Lecture 1
Chapter 1: Introduction to Statistics and Data Analysis

1.3 Measures of Location: The Sample Mean and Median

The aim of this lecture is to explain the following concepts :

- Measures of Location.
- The Sample Mean and Median.
- The Sample Range and Sample Standard Deviation.
- Histogram.

Definition 1 *Suppose that the observations in a sample are* $x_1, x_2, ..., x_n$. *The sample mean, denoted by* $\bar{x}$, *is*

$$\bar{x} = \sum_{i=1}^{n} \frac{x_1 + x_2 + ... + x_n}{n}$$

Definition 2 *Given that the observations in a sample are* $x_1, x_2, ..., x_n$, *arranged in increasing order of magnitude, the sample median is*

$$\bar{x} = \begin{cases} x_{(n+1)/2}, & if\ n\ is\ odd \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & if\ n\ is\ even. \end{cases}$$

Definition 3 *The sample variance, denoted by* s^2, *is given by*

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

The sample standard deviation, denoted by s, *is the positive square root of* s^2, *that is,*

$$s = \sqrt{s^2}$$

Example 1 *An engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH = 7.0). A sample of size 10 is taken, with results given by*
7.07 7.00 7.10 6.97 7.00 7.03 7.01 7.01 6.98 7.08.
Find sample variance and standard deviation.

Solution : *The sample mean $\bar{x}$ is given by*

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + \ + 7.08}{10} = 7.0250.$$

The sample variance s^2 is given by

$$s^2 = \frac{1}{9}[(7.07-7.025)2+(7.00-7.025)2+(7.10-7.025)2+....+(7.08-7.025)2] = 0.001939.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.001939} = 0.044.$$

So the sample standard deviation is 0.0440 with n-1 = 9 degrees of freedom.

Exercises:

1. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.
3.4 2.5 4.8 2.9 3.6
2.8 3.3 5.6 3.7 2.8
4.4 4.0 5.2 3.0 4.8
Assume that the measurements are a simple random sample.
(a) What is the sample size for the above sample?
(b) Calculate the sample mean for these data.
(c) Calculate the sample median.
(d) Compute the 20 trimmed mean for the above data set.
Solution :

(a) sample size = 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + \ + 4.8) = 3.787$

(c) Sample median is the 8th value, after the data is sorted from smallest to largest = 3.6.

(d) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:
2.9 3.0 3.3 3.4 3.6
3.7 4.0 4.4 4.8.
So. the trimmed mean is

$$\bar{x}_{tr20} = \frac{1}{9}(2.9 + 3.0 + \ + 4.8) = 3.678.$$

2. According to the journal Chemical Engineering, an important property of a fiber is its water absorbency. A random sample of 20 pieces of cotton fiber was taken and the absorbency on each piece was measured. The following are the absorbency values:
18.71 21.41 20.72 21.81 19.29 22.43 20.17
23.71 19.44 20.50 18.92 20.33 23.00 22.85
19.25 21.77 22.11 19.77 18.04 21.12
(a) Calculate the sample mean and median for the above sample values.
(b) Compute the 10% trimmed mean.
Solution :
Given sample size = 15.
(a) Mean=20.768 and Median=20.610.
(b) $\bar{x}_{tr10} = 20.743$.

7. Consider the drying time data for Exercise 1.1 on page 13. Compute the sample variance and sample standard deviation.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{15-1}[(3.4-3.787)2+(2.5-3.787)2+(4.8-3.787)2+.....+(4.8-3.787)2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.9428} = 0.971.$$

8. Compute the sample variance and standard deviation for the water absorbency data of Exercise 1.2 on page 13.
Solution : The sample variance s^2 is given by

$$s^2 = \frac{1}{20-1}[(18.71-20.768)2+(21.41-20.768)2+....+(21.12-20.768)2] = 0.94284.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{2.5345} = 1.592.$$

Histogram:
A table listing relative frequencies is called a **relative frequency distribution**.
The information provided by a relative frequency distribution in tabular form is easier to grasp if presented **graphically**.
Using the midpoint of each interval and the corresponding relative frequency, we construct a **relative frequency histogram**.

Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

Figure 1: Relative Frequency Distribution of Battery Life

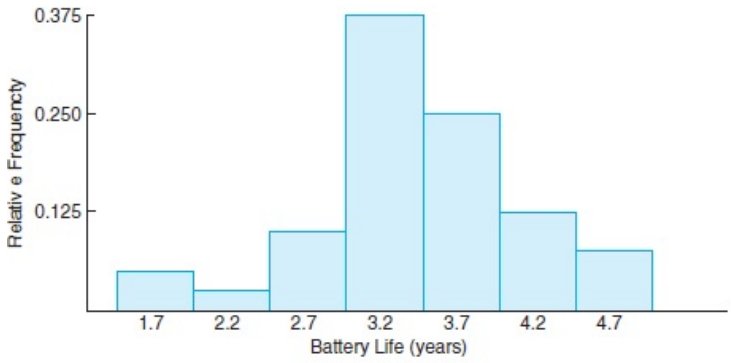


Figure 2: Relative frequency histogram

Lecture2
Chapter 2: Probability

2.1 Sample Space, 2.2 Events, 2.3 Counting Sample Points

The aim of this lecture is to explain the following concepts:

- Sample Space.
- Event.
- Counting Sample Points.

2.1 Sample Space:

Definition 1 *The set of all possible outcomes of a statistical experiment is called the* ***sample space*** *and is represented by the symbol S.*

Notes:

- Each outcome in a sample space is called an element or a member of the sample space, or simply a sample point.
- If the sample space has a finite number of elements, we may list the members separated by commas and enclosed in braces.
- Thus, the sample space S, of possible outcomes when a coin is flipped, may be written $S = \{H, T\}$, where H and T correspond to heads and tails, respectively.

Example 1 *Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is* $S_1 = \{1, 2, 3, 4, 5, 6\}$.
If we are interested only in whether the number is even or odd, the sample space is simply $S_2 = \{even, odd\}$.

Example 2 *An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.1. The sample space is* $S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$.

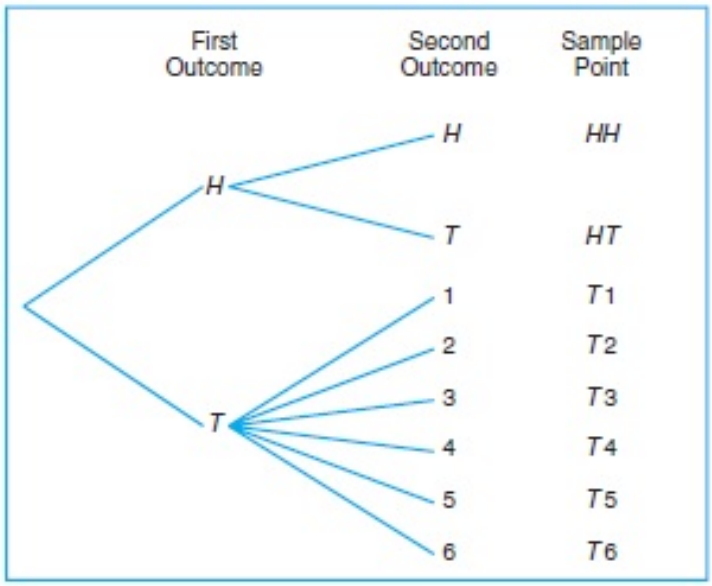


Figure 1: Tree diagram for Ex. 2

Example 3 *Suppose that three items are selected at random from a manufacturing process. Each item is inspected and classified defective, D, or nondefective, N. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.2. The sample space is*

$$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$$

Suppose the experiment is to sample items randomly until one defective item is observed. The sample space for this case is $S = \{D, ND, NND, NNND, ...\}$

2.2 Events:

Definition 2 *An* ***event*** *is a subset of a sample space.*

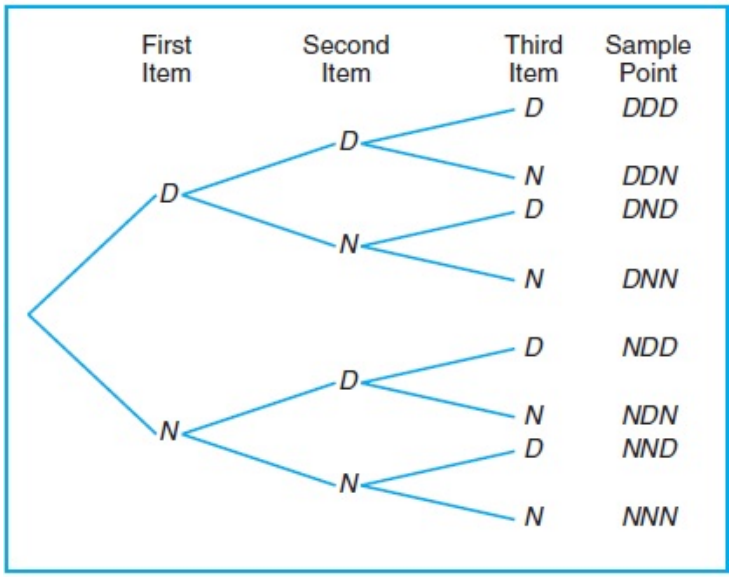


Figure 2: Tree diagram for Ex. 3

Example 4 *The event A that the outcome when a die is tossed is divisible by 3. This will occur if the outcome is an element of the subset$S_1 = \{3, 6\}$ of the sample space S_1 in Example 1.*
In the event B that the number of defectives is greater than 1 in Example 3. This will occur if the outcome is an element of the subset $S = \{DDD, DDN, DND, NDD\}$ of the sample space S.

Definition 3 *The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.*

Definition 4 *The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.*

Definition 5 *Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$, that is, if A and B have no elements in common.*

Definition 6 *The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.*

Exercises:

3. Which of the following events are equal?

(a) $A = \{1, 3\}$

(b) $B = \{x \mid x \text{ is a number on a die}\}$

(c) $C = \{x \mid x^2 - 4x + 3 = 0\}$

(d) $D = \{x \mid x \text{ is the number of heads when six coins are tossed}\}$

Solution :

(a) $A = \{1, 3\}$

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

(c) $C = \{x \mid x^2 - 4x + 3 = 0\} = \{x \mid (x-1)(x-3) = 0\} = \{1, 3\}$

(d) $D = \{0, 1, 2, 3, 4, 5, 6\}$

Clearly, A = C.

7. Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space S_1, using the letter M for male and F for female. Define a second sample space S_2 where the elements represent the number of females selected.
Solution:

$S_1 = \{MMMM, MMMF, MMFM, MFMM, FMMM, MMFF, MFMF, MFFM, FMFM, FF$
$S_2 = \{0, 1, 2, 3, 4\}$

2.3 Counting Sample Points:
Multiplication Rule: If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.

Example 5 *How many sample points are there in the sample space when a pair of dice is thrown once?*
Solution : *The first die can land face-up in any one of $n_1 = 6$ ways. For each of these 6 ways, the second die can also land face-up in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.*

Generalized Multiplication Rule: If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 .....n_k$ ways.

Example 6 *Sam is going to assemble a computer by himself. He has the choice of chips from two brands, a hard drive from four, memory from three, and an accessory bundle from five local stores. How many different ways can Sam order the parts?*
Solution : *Since* $n_1 = 2, n_2 = 4, n_3 = 3, and n_4 = 5$*, there are* $n_l \times n_2 \times n_3 \times n_4 = 2 \times 4 \times 3 \times 5 = 120$ *different ways to order the parts.*

Definition 7 *A* ***permutation*** *is an arrangement of all or part of a set of objects.*

Notes:

- For any non-negative integer n, $n!$, called **n factorial**, is defined as $n! = n(n-1)(2)(1)$, with special case $0! = 1$.
- The number of permutations of n objects is $n!$.
- The number of permutations of n distinct objects taken r at a time is

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

.

Example 7 *In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?*
Solution : *Since the awards are distinguishable, it is a permutation problem. The total number of sample points is*

$$^{25}P_3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = (25)(24)(23) = 13,800$$

.

Notes:

- The number of permutations of n objects arranged in a circle is $(n-1)!$.
- The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind,....., n_k of a kth kind is

$$\frac{n!}{n_1!\ n_2!\ ....\ n_k!}$$

.

Example 8 *In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?*
Solution : *We find that the total number of arrangements is*

$$\frac{10!}{1!\ 2!\ 4!\ 3!} = 12,600$$

.

