\frac{1}{2}$

$\therefore\ q=1-p$

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

$=\frac{1}{2}$

Given that $n=5$ and $N=96$

$\therefore$ The probability mass function of the Binomial distribution is

$P(X=x)={}^nC_x p^x q^{n-x},\ x=0,1,2,3,\cdots,n$

(i.e.) $P(X=x)={}^5C_x\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x},\ x=0,1,2,3,4,5$

(i.e.) $P(X=x)={}^5C_x\left(\frac{1}{2}\right)^5,\ x=0,1,2,3,4,5$

(i.e.) $P(X=x)=\frac{1}{32}\,{}^5C_x,\ x=0,1,2,3,4,5$

The recurrence formula of Binomial distribution is

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

The following table shows the probability of the distribution and the expected frequencies.

x	$P(X=x)=p(x)$	$N\cdot p(x)$
0	$\frac{1}{32}\,{}^{5}C_0=\frac{1}{32}$	$96\cdot\frac{1}{32}=3$
1	$\left(\frac{5-0}{0+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{1}{32}\right)=\frac{5}{32}$	$96\cdot\frac{5}{32}=15$
2	$\left(\frac{5-1}{1+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{5}{32}\right)=\frac{10}{32}$	$96\cdot\frac{10}{32}=30$
3	$\left(\frac{5-2}{2+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{5}{32}\right)=\frac{10}{32}$	$96\cdot\frac{10}{32}=30$
4	$\left(\frac{5-3}{3+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{1}{32}\right)=\frac{5}{32}$	$96\cdot\frac{5}{32}=15$
5	$\left(\frac{5-4}{4+1}\right)\left(\frac{1/2}{1/2}\right)\left(\frac{1}{32}\right)=\frac{1}{32}$	$96\cdot\frac{1}{32}=3$

The second column of the above table shows the probability of getting $0,1,2,3,4,5$ heads and the last column of the table shows the expected frequencies of the experiment.

Check Your Progress

(1) Five fair coins are tossed and number of tails noted. The experiment is repeated 256 times and the following distribution is obtained. Fit a Binomial distribution and find the expected frequencies.

Space for Hint

Space for Hint

No. of heads	0	1	2	3	4	5
Frequencies	12	24	96	104	16	4

(Answer : The expected frequencies are

No. of heads	0	1	2	3	4	5
Frequencies	8	40	80	80	40	8

)

(2) An insurance company accepts policies of 5 persons all of identical age and in good health. The probability of a person of this age will be alive 35 years hence is 0.6. Find the probability that in 35 years (1) all five persons, (ii) at least two persons and (iii) at most three persons will alive.

(3) Six dice are thrown 729 times. How many times do you expect at least two dice shows one or two.

(Answer : 233)

8.4 Poisson distribution

A second important probability distribution is the Poisson distribution, named after the French mathematician S.Poisson in 1837.

The characteristics of the Poisson distribution are :

(i) the occurrence of the events is independent. That is, the occurrence of an event in an interval of time has no effect on the probability of a second occurrence of the event in the same, or any interval of time.

(ii) Theoretically, an infinite number of occurrences of the event must be possible in the interval of time.

(iii) The probability of single occurrence of the event in a given interval is proportional to the length of the interval.

(iv) In any infinitesimal portion of interval, the probability of two or more occurrences of the event is negligible.

Definition :

A random variable X is said to have a Poisson distribution if its probability mass function is given by $p(x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$

Space for Hint

Note :

(1) In Poisson distribution λ are called parameters

(2) Poisson distribution with parameter λ is denoted by $P(\lambda)$

The Poisson distribution is attributable in the case of the rare events. The following are some instances where Poisson distribution may be employed.

(1) Number of deaths due to rare disease such as snake bite, cancer.

(2) The number of defective articles in a packing manufactured well reputed company.

(3) The number children born blind per year in a city.

(4) The number of phone calls received in particular time interval.

Example 8. 23 :

Find the moment generating function of Poisson distribution. Also find is means and variance.

Solution :

Step 1 : Let X have a Poisson distribution.

(i.e.) the probability density function of Poisson distribution is

$$p(x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

The moment generating function of x is $M(t)=E\left(e^{tX}\right)$

$$= \sum_{x=-\infty}^{\infty} e^{tx}\, \mathrm{p}(x)$$

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

Space for Hint

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

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

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

(i.e.) $M_X(t)=e^{\lambda\left(e^{t}-1\right)}$ is the moment generating function of Poisson distribution.

Step 2 : To find the mean and variance

We know that $M_X(t)=e^{\lambda\left(e^{t}-1\right)}$ is the moment generating function of Poisson distribution.

$\therefore\ \log\left(M_X(t)\right)=\lambda\left(e^{t}-1\right)$

Differentiate $M_x(t)$ twice with respect to t, we have,

$$\frac{1}{M_X(t)}M'_X(t)=\lambda e^{t}$$

(i.e.) $M'_X(t)=\lambda e^{t}M_X(t)$

and $M''_X(t)=me^{t}M'_X(t)+me^{t}M_X(t)$

(i.e.) $M''_X(t)=me^{t}me^{t}M_X(t)+me^{t}M_X(t)$.

$\therefore\ M_X(0)=e^{\lambda(1-1)}=1$

and $M'_X(0)=\lambda e^{0}M_X(0)=\lambda$

and $M''_X(0)=\lambda^2+\lambda$.

$\therefore \sigma^2=M''_X(0)-\left(M'_X(0)\right)$

$= \cancel{\lambda^2}+\lambda-\cancel{\lambda^2}$

$=\lambda$

Thus mean and variance of the Poisson distribution are same and it is equal to λ

(i.e) $\boxed{\mu=\sigma^2=\lambda}$

Example 8. 24 :

State and prove the recurrence relation of probability density function in Poisson distribution.

Solution :

Statement : If X follows Poisson distribution then $p(x+1)=\left(\frac{\lambda}{x+1}\right)p(x)$

Proof :

Let X have a Poisson distribution.

$\therefore$ The probability density function of Poisson distribution is

$$p(x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

$$\text{Now } \frac{p(x+1)}{p(x)} = \frac{\dfrac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}}{\dfrac{e^{-\lambda}\lambda^{x}}{x!}}$$

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

$$= \frac{\cancel{e^{-\lambda}}\,\cancel{\lambda^{x}}\,\lambda}{(x+1)\cancel{x!}}\times\frac{\cancel{x!}}{\cancel{e^{-\lambda}}\,\cancel{\lambda^{x}}}$$

$$= \frac{\lambda}{x+1}$$

Hence $p(x+1)=\left(\frac{\lambda}{x+1}\right)p(x)$.

This proves the recurrence relation for $p(x)$ in Binomial distribution.

Example 8. 25 :

Find the mode of the Poisson distribution.

Solution :

Let X have a Poisson distribution

$\therefore$ The probability density function of Poisson distribution is

$$p(x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

Space for Hint

Space for Hint

Let x be the mode of the distribution.

$\therefore\ p(x-1) \le p(x) \ge p(x+1)$ -- (8.4)

$$\text{Now } \frac{p(x+1)}{p(x)} = \frac{\dfrac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}}{\dfrac{e^{-\lambda}\lambda^{x}}{x!}}$$

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

$$= \frac{\cancel{e^{-\lambda}}\cancel{\lambda^{x}}\lambda}{(x+1)\cancel{x!}} \times \frac{\cancel{x!}}{\cancel{e^{-\lambda}}\cancel{\lambda^{x}}}$$

$$= \frac{\lambda}{x+1}$$

Now from (8.4) $p(x) \ge p(x+1)$

$$\Rightarrow \frac{p(x+1)}{p(x)} \le 1$$

$$\Rightarrow \frac{\lambda}{x+1} \le 1$$

$$\Rightarrow \lambda \le x+1$$

$\Rightarrow \lambda - 1 \le x$ -- (8.5)

Similarly $p(x-1) \le p(x) \Rightarrow x \le \lambda$ -- (8.6)

From (8.5) and (8.6), we have, $\lambda - 1 \le x \le \lambda$ -- (8.7)

If λ in not an integer then the integral part of λ is the mode of the distribution.

If λ is an integer then λ and $\lambda - 1$ are the modes of the distribution.

Example 8. 26 :

State and prove addition property of Poisson distribution.

Statement : If X and Y are two independent Poisson variates with parameters λ_1 and λ_2 then $X+Y$ is also a Poisson variate with parameter $\lambda_1 + \lambda_2$

Proof :

Let X and Y are two independent Poisson variates with parameters λ_1 and λ_2

Let $M_X(t)$ and $M_Y(t)$ be the moment generating functions of X and Y respectively.

Space for Hint

Thus $M_X(t)=e^{\lambda_1(e^t-1)}$ and $M_Y(t)=e^{\lambda_2(e^t-1)}$

Hence $M_{X+Y}(t)=M_X(t)M_Y(t)$ {since X and Y are independent random variables}

(i.e.) $M_{X+Y}(t)=e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)}$

(i.e.) $M_{X+Y}(t)=e^{\lambda_1(e^t-1)+\lambda_2(e^t-1)}$

(i.e.) $M_{X+Y}(t)=e^{(\lambda_1++\lambda_2)(e^t-1)}$

which is the moment generating function of Poisson distribution with parameter $\lambda_1+\lambda_2$.

(i.e.) $X+Y$ is also a Poisson variate with parameter $\lambda_1+\lambda_2$.

This proves the problem.

Example 8. 27 :

Find the cumulants of a Poisson distribution.

Solution :

Let X have a Poisson distribution

$\therefore$ The probability density function of Poisson distribution is

$$p(x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!} & ; x=0,1,2,3,\cdots\\ 0 & ; elsewhere\end{cases}$$

Let $M_X(t)$ be the moment generating function of X.

We know that $M_X(t)=e^{\lambda(e^t-1)}$

Thus $K_X(t)=\log M_X(t)$

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

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

$$= \lambda\left(1+\frac{t}{1!}+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots\right)$$

We know that $\kappa_r = r^{th}$ cumulant

$$= \text{coefficient of } \frac{t^r}{r!} \text{ in } K_X(t)$$

Space for Hint

$= \lambda$

Thus $\kappa_r = \lambda$ for $r = 1, 2, 3, \cdots$.

Example 8. 28 :

State and prove the recurrence relation for the moments of the Poisson distribution.

Statement : If X be Poisson variate then $\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\dfrac{d\mu_r}{d\lambda}$.

Proof :

Let X be Poisson variate.

We know that the mean of Poisson distribution is λ.

Then $\mu_r = E(X-\mu)^r$

$$= E(X-\lambda)^r$$

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

Now $\dfrac{d\mu_r}{d\lambda}$

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

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

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

$$\therefore\ \lambda\frac{d\mu_r}{d\lambda} = -r\lambda\mu_{r-1} + \sum_{x=0}^{\infty}\frac{(x-\lambda)^{r+1}\lambda^x}{x!}$$

(i.e.) $\lambda\dfrac{d\mu_r}{d\lambda} = -r\lambda\mu_{r-1} + \mu_{r+1}$

Thus $\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\dfrac{d\mu_r}{d\lambda}$

This proves the recurrence relation for the moments of the Poisson distribution.

Example 8. 29 :

If X is a Poisson variate such that $P(X=1)=P(X=2)$ then find $P(X=4)$.

Proof :

Given that X is Poisson variate

$\therefore$ The probability mass function of X is

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

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

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

$\Rightarrow 2\lambda=\lambda^2$

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

$\Rightarrow \lambda=0$ or $\lambda=2$

Since $\lambda>0$, we have $\lambda=2$.

Thus the probability mass function of Poisson distribution is

$$P(X=x)=\begin{cases}\dfrac{e^{-2}2^x}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

Hence $P(X=4)=\dfrac{e^{-2}2^4}{4!}$

$$=(0.1353)\left(\frac{16}{24}\right)$$

$$=0.0902.$$

Example 8. 30 :

If X is a Poisson variate with $P(X=0)$ is 10%, dind the mean of the distribution.

Proof :

Given that X is Poisson variate

Space for Hint

Space for Hint

$\therefore$ The probability mass function of X is

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

Given that $P(X=0)=10\%$

$\Rightarrow \dfrac{e^{-\lambda}\lambda^{0}}{0!}=0.1$

$\Rightarrow e^{-\lambda}=0.1$

$\Rightarrow -\lambda=-2.3023$

$\Rightarrow \lambda=2.3023$

Therefore the mean of the distribution is 2.3023.

Example 8. 31 :

If 5% of the electric bulb manufactured by a company are defective, use Poisson distribution find the probability that in a sample space of 100 bulbs (i) none is defective, (ii) 5 bulbs will be defective and (iii) at least 2 bulbs will be defective.

Solution :

Given that probability of a bulb is defective = $p=5\%$

(i.e.) $p=\dfrac{5}{100}$

and sample size = $n=100$.

Hence mean of the Poisson distribution = np

(i.e.) $\lambda=\cancel{100}\times\dfrac{5}{\cancel{100}}$

(i.e.) $\lambda=5$

Thus the probability mass function of Poisson distribution is

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}$$

(i.e.) $$P(X=x)=\begin{cases}\dfrac{e^{-5}5^{x}}{x!} & ; x=0,1,2,3,\cdots \\ 0 & ; elsewhere\end{cases}.$$

Space for Hint

(i) The probability that none of the bulb is defective

$= P(X=0)$

$= \dfrac{e^{-5}5^0}{0!}$

$= e^{-5}$

$= 0.0067$

and (ii) The probability that five bulbs will be defective

$= P(X=5)$

$= \dfrac{e^{-5}5^5}{5!}$

$= 0.1745$

and (iii) The probability that at least 2 bulbs will defective

$= P(X \geq 2)$

$= 1 - P(X < 2)$

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

$= 1 - \left[\dfrac{e^{-5}5^0}{0!} + \dfrac{e^{-5}5^1}{1!}\right]$

$= 1 - [0.0067 + 0.0337]$

$= 1 - 0.0404$

$= 0.9596$

Example 8. 32 :

The probability that a person aged 50 years will die within a year is 0.01125. What is the probability that out of 12 such persons at least eleven will reach their 51^{st} birthday?

Solution :

Given that probability that a person will within a year = 0.01125

$\therefore$ the probability that a person will after a year $= 1 - 0.01125$

(i.e.) $p = 0.9888$

and number person chosen = $n = 12$

Hence mean of the Poisson distribution = np

(i.e.) $\lambda = 0.9888 \times 12$

Space for Hint

(i.e.) $\lambda = 11.865$

Thus the probability mass function of Poisson distribution is

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & ;\ x = 0,1,2,3,\cdots \\ 0 & ;\ elsewhere \end{cases}$$

(i.e.) $$P(X = x) = \begin{cases} \dfrac{e^{-11.865}(11.865)^x}{x!} & ;\ x = 0,1,2,3,\cdots \\ 0 & ;\ elsewhere \end{cases}$$

Probability that at least eleven live after a year

$$= P(X \geq 11)$$

$$= P(X = 11) + P(X = 12)$$

$$= \frac{e^{-11.865}(11.865)^{11}}{11!} + \frac{e^{-11.865}(11.865)^{12}}{12!}$$

$$= 0.1156 + 0.1143$$

$$= 0.2299$$

Example 8. 33 :

A manufacturer of cotter pins knows that 5% of his products is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 4 pins will be defective, what is the probability that a box will fail to meet the guaranteed quality?

Solution :

Given that probability that a cotter pin is defective = 5%

(i.e.) $p = \dfrac{5}{100}$

and number of cotter pins in a box = $n = 100$

Hence mean of the Poisson distribution = np

(i.e.) $\lambda = \cancel{100} \times \dfrac{5}{\cancel{100}}$

(i.e.) $\lambda = 5$

Thus the probability mass function of Poisson distribution is

Space for Hint

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & ;\ x=0,1,2,3,\cdots\\ 0 & ;\ elsewhere\end{cases}$$

(i.e.) $$P(X=x)=\begin{cases}\dfrac{e^{-5}5^{x}}{x!} & ;\ x=0,1,2,3,\cdots\\ 0 & ;\ elsewhere\end{cases}$$

Probability that a box will fail to meet the guarantee

$= P(X \geq 5)$

$= 1 - P(X < 5)$

$= 1 - [P(X=1)+P(X=2)+P(X=3)+P(X=4)]$

$= 1 - e^{-5}\left[\dfrac{5^0}{0!}+\dfrac{5^1}{1!}+\dfrac{5^2}{2!}+\dfrac{5^3}{3!}+\dfrac{5^4}{4!}\right]$

$= 0.4405$

Example 8. 34 :

The distribution of typing mistakes committed by a typist is given below. Assuming to be a Poisson model find the expected frequencies.

Mistakes per page	0	1	2	3	4	5
Number of pages	142	156	69	27	5	1

Solution :

Step 1 : First we shall find the mean of the distribution.

We know that mean $= \overline{x} = \dfrac{\sum fx}{\sum f}$

(i.e.) $\overline{x} = \dfrac{400}{400}$

(i.e.) $\overline{x} = 1$

The recurrence formula of Poisson distribution is

$$p(x+1)=\left(\frac{\lambda}{x+1}\right)p(x).$$

Space for Hint

The following table shows the probability of the distribution and the expected frequencies.

x	f	fx	$P(X=x)=p(x)$	$N\cdot p(x)$
0	142	0	$e^{-1}=0.3679$	$147.15\approx 147$
1	156	156	$\frac{1}{0+1}(0.3679)=0.3679$	147
2	69	138	$\frac{1}{1+1}(0.3679)=0.184$	$73.5\approx 74$
3	27	81	$\frac{1}{2+1}(0.184)=0.0613$	$24.5\approx 25$
4	5	20	$\frac{1}{3+1}(0.0163)=0.0153$	$6.13\approx 6$
5	1	5	$\frac{1}{4+1}(0.0153)=0.0031$	$1.224\approx 1$
Total	400	400		

The last column of the table shows the expected frequencies of the experiment.

Example 8. 35 :

In 1000 sets of trials fro an event comparatively small probability the frequencies of the number of success are given below.

Space for Hint

Successes	0	1	2	3	4	5	6	7
Frequency	305	365	210	80	28	9	2	1

Fit a Poisson distribution to the above data and calculate the expected frequencies.

Solution :

The following table shows the probability of the distribution and the expected frequencies.

x	f	fx	$P(X=x)=p(x)$	$N\cdot p(x)$
0	305	0	0.3012	$301.2 \approx 301$
1	365	365	$\frac{1.2}{0+1}(0.3012)=0.3614$	$361.4 \approx 361$
2	210	420	$\frac{1.2}{1+1}(0.3614)=0.2169$	$216.9 \approx 217$
3	80	240	$\frac{1.2}{2+1}(0.2169)=0.0867$	$86.7 \approx 87$
4	28	112	$\frac{1.2}{3+1}(0.0.0867)=0.0260$	26
5	9	45	$\frac{1.2}{4+1}(0.0260)=0.0062$	$6.2 \approx 6$
6	2	12	$\frac{1.2}{3+1}(0.0062)=0.0012$	$1.2 \approx 1$
7	1	7	$\frac{1.2}{4+1}(0.0012)=0.0002$	$0.2 \approx 0$
Total	1000	1201		

Space for Hint

Step 1 : First we shall find the mean of the distribution.

We know that mean $= \bar{x} = \dfrac{\Sigma fx}{\Sigma f}$

(i.e.) $\bar{x} = \dfrac{400}{400}$

(i.e.) $\bar{x} = 1$

The recurrence formula of Poisson distribution is

$$p(x+1) = \left(\frac{\lambda}{x+1}\right) p(x).$$

The last column of the table shows the expected frequencies of the experiment.

Check Your Progress

(1) Between 2 p.m and 4 p.m the average number of phone calls per minute coming into the switch board of a company is 2.35. Find the probability that during the particular minute there will be at most 2 phone calls.

(Answer : 0.583)

(2) Assuming that one in 80 births is a case of twins, calculate the probability of 2 or more births of twins on a day when 30 births occur.

(Answer : 0.055)

(3) In a certain factory producing razor blades there is small chance of $\dfrac{1}{500}$ for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution, calculate the approximate number of packets containing (i) no defective, (ii) one defective and (iii) two defective blades respectively in a consignment of 10000 packets.

(answer : (i) 9802, (ii) 196, (iii) 2)

(4) Fit a Poisson distribution to the following data.

x	0	1	2	3	4
f	142	156	69	27	5

(Answer : The expected frequencies are 121, 61, 15, 3, 1)

(5) Fit a Poisson distribution to the following data.

Births	0	1	2	3	4
Frequency	122	60	15	2	1

(Answer : The expected frequencies are 122, 61, 15, 2, 0)

8. 5 Normal distribution

The Normal distribution ws discovered by De Moivre as the limiting case Binomial model in 1733. It was also known to Laplace no longer than 1774, but through a historical error it has been credited to Gauss, who first made reference to it in 1809. Throughout the 18^{th} and 19^{th} centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables, the name Normal. These efforts failed because of the false premises. The normal model has become the most important probability model in statistical analysis.

Definition :

Let X be a continuous random variable having the probability density function $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $-\infty < x < \infty$ where $\sigma > 0$.

Example 8. 36 :

Find the moment generating function of the normal distribution and also find mean and variance.

Solution :

Let X be a Normal distribution.

Space for Hint

Space for Hint

$\therefore$ the probability density function of Normal distribution is

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

The moment generating function of X

$$= M_X(t)$$

$$= E\left(e^{tX}\right)$$

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

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

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2}\left(\left(\frac{x-\mu}{\sigma}\right)^2-2tx\right)} dx \quad \text{-------------------- (8.8)}$$

Now $\left(\frac{x-\mu}{\sigma}\right)^2-2tx$

$$= \frac{1}{\sigma^2}\left[x^2-2\mu x+\mu^2-2t\sigma^2 x\right]$$

$$= \frac{1}{\sigma^2}\left\{[x-(\mu+\sigma^2 t)]^2-2\mu\sigma^2 t-t^2\sigma^2\right\}$$

$$= \frac{1}{\sigma^2}\left\{[x-(\mu\sigma^2 t)]^2-2\mu\sigma^2\left(\mu t+\sigma^2\frac{t^2}{2}\right)\right\}$$

$$= \frac{1}{\sigma^2}\left(x-(\mu\sigma^2 t)\right)^2-2\left(\mu t+\sigma^2\frac{t^2}{2}\right)$$

$\therefore (8.8) \Rightarrow M_X(t)$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left\{\frac{1}{\sigma^2}\left(x-(\mu\sigma^2 t)\right)^2-2\left(\mu t+\sigma^2\frac{t^2}{2}\right)\right\}\right)dx$$

Space for Hint

$$= \int_{-\infty}^{\infty} e^{\mu t + \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi b}} e^{-\frac{1}{2}\left[\frac{x-(\mu+\sigma^2 t)}{\sigma}\right]^2} dx$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi b}} e^{-\frac{1}{2}\left[\frac{x-(\mu+\sigma^2 t)}{\sigma}\right]^2} dx$$

$= e^{at + \frac{b^2 t^2}{2}}$ (1) ($\because$ the integrand can be thought of as a normal probability density function with a mean $a + b^2 t$)

$\therefore \; M_X(t) = e^{at + \frac{b^2 t^2}{2}}$

which is the required moment generating function of normal distribution.

Step 2 : To find the mean & variance

Now $M_X(t) = e^{at + \frac{b^2 t^2}{2}}$

Differentiate $M_X(t)$ twice with respect to t, we get,

$\therefore \; M'_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} (\mu + \sigma^2 t)$

(i.e.) $M'_X(t) = M_X(t)(\mu + \sigma^2 t)$

and $M''_X(t) = M'_X(t)(\mu + \sigma^2 t) + M_X(t)(\sigma^2)$

Thus $M_X(0) = 1$

$M'_X(0) = \mu$ and

$M''_X(0) = \mu^2 + \sigma^2$

$\therefore$ Mean = μ = $M'_X(0) = \mu$

and variance = σ² = $M''_X(0) - (M'_X(0))^2$

$= \cancel{\mu^2} + \sigma^2 - \cancel{\mu^2}$

$= \sigma^2$

Note : A Normal distribution with μ and variance σ^2 is denoted by $N(\mu, \sigma^2)$ or $n(\mu, \sigma^2)$.

Space for Hint

Standard Normal variate

If the random variable X having mean 0 and variance 1 is called standard normal variate.

Thus the probability density function of standard normal distribution is

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \quad ; -\infty < z < \infty$$

The relationship between Normal variate X and standard normal variate Z is $z=\frac{x-\mu}{\sigma}$.

Note : We know that moment generating function of a Normal distribution $X \sim N(\mu,\sigma^2)$ is $e^{\mu t+\frac{1}{2}\sigma^2 t^2}$.

Thus moment generating function of a standard normal distribution $X \sim N(0,1)$ is $e^{\frac{1}{2}t^2}$

Example 8. 37 :

Find the mode of the Normal distribution.

Solution :

Let $X \sim N(\mu,\sigma^2)$

Let X be a Normal distribution.

$\therefore$ the probability density function of Normal distribution is

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

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

Differentiate $f(x)$ with respect to x twice, we have,

$$\frac{f'(x)}{f(x)}=-\frac{1}{\cancel{2}}\cancel{2}\left(\frac{x-\mu}{\sigma}\right)\left(\frac{1}{\sigma}\right)$$

(i.e.) $f'(x)=-\left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)f(x)$

and $f''(x)=-\left(\frac{1}{\sigma^2}\right)f(x)-\left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)f'(x)$

Space for Hint

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right) f(x)\left[1 + (x-\mu)\frac{f'(x)}{f(x)}\right]$

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right) f(x)\left[1 + (x-\mu)\left(\frac{x-\mu}{\sigma}\right)\left(\frac{1}{\sigma}\right)\right]$

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right) f(x)\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$.

For maximum or minimum $f(x)$, put $f'(x) = 0$

(i.e.) $-\left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right) f(x) = 0$

$\Rightarrow x = \mu$ since $\sigma > 0$ and $f(x) \neq 0$.

When $x = \mu$ then $f''(\mu) = -\left(\frac{1}{\sigma^2}\right) f(\mu) < 0$

(i.e.) $f(x)$ attains its maximum value at $x = \mu$

Hence mode of the Normal distribution is $x = \mu$.

Note : since $f(x)$ is symmetrical about $x = \mu$ then the median is $x = \mu$.

Hence for a Normal distribution mean, median and mode are equal to $x = \mu$.

Example 8. 38 :

Find the points of inflexion for a Normal distribution.

Solution :

Let $X \sim N(\mu, \sigma^2)$

$\therefore$ the probability density function of Normal distribution is

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

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

Differentiate $f(x)$ with respect to x thrice, we have,

$\frac{f'(x)}{f(x)} = -\frac{1}{\cancel{2}}\cancel{2}\left(\frac{x-\mu}{\sigma}\right)\left(\frac{1}{\sigma}\right)$

Space for Hint

(i.e.) $f'(x) = -\left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)f(x)$

and $f''(x) = -\left(\frac{1}{\sigma^2}\right)f(x) - \left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)f'(x)$

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right)f(x)\left[1+(x-\mu)\frac{f'(x)}{f(x)}\right]$

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right)f(x)\left[1+(x-\mu)\left(\frac{x-\mu}{\sigma}\right)\left(\frac{1}{\sigma}\right)\right]$

(i.e.) $f''(x) = -\left(\frac{1}{\sigma^2}\right)f(x)\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]$.

and $f'''(x) = -\left(\frac{1}{\sigma^2}\right)\left\{f'(x)\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]+f(x)\left[\frac{2(x-\mu)}{\sigma^2}\right]\right\}$

(i.e.) $f'''(x) = -\left(\frac{1}{\sigma^2}\right)\left\{-\left(\frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)f(x)\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]+f(x)\left[\frac{2(x-\mu)}{\sigma^2}\right]\right\}$

(i.e.) $f'''(x) = \left(\frac{1}{\sigma^2}\right)\left(\frac{x-\mu}{\sigma^2}\right)f(x)\left\{3-\frac{(x-\mu)^2}{\sigma^2}\right\}$

For points of inflexion put $f''(x) = 0$

(i.e.) $-\left(\frac{1}{\sigma^2}\right)f(x)\left[1-\frac{(x-\mu)^2}{\sigma^2}\right] = 0$

(i.e.) $1-\frac{(x-\mu)^2}{\sigma^2} = 0$

(i.e.) $x = \mu \pm \sigma$

Thus at $x = \mu \pm \sigma$, $f'''(x) \neq 0$

Thus the points of inflexion of the Normal distribution are $x = \mu \pm \sigma$

Example 8. 39 :

Find the mean deviation about mean for a Normal distribution.

Solution :

Let $X \sim N(\mu, \sigma^2)$

$\therefore$ the probability density function of Normal distribution is

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

We know that mean of the Normal distribution is μ.

Now the mean deviation about mean =

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

$$= \int_{x=-\infty}^{\infty} |x-\mu| \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \text{ ------------------(8.9)}$$

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

Then $dz = \frac{dx}{\sigma}$

Now (8.9) $\Rightarrow$ M.D about mean $= \int_{z=-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{z^2}{2}} \sigma dz$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_{z=0}^{\infty} |z| e^{-\frac{1}{2}z^2} dz \text{ (since } |z| e^{-\frac{z^2}{2}} \text{ is even)}$$

$$= \sigma\sqrt{\frac{2}{\pi}} \int_{z=0}^{\infty} z e^{-\frac{1}{2}z^2} dz$$

$$= \sigma\sqrt{\frac{2}{\pi}} \int_{t=0}^{\infty} e^{-t} dt \text{ (by putting } \frac{z^2}{2} = t)$$

$$= \sigma\sqrt{\frac{2}{\pi}} \left[e^{-t}\right]_0^{\infty}$$

$$= \sigma\sqrt{\frac{2}{\pi}}$$

Example 8. 40 :

Find μ_{2n+1} and μ_{2n} for a Normal distribution.

Solution :

Let $X \sim N(\mu, \sigma^2)$

We know that the moment generating function about mean is $M_{X=\mu}(t)$

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

Space for Hint

Space for Hint

$$= 1+\frac{1}{1!}\left(\frac{\sigma^2t^2}{2}\right)+\frac{1}{2!}\left(\frac{\sigma^2t^2}{2}\right)^2+\frac{1}{3!}\left(\frac{\sigma^2t^2}{2}\right)^3+\cdots \text{ --------------- } (8.10)$$

We know that μ_r = coefficient of $\frac{t^r}{r!}$ in the moment generating function about mean.

Thus for $r=2n+1$, μ_{2n+1} = coefficient of $\frac{t^{2n+1}}{(2n+1)!}$ in (8.10)

(i.e.) $\mu_{2n+1}=0$

Again μ_{2n} = coefficient of $\frac{t^{2n}}{(2n)!}$ in (8.10)

$$= \frac{\sigma^{2n}(2n)!}{2^n n!}$$

$$= \frac{\sigma^{2n}}{2^n n!}[1\cdot 2\cdot 3\cdots(2n-1)\cdot 2n]$$

$$= \frac{\sigma^{2n}}{2^n n!}[1\cdot 3\cdot 5\cdots(2n-1)]\cdot[2\cdot 4\cdot 6\cdots 2n]$$

$$= \frac{\sigma^{2n}}{\cancel{2^n} n!}\left[1\cdot 3\cdot 5\cdots(2n-1)\right]\cdot \cancel{2^n}\left[1\cdot 2\cdot 3\cdots n\right]$$

$$= \frac{\sigma^{2n}}{\cancel{n!}}[1\cdot 3\cdot 5\cdots(2n-1)]\cdot \cancel{n!}$$

$$= \sigma^{2n}[1\cdot 3\cdot 5\cdots(2n-1)]$$

Hence $\mu_{2n+1}=0$ and $\mu_{2n}=\sigma^{2n}[1\cdot 3\cdot 5\cdots(2n-1)]$.

Note : From the above example $\mu_1=0$, $\mu_3=0$, $\mu_2=\sigma^2$ and $\mu_4=3\sigma^4$.

Thus $\beta_1=\frac{\mu_3^2}{\mu_2^3}$

(i.e.) $\beta_1=0$

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

(i.e.) $\beta_2=\frac{3\cancel{\sigma^4}}{\cancel{\sigma^4}}$

(i.e.) $\beta_2=3$.

Theorem 8. 1 :

Let $X_1, X_2, X_3, \cdots, X_n$ be mutually stochastically independent random variables having, respectively, the normal distributions $n\left(\mu_i, \sigma_i^2\right)$, $i = 1,2,3,\cdots,n$

Let $Y = k_1X_1 + k_2X_2 + k_3X_3 + \cdots + k_nX_n$ where $k_1, k_2, k_3, \cdots, k_n$ are constants. Prove that Y is normally distributed with mean $k_1\mu_1 + k_2\mu_2 + \cdots + k_n\mu_n$ and $k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + k_3^2\sigma_3^2 + \cdots + k_n^2\sigma_n^2$.

Proof

Given $X_1, X_2, X_3, \cdots, X_n$ are mutually stochastically independent and $X_i \sim n\left(\mu_i, \sigma_i^2\right)$, $i = 1,2,3,\cdots,n$.

$\therefore$ Moment generating function of X_i

$$= M_{X_i}(t)$$

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

$\therefore$ The moment generating function of Y

$$= M_Y(t)$$

$$= E\left(e^{tY}\right)$$

$$= E\left(e^{t(k_1 x_1 + k_2 x_2 + \cdots + k_n x_n)}\right)$$

$$= E\left(e^{tk_1x_1}\right)E\left(e^{tk_2x_2}\right)\cdots E\left(e^{tk_nx_n}\right)$$

(since $X_i's$ are stochastically independent.)

$$= e^{\mu_1 tk_1 + \frac{1}{2}\sigma_1^2 t^2 k_1^2} . e^{\mu_2 tk_2 + \frac{1}{2}\sigma_2^2 t^2 k_2^2} \cdots e^{\mu_1 tk_n + \frac{1}{2}\sigma_n^2 t^2 k_n^2}$$

$$= e^{t\left(k_1\mu_1 + k_2\mu_2 + \cdots + k_n\mu_n\right) + \frac{1}{2}t^2\left(k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2\right)}$$

Thus Y follows Normal distribution with mean $k_1\mu_1 + k_2\mu_2 + \cdots + k_n\mu_n$ and variance $k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + k_3^2\sigma_3^2 + \cdots + k_n^2\sigma_n^2$.

This proves the theorem.

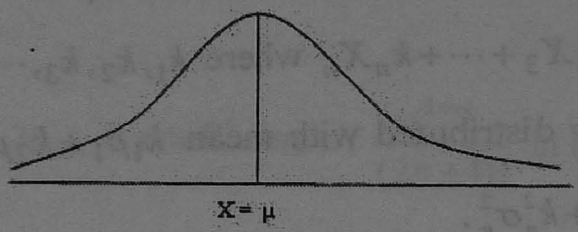
Space for Hint

Space for Hint

Properties of normal distribution

The following are important properties of the normal curve and the Normal distribution.

(1) The normal curve is symmetrical about the mean.



(2) The height of the normal curve is at its maximum at the mean.

(3) The mean, median and mode of the Normal distribution are coincide.

(4) Normal distribution has unimodal.

(5) The point of inflexion occur at $x = \mu \pm \sigma$, $y = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}$.

(6) The first and third quartile of the Normal distribution are equidistant from median.

(7) Mean deviation about mean is $\dfrac{4}{5}\sigma$.

(8) Linear combination of independent Normal distribution is again a Normal distribution.

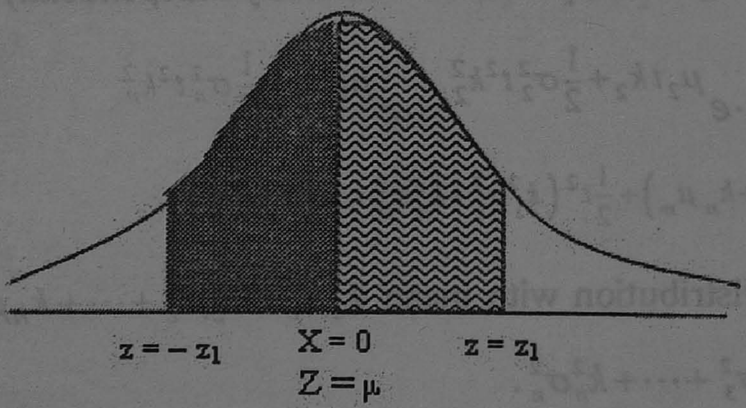
(9) All odd moments of the Normal distribution are zero.

(10) For the Normal distribution $\beta_1 = 0$ and $\beta_2 = 3$.

(11) $P(-\infty < X < \infty) = 1$.

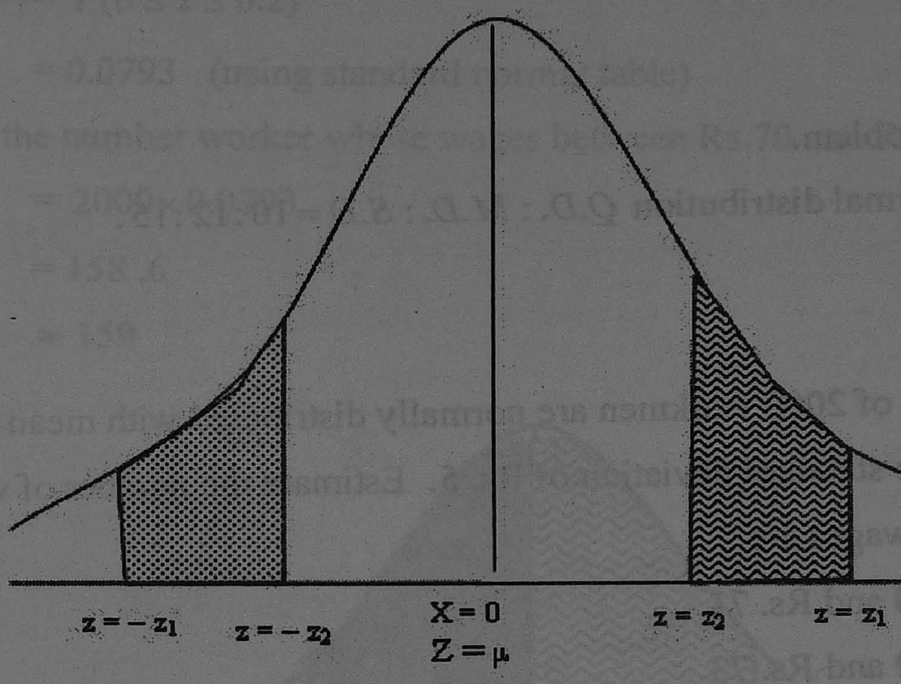
(12) $P(-\infty < z < 0) = P(0 < z < \infty) = 0.5$

(13) $P(-z_1 < z < 0) = P(0 < z < z_1)$



Space for Hint

(14) $P(-z_1 < z < -z_2) = P(z_2 < z < z_1)$



Example 8. 41 :

Prove that the quartile deviation ofr the Normal distribution is $\frac{2}{3}\sigma$.

Solution :

Let $X \sim N(\mu, \sigma^2)$

Let Q_1 and Q_3 be the lower and upper quartiles of the Normal distribution respectively.

We know that $Q.D = \dfrac{Q_3 - Q_1}{2}$

By the definition of lower quartile, $P(X \le Q_1) = 0.25$ and $P(X \le Q_3) = 0.75$

Let $z = \dfrac{x - \mu}{\sigma}$, $z_1 = \dfrac{Q_3 - \mu}{\sigma}$.

Then $-z_1 = \dfrac{Q_1 - \mu}{\sigma}$

Thus $\dfrac{Q_3 - Q_1}{\sigma} = 2z_1$

Hence $\dfrac{Q_3 - Q_1}{2} = \sigma z_1$

(i.e.) $Q.D. = \sigma z_1$

Space for Hint

Again from the normal table, $P(0 < z < z_1) = 0.25 \Rightarrow z_1 = 0.67$

Thus $Q.D. = 0.67\sigma$

(i.e.) $Q\,D. = \frac{2}{3}\sigma$

This proves the problem.

Note : For the Normal distribution $Q.D. : M.D. : S.D = 10:12:15$.

Example 8. 42 :

The weekly wages of 2000 workmen are normally distributed with mean wage of Rs. 70 and wage standard deviation of Rs. 5. Estimate the number of workers whose weekly wages are

(i) between Rs. 70 and Rs. 71

(ii) between Rs. 69 and Rs. 73

(iii) more than Rs. 72

(iv) less than Rs. 65 and

the lowest weekly wages of the 100 highest paid workers.

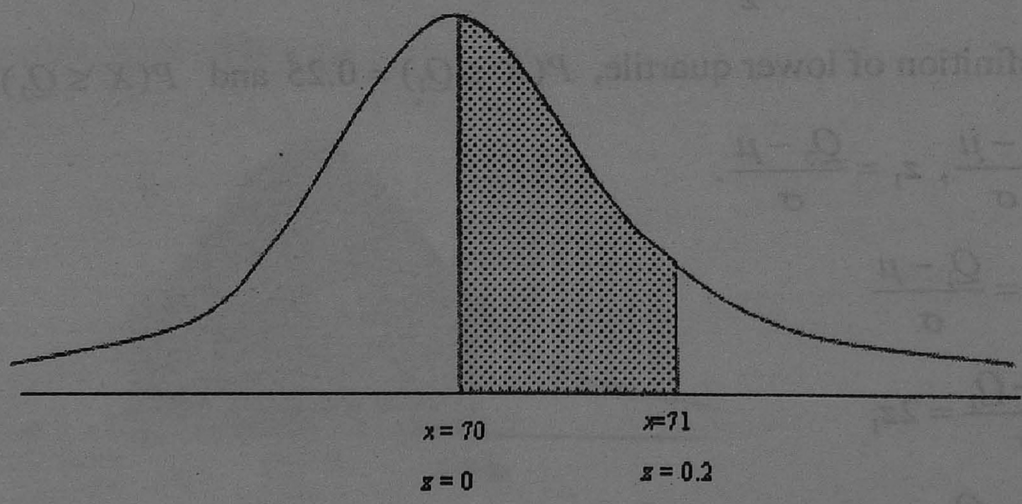
Solution : Given that mean = μ = Rs. 70

and standard deviation = σ = Rs. 5

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

(i.e) $z = \frac{x-70}{5}$

(i)



When $x = 71$ then $z = \frac{71-70}{5} = \frac{1}{5} = 0.2$

Space for Hint

$\therefore$ The probability that a worker has a weekly wage between Rs. 70 and Rs. 71 is $P(70 \le X \le 71)$

$= P(0 \le z \le 0.2)$

$= 0.0793$ (using standard normal table)

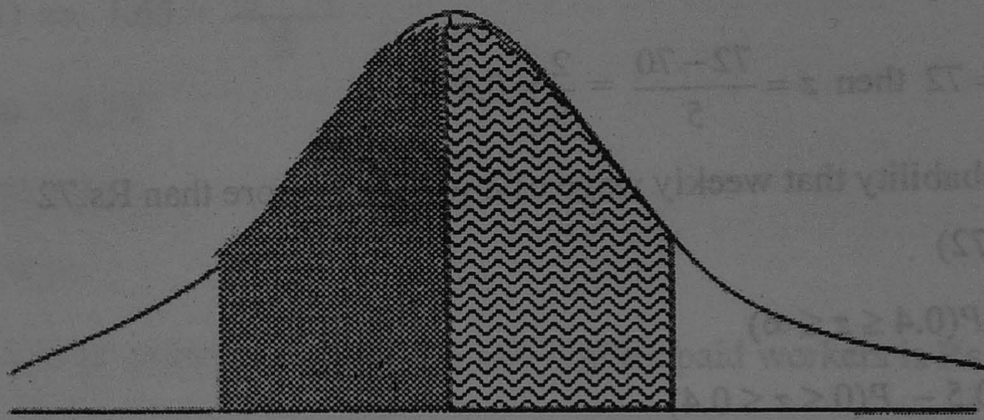
Thus the number worker whose wages between Rs.70 and Rs. 71