Notes:

- The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1,\ n_2, ..., n_r} = \frac{n!}{n_1!\ n_2!\ ....\ n_r!},\ where\ n_1 + n_2 + ... + n_r = n$$

.

- The number of combinations of n distinct objects taken r at a time is

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

Lecture 3
2.4 Probability of an Event
2.5 Additive Rules

The aim of this lecture is to explain the following concepts:

- Probability of an Event.
- Additive Rules.

2.4 Probability of an Event:

Definition 1 *The probability of an event A is the sum of the weights of all sample points in A. Therefore,* $0 \leq P(A) \leq 1, P(\phi) = 0, and P(S) = 1$. *Furthermore, if* $A_1, A_2, A_3, ...$ *is a sequence of mutually exclusive events, then* $P(A_1 \cup A_2 \cup A_3 \cup .....) = P(A_1) + P(A_2) + P(A_3) + ...$

Example 1 *A coin is tossed twice. What is the probability that at least 1 head occurs?*
Solution : *The sample space for this experiment is* $S = \{HH, HT, TH, TT\}$. *If the coin is balanced, each of these outcomes is equally likely to occur. If A represents the event of at least 1 head occurring, then* $A = \{HH, HT, TH\}$ *and* $P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$

Note: If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is $P(A) = \frac{n}{N}$.

2.5 Additive Rules:

Theorem 0.1 *If A and B are two events, then*

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

.

Corrolary 0.1 *If $A_1, A_2, ..., A_n$ are mutually exclusive, then*

$$P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

.

Corrolary 0.2 *If $A_1, A_2, ..., A_n$ is a partition of sample space S, then*

$$P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n) = P(S) = 1$$

.

Theorem 0.2 *For three events A, B, and C, then*

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

.

Theorem 0.3 *If A and $A^{'}$ are complementary events, then*

$$P(A) + P(A^{'}) = 1$$

.

Exercises:

50. Assuming that all elements of S in Exercise 2.8 on page 42 are equally likely to occur, find

(a) the probability of event A.

(b) the probability of event C.

(c) the probability of event $A \cap C$.

Solution :

(a) $P(A) = \dfrac{5}{18}$.

(b) $P(C) = \dfrac{1}{3}$.

(c) $P(A \cap C) = \dfrac{7}{36}$.

53. The probability that an American industry will locate in Shanghai, China, is 0.7, the probability that it will locate in Beijing, China, is 0.4, and the probability that it will locate in either Shanghai or Beijing or both is 0.8. What is the probability that the industry will locate

(a) in both cities?

(b) in neither city?

Solution : Consider the events
S: industry will locate in Shanghai.
B: industry will locate in Beijing.

(a) $P(S \cap B) = P(S) + P(B) - P(S \cap B) = 0.7 + 0.4 - 0.8 = 0.3.$

(b) $P(S' \cap B') = 1 - P(S \cup B) = 1 - 0.8 = 0.2$

58.A pair of fair dice is tossed. Find the probability of getting

(a) a total of 8.

(b) at most a total of 5.

Solution :

(a) Of the (6)(6) = 36 elements in the sample space, only 5 elements (2,6), (3,5), (4,4), (5,3), and (6,2) add to 8. Hence the probability of obtaining a total of 8 is then $\frac{5}{36}$.

(b) Ten of the 36 elements total at most 5. Hence the probability of obtaining a total of at most is $\frac{10}{36} = \frac{5}{18}$.

59. In a poker hand consisting of 5 cards, find the probability of holding

(a) 3 aces.

(b) 4 hearts and 1 club.

Solution :

(a) $P(3\ aces) = \frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} = \frac{94}{54145}$

(b) $P(4heartsand1club) = \frac{\binom{13}{4}\binom{13}{1}}{\binom{52}{5}} = \frac{143}{39984}$

65. Consider the situation of Exercise 2.64. Let A be the event that the component fails a particular test and B be the event that the component displays strain but does not actually fail. Event A occurs with probability 0.20, and event B occurs with probability 0.35.

(a) What is the probability that the component does not fail the test?

(b) What is the probability that the component works perfectly well (i.e., neither displays strain nor fails the test)?

(c) What is the probability that the component either fails or shows strain in the test?

Solution : P(A) = 0.2 and P(B) = 0.35
(a) $P(A^{'}) = 1 - 0.2 = 0.8$
(b) $P(A^{'} \cap B^{'}) = 1 - P(A \cup B) = 1 - 0.2 - 0.35 = 0.45$
(c) $P(A \cup B) = 0.2 + 0.35 = 0.55$.

68. Interest centers around the nature of an oven purchased at a particular department store. It can be either a gas or an electric oven. Consider the decisions made by six distinct customers.

(a) Suppose that the probability is 0.40 that at most two of these individuals purchase an electric oven. What is the probability that at least three purchase the electric oven?

(b) Suppose it is known that the probability that all six purchase the electric oven is 0.007 while 0.104 is the probability that all six purchase the gas oven. What is the probability that at least one of each type is purchased?

Solution : (a) 1-0.40 = 0.60.
(b) The probability that all six purchasing the electric oven or all six purchasing the gas oven is 0.007 + 0.104 = 0.111.
So the probability that at least one of each type is purchased is 1-0.111 = 0.889.

72. Prove that

$$P(A^{'} \cap B^{'}) = 1 + P(A \cap B) - P(A) - P(B)$$

. **Solution :**
$P(A^{'} \cap B^{'})$
$= 1 - P(A \cup B)$
$= 1 - (P(A) + P(B) - P(A \cap B))$
$= 1 + P(A \cap B) - P(A) - P(B)$

Lecture 4
2.6 Conditional Probability, Independence, and the Product Rule

The aim of this lecture is to explain the following concepts:

- Conditional Probability.
- Independence.
- The Product Rule

Definition 1 *The **conditional probability** of B, given A, denoted by $P(B|A)$, is defined by*

$$(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \text{ provided } P(A) > 0$$

.

Definition 2 *Two events A and B are **independent** if and only if*

$$P(B|A) = P(B) \text{ or } P(A|B) = P(A),$$

assuming the existences of the conditional probabilities. Otherwise, A and B are dependent.

The Product Rule, or the Multiplicative Rule:

Theorem 0.1 *If in an experiment the events A and B can both occur, then*

$$P(A \cap B) = P(A)P(B|A), \quad \text{provided } P(A) > 0$$

.

Theorem 0.2 *Two events A and B are independent if and only if*

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

. Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

Theorem 0.3 *If, in an experiment, the events $A_1, A_2, ..., A_k$ can occur, then*

$$P(A_1 \cap A_2 \cap .... \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).....P(A_k|A_1 \cap A_2 \cap ..... \cap A_{k-1}).$$

If the events $A_1, A_2, ..., A_k$ are independent, then

$$P(A_1 \cap A_2 \cap .... \cap A_k) = P(A_1)P(A_2).....P(A_k).$$

Definition 3 *A collection of events $A = \{A_1, ..., A_n\}$ are* ***mutually independent*** *if for any subset of $A, A_{i1}, ..., A_{ik}$, for $k \leq n$, we have*

$$P(A_{i1} \cap ..... \cap A_{ik}) = P(A_{i1}).....P(A_{ik}).$$

Exercises:

74. A class in advanced physics is composed of 10 juniors, 30 seniors, and 10 graduate students. The final grades show that 3 of the juniors, 10 of the seniors, and 5 of the graduate students received an A for the course. If a student is chosen at random from this class and is found to have earned an A, what is the probability that he or she is a senior?
Solution :
$P(S|A) = 10/18 = 5/9.$

75. A random sample of 200 adults are classified below by sex and their level of education attained.

Education	*Male*	*Female*
Elementary	38	45
Secondary	28	50
College	22	17

If a person is picked at random from this group, find the probability that

(a) the person is a male, given that the person has a secondary education;

(b) the person does not have a college degree, given that the person is a female.

Solution : Consider the events:
M: a person is a male;
S: a person has a secondary education;
C: a person has a college degree.

(a) $P(M|S) = \dfrac{28}{78} = \dfrac{14}{39}$.

(b) $P(C'|M') = \dfrac{95}{112}$.

77. In the senior year of a high school graduating class of 100 students, 42 studied mathematics, 68 studied psychology, 54 studied history, 22 studied both mathematics and history, 25 studied both mathematics and psychology, 7 studied history but neither mathematics nor psychology, 10 studied all three subjects, and 8 did not take any of the three. Randomly select a student from the class and find the probabilities of the following events.

(a) A person enrolled in psychology takes all three subjects.

(b) A person not taking psychology is taking both history and mathematics.

Solution :

(a) $P(M \cap P \cap H) = \dfrac{10}{68} = \frac{5}{34}$.

(b) $P(H \cap M|P') = \dfrac{P(H \cap M \cap P')}{P(P')} = \dfrac{22-10}{100-68} = \dfrac{12}{32} = \dfrac{3}{8}$.

80. The probability that an automobile being filled with gasoline also needs an oil change is 0.25; the probability that it needs a new oil filter is 0.40; and the probability that both the oil and the filter need changing is 0.14.

(a) If the oil has to be changed, what is the probability that a new oil filter is needed?

(b) If a new oil filter is needed, what is the probability that the oil has to be changed?

Solution : Consider the events:
C: an oil change is needed,
F: an oil filter is needed.

(a) $P(F|C) = \dfrac{P(F \cap C)}{P(C)} = \dfrac{0.14}{0.25} = 0.56.$

(b) $P(C|F) = \dfrac{P(C \cap F)}{P(F)} = \dfrac{0.14}{0.40} = 0.35.$

89. A town has two fire engines operating independently. The probability that a specific engine is available when needed is 0.96.

(a) What is the probability that neither is available when needed?

(b) What is the probability that a fire engine is available when needed?

Solution : Let A and B represent the availability of each fire engine.

(a) $P(A' \cap B') = P(A')P(B') = (0.04)(0.04) = 0.0016.$

(b) $P(A \cup B) = 1 - P(A' \cap B') = 1 - 0.0016 = 0.9984.$

91. Find the probability of randomly selecting 4 good quarts of milk in succession from a cooler containing 20 quarts of which 5 have spoiled, by using

(a) the first formula of Theorem 2.12 on page 68

(b) the formulas of Theorem 2.6 and Rule 2.3 on pages 50 and 54, respectively.

Solution :

(a) $P(Q_1 \cap Q_2 \cap Q_3 \cap Q_4) = P(Q_1)P(Q_2|Q_1)P(Q_3|Q_1 \cap Q_2)P(Q_4|Q_1 \cap Q_2 \cap Q_3) = (15/20)(14/19)(13/18)(12/17) = 91/323.$

(b) Let A be the event that 4 good quarts of milk are selected. Then $P(A) = \dfrac{\binom{15}{4}}{\binom{20}{4}} = \dfrac{91}{323}.$

Lecture 5
2.7 Bayes' Rule

The aim of this lecture is to explain the following concepts:

- Total Probability.
- Bayes' Rule

Theorem of total probability or the rule of elimination:

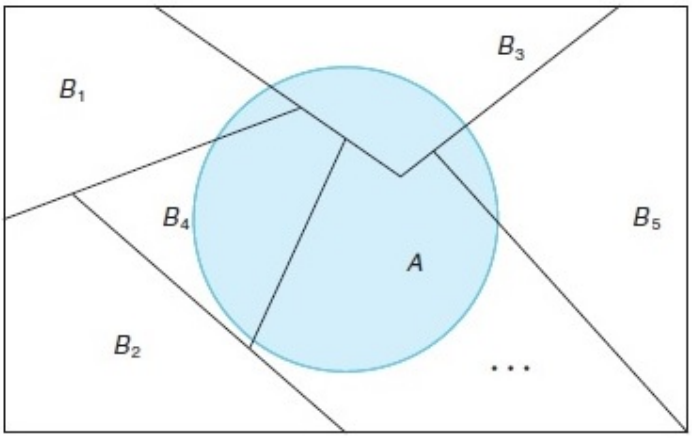


Figure 1: Partitioning the sample space S.

Theorem 0.1 *If the events $B_1, B_2, ..., B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for i = 1,2,.....,k, then for any event A of S,*

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

Theorem 0.2 *If the events $B_1, B_2, ..., B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for i = 1,2,.....,k, then for any event A of S,*

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \quad for\ r = 1, 2, ..., k.$$

Exercises:

95. In a certain region of the country it is known from past experience that the probability of selecting an adult over 40 years of age with cancer is 0.05. If the probability of a doctor correctly diagnosing a person with cancer as having the disease is 0.78 and the probability of incorrectly diagnosing a person without cancer as having the disease is 0.06, what is the probability that an adult over 40 years of age is diagnosed as having cancer?