$= 2000 \times 0.0793$

$= 158.6$

≈ 159

(ii)



When $x = 69$ then $z = \dfrac{69-70}{5} = -\dfrac{1}{5} = -0.2$

When $x = 73$ then $z = \dfrac{73-70}{5} = \dfrac{3}{5} = 0.6$

$\therefore$ the probability that a worker has a weekly wage between Rs. 69 and Rs. 73 is $P(69 \le X \le 73)$

$= P(-0.2 \le z \le 0.6)$

$= P(-0.2 \le z \le 0) + P(0 \le z \le 0.6)$

$= P(0 \le z \le 0.2) + P(0 \le z \le 0.6)$

$= 0.0793 + 0.2257$ (using standard normal table)

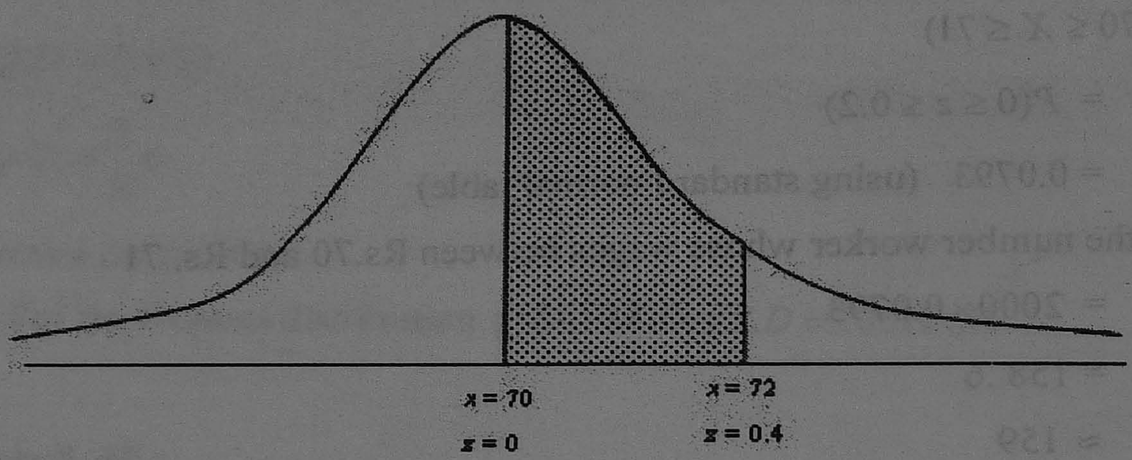
$= 0.3050$

Thus the number worker whose wages between Rs.69 and Rs. 73 = 2000×0.3050

$= 610$

Space for Hint

(iii)



When $x = 72$ then $z = \dfrac{72-70}{5} = \dfrac{2}{5} = 0.4$

$\therefore$ the probability that weekly wage of a worker is more than Rs.72

$= P(X \geq 72)$

$= P(0.4 \leq z \leq \infty)$

$= 0.5 - P(0 \leq z \leq 0.4)$

$= 0.5 - 0.1554$ (using standard normal table)

$= 0.2446$

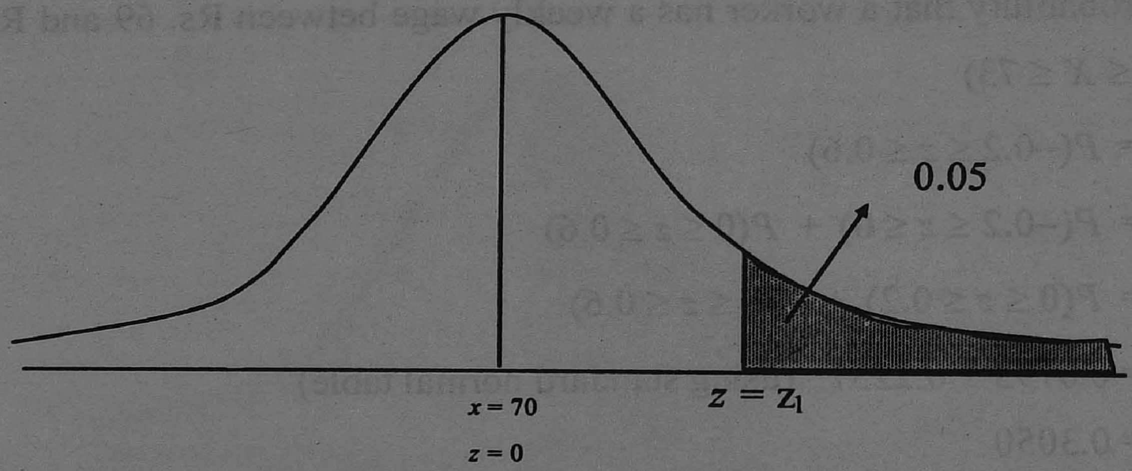
Thus the number worker whose wages between Rs.70 and Rs. 71

$= 2000 \times 0.2446$

$= 489.2$

$\approx 489.$

(v)



Let x_1 be the lowest weekly wages of the 100 highest workers.

Let $z_1 = \dfrac{x_1 - 70}{5}$ when $x = x_1$ -- (8.11)

Space for Hint

Given that $P(X \geq x_1) = \dfrac{100}{2000}$

(i.e) $P(z_1 \leq z < \infty) = 0.05$

(i.e) $0.5 - P(0 \leq z \leq z_1) = 0.05$

(i.e) $P(0 \leq z \leq z_1) = 0.5 - 0.05$

(i.e) $P(0 \leq z \leq z_1) = 0.45$

(i.e) $z_1 = 1.65$ (using standard normal table)

Thus (8.11) $\Rightarrow$ $1.65 = \dfrac{x_1 - 70}{5}$

(i.e) $x_1 - 70 = 8.25$

(i.e) $x_1 = 70 + 8.25$

(i.e) $x_1 = 78.25$

Hence the lowest weekly wages of the 100 highest paid workers is Rs.78.25

Example 8. 43 :

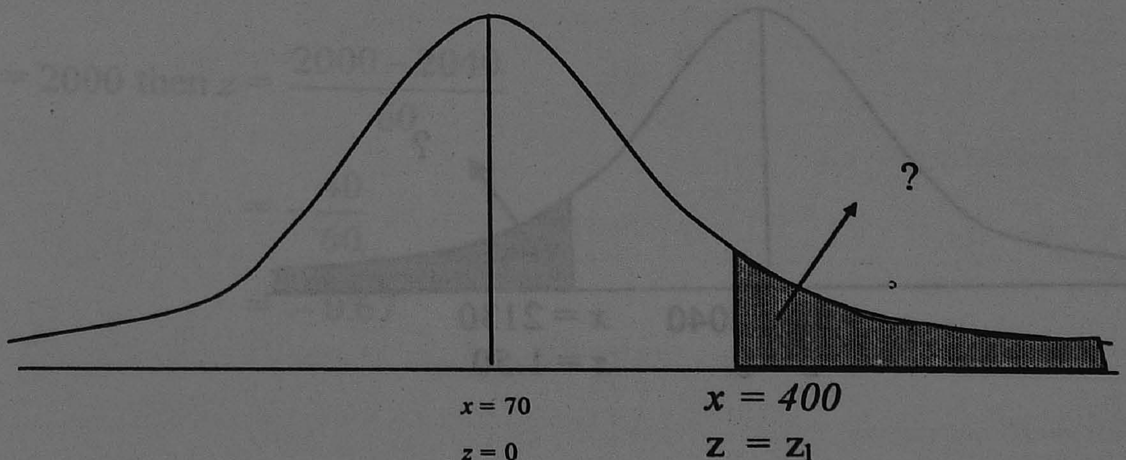
It is known that from the past experience that the number of telephone calls made daily in a certain city between 10 am and 11 am has a mean of 352 and a standard deviation of 31. What percentage of times will there be more than 400 telephone calls made in this locality between 10 am and 11 am?

Solution : Given that mean = $\mu = 352$

and standard deviation = $\sigma = 31$

Let $z = \dfrac{x - \mu}{\sigma}$

(i.e) $z = \dfrac{x - 352}{31}$



Space for Hint

When $x = 400$ then $z = \dfrac{400-352}{31}$

$= \dfrac{48}{31}$

$= 1.55$

$\therefore$ The probability that there will be more than 400 calls

$= P(X > 400)$

$= P(1.55 < z < \infty)$

$= 0.5 - P(0 < z < 1.55)$

$= 0.5 - 0.4394$ (using standard normal table)

$= 0.606$

Hence the percentage of days on which the number of calls will exceed 400 is 60.6%.

Example 8. 44 :

As a result of tests on 20000 electric fans manufactured by a company, it was found that lifetime of the fans was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. On the basis of the information, estimate the number of fans that is expected to run for (i) more than 2130 hours and (ii) less than 2000 hours ?

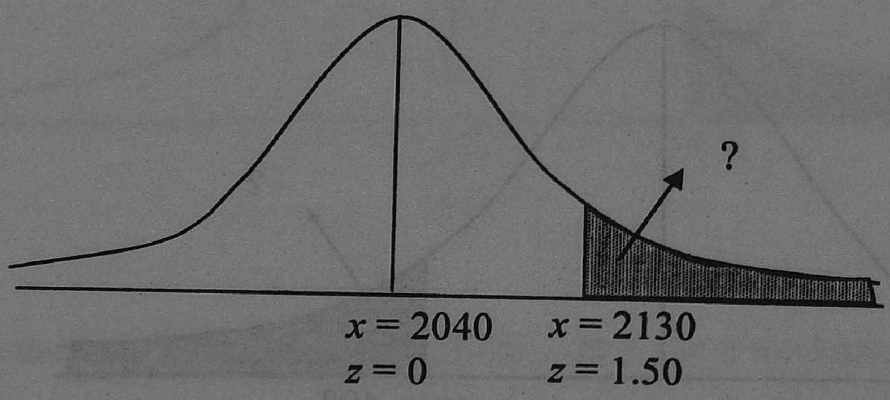
Solution : Given that mean = $\mu = 2040$

and standard deviation = $\sigma = 60$

Let $z = \dfrac{x-\mu}{\sigma}$

(i.e) $z = \dfrac{x-2040}{}$

(i)



When $x = 2130$ then $z = \dfrac{2130 - 2040}{60}$

$= \dfrac{90}{60}$

$= 1.50$

Now the probability that a fan is expected to run for more than 2130 hours = $P(X > 2130)$

$= P(1.50 < z < \infty)$

$= 0.5 - P(0 < z < 1.50)$

$= 0.5 - 0.4332$ (using standard normal table)

$= 0.0668$

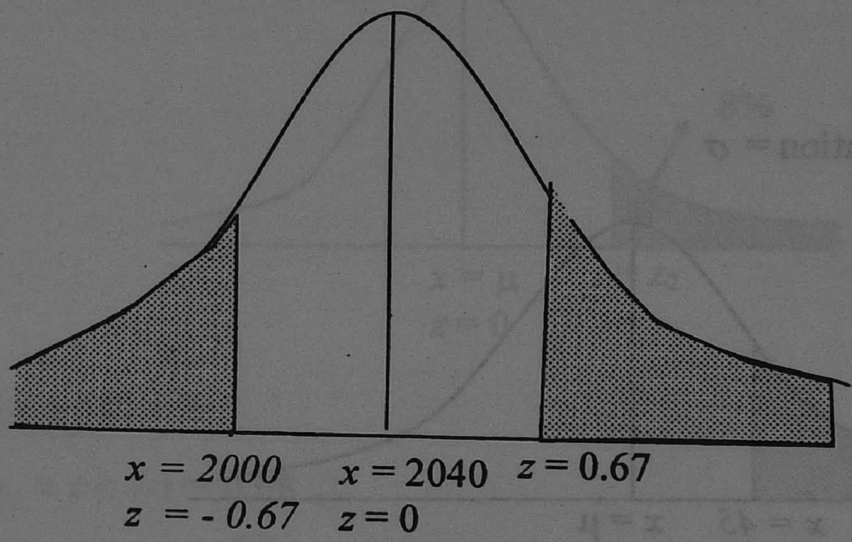
$\therefore$ The number of fans is expected to run for more than 2130 hours

$= 20000 \times P(X > 2130)$

$= 20000 \times 0.0668$

$= 1336$

(ii)



When $x = 2000$ then $z = \dfrac{2000 - 2040}{60}$

$= -\dfrac{40}{60}$

$= -0.67$

Space for Hint

Space for Hint

Now the probability that a fan is expected to run for less than 2000 hours

$= P(X < 2000)$

$= P(-\infty < z < -0.67)$

$= P(0.67 < z < \infty)$

$= 0.5 - P(0 < z < 0.67)$

$= 0.5 - 0.2486$

$= 0.2514$

$\therefore$ The number of fans is expected to run for less than 2000 hours

$= 20000 \times P(X < 2000)$

$= 20000 \times 0.2514$

$= 5028$

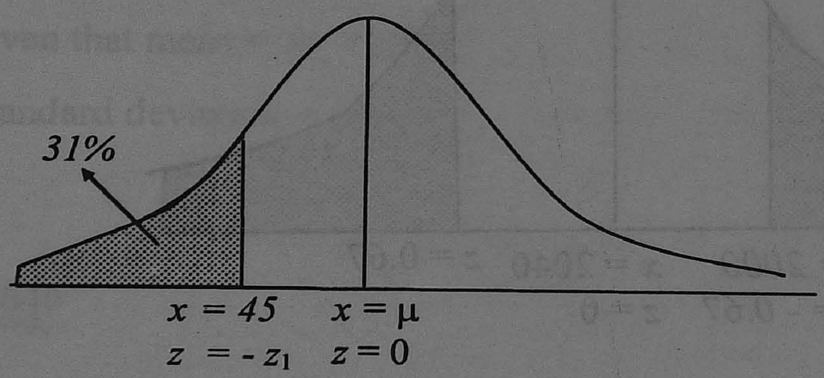
Example 8. 45 :

In a Normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Solution :

Let mean $= \mu$

and standard deviation $= \sigma$



Let $z = \dfrac{x-\mu}{\sigma}$

When $x = 45$, let $z = -z_1$

(i.e) $-z_1 = \dfrac{45-\mu}{\sigma}$ -- (8.12)

Given that 31% of the items are under 45 and therefore the ordinate of $x = 45$ lies left side of $x = \mu$.

(i.e) $P(X < 45) = 31\%$

Space for Hint

(i.e) $P(-\infty < X < -z_1) = 0.31$

(i.e) $P(z_1 < z < \infty) = 0.31$

(i.e) $0.5 - P(0 < z < z_1) = 0.31$

(i.e) $P(0 < z < z_1) = 0.5 - 0.31$

(i.e) $P(0 < z < z_1) = 0.19$

Thus $z_1 = 0.5$ (from the standard normal table)

Hence (8.12) $\Rightarrow -0.5 = \dfrac{45-\mu}{\sigma}$

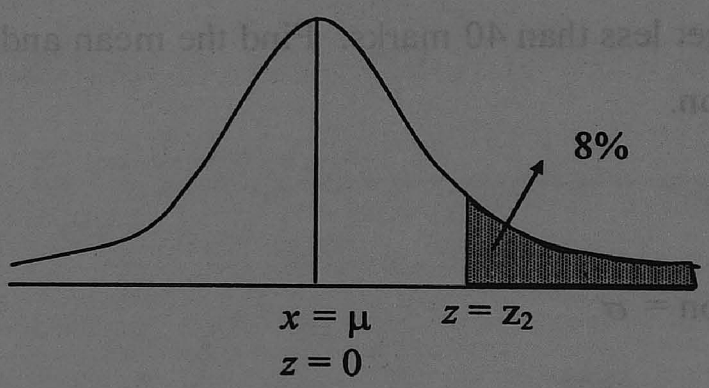
(i.e) $-0.5\sigma = 45 - \mu$

(i.e) $\mu = 45 + 0.5\sigma$ --(8.13)

Again 8% of the items are over 64 and therefore the ordinate of $x = 64$ lies right side of $x = \mu$.

Thus when $x = 64$, let $z = z_2$

(i.e) $z_2 = \dfrac{64-\mu}{\sigma}$ -- (8.14)



Now $P(X > 64) = 8\%$

(i.e) $P(z_2 < z < \infty) = 0.08$

(i.e) $0.5 - P(0 < z < z_2) = 0.08$

(i.e) $P(0 < z < z_2) = 0.5 - 0.08$

(i.e) $P(0 < z < z_2) = 0.42$

Thus $z_2 = 1.4$ (from the standard normal table)

Hence (8.14) $\Rightarrow 1.4 = \dfrac{64-\mu}{\sigma}$

(i.e) $1.4\sigma = 64 - \mu$ --(8.15)

Space for Hint

Thus (8.15)(4) $\Rightarrow$ $1.4\sigma = 64-(45+0.5\sigma)$ (using (8.13))

(i.e) $1.4\sigma = 64-45-0.5\sigma)$

(i.e) $1.4\sigma + 0.5\sigma = 19$

(i.e) $1.9\sigma = 19$

(i.e) $\sigma = \dfrac{19}{1.9}$

(i.e) $\sigma = 10$

Hence (8.13) $\Rightarrow$ $\mu = 45 + 0.5 \times 10$

(i.e) $\mu = 45 + 5$

(i.e) $\mu = 50$

Hence mean and standard deviation of the distribution are 50 and 10 respectively.

Example 8. 46 :

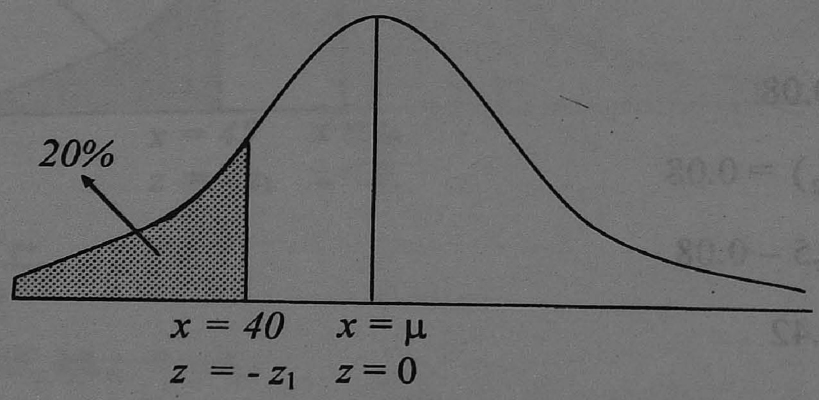
The marks of the students are normally distributed. 10% get more than 75 marks and 20% get less than 40 marks. Find the mean and standard deviation of the distribution.

Solution :

Let mean = μ

and standard deviation = σ

Let $z = \dfrac{x-\mu}{\sigma}$



When $x = 40$, let $z = -z_1$

(i.e) $-z_1 = \dfrac{40-\mu}{\sigma}$ -- (8.16)

Given that 20% of the items are under 40 and therefore the ordinate of $x = 45$ lies left side of $x = \mu$.

(i.e) $P(X < 40) = 20\%$

(i.e) $P(-\infty < X < -z_1) = 0.20$

(i.e) $P(z_1 < z < \infty) = 0.20$

(i.e) $0.5 - P(0 < z < z_1) = 0.20$

(i.e) $P(0 < z < z_1) = 0.5 - 0.20$

(i.e) $P(0 < z < z_1) = 0.30$

Thus $z_1 = -0.84$ (from the standard normal table)

Hence (8.16) $\Rightarrow$ $-0.84 = \dfrac{40-\mu}{\sigma}$

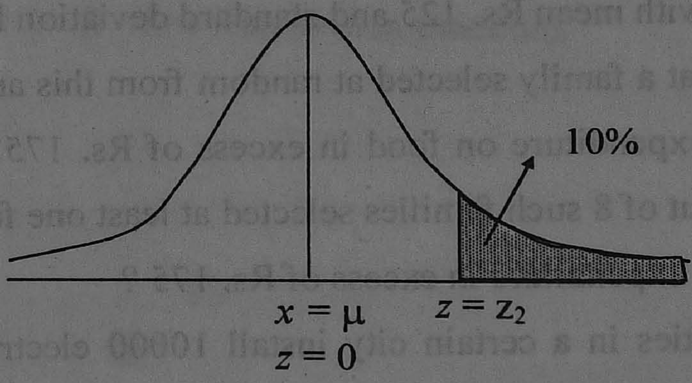
(i.e) $-0.84\sigma = 40 - \mu$

(i.e) $\mu = 40 + 0.84\sigma$ --(8.17)

Again 10% of the items are over 75 and therefore the ordinate of $x = 75$ lies right side of $x = \mu$.

Thus when $x = 75$, let $z = z_2$

(i.e) $z_2 = \dfrac{75-\mu}{\sigma}$ -- (8.18)



Now $P(X > 75) = 10\%$

(i.e) $P(z_2 < z < \infty) = 0.10$

(i.e) $0.5 - P(0 < z < z_2) = 0.10$

(i.e) $P(0 < z < z_2) = 0.5 - 0.10$

(i.e) $P(0 < z < z_2) = 0.40$

Thus $z_2 = 1.28$ (from the standard normal table)

Space for Hint

Space for Hint

Hence (8.18) $\Rightarrow$ $1.28 = \dfrac{75-\mu}{\sigma}$

(i.e) $1.28\sigma = 75-\mu$ -- (8.19)

Thus (8.19) $\Rightarrow$ $1.28\sigma = 75-(40+0.84\sigma)$ (using (8.17))

(i.e) $1.28\sigma + 0.84\sigma = 75-40$

(i.e) $2.12\sigma = 35$

(i.e) $\sigma = \dfrac{35}{2.12}$

(i.e) $\sigma = 16.5\dot{1}$

Hence (8.17) $\Rightarrow$ $\mu = 40 + 0.84 \times 16.51$

(i.e) $\mu = 53.87$

Hence mean and standard deviation of the distribution are 53.87 and 16.51 respectively.

Check Your Progress

(1) The incomes of a group of 5000 persons were found to be normally with mean Rs. 900 and standard deviation Rs. 75. What was the higher income among the poorest 200 ?