Solution :

Consider the events:
C: an adult selected has cancer,
D: the adult is diagnosed as having cancer.
P(C) = 0.05,
$P(D|C) = 0.78$,
$P(C^{'}) = 0.95$
and $P(D|C^{'}) = 0.06$.
So, $P(D) = P(C \cap D) + P(C^{'} \cap D)$
$= (0.05)(0.78) + (0.95)(0.06)$
$= 0.096$.

96. Police plan to enforce speed limits by using radar traps at four different locations within the city limits. The radar traps at each of the locations L1, L2, L3, and L4 will be operated 40%, 30%, 20% ,and 30% of the time. If a person who is speeding on her way to work has probabilities of 0.2, 0.1, 0.5,

and 0.2, respectively, of passing through these locations, what is the probability that she will receive a speeding ticket?
Solution : Let $S_1, S_2, S_3,$ *and* S_4 represent the events that a person is speeding as he passes through the respective locations and
let R represent the event that the radar traps is operating resulting in a speeding ticket.
Then the probability that he receives a speeding ticket:
$P(R) = \sum_{i=1}^{4} P(R|S_i)P(S_i)$
$= (0.4)(0.2) + (0.3)(0.1) + (0.2)(0.5) + (0.3)(0.2)$
$= 0.27.$

97. Referring to Exercise 2.95, what is the probability that a person diagnosed as having cancer actually has the disease?
Solution : $P(C|D) = \dfrac{P(C \cap D)}{P(D)} = \dfrac{0.039}{0.096} = 0.40625.$

98. If the person in Exercise 2.96 received a speeding ticket on her way to work, what is the probability that she passed through the radar trap located at L2?
Solution : $P(S_2|R) = \dfrac{P(R \cap S_2)}{P(R)} = \dfrac{0.03}{0.27} = 1/9.$

RANDOM VARIABLE AND PROBABILITY DISTRIBUTIONS

LECTURE-6

1 Concept of Random Variable:-

Definition 1.1. A Random variable is a function that association a real number with each element in the sample space.

Example 1.1.
Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Definition 1.2. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Example 1.2.
Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Definition 1.3. If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**, (i.e. The set of possible outcomes is countable).

Definition 1.4. If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**,(i.e. The set of possible outcomes is uncountable).

2 Discrete Probability Distributions:-

Definition 2.1. The set of ordered pairs $(x, f(x))$ is a **probability function, probability mass function, or probability distribution of the discrete random variable** X if, for each possible outcome x,
1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x)$.

Definition 2.2. The cumulative distribution function $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is
$F(x) = P(X < x) = \sum_{t \leq x} f(t), \text{for} -\infty < x < \infty.$

3 Continuous Probability Distributions

Definition 3.1.

The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if
1. $f(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < x < b) = \int_a^b f(x)dx.$

Definition 3.2. The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

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

where $P(a < x < b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$.

LECTURE-7

Problem 3.3. 7).The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

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

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours,

(b) between 50 and 100 hours.

sols:-a)

$$\begin{aligned} P(X < 1.2) &= \int_{-\infty}^{1.2} f(x)dx \\ &= \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{1.2} f(x)dx \\ &= 0 + \int_{0}^{1} xdx + \int_{1}^{1.2} (2 - x)dx \\ &= \frac{1}{2} + 2(1.2 - 1) - \frac{1}{2}((1.2)^2 - 1) \\ &= \frac{1}{2} + 0.4 - 0.22 \\ &= 0.68. \end{aligned}$$

b)

$$\begin{aligned} P(0.5 < X < 1) &= \int_{0.5}^{1} xdx \\ &= \frac{x^2}{2}\Big]_{0.5}^{1} \\ &= \frac{1}{2}(1 - 0.25) \\ &= 0.375. \end{aligned}$$

Problem 3.4. 10). Find a formula for the probability distribution of the random variable X representing the outcome when a single die is rolled once.
solution:-

$$P(X=x)=\begin{cases}\frac{1}{6},\ X=1,2,3,4,5,6,\\ 0,\ elesewhere.\end{cases}$$

Problem 3.5. 11 A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets purchased by the hotel, find the probability distribution of X.
solution:-

$$f(x)=\frac{^2C_x\ *\ ^5C_{3-x}}{^7C_3},\ x=0,1,2$$

When'$x=0,\ f(x)=\frac{2}{7},\ x=1,\ then f(x)=\frac{4}{7}$ similarly $x=2,\ f(x)=\frac{1}{7}$.

Problem 3.6. 12). An investment firm offers its customers municipal bonds that mature after varying numbers of years. Given that the cumulative distribution function of T, the number of years to maturity for a randomly selected bond, is

$$F(t)=\begin{cases}0, & t<1\\ \frac{1}{4}, & 1\le t<3\\ \frac{1}{2}, & 3\le t<5\\ \frac{3}{4}, & 5\le t<7\\ 1, & t\ge 7\end{cases}$$

find (a) $P(T=5)$
(b) $P(T>3)$
(c) $P(1.4<T<6)$
(d) $P(T\le 5\mid T\ge 2)$
solutions:- a) $P(T=5)=F(5)-F(4)=\frac{3}{4}-\frac{1}{2}=\frac{1}{4}$
b) $P(T>3)=1-P(T\le 3)=1-F(3)=1-\frac{1}{2}=\frac{1}{2}$
c)$P(1.4<T<6)=F(6)-F(1.4)=\frac{3}{4}-\frac{1}{4}=\frac{1}{2}$
d)$P(T\le 5\mid T\ge 2)=\frac{P(2\le T\le 5)}{P(T\ge 2)}=\frac{F(5)-F(2)}{1-F(2)}=\frac{\frac{3}{4}-\frac{1}{4}}{1-\frac{1}{4}}=\frac{1}{2}*\frac{4}{3}=\frac{2}{3}$.

Problem 3.7. 14). The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution

function

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-8x}, & x \geq 0 \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders
(a) using the cumulative distribution function of X;
(b) using the probability density function of X.
solutions:- a)$P(X < 0.2) = F(0.2) = 1 - e^{-8*0.2} = 1 - e^{-1.6} = 0.7981.$
b)$f(x) = F'(x) = 8 * e^{-8x}$
$P(X < 0.2) = 8 * \int_0^{0.2} e^{-8x}dx = -e^{-8x}\big]_0^{0.2} = 0.7981.$

Problem 3.8. 21). Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Evaluate k.
(b) Find $F(x)$ and use it to evaluate

$$P(0.3 < X < 0.6)$$

solutions:- a)$\int_{-\infty}^{\infty} f(x)dx = 1$
$\Rightarrow \int_0^1 k\sqrt{x}dx = 1$
$\Rightarrow k\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\Big]_0^1 = 1$
$\Rightarrow \frac{2k}{3} = 1$
$\Rightarrow k = \frac{3}{2}$
b) $F(x) = \int_{-\infty}^{x} f(t)dt$
$= \int_{-\infty}^{0} 0dt + \int_0^x \frac{3}{2}\sqrt{t}dt$
$= x^{\frac{3}{2}},$ **where**

$$F(x) = \begin{cases} 0, & x < 0 \\ x^{\frac{3}{2}}, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

$P(0.3 < X < 0.6) = F(0.6) - F(0.3) = (0.6)^{\frac{3}{2}} - (0.3)^{\frac{3}{2}} = 0.3004.$

Problem 3.9. 29).3.29 An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Verify that this is a valid density function.
(b) Evaluate $F(x)$.
(c) What is the probability that a random particle from the manufactured fuel exceeds4 micrometers?
solution:- a) $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx$
$= \int_1^{\infty} 3x^{-4}dx = \frac{3x^{-3}}{-3}\big]_1^{\infty}$
$= -(\infty^{-3} - 1)$
$= 1$
Hence this is a valid valid density function for $x \geq 1$.
b)$F(x) = \int_{-\infty}^{x} f(t)dt$
$= \int_{-\infty}^{1} 0dt + \int_1^x 3t^{-4}dt$
$= 0 + \frac{3t^{-3}}{-3}\big]_1^x$
$= 1 - x^{-3}$, **As**

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-3}, & x \leq 1. \end{cases}$$

c) $P(X > 4) = 1 - F(4) = 1 - (1 - 4^{-3}) = 4^{-3} = 0.0156.$

Problem 3.10. 30). Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k\left(3 - x^2\right), & -1 \leq x \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.
(b) Find the probability that a random error in measurement is less than 1/2.

(c) For this particular measurement, it is undesirable if the magnitude of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

solutions:- a)$f(x) \geq 0$

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

$\Rightarrow k\int_{-1}^{1}(3-x^2)dx = 1$

$\Rightarrow 2k\int_{0}^{1}(3-x^2)dx = 1$

$\Rightarrow 2k\big[3x - \frac{x^3}{3}\big]_0^1 = 1$

$\Rightarrow \frac{16}{3}k = 1$

$\Rightarrow k = \frac{3}{16}$.

b) for$-1 \leq x \leq 1$

$F(x) = \int_{-1}^{x} \frac{3}{16}(3-t^2)dt = \frac{3}{16}\big[3t - \frac{t^3}{3}\big]_{-1}^{x}$

$= \frac{3}{16}\big[3(x+1) - \frac{1}{3}(x^3+1)\big]$

$= \frac{3}{16}\big(3x + 3 - \frac{x^3}{3} - \frac{1}{3}\big)$

$= \frac{9}{16}x - \frac{x^3}{16} + \frac{1}{2}$

$P(X < \frac{1}{2}) = F(\frac{1}{2}) = \frac{1}{2} + \frac{9}{16}\frac{1}{2} - \frac{1}{16}\frac{1}{8}$

$= 0.773.$

c)$P(|\ X\ | > 0.8) = P(X < -0.8) + P(X > 0.8)$

$= F(-0.8) + 1 - F(0.8)$

$= \frac{1}{2} + \frac{9}{16} * (-0.8) - \frac{1}{16}(-0.8)^3 + 1 - \frac{1}{2} + \frac{9}{16} * 0.8 - \frac{1}{16}(0.8)^3$

$= 0.164.$

Problem 3.11. 35).Suppose it is known from large amounts of historical data that X, the number of cars that arrive at a specific intersection during a 20 -second time period, is characterized by the following discrete probability function:

$$f(x) = e^{-6}\frac{6^x}{x!}, \text{ for } x = 0, 1, 2, \ldots$$

(a) Find the probability that in a specific 20 -second time period, more than 8 cars arrive at the intersection.

(b) Find the probability that only 2 cars arrive.

solutions:- a)$P(X > 8) = 1 - P(X \leq 8)$

$= 1 - \sum_{x=0}^{8} e^{-6}\frac{6^x}{x!}$

$= 1 - \big(e^{-6}\frac{6^0}{0!} + e^{-6}\frac{e^1}{1!} + \cdots + e^{-6}\frac{6^8}{8!}\big)$

$= 0.1528.$

b) $P(X = 2) = f(2) = e^{-6}\frac{6^2}{2!} = 0.0446.$

LECTURE-8

4 Joint Probability Distributions

Definition 4.1. The function $f(x, y)$ is a joint probability distribution or probability mass function of the discrete random variable X and Y if
1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$
for any region A in the xy-plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Definition 4.2. The function $f(x, y)$ is a joint density function of the continuous random variables X and Y if
1. $f(x, y) \geq 0$, for all (x, y)
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$
3. $P[(X, Y) \in A] = \iint_A f(x, y)dxdy$, for any region A in the xy plane.

Definition 4.3. The marginal distributions of X alone and of Y alone are

$$g(x) = \sum_y f(x, y) \quad \text{and} \quad h(y) = \sum_x f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

for the continuous case.