(2) The average weekly food expenditure of families in certain area as a Normal distribution with mean Rs. 125 and standard deviation Rs. 25. What is the probability that a family selected at random from this area will have an average weekly expenditure on food in excess of Rs. 175 ? What is the probability that out of 8 such families selected at least one family will have their weekly food expenditure in excess of Rs. 175 ?

(3) The local authorities in a certain city install 10000 electric bulbs in the streets of the city. If these bulbs have an average life of 1000 hours burning hours with a standard deviation of 200 hours, assuming normality, what number of bulbs might be expected to fail in the first 800 hours?

(4) Assume that examination marks from a university examination are normally distributed with a mean 450 and a standard deviation 100.

(i) What percentage of the students taking the examination between 400 and 500.

(ii) Suppose someone received a mark of 630. What percentage of the students taking examination marks better? What percentage marks worse?

(5) It is estimataed that if calculating machines are installed in a departmental store. It is likely to save, on an average 100 hours of lobour time per week of those employed at the cash counters. The probability that they will save less than 80 hours is 0.25. On the basis of a Normal distribution, what is the probability that they will save (i) more than 135 hours and (ii) less than 75 hours?

(answer : (i) 0.1210, (ii) 0.2033

(6) Suppose that life of a gas cylinder is normally distributed with a mean of 40 days and a standard deviation of 5 days. If, at a time, 10000 cylinders are issued to customers, how many will need replacement after 35 days?

(Answer : 3413)

Space for Hint

SUMMARY

In this unit, we have discussed moment generating function, cumulants, Binomial distribution, Poisson distribution and Normal distribution with examples and how to fit these distribution functions

Space for Hint

Unit IX

Large samples

Objectives

In this unit we shall discuss how to test for proportion, test for means, test for equality of means, test for standard deviation and test for correlation.

Introduction :

The need and reliable data is ever increasing for taking wise decision in different fields of human activity and business is no exception to it. There are two ways in which the required information may be obtained.

(1) Complete enumeration survey or census method, and

(2) sampling method.

Under complete enumeration method survey, data are collected for each and every unit belonging to the population or universe which is the complete set of items which are of interest in any particular situation.

9. 1 Sampling

Definition :

A finite subset of a population is called a *sample* and the number of objects in a sample is called the *sample size*.

Definition :

In order to determine some population characteristics, the objects in the sample are observed and the sample characteristics are used to approximately estimate the same for the entire population. The inherent and unavoidable error in any such approximation is called *sample error* .

Space for Hint

Sampling is common in day-to-day life. Some of the important types of sampling are (i) purposive sampling, (ii) random sampling, (iii) simple sampling and (iv) stratified sampling.

(i) Purposive sampling :

If the sample elements are selected with a definite purpose in mind then the sample selected is called *purposive sample*. Purposive sampling yields favoritism and nepotism in the selection of individuals in the sample. Hence this type of sampling does not fully represent the population.

(ii) Ransdom sampling :

Random sampling refers to the sampling technique in which each and every item of the population is given an equal chance of being included in the sample. The selection is thus free from personal bias because the investigator does not exercise his discretion of preference in the choice of items. Since selection of items in the sample depends entirely on chance, this method is also known as the method of chance selection. Some people believe that randomness of selection can be achieved by unsystematic and haphazard procedures. But this is quite wrong. However, the point to be emphasized is that unless precaution is taken to avoid bias and a conscious effort is made to ensure the operation of chance factors, the resulting sample shall not be a random sample.

(iii) Simple sampling :

Simple sampling is a special type of random sampling in which each element of the population has an equal and independent chance of being include in the sample.

(iv) Stratified sampling :

Stratified random sampling is one of the restricted random methods which, by using available information concerning the data attempts to design a more efficient than that obtained by the simple random procedure. The purpose of stratification requires that the population may be divided into homogenous groups or classes called *strata*. Then a sample may be taken from each group by simple random method, and the resulting sample is called stratified sample. A stratified sample may be may be either proportional or disproportional. In proportional stratified sampling plan, the number of items drawn from each stratum is proportional to the size of the strata.

Space for Hint

9. 2 Sampling distribution

The statistical constants of any population such as mean, variance, mode, median, etc., are referred to as *parameters* of the population and are usually represent by the Greek letters such as μ, σ^2 etc.,

Statistical measures computed from a sample of the population are called *statistic.* For, example sample mean is denoted by $\bar{x}$ and the sample variance is denoted by s^2 are the statistic of the population.

Definition :

The standard deviation of the sampling distribution of a statistic is known as ***standard error.***

Standard errors for some sampling distributions

Sampling statistic	Standard error
Mean $\bar{x}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Proportion P	$\sigma_P = \sqrt{\frac{PQ}{n}}$
Standard deviation s	$\sigma_s = \frac{\sigma}{\sqrt{2n}}$
Variance s^2	$\sigma_{s^2} = \sqrt{\frac{2}{n}}\sigma^2$
Correlation coefficient r	$\sigma_r = \frac{1-\rho^2}{\sqrt{n}}$

Note 1 : Standard error plays a vital role in the theory of large samples and forms the basis of the testing of hypothesis.

Space for Hint

Note 2 : If t is any statistic, then for large samples $z = \dfrac{t - E(t)}{\sqrt{\text{var}(t)}}$ tends to $N(0,1)$ asymptotically as $n \to \infty$ and hence $z = \dfrac{t - E(t)}{S.E} \sim N(0,1)$ asymptotically as $n \to \infty$.

9. 3 Testing of hypothesis

A hypothesis is an assumption to be tested. The statistical testing of hypothesis is the most important technique in statistical inference. Hypothesis tests are widely used in business and industry for making decisions. It is here that probability and sampling theory plays an ever increasing role in constructing the criteria on which business decisions are made. Very often in practice we are called upon to make decisions about population on the basis of sample information. For example, we may wish to decide on the basis of sample data whether a new medicine is really effective in curing a disease, whether one training procedure is better than another, etc. Such decisions are called statistical decisions.

Procedure of Hypothesis Testing

The general procedure followed in testing hypothesis comprises the following steps:

Set up a hypothesis. The first step in hypothesis testing is to establish the hypothesis to be tested. Since statistical hypotheses are usually assumptions about the value of some unknown parameter, the hypothesis specifies a numerical value or range of values for the parameter. The conventional approach to hypothesis testing is not to construct single hypothesis about the population parameter, but rather to set up two different hypotheses. These hypotheses are normally referred to as (i) null hypothesis denoted by H_0 and (ii) alternative hypothesis denoted by H_1.

The null hypothesis asserts that there is no true difference in the sample statistic and population parameter under consideration and the difference found accidental arising out of fluctuations of sampling.

Space for Hint

Set up a suitable significance level :

Having set up a hypothesis, the next step is to select a suitable level of significance. The confidence with which an experimenter rejects or retains null hypothesis depends on the significance level adopted. The level of significance, usually denoted by "α" is generally specified before any samples are drawn, so that results obtained will not influence our choice.

Determination of a suitable test statistic :

The third step is to determine a suitable test statistic and its distribution. Many of the test statistics that we shall encounter will be of the following form:

$$Test\ statistic = \frac{Sample\ statistic - Hypothesised\ population\ parameter}{S\tan dard\ error\ of\ the\ sample\ statistic}$$

Determine the critical region :

It is important to specify, before the sample is taken, which values of the test statistic will lead to a rejection of H_0 and which lead to acceptance of H_0. The former is called the critical region. The value of α, the level of significance, indicates the importance that one attaches to the consequences associated with incorrectly rejecting H_0. It can be shown that when the level of significance is α, the optimal critical region for a two-sided test consists of that $\frac{\alpha}{2}$% of the area in the right-hand tail of the distribution plus that $\frac{\alpha}{2}$% in the left-hand tail.

Doing computations :

The fifth step in testing hypothesis is the performance of various computations from a random sample of size n, necessary for the test statistic. Then we need to see whether sample result falls in the critical region or in the acceptance regions.

Making Decisions :

Finally, we may draw statistical conclusions and the management may take decisions. A statistical decision or conclusion comprises either accepting the null hypothesis or rejecting it. The decision will depend on whether the computed value of the test criterion falls in the region of rejection or the region of acceptance.

Space for Hint

Type I and Type II Errors

When a statistical hypothesis is tested, there are four possible results:

(1) The hypothesis is true but our test rejects it.

(2) The hypothesis is false but our test accepts it.

(3) The hypothesis is true and our test accepts it.

(4) The hypothesis is false and our test rejects it.

Obviously, the first two possibilities lead to errors. If we reject a hypothesis when it should be accepted (possibility No. 1) we say that a Type I error has been made. On the other hand, if we accept a hypothesis when it should be rejected (possibility No. 2) we say that a Type II error has been made. In either case a wrong decision or error in judgment has occurred.

TWO KINDS OF ERRORS IN HYPOTHESIS TESTING

Condition / Decision	H_0 : True	H_0 : False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

The probability of committing a *type I error* is denoted by α and is called the ***level of significance.***

Thus $\alpha = P(Type\,I\,error)$

$= P(\text{Rejecting } H_0 / H_0 \text{ is true})$

and $1-\alpha = P(Accepting\ H_0\ /\ H_0\ is\ true)$

the $(1-\alpha)$ corresponds to the concept of $100(1-\alpha)\%$ confidence interval.

Similarly the probability of committing a *type II error* is denoted by β

Space for Hint

Thus $\beta = P(Type\ II\ error)$

$= P(\text{Accepting } H_0 / H_0 \text{ is false})$

and $1-\beta = P(\text{Rejecting } H_0 / H_0 \textit{ is false})$,

and the probability $(1-\beta)$ is known as the *power* of a statistical test.

One – Tailed and Two – Tailed tests

Basically, there three kinds of problems of test of hypothesis.

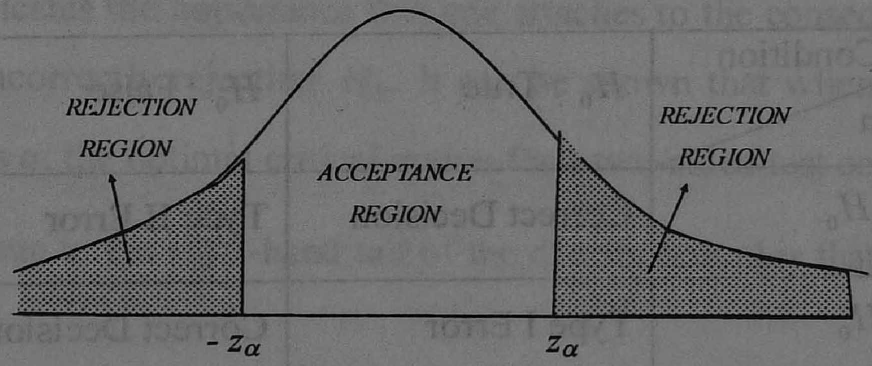
(1) Two tailed test,

(2) Left tailed test,

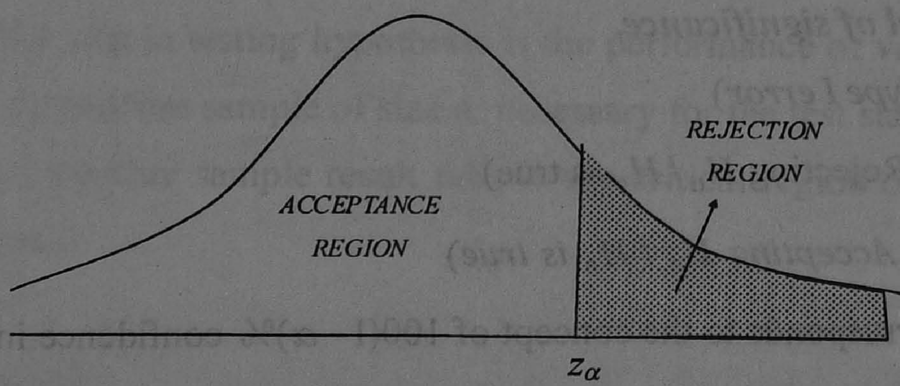
(3) Right tailed test.

Two tailed test is that where the hypothesis about the population mean is rejected for value of falling into either tail of the sampling distribution. When the hypothesis about population mean is rejected only for value of falling into one of the tails of the sampling distribution, then it is known as one tailed test. If it is right tail then it is called right tail test and the other is called left tail test.

The following picture gives an idea about two tailed test and one tail test.

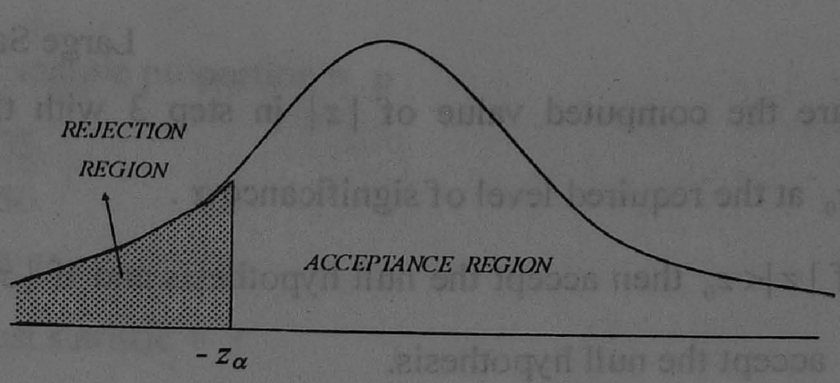


TWO TAILED TEST



RIGHT TAILED TEST

Space for Hint



LEFT TAILED TEST

The following table gives critical values of z for both one-tailed test and two-tailed tests at various levels of significance. Critical values of z for other level of significance are found by the use of the table of normal curve area.

level of significance	0.1	0.05	0.01
critical value of z for one-tailed test	-1.28 or 1.28	-1.28 or 1.28	-1.28 cr 1.28
critical value of z for two-tailed test	-1.645 and 1.645	-1.96 and 1.96	-1.58 and 1.58

The following is the procedure for testing of a statistical hypothesis.

Step 1 : Set the null hypothesis H_0

Step 2 : Set the alternative hypothesis H_1. (This will guide us to use right tailed or left tailed test or two tailed test)

Step 3 : Compute the test statistic $z = \dfrac{t - E(t)}{S.E.}$ under the null hypothesis.

Step 4 : Choose the appropriate level of significance α (If it is not given in the problem, it is taken as either 1% or 5%)

Space for Hint

Step 5 : Compare the computed value of $|z|$ in step 3 with the critical value z_0 at the required level of significance α .

Conclusion : If $|z| < z_0$ then accept the null hypothesis and If $|z| < z_0$ then accept the null hypothesis.

9.3.1 Test the significance for proportions

Test of significance for single proportion

If X is the number of success in independent trails with constant probability of success P for each trail we have $E(X) = nP$ and $V(X) = nPQ$.

Then the test statistic is $\boxed{z = \dfrac{X - nP}{\sqrt{nPQ}}}$.

The confidential limit for single proportion is given by $\boxed{P \pm 3\sqrt{\dfrac{PQ}{n}}}$

If P is not known then the probable limits for the proportion in the population are $\boxed{p \pm 3\sqrt{\dfrac{pq}{n}}}$

Example 9. 1 :

A sales clerk in the departmental store claims that 60% of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of them left without buying anything. Are these sample results consistent with the claim of the sales clerk? Use the level of significance of 0.05.

Solution :

Given that $P = 0.60$.

$\therefore Q = 1 - P$

$= 1 - 0.60$

$= 0.40$

Set $H_0 : 0.60$

$\therefore\ H_1 \neq 0.60$

Space for Hint

Given that sample proportion = p

$$= \frac{35}{50}$$

$$= 0.70$$

Now the test statistic = z

$$= \frac{X - nP}{\sqrt{nPQ}}$$

$$= \frac{35 - 50(0.6)}{\sqrt{50 \times (0.6) \times (0.4)}}$$

$$= \frac{5}{\sqrt{12}}$$

$$= 1.44$$

At 5% value of level of significance is 1.64

Now the computed $|z|$ value is less than the critical value and therefore accept the null hypothesis.

(i.e.) based on the sample data, we cannot reject the claim of the sales clerk.

Example 9. 2 :

A dice is thrown 49125 times and of these 25145 yielded either 1 or 3 or 5. Is this consistent with the hypothesis that the dice must be unbiased?

Solution :

Set H_0 : The dice is unbiased.

Thus H_1 : The dice is biased.

Given that $X = 25145$ and $n = 49152$.

$$\therefore\ P = \frac{25145}{49152}$$

(i.e.) $P = 0.512$

and $Q = 1 - P$

$$= 1 - 0.512$$

$$= 0.488$$

Given that sample proportion = p

$$= \frac{1}{2} = 0.50$$

Space for Hint

Now the test statistic $= z$

$$= \frac{X - nP}{\sqrt{nPQ}}$$

$$= \frac{25145 - 49152 \times 0.5)}{\sqrt{49152 \times (0.512) \times (0.488)}}$$

$$= \frac{25145 - 24576}{\sqrt{12280.92}}$$

$$= \frac{569}{110.82}$$

$$= 5.13 > 3$$

Therefore reject the null hypothesis.

(i.e.) based on the sample data, the dice is biased.

Example 9. 3 :

A manufacturer of bulb claims that on the average 2 percent or less of all bulbs manufactured by his firm are defective. A random sample of 400 bulbs contained 13 defective bulbs. On the evidence of this sample do you support the manufacture's claim? Assume that the maximum risk you wish to run of falsely rejecting the manufacturer's claim has been set at 5%.

Solution :

Given that $X = 13$ and $n = 400$.

$\therefore P = 2\%$

(i.e.) $P = 0.02$

and $Q = 1 - P$

$$= 1 - 0.02$$

$$= 0.98$$

Set $H_0 : P \leq 2\%$.

Thus $H_1 : P > 2\%$.

Given that sample proportion $= p$

$$= \frac{13}{400}$$

$$= 0.0325$$

Now the test statistic = z

$$= \frac{X - nP}{\sqrt{nPQ}}$$

$$= \frac{13 - 400 \times (0.02)}{\sqrt{400 \times 0.02 \times 0.98}}$$

$$= \frac{13 - 8}{\sqrt{7.84}}$$

$$= \frac{5}{2.8}$$

$$= 1.786$$

At 5% level of significance value for the right tail test is 1.645

Now the calculated z value is greater than the critical value at 5% level of significance and therefore reject the null hypothesis.

(i.e.) based on the sample data, the manufacturers' claim is not true.

Example 9. 4 :

A social worker believes that fewer than 25% of the couples in a certain are ever used any form of birth control. A random sample of 120 couples was contacted. Twenty of them said they had used some method of birth control. Comment on the social worker's belief.

Solution :

Given that $X = 20$ and $n = 120$.

$\therefore P = 25\%$

(i.e.) $P = 0.25$

and $Q = 1 - P$

$= 1 - 0.25$

$= 0.75$

Set $H_0 : P \leq 25\%$.

Thus $H_1 : P > 25\%$.

Given that sample proportion = p

$$= \frac{20}{120}$$

$$= 0.167$$

Space for Hint

Now the test statistic = z

$$= \frac{X - nP}{\sqrt{nPQ}}$$

$$= \frac{20 - 120 \times (0.25)}{\sqrt{120 \times 0.25 \times 0.75}}$$

$$= \frac{20 - 30}{\sqrt{22.5}}$$

$$= \frac{-10}{4.74}$$

$$= -2.108$$

At 5% level of significance value for the left tail test is -1.645.

Now the calculated z value is greater than the critical value at 5% level of significance and therefore accept the null hypothesis.

(i.e.) based on the sample data, the social worker's belief is correct.

Example 9. 5 :

500 eggs are taken from a large consignment and 50 are found to be bad. Estimate the percentage of bad eggs in the consignment and assign the limits within which the percentage probably lies.

Solution :

Given that $n = 500$.

$$\therefore\ p = \frac{50}{500}$$

and $q = 1 - p$

$$= 1 - \frac{50}{500}$$

$$= \frac{450}{500}$$

The probable limits are $p \pm 3\sqrt{\frac{PQ}{n}}$

$$= \frac{50}{500} \pm 3 \times \sqrt{\frac{\frac{50}{500} \times \frac{450}{500}}{500}}$$

Space for Hint

$= 0.1 \pm 3 \times \sqrt{0.00018}$

$= 0.1 \pm 0.040249$

$= (0.059751, 0.140249)$

Thus the required probable limits for percentage of bad eggs are 5.98% and 14.03%

Example 9. 6 :

A random sample of 160 people is taken and 120 were in favour of liberalizing licensing regulations. With 95% confidence, what proportion of all people are in favour?

Solution :

Given that $n = 160$.

$\therefore\ p = \dfrac{120}{160}$

(i.e.) $p = 0.75$

and $q = 1 - p$

$= 1 - 0.75$

$= 0.25$

Now $S.E. = \sqrt{\dfrac{pq}{n}}$

$= \sqrt{\dfrac{0.75 \times 0.25}{160}}$

$= 0.034$

The probable limits are $p \pm 1.96\sqrt{\dfrac{PQ}{n}}$

$= 0.75 \pm 1.96 \times 0.034$

$= (0.684, 0.816)$

Thus the proportion of people in favour of liberalizing regulations lies between 0.684 and 0.816

Space for Hint

Example 9. 7 :

In a locality of 16000 families, a simple sample of 860 families were selected. Of these 860 families, 215 families were found to have a monthly income of Rs. 1000 or less. Find the limits of the number of families out of 16000 having monthly income RS. 1000 or less.

Solution :

Given that $n = 860$, $N = 16000$.

$$\therefore p = \frac{215}{860}$$

(i.e.) $p = 0.25$

and $q = 1 - p$

$$= 1 - 0.25$$

$$= 0.75$$

Now $S.E. = \sqrt{\frac{pq}{n}}$

$$= \sqrt{\frac{0.75 \times 0.25}{860}}$$

$$= 0.015$$

The probable limits are $p \pm 3\sqrt{\frac{pq}{n}}$

$$= 0.25 \pm 3 \times 0.034$$

$$= (0.205, 0.295)$$

Thus the limits of the number of families out of 16000 having monthly income RS. 1000 or less is $(16000 \times 0.205, 16000 \times 0.295)$

$= (3280, 4720)$

Check Your Progress

(1) An advertising company claims that 40% of the people who saw an advertisement on their television set remembered the name of the product advertised 24 hours after they have seen the show. In a sample survey conducted 24 hours after the show 152 out of 400 persons remembered the name of the product advertised. Test if the claim of the company can be accepted at a level of 1%.