Definition 4.4. Let X and Y be two random variables, discrete or continuous. The conditional distribution of the random variable Y given that $X = x$ is

$$f(y \mid x) = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x \mid y) = \frac{f(x, y)}{h(y)}, \text{ provided } h(y) > 0.$$

Definition 4.5. Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Problem 4.6. 38. If the joint probability distri is given by

$$f(x,y) = \frac{x+y}{30}, \quad \text{for } x = 0,1,2,3$$

find

(a) $P(X \leq 2, Y = 1)$

(b) $P(X > 2, Y \leq 1)$

(c) $P(X > Y)$

(d) $P(X + Y = 4)$

Solution:- a) $P(X \leq 2, Y = 1) = f(0,1) + f(1,1) + f(2,1)$

$= \frac{1}{3} + \frac{2}{30} + \frac{3}{30} = \frac{1}{5}$

b)$4P(X > 2, Y \leq 1) = f(3,0) + f(3,1)$

$= \frac{3}{30} + \frac{4}{30} = \frac{7}{30}$

c) $P(X > Y) = f(1,0) + f(2,0) + f(2,1) + f(3,0) + f(3,1) + f(3,2)$

$= \frac{1}{30} + \frac{2}{30} + \frac{3}{30} + \frac{3}{30} + \frac{4}{30} + \frac{5}{30}$

$= \frac{9}{15}$

d)$P(X + Y = 4) = f(2,2) + f(3,1) = \frac{4}{30} + \frac{4}{30} = \frac{4}{15}$

Problem 4.7. 42. Let X and Y denote the lengths of life, in years, of two components in an electronic system. If the joint density function of these variables is

$$f(x,y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Solution:- $f(x \mid y) = \frac{f(x,y)}{h(y)}$, provided $h(y) > 0$

$Now \;\; h(y) = \int_{-\infty}^{\infty} f(x,y)dx$

$= \int_0^{\infty} e^{-(x+y)}dx$

$= e^{-y} \int_0^{\infty} e^{-x}dx = e^{-y}\big[-e^{-x}\big]_0^{\infty}$

$= -e^{-y}(e^{-\infty} - e^0)$

$= e^y > 0, \text{ for all } y \in \mathbb{R}$

$f(x \mid y)$

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

$P(0 < Y < 1 \mid Y = 2) = \int_0^1 e^{-x}dx$

$= -e^{-x}\big]_0^1$

$= -(e^{-1} - 1)$

$= 1 - \frac{1}{e}.$

LECTURE-9

Problem 4.8. 49). Let X denote the number of times a certain numerical control machine will malfunction: $1, 2$, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

		x		
$f(x,y)$	1	2	3	
	1	0.05	0.05	0.10
	3	0.05	0.10	0.35
	5	0.00	0.20	0.10

(a) Evaluate the marginal distribution of X.
(b) Evaluate the marginal distribution of Y.
(c) Find $P(Y = 3 \mid X = 2)$
Solution:- a).
We have $g(x) = \sum_y f(x, y)$
where $g(1) = \sum_y f(1, y) = f(1, 1) + f(1, 3) + f(1, 5)$
$= 0.05 + 0.05 + 0.00 = 0.1$
$g(2) = \sum_y f(2, y) = f(2, 1) + f(2, 3) + f(2, 5)$
$= 0.05 + 0.10 + 0.20 = 0.35$
$g(3) = \sum_y f(3, y) = f(3, 1) + f(3, 3) + f(3, 5)$
$= 0.10 + 0.35 + 0.10 = 0.55$
Here these are the margional distributions of X.
b). we have $h(y) = \sum_y f(x, y)$
$h(1) = f(1, 1) + f(2, 1) + f(3, 1) = 0.05 + 0.05 + 0.10 = 0.20$
$h(2) = f(1, 3) + f(2, 3) + f(3, 3) = 0.05 + 0.10 + 0.35 = 0.50$
$h(5) = f(1, 5) + f(2, 5) + f(3, 5) = 000 + 0.20 + 0.10 = 0.30$
Here these are the marginonal distribution of Y
c)we have $f(y \mid x) = \frac{f(x,y)}{g(x)}$, provided $g(x) > 0$
'Hence $P(Y = 3 \mid X = 2) = \frac{f(2,3)}{g(2)} = \frac{0.10}{0.35} = \frac{10}{35} = \frac{2}{7.}$

Problem 4.9. 56). The joint density function of the random variables X and Y

is

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1, 0 < y < 1 - x \\ 0, & \text{elsewhere} \end{cases}$$

(a) Show that X and Y are not independent.

(b) Find $P(X > 0.3 \mid Y = 0.5)$.

Solutions:- a) we have the X and Y are independent random variable if $f(x,y) = g(x)h(y)$ otherwise not independent

Now $g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{1-x} 6xdy = 6x\big[y\big]_0^{1-x} = \begin{cases} 6x(1-x),\ 0 < x < 1 \\ 0,\ elsewhere. \end{cases}$

$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 6xdx = 6\frac{x^2}{2}\big]_0^1 = \begin{cases} 3,\ 0 < y < 1 - x \\ 0,\ elsewhere. \end{cases}$

as $g(x)h(y) = 6x(1-x) * 3 = 18x(1-x) \neq f(x,y)$

Hence X and Y are not independent.

b)we have $f(x \mid y) = \frac{f(x,y)}{h(y)},\ provided\ h(y) > 0$

$= \frac{6x}{3} = 2x,$

$P(X > 0.3 \mid Y = 0.5) = \int_{0.3}^1 2xdx = 2\frac{x^2}{2}\big]_{0.3}^1 = 1 - 0.09 = 0.91$

LECTURE - 10

CHEPTER-4

4.1 Mean of random variable

If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4)+(1)(7)+(2)(5)}{16} = 1.06$$

This can be written as

$$(0)(\tfrac{4}{16}) + (1)((\tfrac{7}{16}) + (2)((\tfrac{5}{16}) = 1.06$$

Here $\frac{4}{16}, \frac{7}{16}, and \frac{5}{16}$ are probabilities of getting 0 head, one head, two heads in tossing of two coin respectively . This average value is the mean of the random variable X or the mean ofthe probability distribution of X and write it as μ_x or simply as μ.

It is also common among statisticians to refer to this mean as the mathematical expectation, or the expected value of the random variable X, and denote it as E(X).

Definition 4.1

Mathematical Expectation E(X)

Let X be a random variable with probability distribution f(x). The mean, or expected value of X is

$$\mu = \text{E(X)} = \sum_x xf(x)$$

if X is discrete

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

if X is continuous.

Example-4.1

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Solution : Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \text{ x=0,1,2,3}$$

So $f(0) = \frac{1}{35}, f(1) = \frac{12}{35}, f(2) = \frac{18}{35} and\ f(3) = \frac{4}{35}$

Therefore $E(X) = (0)(\frac{1}{35}) + (1)(\frac{12}{35}) + (2)(\frac{18}{35}) + (3)(\frac{4}{35}) = \frac{12}{7}$

Exercise-4.4
A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

Solution: Let X denotes the number of tails. So X takes the values 0,1,2. Here a head is three times as likely to occur as a tail.
So p(H)=$\frac{3}{4}$ and p(T)=$\frac{1}{4}$.

Now f(0)=$\frac{9}{16}$, f(1)=$\frac{1}{16}$ and f(2)=$\frac{1}{16}$.

Therefore E(X)=(0)($\frac{9}{16}$)+(1)($\frac{1}{16}$)+(2)($\frac{1}{16}$)=$\frac{3}{16}$

Exercise-4.7
By investing in a particular stock, a person can make a profit in one year of $4000 with probability 0.3 or take a loss of $1000 with probability 0.7. What is this person's expected gain?
solution:
Let the profit variable is X

The person's expected gain
$E(X) = \sum_x xf(x) = (4000)(0.3) + (1000)(0.7) = \$1900$

Theorem-4.1
Let X be a random variable with probability distribution f(x). The expected value of the random variable g(X) is

$$\mu_g(X) = \text{E[g(X)]} = \sum_x g(x)f(x)$$

if X is discrete, and

$$\mu_g(X) = \text{E[g(X)]} = \int_{-\infty} \infty g(x)f(x)dx$$

if X is continuous.

Example-4.4
Suppose that the number of cars X that pass through a car wash between

4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

X	4	5	6	7	8	9
P(X=x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let g(X)=2X−1 represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution : The attendant can expect to receive
E(g(X))=E(2X-1)=$\sum_4^9(2x-1)f(x)$

$=(7)(\frac{1}{12}+(9)(\frac{1}{12})+(11)(\frac{1}{4})+(13)(\frac{1}{4})+(15)(\frac{1}{6})+(17)(\frac{1}{6})$

$=\$12.67$

Exercise-4.12

If a dealer's profit, in units of \$5000, on a newautomobile can be looked upon as a random variable X having the density function

$$f(x)=\begin{cases}2(1-x) & 0\le x\le 1\\ 0 & elsewhere\end{cases}$$

Find the average profit per automobile.

Solution : E(X)=$\int_0^1 xf(x)dx=x^2-\frac{2x^3}{3}\Big|_0^1=\frac{1}{3}$
The average profit per automobile($\frac{1}{3}$)(5000)=\$ $\frac{5000}{3}$

Definition 4.2

Let X and Y be random variables with joint probability distribution f(x, y). The mean, or expected value, of the random variable g(X, Y) is

$$\mu_{g(X,Y)}=E[g(X,Y)]=\sum_x\sum_y g(x,y)f(x,y)$$

if X and Y are discrete and

$$\mu_{g(X,Y)}=E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)dxdy$$

if X and Y are continuous

Exercise-4.10

Two tire-quality experts examine stacks of tires and assign a quality rating to each tire on a 3-point scale. Let X denote the rating given by expert A and Y denote the rating given by B. The following table gives the joint distribution for X and Y .

		y		
	f(x,y)	1	2	3
	1	0.10	0.05	0.02
x	2	0.10	0.35	0.05
	3	0.10	0.35	0.05

Solution :

		y			row total
	f(x,y)	1	2	3	g(x)
	1	0.10	0.05	0.02	0.17
x	2	0.10	0.35	0.05	0.50
	3	0.03	0.10	0.20	0.33
column total	h(y)	0.23	0.50	0.27	1

$\mu_X = \sum_x xg(x) = (1)(0.17) + (2)(0.50) + (3)(0.33) = 2.16$
$\mu_Y = \sum_y yh(y) = (1)(0.23) + (2)(0.50) + (3)(0.27) = 2.04$

Exercise-4.20
A continuous random variable X has the density function

$$\text{f(x)}=\begin{cases} e^{-x} & x > 1 \\ 0 & elsewhere \end{cases}$$

Find the expected value of g(X) = $e^{\frac{2X}{3}}$
Solution :
E(g(X))=$\int_1^\infty g(x)f(x)dx$

$= \int_1^\infty (e^{\frac{2x}{3}})(e^{-x})dx = \int_1^\infty e^{\frac{-x}{3}}dx = -3(e^{\frac{-x}{3}})\Big|_1^\infty$

$= 3e^{\frac{-1}{3}}$

Exercise-4.23
Suppose that X and Y have the following joint probability function:

		x	
	f(x,y)	2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

(a) Find the expected value of g(X, Y) = XY^2.
(b) Find μ_X and μ_Y.
Solution :

		x		Row total h(y)
	f(x,y)	2	4	
	1	0.10	0.15	0.25
y	3	0.20	0.30	050
	5	0.10	0.15	0.25
Column total	g(x)	0.40	0.60	1

(a) E(g(X,Y))=$\sum_x \sum_y g(x,y)f(x,y) = \sum_x \sum_y xy^2 f(x,y)$

=(2)(1)(0.10)+(4)(1)(0.15)+(2)(9)(.20)+(4)(9)(0.30)+(2)(25)(0.10)

+(4)(25)(0.15)

=0.20+0.60+3.60+10.80+5.00+15.00=35.20

(b)

($\mu_X = \sum_x xg(x) = (2)(0.40) + (4)(0.60) = 3.20$

$\mu_Y = \sum_y yh(y) = (1)(0.25) + (3)(0.50) + (5)(0.25) = 3$

*****Completed*****

LECTURE - 11

CHEPTER-4

4.2 Variance and Covariance of random variables

The most important measure of variability of a random variable X is the variance of the random variable X or the variance of the probability distribution of X and is denoted by Var(X) or the symbol σ_X^2, or simply by σ^2

Definition 4.3

Let X be a random variable with probability distribution f(x) and meanμ. The variance of X is

$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$, if X is discrete, and

$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$, if X is continuous.

The positive square root of the variance σ is called the standard deviation of X.

Theorem-4.2:

The variance of a random variable X is $\sigma^2 = E(X^2) - \mu^2$.

Theorem-4.3:

Let X be a random variable with probability distribution f(x). The variance of the random variable g(X) is

$\sigma^2_{g(X)} = E[g(X) - \mu_{g(X)}]^2 = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$

if X is discrete, and

$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)dx$

if X is continuous.