(2) A coin was thrown 400 times and head turned 160 times. Find the *standard error* of the observed proportion of heads. Show that the probability of getting a head in a throw of the coin lies almost certainly between 0.53 and 0.67.

(3) Out of 20000 customers' ledger accounts, a sample of 600 accounts was taken to test the accuracy of posting and balancing wherein 45 mistakes were found. Assign limits within which the number of defective cases can be expected at 95% level.

Space for Hint

9.3.2 Test the significance for difference of proportions

Test of significance for difference of proportion

Suppose we want to compare distinct populations with regard to possession of an attributes. Let a sample of size n_1 be chosen from the first population and another sample of size n_2 be chosen from the second population. Let x_1 be the number of items possessing the attribute A in the first sample and x_2 be the number of items possessing the attribute B in the second sample.

Let $p_1 = \frac{x_1}{n_1}$ and $p_2 = \frac{x_2}{n_2}$

Under the null hypothesis $H_0 : P_1 = P_2 = P$, the test statistic is

$$\boxed{z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$ where $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

Suppose the population proportions P_1 and P_2 are given to be different, then by setting the null hypothesis $H_0 : P_1 \neq P_2$ and then the test statistic is

$$\boxed{z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}}$$

Space for Hint

If the sample proportion are not given then the null hypothesis is $H_0 : p_1 \neq p_2$ and then the test statistic is $\boxed{z = \dfrac{p_1 - p2}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}}$

Example 9. 8 :

A company is considering two different television advertisements for promotion of a new product. Management believes that advertisement A is more effective than advertisement B. Two test markets areas with virtually identified consumer characteristics are selected : advertisement A is used in one area and advertisement B in the other area. In a random sample of 60 customers who saw advertisement A. 18 tried the product. In a random sample of 100 customers who saw advertisement B, 22 tried the product. Does this indicate that advertisement A is more effective than advertisement B, if a 5% level of significance is used?

Solution :

Set $H_0 : P_1 = P_2$

(i.e.) there is no significant difference in the effectiveness of the two advertisements A and B.

$\therefore\ H_1 : P_1 \neq P_2$

Given that $X_1 = 18,\ n_1 = 60,\ X_2 = 22,\ n_2 = 100$.

Thus $p_1 = \dfrac{x_1}{n_1}$

(i.e.) $p_1 = \dfrac{18}{60} = 0.30$

and $p_2 = \dfrac{x_2}{n_2}$

(i.e.) $p_2 = \dfrac{22}{100} = 0.22$

Now $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

(i.e.) $P = \dfrac{x_1 + x_2}{n_1 + n_2}$

Space for Hint

(i.e.) $P = \dfrac{18+22}{60+100}$

(i.e.) $P = \dfrac{40}{160}$

(i.e.) $P = 0.25$

$\therefore\ Q = 1 - P$

$= 1 - 0.25$

$= 0.75$

Now the test statistic is $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

(i.e.) $z = \dfrac{0.30 - 0.22}{\sqrt{(0.25)(0.75)\left(\dfrac{1}{60} + \dfrac{1}{100}\right)}}$

$= \dfrac{0.08}{\sqrt{0.005}}$

$= \dfrac{0.08}{0.071}$

$= 1.13$

At 5% level of significance the critical value is 1.96.

Here the calculated z value is less than the critical value and therefore null hypothesis H_0 will be accepted at 5% level of significance.

That is there is no significant difference in the effectiveness of the two advertisements A and B.

Example 9. 9 :

A sample survey results show that out of 800 literate people 480 are employed whereas out of 600 illiterate people only 350 are employed. Can the difference between two proportions of employed persons be ascribed due to sampling fluctuations?

Solution :

Set $H_0 : P_1 = P_2$

(i.e.) there is no significant difference between two proportions of employed persons be ascribed due to sampling fluctuations.

Space for Hint

$\therefore\ H_1 : P_1 \neq P_2$

Given that $X_1 = 480$, $n_1 = 800$, $X_2 = 350$, $n_2 = 600$.

Thus $p_1 = \dfrac{x_1}{n_1}$

(i.e.) $p_1 = \dfrac{480}{600} = 0.60$

and $p_2 = \dfrac{x_2}{n_2}$

(i.e.) $p_2 = \dfrac{350}{600} = 0.583$

Now $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

(i.e.) $P = \dfrac{x_1 + x_2}{n_1 + n_2}$

(i.e.) $P = \dfrac{480 + 350}{800 + 600}$

(i.e.) $P = \dfrac{830}{1400}$

(i.e.) $P = 0.59$

$\therefore\ Q = 1 - P$

$= 1 - 0.59$

$= 0.41$

Now the test statistic is $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

(i.e.) $z = \dfrac{0.60 - 0.583}{\sqrt{(0.59)(0.41)\left(\dfrac{1}{800} + \dfrac{1}{600}\right)}}$

$= \dfrac{0.017}{\sqrt{0.000704}}$

$= \dfrac{0.017}{0.0265}$

$= 0.642$

At 5% level of significance the critical value is 1.96.

Here the calculated z value is less than the critical value and therefore null hypothesis H_0 will be accepted at 5% level of significance.

That is there is no significant difference between two proportions of employed persons be ascribed due to sampling fluctuations.

Space for Hint

Example 9. 10 :

Random samples of 400 men and 600 women in a locality were asked whether they would like to have a bus stand near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposals are same in the male and female. Discuss at 5% level of significance.

Solution :

Set $H_0 : P_1 = P_2$

(i.e.) there is no significant difference between two proportions that men and women in favour of the proposals.

$\therefore\ H_1 : P_1 \neq P_2$

Given that $X_1 = 200$, $n_1 = 600$, $X_2 = 325$, $n_2 = 600$.

Thus $p_1 = \dfrac{x_1}{n_1}$

(i.e.) $p_1 = \dfrac{200}{400} = 0.50$

and $p_2 = \dfrac{x_2}{n_2}$

(i.e.) $p_2 = \dfrac{325}{600} = 0.542$

Now $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

(i.e.) $P = \dfrac{x_1 + x_2}{n_1 + n_2}$

(i.e.) $P = \dfrac{200+325}{400+600}$

(i.e.) $P = \dfrac{525}{1000}$

Space for Hint

(i.e.) $P = 0.525$

$\therefore \; Q = 1 - P$

$= 1 - 0.525$

$= 0.475$

Now the test statistic is $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

(i.e.) $z = \dfrac{0.50 - 0.542}{\sqrt{(0.525)(0.475)\left(\dfrac{1}{400} + \dfrac{1}{600}\right)}}$

$= \dfrac{-0.04167}{\sqrt{0.001039}}$

$= \dfrac{-0.04167}{0.032234}$

$= -1.29261$

At 5% level of significance the critical value is 1.96.

Here the calculated $|z|$ value is less than the critical value and therefore null hypothesis H_0 will be accepted at 5% level of significance.

That is there is no significant difference between two proportions that men and women in favour of the proposals.

Example 9. 11 :

In a random sample of 500 persons from Tamil Nadu 200 are found to be consumer of a olive oil. In another sample of 400 persons from Kerala 200 are found to be consumer of olive oil. Discuss whether the data reveal a significant difference between Tamil Nadu and Kerala so far as proportion of olive oil consumers if concerned.

Solution :

Set $H_0 : P_1 = P_2$

(i.e.) there is no significant difference between Tamil Nadu and Kerala so far as proportion of olive oil consumers.

$\therefore \; H_1 : P_1 \neq P_2$

Given that $X_1 = 200$, $n_1 = 500$, $X_2 = 200$, $n_2 = 400$.

Thus $p_1 = \dfrac{x_1}{n_1}$

(i.e.) $p_1 = \dfrac{200}{500} = 0.4$

Thus $q_1 = 1 - p_1$

$= 1 - 0.4$

$= 0.6$

and $p_2 = \dfrac{x_2}{n_2}$

(i.e.) $p_2 = \dfrac{200}{400} = 0.5$

Thus $q_2 = 1 - p_2$

$= 1 - 0.5$

$= 0.5$

Now the test statistic is $z = \dfrac{p_1 - p_2}{\sqrt{\left(\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}\right)}}$

(i.e.) $z = \dfrac{0.4 - 0.5}{\sqrt{\dfrac{(0.4)(0.6)}{500} + \dfrac{(0.5)(0.5)}{400}}}$

$= \dfrac{-0.1}{\sqrt{0.001025}}$

$= \dfrac{-0.1}{0.032016}$

$= -3.12348$

At 5% level of significance the critical value is 1.96.

Here the calculated $|z|$ value is less than the critical value and therefore null hypothesis H_0 will be rejected at 5% level of significance.

That is there is significant difference between Tamil Nadu and Kerala so far as proportion of olive oil consumers.

Space for Hint

Space for Hint

Check Your Progress

(1) A machine puts out 16 imperfect articles in a sample of 500 articles. After the machine is overhauled it puts out 3 defectives articles in a sample of 100. Has the machine improved?

(2) In a sample of 600 students of certain college, 400 are found to use dot pens. In another college from a sample of 900 students 450 were found to use dot pens. Test whether two colleges are significantly different with respect to the habit of using dot pens.

(3) Two samples of sizes 1200 and 900 respectively are drawn from two large populations. In the two large populations there are 30% and 25% respectively of fair haired people. Test whether these two samples will reveal the difference in the population proportions.

(4) 500 units from a factory are inspected and 12 are found to be defective, 800 units from another factory are inspected and 12 are found to be defective. Can it be concluded at 5% level of significance that production at second factory is better than in first factory?

9.3.3 Test the significance for single mean

To test the significance for single mean the test statistics is $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

where $\bar{x}$ is the sample mean and σ is the population standard deviation.

If the population standard deviation is not known then the test statistic becomes $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$.

(1) 95% confidence limits for μ is $\bar{x} \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$

(2) 98% confidence limits for μ is $\bar{x} \pm 2.33 \dfrac{\sigma}{\sqrt{n}}$

(3) 99% confidence limits for μ is $\bar{x} \pm 2.58 \dfrac{\sigma}{\sqrt{n}}$

Space for Hint

Example 9. 12 :

The mean I.Q. of a sample of 1600 children was 99. Is it likely that this was a random sample from a population with mean I.Q. 100 and standard deviation 15?

Solution :

Set $H_0 : \mu = 100$

(i.e.) The sample has been drawn from a population with mean $\mu = 100$ and standard deviation $\sigma = 15$

$\therefore\ H_1 : \mu \neq 100$

Given that $n = 1600$, $\overline{x} = 99$, $\mu = 100$ and $\sigma = 15$.

Thus the test statistic is $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{99-100}{15/\sqrt{1600}}$$

$$= \frac{-1}{15/40}$$

$$= -2.67$$

At 5% level of significance the critical value is 1.96

Here the calculated $|z|$ value is greater that the critical value and hence the null hypothesis is rejected.

(i.e.) The sample has not been drawn from a population with mean $\mu = 100$ and standard deviation $\sigma = 15$

Example 9. 13 :

The mean lifetime of a sample of 100 light tubes produced by a company is found to be 1,570 hours with standard deviation of 80 hours. Test the hypothesis that the mean life time of the tubes produced by the company is 1,600 hours.

Solution :

Set $H_0 : \mu = 1600$

(i.e.) The null hypothesis is that there is no difference between the sample mean and hypothetical population mean

Space for Hint

$\therefore \; H_1 : \mu \neq 1600$

Given that $n = 100$, $\bar{x} = 1570$, $s = 80$ and $\mu = 1600$.

Thus the test statistic is $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{1570 - 1600}{80/\sqrt{100}}$$

$$= \frac{-30}{8}$$

$$= -3.75$$

At 5% level of significance the critical value is 1.96

Here the calculated $|z|$ value is greater that the critical value and hence the null hypothesis is rejected.

(i.e.) the mean lifetime of the tubes produced by the company may not be 1,600 hours

Example 9. 14 :

A company markets car tyers. Their lives are normally distributed with a mean of 40000 kilometers and standard deviation of 3000 kilometers. A change in the production process is believed to result in a better product. A test sample of 64 new tyers has a mean life of 41200 kilometers. Can you conclude that the new product is significantly better than the current one?

Solution :

Set $H_0 : \mu \leq 40000$

(i.e.) The null hypothesis is that there is no difference between the sample mean and hypothetical population mean

$\therefore \; H_1 : \mu \geq 40000$

Given that $n = 64$, $\bar{x} = 41200$, $\mu = 1600$ and $\sigma = 3000$.

Thus the test statistic is $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{41200 - 40000}{3000/\sqrt{64}}$$

$$= \frac{1200 \times 8}{3000}$$

$= 3.2$

At 5% level of significance the critical value is 1.645

Here the calculated z value is greater that the critical value and hence the null hypothesis is rejected.

(i.e.) the data do not support the null hypothesis.

(i.e.) the company's claim that the new product is significantly better than the current one is valid.

Space for Hint

Example 9. 15 :

The mean weight of a random sample of size 100 from a student's population is 65.8 kilograms, and the standard deviation is 4 kilograms. Set up 95% confidence limits of the mean weight of the student's population.

Solution :

Given that $n = 100$ $\overline{x} = 65.8$ and $s = 4$.

Here the S.E. = $\frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}}$

(i.e.) S.E. = $\frac{s}{\sqrt{n}}$ (since population standard deviation is not given)

(i.e.) S.E. = $\frac{4}{\sqrt{100}}$

(i.e.) S.E. = 0.4 .

Hence the 95% confidence limits of the mean weight of the population are $\overline{x} \pm 1.96\ S.E(\overline{x})$

(i.e.) $65.8 \pm 1.96(0.4)$

(i.e.) (65.016, 66.584)

Thus the required 95% confidence limits for mean are (65.016, 66.584)

Check Your Progress

(1) The income distribution of the population of a certain village has a mean of Rs. 6000 and variance of Rs. 32000. Could a sample of 64 persons with a mean income of Rs. 5950 belong to this population?

Space for Hint

(2) Suppose that the distribution of heights of men follows a normal distribution with standard deviation 2.48. 100 male students in the Madurai Kamaraj University are measured and their average height is found to be 68.52. Determine the 98% confidence limits for the mean height of the male students of the university.

(3) Mr White wants to determine the average time to complete a certain job. The past records show that the population standard deviation is 10 days. Determine the sample size so that Mr.White may be 95% confident that the sample average remains ± 2 days of the average.

9.3.4 Test the significance for difference of sample means

Consider two different normal populations with mean μ_1, μ_2 and standard deviations σ_1, σ_2 respectively. Let a sample of size n_1 be drawn from the first population and an independent sample of size n_2 be drawn from the second population. Let $\bar{x}_1$, $\bar{x}_2$ be the means of the first sample and second sample respectively. Then the test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$.

Note 1 : If the samples have been drawn from 2 distinct population with common standard deviation σ then the test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$.

Note 2 : If the common standard deviation σ is not known then the test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_2} + \dfrac{s_2^2}{n_1}}}$.

Space for Hint

Note 3 : If the population standard deviations are not known then the test statistic is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$.

Example 9. 16 :

A random sample of 200 villages was taken from Madurai district and the average population per village was found to be 485 with a standard deviation of 50. Another random sample of 200 villages from the same district gave an average population of 510 per village with a standard deviation of 40. Is the difference between averages of the two samples statistically significant ?

Solution :

Set $H_0 : \mu_1 = \mu_2$

(i.e.) there no significant difference between the mean of populations.

$\therefore\ H_1 : \mu_1 \neq \mu_2$

Given that $n_1 = 200$, $\bar{x}_1 = 485$, $s_1 = 50$ and

$n_2 = 200$, $\bar{x}_2 = 510$, $s_2 = 40$.

Since the population standard deviations are not known and the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_2} + \dfrac{s_2^2}{n_1}}}$$

$$\text{(i.e.) } z = \frac{485 - 510}{\sqrt{\dfrac{(50)^2}{200} + \dfrac{(40)^2}{200}}}$$

$$= \frac{-25}{\sqrt{12.5 + 8}}$$

$$= -5.519$$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is greater than the critical value and therefore reject the null hypothesis.

(i.e.) there is significant difference between the mean of populations.

Space for Hint

Example 9. 17 :

Intelligence test given to two groups of boys and girls gave the following information :

	Boys	Girls
Mean score	75	70
Standard deviation	10	12
Number	50	100

Is the difference in the mean scores of boys and girls statistically independent?

Solution :

Set $H_0 : \mu_1 = \mu_2$

(i.e.) there no significant difference between the scores of boys and girls.

$\therefore\ H_1 : \mu_1 \neq \mu_2$

Given that $n_1 = 50$, $\bar{x}_1 = 75$, $s_1 = 10$ and

$n_2 = 100$, $\bar{x}_2 = 70$, $s_2 = 12$.

Since the population standard deviations are not known and the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(i.e.) $$z = \frac{75 - 70}{\sqrt{\frac{(10)^2}{50} + \frac{(12)^2}{100}}}$$

$$= \frac{5}{\sqrt{3.44}}$$

$$= 2.695$$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is greater than the critical value and therefore reject the null hypothesis.

(i.e) there is significant difference between the scores of boys and girls.

Example 9. 18 :

Two types of new cars produced in India are tested for petrol mileage. One group consisting of 36 cars averaged 14 kilometers per liter. While the other group consisting of 72 cars averaged 12.5 kilometers per liter. Further variances of two populations are 1.5 kilometers and 2 kilometers respectively. Test whether there exists a significant difference in the petrol consumption of these two types of cars.

Solution :

Set $H_0 : \mu_1 = \mu_2$

(i.e.) there no significant difference between the petrol consumption of these two types of cars.

$\therefore\ H_1 : \mu_1 \neq \mu_2$

Given that $n_1 = 36$, $\bar{x}_1 = 14$, $\sigma_1^2 = 1.5$ and

$$n_2 = 72,\ \bar{x}_2 = 12.5,\ \sigma_2^2 = 2.$$

Since the population standard deviations are not known and the test statistic is

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

(i.e.) $z = \dfrac{14 - 12.5}{\sqrt{\frac{1.5}{36} + \frac{2}{72}}}$

$$= \frac{1.5}{0.264}$$

$$= 5.68$$

At 1% level of significance the critical value is 2.58

Since the calculated $|z|$ value is greater than the critical value and therefore reject the null hypothesis.

(i.e) there significant difference between the petrol consumption of these two types of cars.

Example 9. 19 :

60 new entrants in a given college are found to have a mean weight of 68.6 kilograms and 50seniors have a mean weight of 69.51 kilograms. Is the

Space for Hint

Space for Hint

evidence conclusive that mean weight of the seniors is greater than that of the new entrants? Assume that standard deviation of weights to be 2.48 kilograms.

Solution :

Set $H_0 : \mu_1 = \mu_2$

(i.e.) there no significant difference between the mean weight of the seniors is greater than that of the new entrants.

$\therefore\ H_1 : \mu_1 > \mu_2$

Given that $n_1 = 60$, $\overline{x}_1 = 68.6$ and

$$n_2 = 50,\ \overline{x}_2 = 69.51,\ \sigma = 2.48.$$

Since the population standard deviations are not known and the test statistic is

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{(i.e.) } z = \frac{68.6 - 69.51}{2.48\sqrt{\frac{1}{60} + \frac{1}{50}}}$$

$$= \frac{-0.91}{2.48\sqrt{0.037}}$$

$$= \frac{-0.91}{2.48 \times 0.192}$$

$$= \frac{-0.91}{0.476}$$

$$= -1.911$$

At 5% level of significance the critical value is – 1.645

Since the calculated z value is smaller than the critical value and therefore reject the null hypothesis (under left tail test)

(i.e) there significant difference between the mean weight of the seniors is greater than that of the new entrants.

Example 9. 20 :

A college conducts both day and night classes intended to be identical. A sample of 100 day students yields examination result, mean 72.4 and standard deviation 14.8. A sample of 200 night students yields examination result mean 73.9 and standard deviation 17.9. Are the two means statistically equal at 5% level?

Solution :

Set $H_0 : \mu_1 = \mu_2$

(i.e.) the two sample means are statistically equal.

$\therefore \; H_1 : \mu_1 \neq \mu_2$

Given that $n_1 = 100$, $\bar{x}_1 = 72.4$, $s_1^2 = 14.8$ and

$n_2 = 200$, $\bar{x}_2 = 73.9$, $s_2^2 = 17.9$.

Since the population standard deviations are not known and the test statistic is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\text{(i.e.) } z = \frac{72.4 - 73.9}{\sqrt{\frac{(14.5)^2}{100} + \frac{(17.9)^2}{200}}}$$

$$= \frac{-1.5}{\sqrt{3.70455}}$$

$$= \frac{-1.5}{1.924721}$$

$$= -0.77933$$

At 5% level of significance the critical value is 1.96.

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis (under two tailed test)

(i.e) the two sample means are statistically equal.

Check Your Progress

(1) In a random sample of 500 the mean is found to be 20. In another independent sample of 400 the mean is 15. could the sample have been drawn from the same population with standard deviation 4?

(2) A random sample of 1000 men from Madurai gives their mean wage to be Rs.30 per day with a standard deviation of Rs.1.50. A sample of 1500 men from Chennai gives a mean wage of Rs.32 per day with standard deviation of Rs.2. Discuss whether the mean rate of wages varies between the two regions.

Space for Hint

Space for Hint

(3) Your are working as a purchase manager for a company. The following information has been supplied to you by two manufacturers of electric bulbs:

	Company A	Company B
Mean life (in hours)	1300	1288
Standard deviation (in hours)	82	93
Sample size	100	100

Which brand of bulbs are you going to purchase if you desire to take a risk of 5% ?

(4) The means of simple samples of 1000 and 2000 are 67.5 and 68.8 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches

9.3.5 Test the significance for single S.D.

If we want to test whether a sample with known standard deviation s could have come from a population with standard deviation σ we can use test statistic as $z = \dfrac{s-\sigma}{\sigma/\sqrt{2n}}$

Example 9. 21 :

From the experience of manufacturer of battery cells according to a technique the standard deviation of the life of battery cells was 150 days. He is interested to introduce a new technique in manufacturing batteries with fewer variations in life of batteries. In his new technique with a sample of 200 batteries he got the standard deviation of the life of battery cells as 140 days. Is the manufacturer justified in changing the technique?