Definition 4.4

Let X and Y be random variables with joint probability distribution f(x, y). The covariance of X and Y is

$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f(x,y)$

if X and Y are discrete, and

$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty} \infty \int_{-\infty} \infty (x-\mu_x)(y-\mu_y) f(x,y)dxdy$

if X and Y are continuous

Theorem-4.4:

The covariance of two random variables X and Y with means μ_X and μ_Y , respectively, is given by

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$.

Example-4.9
Let the random variable X represents the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

Calculate σ^2.
Solution :
$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$

$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$
Therefore,
$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$

Example-4.10
The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

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

Find the mean and variance of X.
Solution :
Calculating E(X) and$E(X^2)$,
we have $\mu = E(X) = 2\int_1^2 x(x-1)dx = \frac{5}{3}$
and
$E(X^2) = 2\int_1^2 x^2(x-1)dx = \frac{17}{6}$
$Therefore$,
$\sigma^2 = \frac{17}{6} - (\frac{5}{3})^2 = \frac{1}{18}.$

Exercise-4.34
Let X be a random variable with the following probability distribution:

x	-2	3	5
f(x)	0.3	0.2	0.5

Find the standard deviation of X.
Solution :

Calculating E(X) and$E(X^2)$,
we have
$\mu = (-2)(0.3) + (3)(0.2) + (5)(0.5) = 2.5$

$E(X^2) = (4)(0.3) + (9)(0.2) + (25)(0.5) = 15.5$
$Therefore$,
$\sigma^2 = 15.5 - 2.5^2 = 9.25.$
The standard deviation $\sigma = \sqrt{9.25} = 3.04138.$

Exercise-4.35
The random variable X, representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
f(x)	0.01	0.25	0.4	0.3	0.04

Solution :
E(X)=$\sum_x xf(x) = 2(0.01) + 3(0.25) + 4(0.4) + 5(0.3) + 6(0.04) = 4.11$

$E(X^2) = \sum_x x^2 f(x) = 4(0.01)+9(0.25)+16(0.4)+25(0.3)+36(0.04) = 17.63$

$\sigma^2 = E(X^2) - (E(X))^2 = 17.63 - (4.11)^2 = 0.7379$

Exercise-4.50
For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

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

Find the variance and standard deviation of X.
Solution :
Calculating E(X) and$E(X^2)$,
we have $\mu = E(X) = 2\int_0^1 x(1-x)dx = \frac{1}{3}$
and
$E(X^2) = 2\int_0^1 x^2(1-x)dx = \frac{1}{6}$
$Therefore$,
$\sigma^2 = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$ and $\sigma = \sqrt{\frac{1}{18}} = 0.2357$

*****Completed*****

LECTURE - 12 and 13

CHEPTER-4

4.3 Means and Variances of Linear Combinations of Random Variables

Theorem-4.5
If a and b are constants, then E(aX + b) = aE(X) + b.
Substituting a=0 ,we get E(b)=b and b=0 we get E(aX)=aE(X)

Example-4.17
Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

X	4	5	6	7	8	9
P(X=x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let g(X)=2X−1 represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.
Solution :
we can write E(2X-1) = 2E(X)-1.
Now
$\mu = E(X) = \sum_{x=4}^{9} xf(x)$

$$=(4)(\tfrac{1}{12}) + (5)(\tfrac{1}{12}) + (6)(\tfrac{1}{4}) + (7)(\tfrac{1}{4}) + (8)(\tfrac{1}{6}) + (9)(\tfrac{1}{6}) = \tfrac{41}{6}.$$

Therefore,
$\mu_{2X-1} = (2)(\frac{41}{6}) - 1 = \$12.67$

Theorem-4.6
The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,
E[g(X) ± h(X)] = E[g(X)] ± E[h(X)].

Exercise-4.57
Let X be a random variable with the following probability distribution:

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

Find E(X) and $E(X^2)$ and then, using these values, evaluate$E[(2X+1)^2]$.
Solution :
$E(X) = (-3)(\frac{1}{6}) + (6)(\frac{1}{2}) + (9)(\frac{1}{3}) = 5.5$
$E(X^2) = (9)(\frac{1}{6}) + (36)(\frac{1}{2}) + (81)(\frac{1}{3}) = 46.5$
$E[(2X+1)^2] = E(4(X^2) + 4(X) + 1) = 4E(X^2) + 4E(X) + 1$
$= 4(46.5) + 4(5.5) + 1 = 209$

Theorem-4.7
The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,
E[g(X, Y) ± h(X, Y)] = E[g(X, Y)] ± E[h(X, Y)].

Theorem-4.8
Let X and Y be two independent random variables.Then E(XY) = E(X)E(Y).
Let X and Y be two independent random variables. Then $\sigma_{XY} = 0$.

Theorem-4.9
If X and Y are random variables with joint probability distribution f(x, y) and a, b, and c are constants, then $\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}$
Setting b = 0, we see that $\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2$
Setting a = 1 and b = 0, we see that $\sigma^2_{X+c} = \sigma^2_X = \sigma^2$.
Setting b = 0 and c = 0, we see that $\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2$.
If X and Y are independent random variables, then $\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$ and $\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y$

Example-4.22
If X and Y are random variables with variances $\sigma^2_X = 2\text{and}\sigma^2_Y = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable Z = 3X-4Y+ 8.
Solution :
$\sigma^2_Z = \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y}$

$= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma XY = (9)(2) + (16)(4) - (24)(-2) = 130$

Example-4.23
Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 3$. Find the variance of the random variable Z = 3X-2Y+ 5.

Solution :
$\sigma_Z^2 = \sigma_{3X-2Y+5}^2 = \sigma_{3X-2Y}^2 = 9\sigma_X^2 + 4\sigma_Y = (9)(2) + (4)(3) = 30$

Exercise-4.58
The total time, measured in units of 100 hours, that a teenager runs her hair dryer over a period of one year is a continuous random variable X that has the density function

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

Evaluate the mean of the random variable Y = $60X^2$ + 39X, where Y is equal to the number of kilowatt hours expended annually.
Solution:
E(Y)=E($60X^2$ + 39X)=60E(X^2)+39E(X)
$E(X) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 (x)(x)dx + \int_1^2 (x)(2-x)dx$

$=\frac{x^3}{3}\Big|_0^1 + (2\frac{x^2}{2} - \frac{x^3}{3})\Big|_1^2 = 1$
$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 (x^2)(x)dx + \int_1^2 (x^2)(2-x)dx$

$=\frac{x^4}{4}\Big|_0^1 + (2\frac{x^3}{3} - \frac{x^4}{4})\Big|_1^2 = \frac{7}{6}$

$E(Y) = (60)(\frac{7}{6}) + (39)(1) = 109$
Total time=(109)(100)=10900 hours.

Exercise-4.60
Suppose that X and Y are independent random variables having the joint probability distribution

		x	
	f(x,y)	2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

Find (a) E(2X-3Y); (b) E(XY).
Solution:

	f(x,y)	x 2	x 4	Row total h(y)
	1	0.10	0.15	0.25
y	3	0.20	0.30	0.50
	5	0.10	0.15	0.25
	Column total g(x)	0.40	0.60	1

E(2X-3Y)=2E(X)-3E(Y)=$2\sum_x xg(x) - 3\sum_y yh(y)$

$=2((2)(0.40) + (4)(0.60)) - 3((1)(0.25) + (3)(0.5) + (5)(0.25))$

$= 6.40 + 9 = 6.40 - 9 = -2.60$

$E(XY) = E(X)E(Y) = (\sum_x xg(x))(\sum_y yh(y)) = (3.20)(3) = 9.60$

4.4 Chebyshev's Theorem
The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1 - 1/k^2$. That is,
$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$.

Example-4.23
A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find (a)P$(-4 < X < 20)$ (b) P$(|\ (X - 8)\ | \geq 6)$.
Solution:
(a)$P(-4 < X < 20)$

$=P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}$.

(b) $P(|\ (X - 8)\ | \geq 6)$

$=1 - P(|\ (X - 8)\ | < 6) = 1 - P(-6 < X - 8 < 6)$

$=1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}$

Exercise-4.75
An electrical firm manufactures a 100-watt light bulb, which, according to specifications written on the package, has a mean life of 900 hours with a standard deviation of 50 hours. At most, what percentage of the bulbs fail to last even 700 hours? Assume that the distribution is symmetric about the mean.
Solution:

X is the random variable define life of the 100-watt bulb
Here μ=900 hours and $\sigma = 50$
To find the probability of$P(X \leq 700)$.
Given that the distribution is symmetric about the mean. According to Chebyshev's theorem
$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$.

$P(X \leq 700) = (0.5)(P(\mid X - 900 \mid \geq 200)) = (0.5)(1 - P(\mid X - 900 \mid \leq 200))$

$=(0.5)(1 - P(900 - (4)(50) < X < 900 + (4)(50))) \leq (0.5)(\frac{1}{4^2}) = 0.03215$

Therefore the percentage of the bulbs fail to last even 700 hours is 3.215%.

Exercise-4.77
A random variable X has a mean$\mu = 10$ and a variance$\sigma^2 = 4$. Using Chebyshev's theorem, find
$(a)P(\mid X - 10 \mid \geq 3)$;
$(b)P((\mid X - 10 \mid < 3)$;
$(c)P(5 < X < 15)$;
(d) the value of the constant c such that $P(\mid X - 10 \mid \geq c) \leq 0.04$.
Solution:
(a)$P(\mid X - 10 \mid \geq 3$

$= 1 - P(\mid (X - 10) \mid < 3) = 1 - P(-3 < X - 10 < 3)$

$=1 - P[10 - (2)(\frac{3}{2}) < X < 10 + (2)(\frac{3}{2})] \leq \frac{4}{9}$

(b) $P((\mid X - 10 \mid < 3)$

$=P(-3 < X - 10 < 3) = P[10 - (2)(\frac{3}{2}) < X < 10 + (2)(\frac{3}{2})] \geq 1 - \frac{1}{(\frac{3}{2})^2} = \frac{5}{9}$

(c)$P(5 < X < 15)$

$=P[10 - (2)(\frac{5}{2}) < X < 10 + (2)(\frac{5}{2})] \geq 1 - \frac{1}{(\frac{5}{2})^2} = \frac{21}{25}$

(d)$P(\mid X - 10 \mid \geq c)$

$=1 - P(\mid X - 10 \mid \leq c) = 1 - P(10 - c < X < 10 + c)$

$= 1 - P(10 - 2(\frac{c}{2}) < X < 10 + 2(\frac{c}{2})) \leq \frac{1}{(\frac{c}{2})^2} = \frac{4}{c^2}$

Given that $P(\mid X-10 \mid \geq c) \leq 0.04$.
Therefore $\frac{4}{c^2} = 0.04 \implies c^2 = 100 \implies c = 10$

Example-4.78
Compute$P(\mu - 2\sigma < X < \mu + 2\sigma)$, where X has the density function

$$\text{f(x)}=\begin{cases} 6x(1-x) & 0 < x < 1 \\ 0 & elsewhere \end{cases}$$

Solution:
E(X)=$\int_0^1 (x)(6x(1-x))dx = \int_0^1 (6x^2 - 6x^3)dx = \frac{1}{2}$

$E(X^2) = \int_0^1 (x^2)(6x(1-x))dx = \int_0^1 (6x^3 - 6x^4)dx = \frac{3}{10}$

$\sigma^2 = E(X^2) - (E(X))^2 = \frac{3}{10} - (\frac{1}{2})^2 = \frac{1}{20} \implies \sigma = \sqrt{(\frac{1}{20})} = 0.223$

$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(0.053 < X < 0.9472) = \int_{0.053}^{0.947} 6x(1-x)dx =$
$\int_{0.053}^{0.947} (6x - 6x^2)dx = 6(\frac{x^2}{2} - \frac{x^3}{3})\Big|_{0.053}^{0.9472} = 0.9838$

By Chebyshev's theorem,
$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{2^2} = 0.75$
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9838 \geq 0.75$
Hence Chebyshev's theorem is verified.

*****Completed*****

Lecture-14

Binomial Distribution

Bernoulli trails:

A Series of trails that satisfies the following assumptions is known as Bernoulli trail.