Solution :

Set $H_0 : s = \sigma$

(i.e.) there no significant difference between the life of batteries produced in old and new techniques.

$\therefore\ H_1 : s \neq \sigma$

Given that $n = 200$, $s = 140$, $\sigma = 150$.

Now $z = \dfrac{s - \sigma}{\sigma/\sqrt{2n}}$

(i.e.) $z = \dfrac{140 - 150}{150/\sqrt{2 \times 200}}$

$= \dfrac{-10}{150/20}$

$= \dfrac{-10}{7.5}$

$= 1.333$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis.

(i.e) there no significant difference between the life of batteries produced in old and new techniques.

(i.e.) the manufacturer need not change the old technique.

9.3.6 Test the significance for equality of standard deviations of two normal populations .

Consider two normal populations with standard deviation σ_1 and σ_2 respectively. Let two independent random samples o large sizes n_1 and n_2 having standard deviations s_1 and s_2 be drawn from the first and second normal populations respectively. Then the test statistic is $z = \dfrac{s_1 - s_2}{\sqrt{\dfrac{\sigma_1^2}{2n_1} + \dfrac{\sigma_2^2}{2n_2}}}$

Space for Hint

Space for Hint

Note : If we want to test whether the two independent samples with known standard deviations s_1 and s_2 have come from the same population wit standard deviation σ. Then the test statistic is $z = \dfrac{s_1 - s_2}{\sigma\sqrt{\dfrac{1}{2n_1} + \dfrac{1}{2n_2}}}$

Example 9. 22 :

The standard deviation of a random sample of 1000 found to be 2.6 and the standard deviation of another random sample of 500 is 2.7. Assuming the samples to be independent discuss whether the two samples could have come from the universe with the same standard deviation.

Solution :

Set $H_0 : \sigma_1 = \sigma_2$

$\therefore\ H_1 : \sigma_1 \neq \sigma_2$

Given that $n_1 = 1000$, $s_1 = 2.6$ and

$n_2 = 500$, $s_2 = 2.7$.

Now $z = \dfrac{s_1 - s_2}{\sqrt{\dfrac{\sigma_1^2}{2n_1} + \dfrac{\sigma_2^2}{2n_2}}}$

(i.e.) $z = \dfrac{2.6 - 2.7}{\sqrt{\dfrac{(2.6)^2}{2\times1000} + \dfrac{(2.7)^2}{2\times500}}}$

$= \dfrac{-0.1}{\sqrt{0.01067}}$

$= \dfrac{-0.1}{0.103296}$

$= 0.968$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis.

(i.e the two samples could have come from the universe with the same standard deviation.

Space for Hint

Example 9. 23 :

The standard deviation of random sample of 900 members is 4.6 and that of another independent sample of 1600 is 4.8. Examine if the standard deviation are significantly different.

Solution :

Set $H_0 : \sigma_1 = \sigma_2$

(i.e.) there is no significant difference between the standard deviations of the populations.

$\therefore\ H_1 : \sigma_1 \neq \sigma_2$

Given that $n_1 = 900$, $s_1 = 4.6$ and

$n_2 = 1600$, $s_2 = 4.8$.

Now $z = \dfrac{s_1 - s_2}{\sqrt{\dfrac{\sigma_1^2}{2n_1} + \dfrac{\sigma_2^2}{2n_2}}}$

(i.e.) $z = \dfrac{4.6 - 4.8}{\sqrt{\dfrac{(4.6)^2}{2\times 900} + \dfrac{(4.8)^2}{2\times 1600}}}$

$= \dfrac{-0.2}{\sqrt{0.018956}}$

$= \dfrac{-0.2}{0.137679}$

$= -1.45265$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis.

(i.e.) there is no significant difference between the standard deviations of the populations.

Check Your Progress

(1) In a survey of incomes of two classes of workers random samples the following details. Examine whether the difference between (i) means and (ii) standard deviation are significant.

Space for Hint

Sample	Size	Mean annual income	Standard deviation in (Rs.)
I	100	582	24
II	100	546	28

9.3.7 Test the significance for correlation coefficient.

For a random sample of size n from the bivariate normal population to test the hypothesis $H_0 : r = \rho$, we take the test statistic as $z = \dfrac{z - z_0}{1/\sqrt{n-3}}$ where

$$z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) \text{ and } z_0 = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)$$

To test the null hypothesis $H_0 : \rho_1 = \rho_2$, then the test statistic is

$$z = \frac{z_1 - z_2}{\sqrt{\dfrac{1}{\sqrt{n_1-3}} + \dfrac{1}{\sqrt{n_2-3}}}}$$ where z_1 and z_2 are obtained from

$$z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_1}\right) \text{ and } z_2 = \frac{1}{2}\log_e\left(\frac{1+r_2}{1-r_2}\right)$$

Example 9. 24 :

The correlation coefficient between the temperature of rice and breakage percentage calculated from two samples 120 and 160 are 0.69 and 0.74 respectively. Do the two estimates differ significantly?

Solution :

Set $H_0 : \rho_1 = \rho_2$

$\therefore\ H_1 : \rho_1 \neq \rho_2$

Given that $n_1 = 120$, $\rho_1 = 0.69$ and

$n_2 = 160$, $\rho_2 = 0.74$.

Space for Hint

Now $z_1 = \frac{1}{2}\log_e\left(\frac{1+r_1}{1-r_1}\right)$

$$= \frac{1}{2}\log_e\left(\frac{1+0.69}{1-0.69}\right)$$

$$= \frac{1}{2}(1.6959)$$

$$= 0.8480$$

and $z_2 = \frac{1}{2}\log_e\left(\frac{1+r_2}{1-r_2}\right)$

$$= \frac{1}{2}\log_e\left(\frac{1+0.74}{1-0.74}\right)$$

$$= \frac{1}{2}(1.9010)$$

$$= 0.9505$$

Now test statistic is $z = \dfrac{z_1 - z_2}{\sqrt{\dfrac{1}{\sqrt{n_1 - 3}} + \dfrac{1}{\sqrt{n_2 - 3}}}}$

$$= \frac{0.8480 - 0.9505}{\sqrt{\dfrac{1}{\sqrt{120-3}} + \dfrac{1}{\sqrt{160-3}}}}$$

$$= \frac{-0.1025}{\sqrt{\dfrac{1}{10.8167} + \dfrac{1}{12.53}}}$$

$$= \frac{-0.1025}{\sqrt{0.0924 + 0.0798}}$$

$$= \frac{-0.1025}{0.4150}$$

$$= -0.2470$$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis.

(i.e.) the two estimates do differ significantly.

Space for Hint

Example 9. 25 :

In random sample of 50 pairs of values the correlation was found to be 0.89. Is this consistent with the assumption that the correlation in the population is 0.84?

Solution :

Set $H_0 : r = \rho$

$\therefore\ H_1 : r_1 \neq \rho$

Given that $n = 50$, $\sqrt{} = 0.89$ and $\rho = 0.84$.

Now $z = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)$

$= \frac{1}{2}\log_e\left(\frac{1+0.89}{1-0.89}\right) = 1.42$

and $z_0 = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)$

$= \frac{1}{2}\log_e\left(\frac{1+0.84}{1-0.84}\right) = 1.22$

Thus $z = \frac{z - z_0}{1/\sqrt{n-3}} = \frac{1.42 - 1.22}{1/\sqrt{50-3}} = \frac{0.20}{1/\sqrt{47}} = 1.37$

At 5% level of significance the critical value is 1.96

Since the calculated $|z|$ value is smaller than the critical value and therefore accept the null hypothesis.

SUMMARY

In this unit we learned how to test for proportion, test for means, test for equality of means, test for standard deviation and test for correlation.

Space for Hint

(2) A sample of 67 boys was taken and Karl Pearson's coefficient of correlation between two attributes x and y was found to be 0.72. Another sample of 39 girls was taken and coefficient of correlation between the same attributes was found to be 0.84. Can these two samples be considered as coming from populations having equal correlation coefficient?

(3) Two groups of children of different ages were given an intelligence test and an arithmetic test and the scores in the two test were found to be correlated. Given the following data, examine whether the correlation in the two groups are significantly different.

Group	Number	Correlation
A	63	0.63
B	80	0.55

SUMMARY

In this unit we learned how to test for proportion, test for means, test for equality of means, test for standard deviation and test for correlation.

Space for Hint

Unit X

Tests of significance of small samples – t, F, χ^2

Objectives

In this unit we shall discuss how to test for small samples using t, F, χ^2 test.

10. 1 Small Sampling Theory-t test

The Student's t-distribution obtained by W.S.Gosset was **published** under the pen name of "Student" in the year 1908. **It is reporter that Gosset** was a statistician for a brewery, **and that the management did not want him to** publish his scholarly theoretical work under his **real name and bring shame to** his employer. Consequently, he selected the pen name of **Student.**

Properties of *t*-Distribution

(1) The *t*-distribution ranges from $-\infty$ to ∞ just as does a normal distribution.

(2) The *t*-distribution like the standard normal distribution is bell-shaped and symmetrical around mean *zero*.

(3) The shapes of the *t*-distribution changes as the number of degrees of freedom changes. Therefore, for different degrees of freedom, the *t*-distribution has a family of *t*-distributions. Hence the degrees of freedom is a parameter of the *t*-distribution.

(4) The variance of the ***t*-distribution is always greater** than one and is defined only when **ν≥3 and is given as** $\mathbf{Var}(t) = \left(\dfrac{\nu}{\nu - 2}\right)$

(5) The *t*-distribution **is more of platykurtic** (less peaked at the centre and higher in tails) than the normal distribution.

(6) The t-distribution has a greater dispersion than the standard normal distribution. As n gets larger the t-distribution approaches the normal form. When n is as large as 30, the difference is very small.

The t-distribution has different shapes depending on the size of the sample. When the sample is quite small, for example, if n equal five, the height of the t-distribution is shorter than the normal distribution and the tails are wider. As n nears 30, however, the t-distribution approaches the normal distribution in shape.

The t-table. The t-table gives over a range of values of v at different levels of significance. By selecting a particular degrees of freedom and level of significance, we dete3rmine the tabular value of t. We establish a nul! hypothesis : and if our computed t is greater than the tabular t, we reject the null hypothesis; if our computed t is smaller than the tabular t we accept the null hypothesis.

Applications [illegible] distribution. The following are some important applications of the t-distribution:

(1) Test of Hypothesis about the population mean.

(2) Test of Hypothesis about the difference between two means.

(3) Test of hypothesis about the difference between two means with dependent samples.

(4) Test of hypothesis about coefficient of correlation.

10.1.1 Test for Hypothesis about the Population Mean

Test for Hypothesis about the Population Mean (σ unknown and sample size is small)

When the population distribution is normal and standard deviation σ is unknown then the t statistic is defined as $t = \dfrac{\overline{x}-\mu}{s/\sqrt{n}}$ follows the Student's t distribution with $n-1$ degrees of freedom, where

$\overline{x}$ = sample mean,

μ = hypothesized population mean,

Space for Hint

Space for Hint

n = sample size,

s = sample standard deviation which obtained from $s = \sqrt{\dfrac{\sum(x-\overline{x})^2}{n-1}}$

Note : If the population distribution is not normal then the t statistic is $t = \dfrac{\overline{x}-\mu}{s/\sqrt{n-1}}$

Confidence interval for the population mean

(1) 95% confidence of fiducial limits are $\left(\overline{x} - \dfrac{s}{\sqrt{n-1}} t_{0.05},\ \overline{x} + \dfrac{s}{\sqrt{n-1}} t_{0.05}\right)$ and

(2) 99% confidence of fiducial limits are $\left(\overline{x} - \dfrac{s}{\sqrt{n-1}} t_{0.01},\ \overline{x} + \dfrac{s}{\sqrt{n-1}} t_{0.01}\right)$

Example 10. 1 :

An automobile tyre manufacture claims that the average life of a particular grade of tyre is more than 20000 kilometers, when used under normal driving conditions. A random sample of 16 tyres was tested and a mean and standard deviation of 22000 and 5000 kilometers respectively, were computed. Assuming the lives of the tyres in kilometers, to be approximately normally distributed, decide whether the manufacturer's product is as good as claimed.

Solution :

Set $H_0 : \mu \leq 20000$

(i.e.) the manufaturer's claim is valid.

$\therefore\ H_1 : \mu > 20000$.

Given that $n = 16$, $\overline{x} = 22000$, $s = 5000$ and $\mu = 20000$

Now $t = \dfrac{\overline{x}-\mu}{s/\sqrt{n-1}}$

$= \dfrac{22000-20000}{5000/\sqrt{16-1}}$

$= \dfrac{2000}{5000/3.873}$

$= 1.55$

Here the degrees of freedom is $n-1 = 15$.

At 5% level of significance for 15 $d.f$ the table value for t is 1.753.

Since the calculate t value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

Thus the manufacturer's claim is valid.

Example 10. 2 :

The mean life time of electric bulbs produced by a company has in the past been 1120 hours with a standard deviation of 125 hours. A sample of 8 electric bulbs recently chosen from a supply of newly produced bulbs showed a mean life time of 1070 hours. Test the hypothesis that the mean life time of the bulbs has not changed.

Solution :

Set $H_0 : \mu = 1120$

(i.e.) the mean life time of the bulbs has not changed.

$\therefore\ H_1 : \mu \neq 120$.

Given that $n = 8$, $\overline{x} = 1070$, $s = 125$ and $\mu = 1120$

$$\text{Now } t = \frac{\overline{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{1070 - 1120}{125/\sqrt{8-1}}$$

$$= \frac{-50}{125/2.646}$$

$$= \frac{-132.3}{125}$$

$$= 1.0584$$

Here the degrees of freedom is $n - 1 = 7$

At 5% level of significance for 7 $d.f$ the table value for t is 1.894.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

Thus the mean life time of the bulbs has not changed.

Example 10. 3 :

Ten specimen of copper wires drawn from a large lot have the following ...king strength in Kg weight 578, 572, 570, 568, 572, 578, 570, 572, 596, 584. Test whether the mean breaking strength of the lot may be taken to be 578 Kg weight.

Space for Hint

Space for Hint

Solution :

Set $H_0 : \mu = 578$

(i.e.) the mean breaking strength of the lot may be taken to be 578 Kg weight.

$\therefore\ H_1 : \mu \neq 578$.

Given that $n = 8$ and $\mu = 578$.

Now we shall find $\bar{x}$ and s

x	$x-\bar{x}$	$(x-\bar{x})^2$
578	2	4
572	-4	16
570	-6	36
568	-8	64
572	-4	16
578	2	4
570	-6	36
572	-4	16
596	20	400
584	8	64
5760		656

Thus $\bar{x} = \dfrac{\Sigma x}{n}$

$= \dfrac{5760}{10}$

$= 576$

and $s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\dfrac{656}{10}}$

$= \sqrt{65.6}$

$= 8.099$

Space for Hint

Now $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n-1}}$

$= \dfrac{576 - 578}{8.099/\sqrt{10-1}}$

$= \dfrac{-2}{8.099/3}$

$= \dfrac{-6}{8.099}$

$= -0.741$

Here the degrees of freedom is $n-1=9$.

At 5% level of significance for 9 $d.f$ the table value for t is 1.833.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

Thus the mean breaking strength of the lot may be taken to be 578 Kg weight.

Example 10. 4 :

Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh in kilograms as follows. 50, 49, 52, 44, 45, 48, 46, 45, 49, 45. Test whether the average packing can be taken to be 50 kilograms.

Solution :

Set $H_0 : \mu = 50$

(i.e.) the average packing can be taken to be 50 kilograms.

$\therefore\ H_1 : \mu \neq 50$.

Given that $n = 10$ and $\mu = 50$.

Now we shall find $\overline{x}$ and s

Thus $\overline{x} = \dfrac{\sum x}{n}$

$= \dfrac{473}{10}$

$= 47.3$

Space for Hint

x	$x-\overline{x}$	$(x-\overline{x})^2$
50	2.7	7.29
49	1.7	2.89
52	4.7	22.09
44	-3.3	10.89
45	-2.3	5.29
48	0.7	0.49
46	-1.3	1.69
45	-2.3	5.29
49	1.7	2.89
45	-2.3	5.29
473		64.1

and $s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n}}$

$= \sqrt{\frac{64.1}{10}}$

$= \sqrt{6.41}$

$= 2.532$

Now $t = \frac{\overline{x}-\mu}{s/\sqrt{n-1}}$

$= \frac{47.3-50}{2.532/\sqrt{10-1}}$

$= \frac{-2.7}{2.532/3}$

$= \frac{-8.1}{2.532}$

$= -3.199$

Here the degrees of freedom is $n-1=9$.

At 5% level of significance for 9 $d.f$ the table value for t is 1.833.

Since the calculate $|t|$ value is greater than the table value of t and therefore at 5% level of significance we reject the null hypothesis.

Thus the average packing cannot be taken to be 50 kilograms.

Example 10. 5 :

A random sample of 16 values from a normal population showed a mean of 41.5 cm and the sum of squares of deviations from their mean equal to 135 sq.cm. Show that the assumption of a mean of 43.5 cm for the population is not reasonable at 5% level of significance, but reasonable at 1% level of significance. Also obtain 95% and 99% fiducial limits for the same.

Solution :

Set $H_0 : \mu = 43.5$

$\therefore\ H_1 : \mu \neq 50$.

Given that $n=16$, $\bar{x}=41.5$, $\mu=43.5$ and $\Sigma(x-\bar{x})^2=135$

$$\text{Thus } s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$=\sqrt{\frac{135}{16}}$$

$$=\sqrt{8.4375}$$

$$= 2.91$$

$$\text{Now } t=\frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$

$$=\frac{41.5-43.5}{2.91/\sqrt{16-1}}$$

$$=\frac{-2}{2.91/3.873}$$

$$=\frac{-2\times 3.873}{2.91}$$

$$= -2.6619$$

Here the degrees of freedom is $n-1=15$.

At 5% level of significance for 15 $d.f$ the table value for t is 2.131.

Space for Hint

Since the calculate $|t|$ value is greater than the table value of t and therefore at 5% level of significance we reject the null hypothesis.

Thus the assumption of a mean of 43.5 cm for the population is not reasonable at 5% level of significance.

At 1% level of significance for 15 $d.f$ the table value for t is 2.947.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

Thus the assumption of a mean of 43.5 cm for the population is reasonable at 1% level of significance.

To find 95% fiducial limits :

We know that the 95% fiducial limits are $\bar{x} \pm \frac{s}{\sqrt{n-1}} t_{0.05}$

$$= 41.5 \pm \frac{2.91}{\sqrt{16-1}} \times 2.131$$

$$= 41.5 \pm 1.6011$$

$$= (39.8989, 43.1011)$$

To find 99% fiducial limits :

We know that the 99% fiducial limits are $\bar{x} \pm \frac{s}{\sqrt{n-1}} t_{0.01}$

$$= 41.5 \pm \frac{2.91}{\sqrt{16-1}} \times 2.947$$

$$= 41.5 \pm 2.2143$$

$$= (39.2857, 43.7143)$$

10.1.2 Test for the difference between Means of two samples

If $\bar{x}_1$ and $\bar{x}_2$ are the means of two independent samples of sizes n_1 and n_2 from a normal population with mean μ and standard deviation σ then the for testing the hypothesis $H_0 : \mu_1 = \mu_2$, test statistic is given by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ where s_1 and s_2 are standard deviations of the samples. Here the $d.f.$ is $n_1 + n_2 - 2$

Note : If the samples are of equal sizes then $n_1 = n_2 = n$ and the test statistic becomes $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2 + s_2^2}{n-1}}}$ and the $d.f.$ is $2n-2$

Space for Hint

Example 10. 6 :

The Intelligence Quotients (IQ) of16 students from Physics Department showed a mean 107 with a standard deviation 10 while the IQ of 14 students of Mathematics Department showed a mean of 112 with a standard deviation 8. Is there a significant difference between the IQs of students Physics and Mathematics Department?

Solution :

Set $H_0 : \mu_1 = \mu_2$.

(i.e.) there a no significant difference between the IQs of students Physics and Mathematics Department.

$\therefore\ H_1 : \mu_1 \neq \mu_2$.

Given that $n_1 = 16$, $\bar{x}_1 = 107$, $s_1 = 10$ and

$n_2 = 14$, $\bar{x}_2 = 112$, $s_2 = 8$

Now $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

$$= \frac{107 - 112}{\sqrt{\frac{16(10)^2 + 14(8)^2}{16 + 14 - 2}\left(\frac{1}{16} + \frac{1}{14}\right)}}$$

$$= \frac{-5}{\sqrt{89.1429 \times 0.1339}}$$

$$= \frac{-5}{\sqrt{11.9688}}$$

Space for Hint

$$= \frac{-5}{3.4549}$$

$$= -1.4472$$

Here the degrees of freedom is $n_1 + n_2 - 2 = 16 + 14 - 2 = 28$.

At 5% level of significance for 28 $d.f$ the table value for t is 2.048.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

Thus there a no significant difference between the IQs of students Physics and Mathematics Department at 5% level of significance.

Example 10. 7 :

In an examination in Mathematics 12 students in one class had mean mark of 78 with a standard deviation of 6 while 15 students in another class had a mean mark 74 with a standard deviation of 8. Determine whether the first group is superior to the second group using a significance level of 5% ?

Solution :

Set $H_0 : \mu_1 = \mu_2$.

(i.e.) there a no significant difference between the mean marks of two groups of students.

$\therefore\ H_1 : \mu_1 > \mu_2$.

Given that $n_1 = 12$, $\bar{x}_1 = 78$, $s_1 = 6$ and

$n_2 = 15$, $\bar{x}_2 = 74$, $s_2 = 8$

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

$$= \frac{78 - 74}{\sqrt{\frac{12(6)^2 + 15(8)^2}{12 + 15 - 2}\left(\frac{1}{12} + \frac{1}{15}\right)}}$$

$$= \frac{4}{\sqrt{55.68 \times 0.15}}$$

$$= \frac{4}{\sqrt{8.352}}$$

Space for Hint

$$= \frac{4}{2.89}$$

$$= 1.3841$$

Here the degrees of freedom is $n_1 + n_2 - 2 = 12 + 15 - 2 = 25$.

At 5% level of significance for 25 $d.f$ the table value for t is 2.060.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

That is there a no significant difference between the mean marks of two groups of students.

(i.e.) first group is not superior to the second group of students.

Example 10. 8 :

Test whether the two sets of observations given below are drawn from the same population.

First set	17	27	18	25	27	29	27	23	17
Second set	16	16	20	16	20	17	15	21	–

Solution :

Step 1 : First we find $\overline{x}_1$ and s_1.

	x	$x-\overline{x}$	$(x-\overline{x})^2$
	17	-6.33	40.11
	27	3.67	13.44
	18	-5.33	28.44
	25	1.67	2.78
	27	3.67	13.44
	29	5.67	32.11
	27	3.67	13.44
	23	-0.33	0.11
	17	-6.33	40.11
Total	210		184

Space for Hint

Now $\bar{x}_1 = \frac{\sum x}{n}$

$= \frac{210}{9}$

$= 23.33$

and $s_1 = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{184}{9}}$

$= 4.52$

Step 2 : Now we shall find $\bar{x}_2$ and s_2.

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	16	-1.63	2.64
	16	-1.63	2.64
	20	2.38	5.64
	16	-1.63	2.64
	20	2.38	5.64
	17	-0.63	0.39
	15	-2.63	6.89
	21	3.38	11.39
Total	141		37.87

Now $\bar{x}_2 = \frac{\sum x}{n}$

$= \frac{141}{8}$

$= 17.63$

and $s_2 = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{37.88}{8}}$

$= 2.18$

Step 3 :

Set $H_0 : \mu_1 = \mu_2$.

$\therefore\ H_1 : \mu_1 \neq \mu_2$.

Given that $n_1 = 9$, $\overline{x}_1 = 78$, $s_1 = 6$ and

$n_2 = 15$, $\overline{x}_2 = 74$, $s_2 = 8$

$$\text{Now } t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{23.33 - 17.63}{\sqrt{\frac{9(4.52)^2 + 8(2.18)^2}{9+8-2}\left(\frac{1}{9} + \frac{1}{8}\right)}}$$

$$= \frac{5.7}{\sqrt{14.7929 \times 0.2361}}$$

$$= \frac{5.7}{\sqrt{3.4926}}$$

$$= \frac{5.7}{1.8688}$$

$$= 3.05$$

Here the degrees of freedom is $n_1 + n_2 - 2\ = 9 + 8 - 2 = 15$.

At 5% level of significance for 15 $d.f$ the table value for t is 2.131.

Since the calculate $|t|$ value is greater than the table value of t and therefore at 5% level of significance we reject the null hypothesis.

The two sets of observations are drawn from the different populations.

Example 10. 9 :

Two kinds of manure applied to 16 one-acre plots, other conditions remaining the same. The yields (in quintals) are given below.

Manure I	18	20	36	50	49	36	34	49	41
Manure II	29	28	26	35	30	44	46	–	–

Examine the significance of the difference between the mean yields due to the application of different kinds of manure.

Space for Hint

Space for Hint

Solution :

Step 1 : First we find $\overline{x}_1$ and s_1.

	x	$x-\overline{x}$	$(x-\overline{x})^2$
	18	-19	361
	20	-17	289
	36	-1	1
	50	13	169
	49	12	144
	36	-1	1
	34	-3	9
	49	12	144
	41	4	16
Total	333		1134

Now $\overline{x}_1 = \dfrac{\sum x}{n}$

$= \dfrac{333}{9}$

$= 37$

and $s_1 = \sqrt{\dfrac{\sum(x-\overline{x})^2}{n}}$

$= \sqrt{\dfrac{1134}{9}}$

$= 11.23$

Step 2 : Now we shall find $\overline{x}_2$ and s_2.

Now $\overline{x}_2 = \dfrac{\sum x}{n}$

$= \dfrac{238}{7}$

$= 34$

and $s_2 = \sqrt{\dfrac{\sum(x-\overline{x})^2}{n}}$

$$= \sqrt{\frac{386}{7}}$$

$$= 7.43$$

	x	$x-\overline{x}$	$(x-\overline{x})^2$
	29	-5	25
	28	-6	[illegible]
	26	-8	64
	35	1	1
	30	-4	16
	44	10	100
	46	12	144
Total	238		386

Now $\overline{x}_2 = \frac{\sum x}{n}$

$$= \frac{141}{8}$$

$$= 17.63$$

and $s_2 = \sqrt{\frac{\sum(x-\overline{x})^2}{n}}$

$$= \sqrt{\frac{37.88}{8}}$$

$$= 2.18$$

Step 3 :

Set $H_0 : \mu_1 = \mu_2$.

(i.e.) there is no significant difference between the mean yields due to the application of two kinds of manures.

$\therefore\ H_1 : \mu_1 \neq \mu_2$.

Given that $n_1 = 9$, $\overline{x}_1 = 37$, $s_1 = 11.23$ and

$n_2 = 7$, $\overline{x}_2 = 34$, $s_2 = 7.43$

Space for Hint

Space for Hint

Now $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

$$= \frac{37-34}{\sqrt{\dfrac{9(11.23)^2 + 7(7.43)^2}{9+7-2}\left(\dfrac{1}{9} + \dfrac{1}{7}\right)}}$$

$$= \frac{3}{\sqrt{108.68 \times 0.25}}$$

$$= \frac{3}{\sqrt{27.60}}$$

$$= \frac{3}{5.25}$$

$$= 0.57$$

Here the degrees of freedom is $n_1 + n_2 - 2 = 9 + 7 - 2 = 14$.

At 5% level of significance for 15 $d.f$ the table value for t is 2.145.

Since the calculate $|t|$ value is smaller than the table value of t and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the mean yields due to the application of two kinds of manures.

Example 10. 10 :

A medicine was administered to 10 patients to ascertain its effect on the heart beat. The heart beats before giving the medicine and after the heart beats of the patients were noted as follows. Did the medicine significantly increase the heart beats of the patient?

Patient	1	2	3	4	5	6	7	8	9	10
Before	70	68	72	71	70	65	68	70	71	72
After	76	74	69	71	75	72	69	70	76	75

Space for Hint

Solution :

Here we are concerned with the same set of patients in the sample. We compute the difference in their heart beats $z = x - y$ and calculate the mean $\overline{z}$ and the standard deviation of z.

Step 1 : First we find x and s.

x	y	$z = x - y$	$z - \overline{z}$	$(z - \overline{z})^2$
70	76	-6	-3	9
68	74	-6	-3	9
72	69	3	6	36
71	71	0	3	9
70	75	-5	-2	4
65	72	-7	-4	16
68	69	-1	2	4
70	70	0	3	9
71	76	-5	-2	4
72	75	-3	0	0
Total		-30		100

Now $\overline{z} = \dfrac{\sum x}{n}$

$= \dfrac{-30}{10}$

$= -3$

and $s = \sqrt{\dfrac{\sum (x - \overline{x})^2}{n}}$

$= \sqrt{\dfrac{100}{10}}$

$= 3.16$

Set $H_0 : \overline{z} = 0$

Space for Hint

(i.e.) the medicine not significantly increase the heart beats of the patient.

$\therefore\ H_1 : \bar{z} < 0$.

$$\text{Thus } t = \frac{\bar{z} - 0}{s/\sqrt{n-1}}$$

$$= \frac{-3}{3.16/\sqrt{10-1}}$$

$$= \frac{-3 \times 3}{3.16}$$

$$= -2.85$$

At 5% level of significance the table t value is 2.261

Since the calculate $|t|$ value is larger than the table value of t and therefore at 5% level of significance we reject the null hypothesis.

(i.e.) the medicine significantly increase the heart beats of the patient.

Check Your Progress

(1) Ten oil tins are taken at random from an automatic filling machine. The mean weight of the tins is 15.8kg and standard deviation is 0.50kg. Does the sample mean differ significantly from the intended weight of 16kg?

(2) Prices of shares of a company on the different days in a month were found to be: 66, 65, 69, 70, 69, 71, 70, 63, 64, 68. Discuss whether the mean price of the shares in the month is 65.

(3) Eleven school boys were given a test in drawing. They were given a month's futher tuition and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefited by the extra coaching?

Students	1	2	3	4	5	6	7	8	9	10	11
Test I	23	20	19	21	18	20	18	17	23	16	19
Test II	24	19	22	18	20	22	20	20	23	20	17

(4) The height in inches of 6 randomly chosen N.C.C students in Sourashtra college are 76, 70, 68, 69, 69, 68. Those of 6 randomly chosen N.S.S. students in the same college have height in inches 68, 64, 65, 69, 72, 64. Discuss the suggestion that N.C.C. students are on the average taller than N.S.S. students.

(5) From a population of college students 10 students were randomly selected. Their weekly pocket money was observed as (in rupees) 20, 22, 21, 15, 25, 19, 18, 20, 21, 22. Test whether the sample supports that on an average the students get Rs. 25 as pocket money.

Space for Hint

10. 2 F - test

The F-distribution is named in honour of R.A. Fisher who first studied it in 1924.This distribution is usually defined in terms of the ratio of the variances of two normally distributed populations. The quantity $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ is distributed as F-distribution with numerator has $\nu_1 = n_1 - 1$ degrees of freedom and denominator has $\nu_2 = n_2 - 1$ degrees of freedom, where $s_1^2 = \frac{\Sigma(x-\bar{x}_1)^2}{n_1 - 1}$ is the unbiased estimator of σ_1^2 and $s_2^2 = \frac{\Sigma(x-\bar{x}_2)^2}{n_2 - 1}$ is unbiased estimator of σ_2^2

Thus the test statistic for F distribution is $F = \frac{n_1 s_1^2/(n_1 - 1)}{n_2 s_2^2/(n_2 - 1)}$

If $\sigma_1^2 = \sigma_2^2$, then the statistic $F = \frac{s_1^2}{s_2^2}$ follows F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

The F-distribution sometimes is also called variance ratio distribution.

Example 10. 11 :

For two samples of sizes 8 and 12 the observed variances are 0.064 and 0.024. Test the hypothesis that the samples came from normal population with equal variances.

Space for Hint

Solution :

Set $H_0 : \sigma_1^2 = \sigma_2^2$.

(i.e.) there is no significant difference between the variances of the populations.

$\therefore \; H_1 : \sigma_1^2 \neq \sigma_2^2$.

Given that $n_1 = 8, \; s_1^2 = 0.064$ and

$$n_2 = 12, \; s_2^2 = 0.024.$$

$$\text{Now } F = \frac{n_1 s_1^2/(n_1 - 1)}{n_2 s_2^2/(n_2 - 1)}$$

$$= \frac{8(0.064)/(8-1)}{12(0.024)/(12-1)}$$

$$= \frac{0.07}{0.03}$$

$$= 2.67$$

At 5% level of significance the table F value for $(11,7)$ $d.f.$ is 3.01

Since the calculate F value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the variances of the populations.

Example 10. 12 :

In a sample of 8 observations the sum of the squared deviations of items from the mean was 94.5. In another sample of 10 observations the value was found to be 101.7. Test whether the difference is significant.

Solution :

Set $H_0 : \sigma_1^2 = \sigma_2^2$.

(i.e.) there is no significant difference between the variances of the populations.

$\therefore \; H_1 : \sigma_1^2 \neq \sigma_2^2$.

Given that $n_1 = 8, \; \Sigma(x - \bar{x}_1)^2 = 94.5$ and

$$n_2 = 10, \; \Sigma(x - \bar{x}_2)^2 = 101.7.$$

Now $s_1^2 = \frac{1}{n_1}\Sigma(x-\bar{x}_1)^2$

$\Rightarrow n_1 s_1^2 = \Sigma(x-\bar{x}_1)^2$

$\Rightarrow n_1 s_1^2 = 94.5$

Now $s_1^2 = \frac{1}{n_2}\Sigma(x-\bar{x}_2)^2$

$\Rightarrow n_2 s_2^2 = \Sigma(x-\bar{x}_2)^2$

$\Rightarrow n_2 s_2^2 = 101.7$

Now $F = \dfrac{n_1 s_1^2/(n_1-1)}{n_2 s_2^2/(n_2-1)}$

$= \dfrac{94.7/(8-1)}{101.7/(10-1)}$

$= \dfrac{13.53}{1.20}$

$= 1.20$

At 5% level of significance the table F value for (7, 9) *d.f.* is 3.29

Since the calculate F value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the variances of the populations.

Example 10. 13 :

In two groups of 10 children each the increase in weight in kilograms due to two different diets over the same period were as follows. Determine whether the variances are significantly different for two groups.

I group	8	5	7	8	3	2	7	6	5	7
II group	3	7	5	6	5	4	4	5	3	6

Solution :

Set $H_0 : \sigma_1^2 = \sigma_2^2$.

Space for Hint

Space for Hint

(i.e.) there is no significant difference between the variances of the populations.

$\therefore\ H_1 : \sigma_1^2 \neq \sigma_2^2$.

Now we shall find the mean and variances of two groups.

First we shall find the mean and variance of the first group.

	x	$x-\bar{x}_1$	$(x-\bar{x}_1)^2$
	8	2.20	4.84
	5	-0.80	0.64
	7	1.20	1.44
	8	2.20	4.84
	3	-2.80	7.84
	2	-3.80	14.44
	7	1.20	1.44
	6	0.20	0.04
	5	-0.80	0.64
	7	1.20	1.44
Total	58		37.60

Thus $\bar{x}_1 = \dfrac{\Sigma x}{n_1}$

$= \dfrac{58}{10}$

$= 5.8$

and $s_1^2 = \dfrac{1}{n_1}\Sigma(x-\bar{x}_1)^2$

$= \dfrac{37.6}{10}$

$= 3.76$

Now we shall find $\bar{x}_2$ and s_2^2 for the second group.

Space for Hint

	x	$x-\overline{x}_2$	$(x-\overline{x}_2)^2$
	3	-1.80	3.24
	7	2.20	4.84
	5	0.20	0.04
	6	1.20	1.44
	5	0.20	0.04
	4	-0.80	0.64
	4	-0.80	0.64
	5	0.20	0.04
	3	-1.80	3.24
	6	1.20	1.44
Total	48		15.60

Thus $\overline{x}_2 = \dfrac{\sum x}{n_2}$

$= \dfrac{48}{10}$

$= 4.8$

and $s_2^2 = \dfrac{1}{n_2}\sum(x-\overline{x}_2)^2$

$= \dfrac{15.6}{10}$

$= 1.56$

Hence $n_1 = 10$, $s_1^2 = 3.76$ and

$n_2 = 10$, $s_2^2 = 1.56$.

Now $n_1 s_1^2 = \sum(x-\overline{x}_1)^2$

$\Rightarrow n_1 s_1^2 = 37.60$

Now $n_2 s_2^2 = \sum(x-\overline{x}_2)^2$

$\Rightarrow n_2 s_2^2 = 15.60$

Now $F = \dfrac{n_1 s_1^2/(n_1-1)}{n_2 s_2^2/(n_2-1)}$

Space for Hint

$$= \frac{37.60/9}{15.60/9}$$

$$= 2.41$$

At 5% level of significance the table F value for (9, 9) *d.f.* is 3.18

Since the calculate F value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the variances of the populations.

Example 10. 14 :

The time taken by workers in performing a job by method I and method II is given below. Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

Method I	20	16	26	27	23	22	–
Method II	27	33	42	35	32	34	38

Solution :

Set $H_0 : \sigma_1^2 = \sigma_2^2$.

(i.e.) there is no significant difference between the variances of the populations.

$\therefore\ H_1 : \sigma_1^2 \neq \sigma_2^2$.

Now we shall find the mean and variances of two groups.

First we shall find the mean and variance of the first group.

Thus $\bar{x}_1 = \dfrac{\Sigma x}{n_1}$

$$= \frac{134}{6}$$

$$= 22.33$$

and $s_1^2 = \dfrac{1}{n_1}\Sigma(x - \bar{x}_1)^2$

$$= \frac{81.33}{6} = 13.56$$

	x	$x-\overline{x}_1$	$(x-\overline{x}_1)^2$
	20	-2.33	5.44
	16	-6.33	40.11
	26	3.67	13.44
	27	4.67	21.78
	23	0.67	0.44
	22	-0.33	0.11
Total	134		81.33

Now we shall find $\overline{x}_2$ and s_2^2 for the second group.

	x	$x-\overline{x}_2$	$(x-\overline{x}_2)^2$
	27	-7.43	55.18
	33	-1.43	2.04
	42	7.57	57.33
	35	0.57	0.33
	32	-2.43	5.90
	34	-0.43	0.18
	38	3.57	12.76
Total	241		133.71

Thus $\overline{x}_2 = \dfrac{\Sigma x}{n_2}$

$= \dfrac{241}{7}$

$= 34.43$

Space for Hint

Space for Hint

and $s_2^2 = \frac{1}{n_2}\Sigma(x - \bar{x}_2)^2$

$= \frac{133.71}{7}$

$= 19.10$

Hence $n_1 = 6$, $s_1^2 = 13.56$ and

$n_2 = 7$, $s_2^2 = 19.10$.

Now $n_1 s_1^2 = \Sigma(x - \bar{x}_1)^2$

$\Rightarrow n_1 s_1^2 = 81.33$

Now $n_2 s_2^2 = \Sigma(x - \bar{x}_2)^2$

$\Rightarrow n_2 s_2^2 = 133.71$

Now $F = \dfrac{n_1 s_1^2/(n_1 - 1)}{n_2 s_2^2/(n_2 - 1)}$

$= \dfrac{133.71/6}{81.33/5}$

$= \dfrac{22.285}{16.266}$

$= 1.37$

At 5% level of significance the table F value for (5, 6) $d.f.$ is 3.97

Since the calculate F value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the variances of the populations.

Example 10. 15 :

Test whether the following two samples have been drawn from the same population.

	Size	Mean	Sum of square of deviation from mean
Sample I	9	68	36
Sample II	10	69	42

Solution :

To test if two independent samples could have drawn from the same normal population we to test (i) the equality of population means (t - test) and (ii) the equality of population variances (F - test)

Given that $n_1 = 9$, $\bar{x}_1 = 68$, $\Sigma(x-\bar{x}_1)^2 = 36$ and

$n_2 = 10$, $\bar{x}_2 = 69$, $\Sigma(x-\bar{x}_2)^2 = 42$

Step 1 : First we shall test the quality of population variances.

Set $H_0 : \sigma_1^2 = \sigma_2^2$

$\therefore\ H_1 : \sigma_1^2 \neq \sigma_2^2$.

Now $n_1 s_1^2 = \Sigma(x-\bar{x}_1)^2$

$\Rightarrow n_1 s_1^2 = 9 \times 36$

$\Rightarrow n_1 s_1^2 = 324$

Now $n_2 s_2^2 = \Sigma(x-\bar{x}_2)^2$

$\Rightarrow n_2 s_2^2 = 10 \times 42$

$\Rightarrow n_2 s_2^2 = 420$

Now $\dfrac{n_1 s_1^2}{n_1 - 1} = \dfrac{324}{8}$

$= 40.5$

and $\dfrac{n_2 s_2^2}{n_2 - 1} = \dfrac{420}{9}$

$= 46.67$

Now $\dfrac{n_2 s_2^2/(n_2 - 1)}{n_1 s_1^2/(n_1 - 1)}$

$= \dfrac{46.67}{40.50}$

$= 1.152$

At 5% level of significance the table F value for (8, 9) $d.f.$ is 3.23

Since the calculate F value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the variances of the populations.

Space for Hint

Space for Hint

Step 2 : To test the equality of population means.

Set $H_0 : \mu_1 = \mu_2$

$\therefore\ H_1 : \mu_1 \neq \mu_2$.

Now the test statistic is $t = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \times \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

$$= \frac{68-69}{\sqrt{\dfrac{36+42}{9+10-2} \times \left(\dfrac{1}{9} + \dfrac{1}{10}\right)}}$$

$$= \frac{-1}{\sqrt{4.58 \times 0.21}}$$

$$= \frac{-1}{\sqrt{0.9618}}$$

$$= \frac{-1}{0.98}$$

$$= -1.02$$

At 5% level of significance the table t value for 17 $d.f.$ is 2.11

Since the calculate $|t|$ value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

(i.e.) there is no significant difference between the means of the populations.

Step 3 : Since both $H_0 : \mu_1 = \mu_2$ and $H_0 : \sigma_1^2 = \sigma_2^2$ are accepted, we may conclude that the given samples could have been drawn from the same population.

Check Your Progress

(1) Two independent samples of 8 and 7 items respectively had the following values of the variables. Do the two estimates of population variances differ significantly?

Sample I	9	11	13	11	15	9	12	14
Sample II	10	12	10	14	9	8	10	–

(2) Two random samples drawn from two normal populations are

Sample I	63	65	68	69	71	–	–	–	–
Sample II	63	62	65	66	69	69	70	71	73

Test whether the two populations have the same variances.

(3) Test whether the following two samples have been drawn from the same population.

	Size	Mean	Sum of square of deviation from mean
Sample I	10	15	90
Sample II	12	14	108

10. 3 Test for significance of an observed sample correlation

Let r correlation coefficient between two samples. To test the correlation coefficient of the populations, the test statistic is $t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$.

Example 10. 16 :

A random sample of 10 observations gave a correlation of 0.2. Is this significant of correlation in the population?

Solution :

Set $H_0 : \rho = 0$

and $H_1 : \rho \neq 0$

Given that $n = 10$, sample correlation $= r = 0.2$

$$\therefore\ t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

$$= \frac{0.2}{\sqrt{1-(0.2)^2}}\sqrt{10-2}$$

Space for Hint

Space for Hint

$$= \frac{0.2}{0.98} \times 2.83$$

$$= 0.58$$

At 5% level of significance the table t value for 8 $d.f.$ is 2.306.

Since the calculate $|t|$ value is smaller than the table value of F and therefore at 5% level of significance we accept the null hypothesis.