1. There are only two possible outcomes for each trail (success and failure)
2. The probability of success is same for each trail
3. The outcomes from different trails are independent

Example-1:

Tossing a Coin 100 times is a Bernoulli trail.

There are only 2 outcomes, Head and Tail. Here getting Head can be considered as success and Tail as failure. Probability of getting Head or success remains same though out the process and the events are independent.

Binomial distribution

Consider a Bernoulli trail which results in success with probability p and a failure with probability $q = 1 - p$.

Let X: the number of success in n trails.

Then the probability distribution of the binomial random variable X is

$$P(X = x) = f(x) = b(x; n, p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x}, \qquad x = 0, 1, 2, ...n$$

Example-2:

If Probability of hitting the target is $\frac{3}{4}$ and three shots are fired, then

(i)Find the probability of hitting the target 2 times.

(ii)Formulate the binomial Probability distribution function

Ans:

Total no. of trails n=3.

Let X = no. of times hitting the target (no. of success), $x = 0, 1, 2, 3$

P (success) = P (hitting the target) =$\frac{3}{4}$

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

Therefore

(i) $P(X = 2) = f(2) = b\left(2; 3, \frac{3}{4}\right) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^{3-2}$

(ii) $P(X = x) = f(x) = b\left(x; 3, \frac{3}{4}\right) = \begin{pmatrix} 3 \\ x \end{pmatrix} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{3-x}, x = 0, 1, 2, .3$

Note: The mean and variance of binomial distribution $\boldsymbol{f(x) = b\,(x; n, p)}$ are $\boldsymbol{\mu = np}$ and $\boldsymbol{\sigma^2 = npq}$

(Q.11) The probability that a patient recovers from a delicate heart operation is 0.9. What is the probability that exactly 5 of the next 7 patients having this operation survive?

Ans:

Let X = no. of patients recovered from the heart operation i.e. $x = 0, 1, 2, ...7$

Here, $n = 7$, $p = 0.9$, $q = 0.1$

hence,

$$P(X = 5) = f(5) = b\,(5; 7, 0.9) = \begin{pmatrix} 7 \\ 5 \end{pmatrix} (0.9)^5 (0.1)^{7-5}$$

$$= 0.1240$$

Binomial distribution Table:

The cumulative distribution for the binomial distribution is pre-calculated and given in the form of a table.

Examples-3:(Use of binomial distribution Table)

We know

$$\begin{aligned} P(X \leq 4) &= F(4) \\ &= P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) \end{aligned}$$

Suppose, $n = 5$, $p = 0.6$

Hence,

$$\begin{aligned} P(X \leq 4) &= F(4) \\ &= \binom{5}{0} (0.6)^0 (0.4)^{5-0} + \binom{5}{1} (0.6)^1 (0.4)^{5-1} + ... + \binom{5}{4} (0.6)^4 (0.4)^{5-4} \\ &= \sum_{x=0}^{4} b(x; 5, 0.6) = 0.9222 \end{aligned}$$

$P(X \leq 4) = B(4; 5, 0.6) = 0.9222$ (from binomial distribution table)

In general $P(X \leq x) = B(x; n, p)$ and $b(x; n, p) = B(x; n, p) - B(x-1; n, p)$

(Q.15) It is known that 60% of mice inoculated with a serum are protected from a certain disease. If 5 mice are inoculated, find the probability that (a) none contracts the disease; (b) fewer than 2 contract the disease; (c) more than 3 contract the disease.

Ans:

Let X = no. of mice from the disease after inoculated, $x = 0, 1, 2, 3, 4, 5$

Here, $n = 5$, $p = 0.4$, $q = 0.6$

(i) $P(X = 0) = f(0) = \binom{5}{0} (0.4)^0 (0.6)^5 = 0.0778$

(ii) $P(X < 2) = P(X \leq 1)$

$$= P(X = 0) + P(X = 1)$$
$$= \begin{pmatrix} 5 \\ 0 \end{pmatrix} (0.4)^0 (0.6)^5 + \begin{pmatrix} 5 \\ 1 \end{pmatrix} (0.4)^1 (0.6)^{5-1} = 0.3370$$

(iii) $P(X > 3) = P(X = 4) + P(X = 5)$

$$= \begin{pmatrix} 5 \\ 4 \end{pmatrix} (0.4)^4 (0.6)^{5-4} + \begin{pmatrix} 5 \\ 5 \end{pmatrix} (0.4)^5 (0.6)^{5-5} = 0.087$$

(Q.16) Suppose that airplane engines operate independently and fail with probability equal to 0.4. Assuming that a plane makes a safe fight if at least one-half of its engines run, determine whether a 4-engine plane or a 2 engine plane has the higher probability for a successful fight.

Ans:

Case-1

The plane has 4 engines; $n = 4$

The plane will make a safe flight if 2 or more engines are working.

$$P(X \geq 2) = 1 - P(X \leq 1)$$
$$= 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - \left[\begin{pmatrix} 4 \\ 0 \end{pmatrix} (0.6)^0 (0.4)^{4-0} + \begin{pmatrix} 4 \\ 1 \end{pmatrix} (0.6)^1 (0.4)^{4-1} \right] = 0.8208$$

Case-2

The plane has 2 engines; $n = 2$

The plane will make a safe flight if 1 or more engines are working.

$$P(X \geq 1) = 1 - P(X = 0)$$
$$= 1 - \left[\begin{pmatrix} 2 \\ 0 \end{pmatrix} (0.6)^0 (0.4)^{2-0} \right] = 0.84$$

Conclusion: comparing the above two cases, a 2 engine flight has higher probability for a successful flight.

Lecture-15

Multinomial Distribution

Multinomial distribution:

Consider a trial which results k outcomes, E_1, E_2, ...E_k with probabilitites p_1, p_2, ..., p_k respectively such that

$$\sum_{i=1}^{k} p_i = 1$$

Let

X_1 = no. of times E_1 occurs in n independent trials

X_2 = no. of times E_2 occurs in n independent trials

.....

X_k = no. of times E_k occurs in n independent trials

Now,

$$P(X_1 = x_1, X_2 = x_2, ...X_k = x_k) = f(x_1, x_2, ...x_k; p_1, p_2, ...p_k)$$
$$= \begin{pmatrix} n \\ x_1, x_2, ...x_k \end{pmatrix} {p_1}^{x_1} {p_2}^{x_2} .....{p_k}^{x_k}$$

With $\sum_{i=1}^{n} x_i = n$, $\sum_{i=1}^{n} p_i = 1$

(Q.19) As a student drives to school, he encounters a trafic signal. This trafic signal stays green for 35 seconds, yellow for 5 seconds, and red for 60 seconds. Assume that the student goes to school each weekday between 8:00 and 8:30 a.m. Let X_1 be the number of times he encounters a green light, X_2 be the number of times he encounters a yellow light, and X_3 be the number of times he encounters a red light. Find the joint distribution of X_1, X_2, and X_3

Ans:

Let

X_1 = no. of times he encounters a green light

X_2 = no. of times he encounters a yellow light

X_3 = no. of times he encounters a red light

Given $p_1 = 0.35, p_2 = 0.05, p_3 = 0.60$

Therefore,

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = f(x_1, x_2, .x_3; n, 0.35, 0.05, 0.60)$$
$$= \begin{pmatrix} n \\ x_1, x_2, ...x_k \end{pmatrix} (0.35)^{x_1} (0.05)^{x_2} (0.60)^{x_3}$$

Where $x_1 + x_2 + x_3 = n$

(Q.22) According to a genetics theory, a certain cross of guinea pigs will result in red, black, and white ofspring in the ratio 8:4:4. Find the probability that among 8 ofspring, 5 will be red, 2 black, and 1 white.

Ans:

Let

X_1 = no. of red guinea pigs

X_2 = no. of black guinea pigs

X_3 = no. of white guinea pigs

It is given that the ratio of red, black, and white guinea pigs is 8:4:4

Hence,

P (guinea pig is red) = 8/16 =0.5

P (guinea pig is black) = 4/16 =0.25

P (guinea pig is white) = 4/16 =0.25

$$P(X_1 = 5, X_2 = 2, X_3 = 1) = f(5, 2, .1; 8, 0.5, 0.25, 0.25)$$
$$= \begin{pmatrix} 8 \\ 5, 2, 1 \end{pmatrix} (0.5)^5 (0.25)^2 (0.25)^1$$
$$= 21/256$$

Lecture-16

Hypergeometric Distribution

General discussion:

Suppose total number of items in a a bag = N

Total number of defective items (out of N) = k

Total number of items selected = n

lets discuss the probability that x out of n ($x \leq n$) items selected is defective.

Now, total number of ways n items can be selected out of N items = $\begin{pmatrix} N \\ n \end{pmatrix}$.

Our requirement is x out of n are defective i.e. remaining $(n-x)$ are non-defective.

Number of ways x defective items can be selected from k defective items = $\begin{pmatrix} k \\ x \end{pmatrix}$.

Number of ways $n-x$ defective items can be selected from $N-k$ defective items = $\begin{pmatrix} N-k \\ n-x \end{pmatrix}$.

Probability of selecting x defectives

$$= \frac{all\ favorable\ cases}{all\ possible\ cases}$$

$$= \frac{\begin{pmatrix} k \\ x \end{pmatrix}\begin{pmatrix} N-k \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

Definition:

Let X = The number of successes in a random sample size n selected from N items of which k are labeled success and $(N-k)$ labeled failure.Then probability distribution of

the above hypergeometric random variable is

$$f(x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Such that $\max\{0, n-(N-k)\} \leq x \leq \min(n, k)$

(Q.30) A random committee of size 3 is selected from 4 doctors and 2 nurses. Write a formula for the probability distribution of the random variable X representing the number of doctors on the committee. Find $P(2 \leq X \leq 3)$.

Ans:

There are 4 doctors and 2 nurses

3 persons will be selected out of 4+2=6 persons

Let X: number of doctors in the committee which consists of 3 persons

So $x = 1, 2, 3$ ($x \neq 0$ why?)

$$P(X = x) = f(x) = \frac{\binom{4}{x}\binom{2}{3-x}}{\binom{6}{3}}$$

Now,

$$P(2 \leq X \leq 3) = P(X = 2) + P(X = 3)$$

$$= \frac{\binom{4}{2}\binom{2}{3-2}}{\binom{6}{3}} + \frac{\binom{4}{3}\binom{2}{3-3}}{\binom{6}{3}} = \frac{4}{5}$$

(Q.32) From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that (a) all 4 will fire? (b)

at most 2 will not fire?

Ans:

Total number of missiles = 10

Total number of defective missiles = 3

Hence, total number of non-defective missiles = 7

4 missiles will be fired

Let X: number of non-defective missiles fired

(a) $P(X=0) = \dfrac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}} = \frac{1}{6}$

(b) At most 2 will not fire means 2 or more will fire

$$P(X \geq 2) = P(X=2) + P(X=3) + P(X=4)$$
$$= \frac{\binom{7}{2}\binom{3}{2}}{\binom{10}{4}} + \frac{\binom{7}{3}\binom{3}{1}}{\binom{10}{4}} + \frac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}} = \frac{29}{30}$$

Multivariate Hyper-geometric Distribution:

If N items can be partitioned into k cells $A_1, A_2, ...A_k$ with $a_1, a_2, ...a_k$elements, respectively then probability distribution of the random variables $X_1, X_2, ...X_k$representing the number of elements selected from $A_1, A_2, ...A_k$in a random sample of size n is

$$f(x_1, x_2, ...x_k; a_1, a_2, ...a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}.....\binom{a_k}{x_k}}{\binom{N}{n}}$$

With $\sum_{i=1}^{n} x_i = n,\ \sum_{i=1}^{n} a_i = N$

(Q.43) A foreign student club lists as its members 2 Canadians, 3 Japanese, 5 Italians, and 2 Germans. If a committee of 4 is selected at random, ?nd the probability that (a)

all nationalities are represented; (b) all nationalities except Italian are represented.

Ans:

Total number of members = 2+3+5+2 12

Total number of members selected = 4

(a) All nationalities represented means one from each country.