Hence the sample could have come from an uncorrelated population.

Check Your Progress

(1) A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0.6. Is this significant of correlation in the population?

(2) A random sample of 18 pairs of observations from a normal population gave a correlation coefficient of 0.52. Is it likely that the variables in the population are uncorrelated?

10. 4 Chi square test

The χ^2 test is one of the simplest and most widely used non-parametric tests in statistical work. It makes no assumptions about the population being sample. The quantity χ^2 describes the magnitude of discrepancy between theory and observation, i.e., with the help of χ^2 test we can know whether a given discrepancy between theory and observation can be attributed to chance or whether it results from the inadequacy of the theory to fit the observed facts. If χ^2 is zero, it means that the observed and expected frequencies completely coincide. The greater the value of χ^2, the greater would be the discrepancy between observed and expected frequencies. The formula for computing chi-square is:

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

where O = observed frequency,

E = expected or theoretical frequency.

The calculate value of χ^2 is compared with the value of χ^2 for given degrees of freedom at specified levels of significance. If the calculated value of χ^2 is greater than the table value, the difference between theory and observation is considered to be significant, i.e., it could not have arisen due to fluctuations of simple sampling. On the other hand, if the calculated value of χ^2 is less than the value, the difference between theory and observation is not considered significant, i.e., it could have arisen due to fluctuations of sampling.

Conditions for the Application of χ^2 Test

The following five basic conditions must be met in order for Chi-square analysis to be applied:

(1) The experimental data (sample observation) must be independent of each other.

(2) The sample data must be drawn at random from the target population.

(3) The data should be expressed in original units for convenience of comparison, and not in percentage or ratio form.

(4) The sample should contain at least 50 observations

10.4.1 χ^2 – test for population variance

To test population variance using chi-square, we use the test statistic as

$\chi^2 = \frac{ns^2}{\sigma_0^2}$, where s^2 is the sample variance.

Note 1 : If the sample size is greater than 50 then we apply Fisher's approximation $\sqrt{2\chi^2} \sim N(\sqrt{2n-1}, 1)$

In this case the test statistic becomes $z = \sqrt{2\chi^2} - \sqrt{2n-1}$.

Example 10. 17 :

A random sample of size 20 from a population gives the sample standard deviation of 6. Test the hypothesis that the sample is from a normal population with standard deviation 9.

Space for Hint

Space for Hint

Solution :

Given that $n = 20$, $s^2 = 6$ and $\sigma_0 = 9$.

Set $H_0 : \sigma^2 = \sigma_0^2$

$\therefore\ H_1 : \sigma^2 \neq \sigma_0^2$

Now $\chi^2 = \dfrac{ns^2}{\sigma_0^2}$

$= \dfrac{20 \times 36}{81}$

$= 8.89$

At 5% level of significance the table χ^2 value for 18 *d.f.* is 28.9

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence the sample is not from a normal population with standard deviation 9.

Example 10. 18 :

Weights in kilograms of 10 students are given as 38, 40, 45, 53, 47, 43, 55, 48, 45, 49. Can we say that variance of the distribution of weight of all students from which the above sample of 10 students was drawn is equal to 20sq. kgs.?

Solution :

Step 1 : First we shall find the variance of the sample.

	x	$x-\overline{x}$	$(x-\overline{x})^2$
	38	-8.30	68.89
	40	-6.30	39.69
	45	-1.30	1.69
	53	6.70	44.89
	47	0.70	0.49
	43	-3.30	10.89
	55	8.70	75.69
	48	1.70	2.89
	45	-1.30	1.69
	49	2.70	7.29
Total	463		254.10

Space for Hint

Now $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{468}{10}$

$= 46.8$

and $s^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

$= \frac{254.1}{10}$

$= 25.41$

Thus $n = 10$, $s^2 = 25.41$ and $\sigma_0^2 = 20$.

Set $H_0 : \sigma^2 = \sigma_0^2$

$\therefore\ H_1 : \sigma^2 \neq \sigma_0^2$

Now $\chi^2 = \frac{ns^2}{\sigma_0^2}$

$= \frac{10 \times 25.41}{50}$

$= 5.082$

At 5% level of significance the table χ^2 value for 8 *d.f.* is 15.5

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence the sample is not from a normal population with standard deviation 9.

Example 10. 19 :

Test the hypothesis that $\sigma = 10$ given that $s = 15$ for a random sample of size 50 from a normal population.

Solution :

Thus $n = 50$, $s = 15$ and $\sigma_0 = 10$.

$\therefore\ H_0 : \sigma^2 = \sigma_0^2$

$\therefore\ H_1 : \sigma^2 \neq \sigma_0^2$

Now $\chi^2 = \frac{ns^2}{\sigma_0^2}$

Space for Hint

$$= \frac{50 \times 225}{100}$$

$$= 112.50$$

Since $n > 30$, the population becomes a normal population.

Thus the test statistic $z = \sqrt{2\chi^2} - \sqrt{2n-1}$

$$= \sqrt{2 \times 112.50} - \sqrt{2 \times 50 - 1}$$

$$= \sqrt{225} - \sqrt{99}$$

$$= 15 - 9.95$$

$$= 5.05 > 3$$

Since the calculate z value is greater than the value of z therefore at 5% level of significance we accept the null hypothesis.

Hence the sample is not from a normal population with standard deviation 10.

Check Your Progress

(1) A sample of 12 values shows the standard deviation to be 11. Does this agree with the hypothesis that the population standard deviation is 10, the population being normal?

(2) Given 10 measurements of an instrument as 2.5, 2.03, 2.4, 2.3, 2.5, 2.7, 2.6, 2.6, 2.7, 2.5. It is believed that the precision of that instrument as measured by the variance is 0.16. Test whether the data are consistent with the hypothesis? (Apply 1% level of significance)

(3) Test the hypothesis that $\sigma = 8$ given that $s = 10$ for a random sample of size 51.

10.4.2 χ^2 – test for goodness of fit

Already we have studied theoretical distribution like Binomial distribution, Poisson distribution and Normal distribution. Now in this section we shall discuss to test the goodness of fit of the distribution. The test statistic is $\chi^2 = \sum \frac{(O-E)^2}{E}$ and the null hypothesis is H_0 : the observed and theoretical frequencies are compatible.

Space for Hint

Example 10. 20 :

In experiment of pea breeding a scientist got the following frequencies of seed, round and yellow : 315; wrinkled yellow : 101; round and green : 108; wrinkled green : 32; total : 556. Theory predicts that the frequencies should be in the proportion 9 : 3 : 3 : 1. Examine the correspondence between theory and experiment.

Solution :

Set H_0 : There exists a correspondence between theory and experiment.

$\therefore$ H_1 : There does not exists a correspondence between theory and experiment.

Now we shall find the χ^2 value.

Nature of seed	O	E	$O-E$	$(O-E)^2$	$\sum\frac{(O-E)^2}{E}$
round and yellow	315	312.75	2.25	5.0625	0.0162
wrinkled yellow	101	104.25	-3.25	10.5625	0.1013
round and green	108	104.25	3.75	14.0625	0.1349
wrinkled green	32	34.75	-2.75	7.5625	0.2176
	Total				0.4700

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

$= 0.47$ (From the last column of the table)

At 5% level of significance the table χ^2 value for 3 $d.f.$ is 7.851

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence There exists a correspondence between theory and experiment.

Example 10. 21 :

The following data represents the monthly sales in rupees of a certain retail shop in a year. Examine if there is any seasonality in the sales.

610, 560, 635, 605, 625, 620, 630, 625, 580, 600, 615, 615

Solution :

Set H_0 : The sales are dependent of seasons.

$\therefore$ H_1 : The sales are independent of seasons.

$$\text{Now } \bar{x} = \frac{610+560+635+605+625+620+630+625+580+600+615+615}{12}$$

$$\text{(i.e.) } \bar{x} = \frac{7320}{12}$$

$$\text{(i.e.) } \bar{x} = 610$$

Now we shall find the χ^2 value.

O	E	$O-E$	$(O-E)^2$	$\sum\frac{(O-E)^2}{E}$
610	610	0	0	0.00
560	610	-50	2500	4.10
635	610	25	625	1.02
605	610	-5	25	0.04
625	610	15	225	0.37
620	610	10	100	0.16
630	610	20	400	0.66
625	610	15	225	0.37
580	610	-30	900	1.48
600	610	-10	100	0.16
615	610	5	25	0.04
615	610	5	25	0.04
	Total			8.44

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

$= 8.44$ (From the last column of the table)

At 5% level of significance the table χ^2 value for 11 *d.f.* is 19.675

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence the sales is independent of season.

Example 10. 22 :

In 120 throws of a single die the following distribution of faces was obtained. Do these data indicate an unbiased dice.

Faces	1	2	3	4	5	6
Frequencies	30	25	18	10	22	15

Solution :

Set H_0 : The dice be unbiased.

$\therefore$ H_1 : The dice be biased.

Now total frequency $= 30+25+18+10+22+15 = 120$

Thus each face be turned $\frac{120}{6} = 20$ times.

Now we shall find the χ^2 value.

O	E	$O-E$	$(O-E)^2$	$\sum \frac{(O-E)^2}{E}$
30	20	10	100	5.00
25	20	5	25	1.25
18	20	-2	4	0.20
10	20	-10	100	5.00
22	20	2	4	0.20
15	20	-5	25	1.25
		Total		12.90

Space for Hint

Space for Hint

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

$= 12.90$ (From the last column of the table)

At 5% level of significance the table χ^2 value for 5 *d.f.* is 11.07

Since the calculate χ^2 value is greater than the table value of χ^2 and therefore at 5% level of significance we reject the null hypothesis.

Hence the dice is biased.

Example 10. 23 :

A survey of 320 families with 5 children each revealed the following distribution. Is this result consistent with the hypothesis that male and female births are equally probable?

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	14	56	110	88	40	12

Solution :

Set H_0 : Male and female births are equally probable.

$\therefore$ H_1 : Male and female births are not equally probable.

Step 1 : First we shall find the expected frequencies using Binomial distribution.

The probability of male birth = $p = \frac{1}{2}$

$\therefore \; q = 1 - p$

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

$= \frac{1}{2}$

Thus the probability mass function of Binomial distribution is

$P(X = x) = {}^nC_x p^x q^{n-x}$, $x = 0, 1, 2, 3, \cdots, n$

(i.e.) $P(X = x) = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$, $x = 0, 1, 2, 3, \cdots, 5$.

(i.e.) $P(X=x)=\left(\frac{1}{2}\right)^5 {}^5C_x$, $x=0,1,2,3,\cdots,5$.

(i.e.) $P(X=x)=\frac{1}{32}{}^5C_x$, $x=0,1,2,3,\cdots,5$.

x	$P(X=x)=p(x)$	$N\cdot p(x)$
0	$\frac{1}{32}{}^5C_0=\frac{1}{32}$	$320\cdot\frac{1}{32}=10$
1	$\frac{1}{32}{}^5C_1=\frac{5}{32}$	$320\cdot\frac{5}{32}=50$
2	$\frac{1}{32}{}^5C_2=\frac{10}{32}$	$320\cdot\frac{10}{32}=100$
3	$\frac{1}{32}{}^5C_3=\frac{10}{32}$	$320\cdot\frac{10}{32}=100$
4	$\frac{1}{32}{}^5C_4=\frac{5}{32}$	$320\cdot\frac{5}{32}=50$
5	$\frac{1}{32}{}^5C_5=\frac{1}{32}$	$320\cdot\frac{1}{32}=10$

Now we shall find the χ^2 value.

x	O	E	$O-E$	$(O-E)^2$	$\sum\frac{(O-E)^2}{E}$
0	12	10	2	4	0.40
1	40	50	-10	100	2.00
2	88	100	-12	144	1.44
3	110	100	10	100	1.00
4	56	50	6	36	0.72
5	14	10	4	16	1.60
		Total			7.16

Space for Hint

Space for Hint

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

= 7.16 (From the last column of the table)

At 5% level of significance the table χ^2 value for 5 *d.f.* is 11.07

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence male and female births are equally probable.

Check Your Progress

(1) A sample analysis of examination results of 500 students was made. It was found that 220 had failed; 170 had secured a third class; 90 were placed in second class and 20 got first class. Are these results commensurate with the general examination results which is in the ratio 4 : 3 : 2 : 1 for the above said categories respectively?

(2) The following table gives the frequency of occurrence of the digits $0,1,2,\cdots,9$ in the last place in four digits of a random number table. Examine if there is any particularity.

Digits	0	1	2	3	4	5	6	7	8	9	Total
Frequency	6	16	15	10	12	12	3	2	9	5	90

10.4.3 χ^2 – test the independence of attributes

Theorem 10.1 :

For the 2×2 contingency table $\boxed{\begin{matrix} a & b \\ c & d \end{matrix}}$, χ^2- test of independence is

$$\chi^2 = \frac{N(ad-bc)}{(a+c)(b+d)(a+b)(c+d)} \text{ where } N = a+b+c+d.$$

Proof :

Let two attributes be A and B.

Then the 2×2 contingency table is

Attributes	B	β	Total
A	a	b	$a+b=(A)$
α	c	d	$c+d=(\alpha)$
Total	$a+c=(B)$	$b+d=(\beta)$	$N=a+b+c+d$

Now the observed frequencies are $o(AB)=a$, $o(A\beta)=b$, $o(B\alpha)=c$, $o(\alpha\beta)=d$.

And the expected frequencies are

$e(AB)=\dfrac{(a+b)(a+c)}{N}$, $\quad e(A\beta)=\dfrac{(a+b)(b+d)}{N}$, $\quad e(\alpha B)=\dfrac{(c+d)(a+c)}{N}$,

$e(\alpha\beta)=\dfrac{(c+d)(b+d)}{N}$.

$$\text{Now } o(AB)-e(AB) = a-\frac{(a+b)(a+c)}{N}$$

$$= \frac{Na-(a+b)(a+c)}{N}$$

$$= \frac{a(a+b+c+d)-(a+b)(a+c)}{N}$$

$$= \frac{\cancel{a^2}+\cancel{ab}+\cancel{ac}+ad-\cancel{a^2}-\cancel{ac}-\cancel{ab}-bc}{N}$$

$$= \frac{ad-bc}{N}.$$

Similarly, $o(A\beta)-e(A\beta) = \dfrac{ad-bc}{N}$,

$$o(\alpha B)-e(\alpha B) = \frac{ad-bc}{N}$$

$$o(\alpha\beta)-e(\alpha\beta) = \frac{ad-bc}{N}$$

Space for Hint

Space for Hint

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

$$= \left(\frac{ad-bc}{N}\right)^2 \left[\frac{1}{e(AB)} + \frac{1}{e(A\beta)} + \frac{1}{e(\alpha B)} + \frac{1}{e(\alpha\beta)}\right]$$

$$= \left(\frac{ad-bc}{N}\right)^2 \left[\frac{N}{(a+b)(a+c)} + \frac{N}{(a+b)(b+d)} + \frac{N}{(c+d)(a+c)} + \frac{N}{(c+d)(b+d)}\right]$$

$$= \frac{(ad-bc)^2}{N} \left[\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)} + \frac{1}{(c+d)(a+c)} + \frac{1}{(c+d)(b+d)}\right]$$

$$= \frac{(ad-bc)^2}{N} \left[\frac{b+d+a+c}{(a+b)(a+c)(b+d)} + \frac{b+d+a+c}{(c+d)(a+c)(b+d)}\right]$$

$$= (ad-bc)^2 \left[\frac{1}{(a+b)(a+c)(b+d)} + \frac{1}{(c+d)(a+c)(b+d)}\right]$$

$$= (ad-bc)^2 \left[\frac{a+b+c+d}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$= \left[\frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}\right]$$

Thus $\chi^2 = \left[\frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}\right]$

This prove the theorem.

Note 1 : In 2×2 contingency table the *d.f.* is (2-1)(2-1) = 1.

Note 2 : In $m \times n$ contingency table the *d.f.* is $(m-1)(n-1)$.

Note 3 : If any one or more of the expected frequencies are less than 5 then in applying χ^2 - test we have also subtract the *d.f.* lost in pooling these frequencies with the preceding or succeeding frequency or frequencies.

Example 10. 24 :

The table given below shows the data obtained during an epidemic of cholera. Test the effectiveness of inoculation in preventing the attack of cholera.

	Attacked	Not attacked	Total
Inoculated	31	469	500
Not inoculated	185	1315	1500
Total	216	1784	2000

Space for Hint

Solution :

Set H_0 : Inoculation not effective to prevent the attack of cholera.

$\therefore$ H_1 : Inoculation effective to prevent the attack of cholera.

Given that

	Attacked	Not attacked	Total
Inoculated	31	469	500
Not inoculated	185	1315	1500
Total	216	1784	2000

Step 1 : To find the expected frequencies.

	Attacked	Not attacked	Total
Inoculated	54	446	500
Not inoculated	162	1338	1500
Total	216	1784	2000

Space for Hint

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

x	O	E	$O-E$	$(O-E)^2$	$\sum \frac{(O-E)^2}{E}$
Attacked and inoculated	31	54	-23	529	9.80
Not attacked and inoculated	469	446	23	529	1.19
Attacked and not inoculated	185	162	23	529	3.27
Not attacked and not inoculated	1315	1338	-23	529	0.40
Total					14.64

At 5% level of significance the table χ^2 value for 1 *d.f.* is 3.841

Since the calculate χ^2 value is greater than the table value of χ^2 and therefore at 5% level of significance we reject the null hypothesis.

Hence inoculation effective to prevent the attack of cholera.

Example 10. 25 :

1000 college students were classified according to their intelligence and economic conditions. Test whether there is any association between intelligence and economic conditions.

Economic conditions	Intelligence				Total
	Excellent	Good	Medium	Dull	
Good	50	200	150	80	480
Not good	80	190	160	90	520
Total	130	390	310	170	1000

Solution :

Set H_0 : There no association between intelligence and economic conditions.

$\therefore$ H_1 : There association between intelligence and economic conditions.

Economic conditions	Intelligence				Total
	Excellent	Good	Medium	Dull	
Good	50	200	150	80	480
Not good	80	190	160	90	520
Total	130	390	310	170	1000

Step 1 : To find the expected frequencies.

Economic conditions	Intelligence				Total
	Excellent	Good	Medium	Dull	
Good	62.4	187.2	148.8	81.6	480
Not good	67.6	202.8	161.2	88.4	520
Total	130	390	310	170	1000

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

Space for Hint

Space for Hint

O	E	$O-E$	$(O-E)^2$	$\sum\frac{(O-E)^2}{E}$
50	62.4	-12.4	153.76	2.46
200	187.2	12.8	163.84	0.88
150	148.8	1.2	1.44	0.01
80	81.6	-1.6	2.56	0.03
80	67.6	12.4	153.76	2.27
190	202.8	-12.8	163.84	0.81
160	161.2	-1.2	1.44	0.01
90	88.4	1.6	2.56	0.03
Total				6.50

At 5% level of significance the table χ^2 value for 3 *d.f.* is 7.851

Since the calculate χ^2 value is smaller than the table value of χ^2 and therefore at 5% level of significance we accept the null hypothesis.

Hence/there no association between intelligence and economic conditions.

Check Your Progress

(1) Two investigators draw samples from the same town in order to estimate the number of persons falling in the income groups *poor*, *middle class*, and *well-to-do* their results are as follows. Test whether the sampling technique of the investigators are significantly dependent of the income groups of people

Investigator	Income group			
	Poor	Middle class	Well-to-do	Total
A	140	100	15	255
B	140	50	20	210
Total	280	150	35	465

(2) From the following data using chi square test find whether recreation depends on sex.

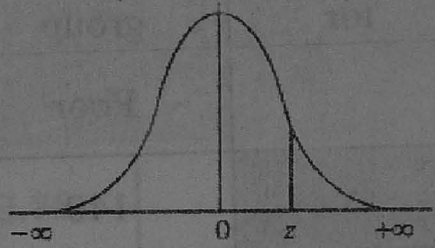
		Sex		Total
		Male	Female	
Recreation	Television	56	31	87
	Radio	18	6	24
Total		74	37	111

SUMMARY

In this unit we learned how to use t, F and χ^2- tests for small samples.

Space for Hint

NORMAL DISTRIBUTION TABLE



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

F Values for $\alpha = 0.10$

d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	39.86	49.5	53.59	55.83	57.24	58.2	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.3	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
inf	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

F Value for $\alpha = 0.10$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	60.19	60.71	61.22	61.74	62	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.40	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.80	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
inf	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

F Values for $\alpha = 0.05$

d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

F Values for $\alpha = 0.05$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.4	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.5
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.10	1.55	1.50	1.43	1.35	1.25
inf	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F Values for $\alpha = 0.01$

d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.2	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.14
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
inf	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

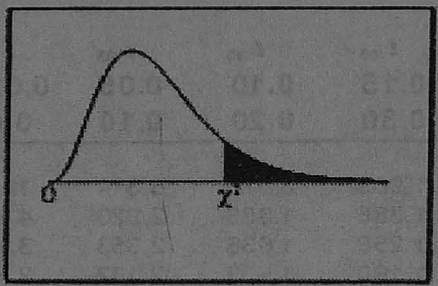
F Values for $\alpha = 0.01$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
inf	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.67	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

STATISTICS

Unit I :

Measures of averages – Measures of dispersion – Skewness based on moments.

Unit 2 :

Correlation and Regression – Rank correlation coefficient.

Unit 3:

Index numbers and Time series.

Unit 4 :

Curve fitting (All types of curves)

Unit -5

Theory of Attributes.

Unit 6 :

Theory of probability – sample space- probability function – Laws of Addition – Boole's inequality – Law of Multiplication – Problems – Baye's theorem – Problems.

Unit 7 :

Random variables – Distribution function – Discrete and Continuous random variables – Probability density function – Mathematical Expectations (One dimension only).

Unit 8 :

Moment generating function – cumulants – Theoretical distributions - binomial, Poisson, Normal.

Unit 9 :

Tests of significance of Large Samples.

Unit 10 :

Tests of significance of small samples – t, F, χ^2.

Text Book :

Statistics by Dr. S.Arumugam – Sci-Tech Publications, 2006.