One person can be selected from 2 Canadians in $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ways

One person can be selected from 3 Japanies in $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ ways

One person can be selected from 5 Italians in $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$ ways

One person can be selected from 2 Germans in $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ways

Four persons can be selected from 12 persons in $\begin{pmatrix} 12 \\ 4 \end{pmatrix}$ ways

$$P(\textit{all nationalities are represented}) = \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} = \frac{4}{33}$$

(b) All nationalities except Italians are represented, then 3 cases arise;

Case-I: 2 Canadians + 1 Japanies+ 0 Italian + 1 German

Case-2: 1 Canadians + 2 Japanies+ 0 Italian + 1 German

Case-3: 1 Canadians + 1 Japanies+ 0 Italian + 2 German

$$P(\textit{all nationalities are represented}) = \frac{\begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} + \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} + \frac{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix}}{\begin{pmatrix} 12 \\ 4 \end{pmatrix}} = \frac{8}{165}$$

(Q.44) An urn contains 3 green balls, 2 blue balls, and 4 red balls. In a random sample of 5 balls, find the probability that both blue balls and at least 1 red ball are selected.

Ans:

Total number of balls = 3+2+4 = 9

Total number of balls selected = 5

Number of blue balls = 2, green balls = 3, red balls = 4

$$P(2\,blue\ balls\ and\ atleast\ 1\ red\ ball)$$
$$= \frac{\binom{2}{2}\binom{4}{1}\binom{3}{2}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{2}\binom{3}{1}}{\binom{9}{5}} + \frac{\binom{2}{2}\binom{4}{3}\binom{3}{0}}{\binom{9}{5}} = \frac{17}{63}$$

Negative Binomial Distribution:

Let repeated independent trials results in a success with probability p and a failure with probability $q = 1 - p$.

Where X: number of trials in which the kth success occurs,

Then,

$$P(X = x) = f(x) = b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}$$

Where $x = k, k+1, k+2....$

Geometric Distribution:

A particular case of negative binomial distribution for $k = 1$ is known as geometric distribution.

Here, X= the number of trials on which the first success occurs.

$$P(X = x) = f(x) = g(x; p) = pq^{x-1}; \quad x = 1, 2, 3.....$$

Where $q = 1 - p$

(Q.49) The probability that a person living in a certain city owns a dog is estimated to be 0.3. Find the probability that the tenth person randomly interviewed in that city is the fifth one to own a dog.

Ans:

Here, $p = 0.3, q = 1 - p = 0.7$

X= The number of persons interviewed in which kth person own a dog.

Given $x = 10, k = 5$

$$b^*(10; 5, 0.3) = \begin{pmatrix} 10-1 \\ 5-1 \end{pmatrix} p^5 q^{10-5}$$
$$= \begin{pmatrix} 9 \\ 4 \end{pmatrix} (0.3)^5 (0.7)^{10-5} = 0.0515$$

(Q.50) Find the probability that a person flipping a coin gets (a) the third head on the seventh flip; (b) the first head on the fourth flip.

Ans:

Here, $p = 0.5, q = 1 - p = 0.5$

X= The number of trials in which kth head occurs.

(a)Third head in 7^{th} flip means $x = 7, k = 3$

Hence,

$$b^*(7; 3, 0.5) = \begin{pmatrix} 7-1 \\ 3-1 \end{pmatrix} (0.5)^3 (0.5)^4 = 0.1172$$

(b)First head in the fourth flip means$x = 4, k = 1$

Using negative binomial or geometric distribution

$$g(x; p) = g(4, 0.5) = 0.5(1 - 0.5)^{4-1} = (0.5)^4$$

Lecture-17

Poisson Distribution

Poisson Distribution:

The probability distribution of the Poisson random variable X, representing the number of outcomes occurring a given time interval t is

$$P(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \qquad x = 0, 1, 2, ...$$

For $t = 1$

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, ...$$

The Poission random variable X, has the mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$.

Note: For the purpose of minimizing calculation, refer Poisson distribution table in the problems.

(Q.58) A certain area of the eastern United States is, on average, hit by 6 hurricanes a year. Find the probability that in a given year that area will be hit by (a) fewer than 4 hurricanes; (b) anywhere from 6 to 8 hurricanes.

Ans:

The average number of hurricane hits in a year is 6 i.e. $\lambda = 6$

Let X : The number of hurricane hits in a year

a.

P(fewer than 4 hurricanes) means

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!} + \frac{e^{-6}6^3}{3!} = 0.1512 \end{aligned}$$

b.

P(anywhere from 6 to 8 hurricanes) means

$$P(6 \leq X \leq 8) = P(X = 6) + P(X = 7) + P(X = 8)$$
$$= \frac{e^{-6}6^6}{6!} + \frac{e^{-6}6^7}{7!} + \frac{e^{-6}6^8}{8!} = 0.4015$$
$$Or$$
$$= F(8) - F(5) = 0.4015 \ (using\ table)$$

(Q.60) The average number of field mice per acre in a 5-acre wheat field is estimated to be 12. Find the probability that fewer than 7 field mice are found (a) on a given acre; (b) on 2 of the next 3 acres inspected.

Ans:

The average number of field mice per acre is 12 i.e. $\lambda = 12$

(a) Let X : The number of mice per acre

Hence,

$$P(X < 7) = P(X \leq 6) = 0.0458 \ \ (using\ table)$$

(b) Let Y : The number of acres of land inspected.

Here Y follows binomial distribution.

Given $n = 3, y = 2$

Hence, $P(Y = 2) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} p^2 q^{3-2}, \ \ with\ p = 0.0458$ (from part a), $q = 1 - p$

(Q.69) The probability that a person will die when he or she contracts a virus infection is 0.001. Of the next 4000 people infected, what is the mean number who will die?

Ans:

Given $p = 0.001, \ n = 4000$

Therefore, $\mu = np = 4000 \times 0.001 = 4$

Lecture-18

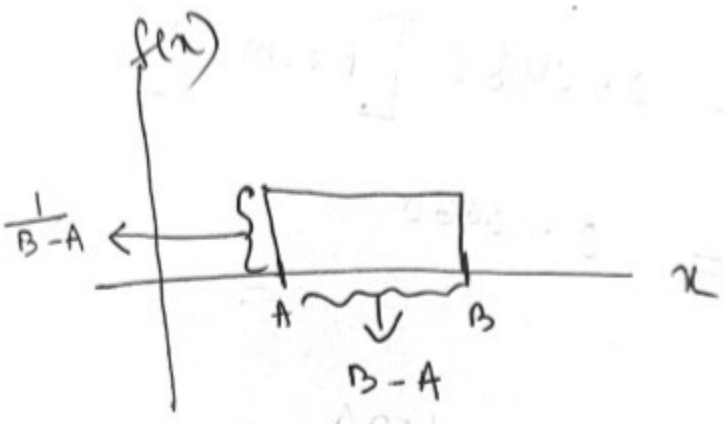
Uniform / Rectangular Distribution

Defination:

The probability density function of the continuous uniform random variable X on the interval $[A, B]$ is

$$f(x; A, B) = f(x) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & otherwise \end{cases}$$

Geometrical interpretation: The density function $f(x)$ forms a rectangle with base



$B - A$ and constant height $\frac{1}{B-A}$ as shown in the figure given above.

Example: The density function of the uniform distribution in the interval $[2, 5]$is

$$f(x) = \begin{cases} \frac{1}{5-2} = \frac{1}{3}, & 2 \leq x \leq 5 \\ 0, & otherwise \end{cases}$$

Theorem-1: Prove that the mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2}, \ \ \sigma^2 = \frac{(B-A)^2}{12}$$

Proof: Here,

mean $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_A^B x\frac{1}{B-A}dx = \frac{A+B}{2}$

Variance $\sigma^2(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} \left(x - \frac{A+B}{2}\right)^2 \frac{1}{B-A}dx = \frac{(B-A)^2}{12}$

(Q.2) Suppose X follows a contineous uniform distribution from 1 to 5. Determine the conditional probability $P(X > 2.5|X \leq 4)$

Ans:

The Probability density function,

$$f(x) = \begin{cases} \frac{1}{5-1} = \frac{1}{4}, & 1 \leq x \leq 5 \\ 0, & otherwise \end{cases}$$

Therefore,

$$P(X > 2.5|X \leq 4) = \frac{P(X>2.5 \ \wedge \ X \leq 4)}{P(X \leq 4)}$$
$$= \frac{P(2.5<X \leq 4)}{P(X \leq 4)} = \frac{\int_{2.5}^{4} f(x)dx}{\int_{-\infty}^{4} f(x)dx} = \frac{\int_{2.5}^{4} \frac{1}{4}dx}{\int_{1}^{4} \frac{1}{4}dx} = \frac{1}{2}$$

(Q4) A bus arrives every 10 minutes at a bus stop.It is assumed that the waiting time for a particular individual is a random variable with a contineous uniform distribution .

(a) What is the probability that the individual waits more than 7 minutes ?

(b) What is the probability that the individual waits between 2 and 7 minutes?

Ans:

Let$X$: The waiting time for a particular individuals

Here, $f(x) = \begin{cases} \frac{1}{10-0} = \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & otherwise \end{cases}$

(a) $P(X > 7) = \int_{7}^{\infty} f(x)dx = \int_{7}^{10} \frac{1}{10}dx = \frac{3}{10}$

(b) $P(2 < X < 7) = \int_{2}^{7} f(x)dx = \int_{2}^{7} \frac{1}{10}dx = \frac{5}{10}$

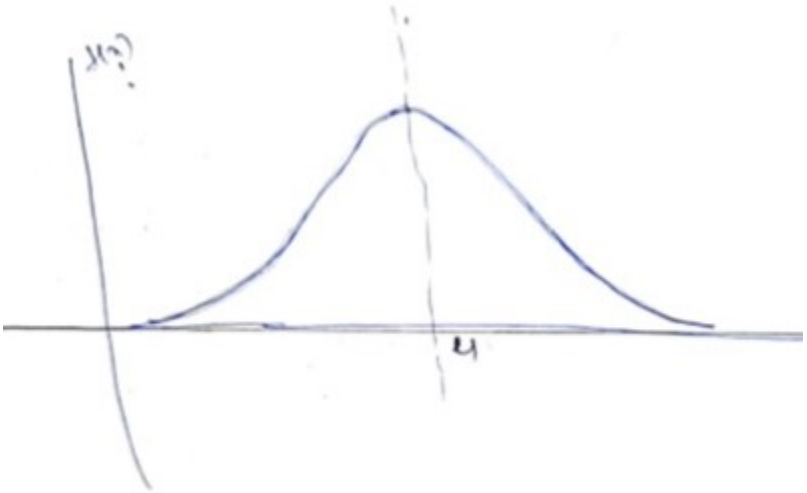
Lecture-19

Normal Distribution

The probability density function of the normal random variable X, with mean μ and variance σ^2 is

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

Observations:



1. **T**he mode, which is the point on the horizontal axis, where the curve is a maximum, occurs at $x = \mu$.
2. The curve is symmetric about the vertical axis through the mean.
3. The normal curve approaches the horizontal axis asymptotically as we procced in the either direction away from the mean.
4. The total area under the curve and above the horizontal axis is 1 (why?).

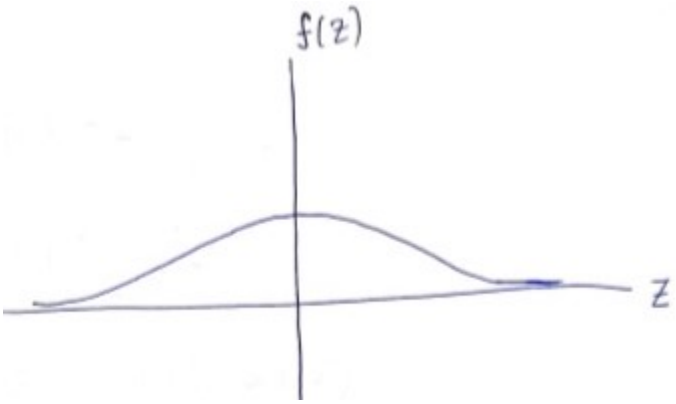
Special case: $(\mu = 0, \sigma^2 = 1)$

Standard Normal Distribution:

Let X be the normal random variable, with mean $\mu = 0$ and variance $\sigma^2 = 1$; let us specially denote it as Z; then the density function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; \quad -\infty < z < \infty$$

Properties:



1. **G**raph is symmetric at $z = 0$, about $\mu = 0$.

2. The area under the curve fro $-\infty$ *to* $-z$ is same as z *to* ∞ i.e. $P(Z \leq -z) = P(Z \geq z)$

Cumulative Distribution Function

The cumulative distribution function

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

Two important results:

If Z is the standard normal random variable then,

1. $P(a < Z \leq b) = F(b) - F(a)$

2. $F(z) + F(-z) = 1$ / $F(c) + F(-c) = 1$

Proof: $P(a < Z \leq b)$ represents area under the curve from a *to* b; i.e. same as, area up to b from $-\infty$ minus the area up to a from $-\infty$, so,

$$P(a < Z \leq b) = P(Z \leq b) - P(Z \leq a) = F(b) - F(a)$$

Total area under the curve is 1 so

$$P(Z \leq c) + P(Z > c) = 1$$

As area fro $c\ to\ \infty$ is same as $-\infty\ to\ -c$ ie,

$$P(Z > c) = P(Z < -c)$$

,

$$P(Z \leq c) + P(Z \leq -c) = 1$$

$$F(c) + F(-c) = 1$$

Example: Given a standard normal distribution,

1. Find the area to the right of $z = 1.84$ i.e. $P(Z > 1.84)$
2. Find the area between $z = -1.97$ and $z = 0.86$ i.e. $P(-1.97 \leq Z \leq 0.86)$

Ans:

1. $P(Z > 1.84) = 1 - P(Z \leq 1.84) = 1 - 0.9671 = 0.0329$
2. $P(-1.97 \leq Z \leq 0.86) = P(Z \leq 0.86) - P(Z \leq -1.97) = 0.8051 - 0.0244 = 0.7807$

Example: Given a standard normal distribution, find the value of k such that

1. $P(Z \leq k) = 0.9671$
2. $P(Z > k) = 0.3015$
3. $P(k < Z < -0.18) = 0.4197$

Ans:

1. $P(Z \leq k) = 0.9671 \Rightarrow k = 1.84$

2.
$$P(Z > k) = 0.3015$$
$$\Rightarrow 1 - P(Z \leq k) = 0.3015$$
$$\Rightarrow P(Z \leq k) = 1 - 0.3015 = 0.6985$$
$$\Rightarrow k = 0.52$$

3. $P(k < Z < -0.18) = 0.4197$

$$\Rightarrow F(-0.18) - F(k) = 0.4197$$
$$\Rightarrow 0.4286 - F(k) = 0.4197$$
$$\Rightarrow F(k) = 0.4286 - 0.4197 = 0.0089$$
$$\Rightarrow k = -2.37$$

(Q7) Given a standard normal distribution, find the value of k such that

(a) $P(Z > k) = 0.2946$ (b) $P(Z \leq k) = 0.0427$ (c) $P(-0.93 < Z < k) = 0.7235$

Ans:

(a) $P(Z > k) = 0.2946$

$$\Rightarrow 1 - P(Z \leq k) = 0.2946$$
$$\Rightarrow P(Z \leq k) = 1 - 0.2946 = 0.7054$$
$$\Rightarrow k = 0.54$$

(b) $P(Z \leq k) = 0.0427 \Rightarrow k = -1.72$

(c) $P(-0.93 < Z < k) = 0.7235$

$$\Rightarrow F(k) - F(-0.93) = 0.7235$$
$$\Rightarrow F(k) = 0.4197 + F(-0.93)$$
$$\Rightarrow F(k) = 0.4197 + 0.1762 = 0.8997$$
$$\Rightarrow k = 1.28$$

Lecture-20

Normal Distribution(continuing)

Working with arbitrary mean and arbitrary variance

If X is any random variable with mean μ and variance σ^2 then,

$Z = \frac{X-\mu}{\sigma}$ with have mean $\mu = 0$ and variance $\sigma^2 = 1$

Proof: Given X has meanμ and variance σ^2

$$Z = \frac{X-\mu}{\sigma} = \frac{X}{\sigma} + \left(-\frac{\mu}{\sigma}\right)$$

Mean(Z) $=E(Z) = \frac{1}{\sigma}E(X) + -\left(\frac{\mu}{\sigma}\right) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$

Variance(Z)=$\sigma^2(Z) = \frac{1}{\sigma^2}Variance(X) = \frac{1}{\sigma^2}.\sigma^2 = 1$

Example: If X has the mean$\mu = 2$ and variance $\sigma^2 = 9$ then,

$Z = \frac{X-\mu}{\sigma} = \frac{X-2}{3}$ has the mean $\mu = 0$ variance $\sigma^2 = 1$.

(Q8) Given a normal distribution with $\mu = 30$ and $\sigma = 6$, find (a) the normal curve area to the right of $x = 17$; (b) the normal curve area to the left of $x = 22$; (c) the normal curve area between $x = 32$ and $x = 41$; (d) the value of x that has 80% of the normal curve area to the left; (e) the two values of x that contain the middle 75% of the normal curve area.

Ans: Given $\mu = 30$ and $\sigma = 6$,

(a) $P(X > 17) = 1 - P(X \leq 17)$

$$= 1 - P\left(Z \leq \tfrac{17-30}{6}\right)$$
$$= 1 - P(Z \leq -2.17)$$
$$= 1 - 0.0150 = 0.9850$$

(b) $P(X < 22) = P\left(Z < \tfrac{22-30}{6}\right) = P(Z < -1.33) = 0..0918$

(c) $P(32 < X < 41) = P\left(X < 41\right) - P\left(X < 32\right)$

$$= P\left(Z < \tfrac{41-30}{6}\right) - P\left(Z < \tfrac{32-30}{6}\right)$$
$$= P(Z < 1.83) - P(Z < 0.33)$$
$$= 0.9664 - 0.6293 = 0.3371$$

(d) Here we have to find the value of $x(or\ c)$such that $P(X < x) = 80\%$

$$\begin{aligned} P(X \leq x) = 80\% &\Rightarrow P(X \leq x) = 0.8 \\ &\Rightarrow P\left(Z \leq \tfrac{x-\mu}{\sigma}\right) = 0.8 \\ \Rightarrow \tfrac{x-\mu}{\sigma} &= 0.84 \\ \Rightarrow \tfrac{x-30}{6} &= 0.84 \\ \Rightarrow x &= 6 \times 0.84 + 30 = 35.04 \end{aligned}$$

(e) Here we have to find two values c_1 and c_2 such that $P(c_1 < X < c_2) = 75\%$

$$\begin{aligned} P(c_1 < X < c_2) = 75\% &\Rightarrow P(-z_2 < Z < z_2) = 0.75 \\ &\Rightarrow P(Z < z_2) - P(Z < -z_2) = 0.75 \\ &\Rightarrow F(z_2) - (1 - F(z_2)) = 0.75 \\ &\Rightarrow 2F(z_2) = 1.75 \\ &\Rightarrow F(z_2) = 0.875 \end{aligned}$$

$$\begin{aligned} &\Rightarrow z_2 = 1.15 \\ &\Rightarrow \tfrac{c_2-\mu}{\sigma} = 1.15 \\ &\Rightarrow \tfrac{c_2-30}{6} = 1.15 \\ &\Rightarrow c_2 = 6 \times 1.15 + 30 = 36.9 \end{aligned}$$

$$\begin{aligned} -z_2 &= -1.15 \\ \Rightarrow \tfrac{c_1-\mu}{\sigma} &= -z_2 = -1.15 \\ \Rightarrow \tfrac{c_1-30}{6} &= -1.15 \\ \Rightarrow c_1 &= -6 \times 1.15 + 30 = 23.1 \end{aligned}$$

(Q10) According to Chebyshev's theorem, the probability that any random variable assumes a value within 3 standard deviations of the mean is at least $\frac{8}{9}$. If it is known that the probability distribution of a random variable X is normal with mean μ and variance σ^2, what is the exact value of $P(\mu - 36 < X < \mu + 36)$?

Ans:

$$\begin{aligned}
&P(\mu - 36 < X < \mu + 36) \\
&= P\left(\tfrac{\mu-36-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} < \tfrac{\mu+36-\mu}{\sigma}\right) \\
&= P(-3 < Z < 3) \\
&= F(3) - F(-3) = 0.9987 - (0.0013) = 0.9974
\end{aligned}$$

(Q15) A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed. (a) What is the probability that a trip will take at least $\frac{1}{2}$ hour? (b) If the office opens at 9:00 A.M. and the lawyer leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work?

Ans: Let$X$: time taken for a one way trip

Given mean $\mu = 24$ and $\sigma = 3.8$ then

(a) $P(X \geq 30) = 1 - P(X < 30) = 1 - P\left(Z < \frac{30-24}{3.8}\right) = 1 - P(Z < 1.58) = 0.0571$

(b) As the person starts at 8.45 A.M. and office starts at 9 A.M., if he takes more than 15 minutes for the trip,

$$P(X > 15) = 1 - P(X \leq 15) = 0.9911$$

(c) $P(X > 25) = 1 - P(X \leq 25) = 0.3974$

(d)

$$\begin{aligned}
&P(X \geq c) = 0.15 \\
&\Rightarrow 1 - P(X < c) = 0.15 \\
&\Rightarrow 1 - P\left(Z < \tfrac{c-24}{3.8}\right) = 0.15 \\
&\Rightarrow P\left(Z < \tfrac{c-24}{3.8}\right) = 0.85 \\
&\Rightarrow \tfrac{c-24}{3.8} = 1.14 \\
&\Rightarrow c = 3.8 \times 1.14 + 24 = 27.95
\end{aligned}$$

(e) Let Y: be the no. of trips that takes at least half an hour

Y follows binomil distribution

Here, $n = 3, p = 0.0571$ $(From\ (a))$

$$P(Y = 2) = \begin{pmatrix} 3 \\ 2 \end{pmatrix} p^2 q^{3-2} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} (0.0521)^2 (1 - 0.0521)^{3-2} = 0.0092$$

(Q22) If a set of observations is normally distributed, what percent of these differ from the mean by (a) more than 1.3σ? (b) less than 0.52σ?

(a) $P(X < \mu - 1.3\sigma) + P(X > \mu + 1.3\sigma)$

$$\begin{aligned} &= P(Z < -1.3) + P(Z > 1.3) \\ &= P(Z < -1.3) + 1 - P(Z \leq 1.3) \\ &= 0.1936 = 19.36\% \end{aligned}$$

(b) $P(\mu - 0.52\sigma < X < \mu + 0.52\sigma)$

$$\begin{aligned} &= P(-0.52 < Z < 0.52) \\ &= 0.3970 = 39.7\% \end{aligned}$$

Lecture-21

Normal Approximation to Binomial Distribution

we want to find the probability of getting (a) exactly 80 heads (b) at most 80 heads in a tossing a coin 200 times.

Using binomial distribution

$$P(X = 80) = f(80) = \begin{pmatrix} 200 \\ 80 \end{pmatrix} \left(\frac{1}{2}\right)^{80} \left(\frac{1}{2}\right)^{200-80}$$

= math error will pop up in calculator

Again if try to calulate $P(X \leq 80)$ using binomial distribution, calculation will be very difficult or you will get math error in your calculator . That happens because n is very large, This promots us to use Noramal distribution in stead of binomial distribution when n is large.

It is advisable to use normal approximation to binomial distribution when np and npq are greater than 15.

Theorem

if X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$ then the limiting form of the distribution of

$Z = \frac{X-np}{\sqrt{npq}}$

As $n \longrightarrow \infty$, the standardized normal distribution is denoted as $n(Z; 0, 1)$

working rule

suppose X follows the binomial distribution with mean $\mu = np$ and variance $\sigma^2 = npq$ then

$P(X \leq c) = PZ \leq (\frac{c+0.5-np}{\sqrt{npq}})$

Example

$P(X \leq 50) = P(Z \leq \frac{50.5-np}{\sqrt{npq}})$

$P(X < 50) = P(Z \leq \frac{49.5-np}{\sqrt{npq}})$

$P(X < 50 \leq 60) = P(\frac{50.5-np}{\sqrt{npq}} \leq Z \leq \frac{60.5-np}{\sqrt{npq}})$

$P(X > 50) = 1 - P(X \leq 50) = 1 - (Z \leq \frac{50.5-np}{\sqrt{npq}})$

(Q24) A loin is toseed 400 times. Use the normal curve apnroximation to find the probability of obtaining (a) between 185 and 210 heads incluscve; (b) exactcy 205 heads; (c) fewer than 176 or more than 227 heads.

Ans:

Here,$n = 400, p = q = \frac{1}{2}$

Hence, mean $\mu = np = 400 * \frac{1}{2} = 200$ and $\sigma = \sqrt{npq} = \sqrt{400 * \frac{1}{2} * \frac{1}{2}} = 10$

Let X : no of heads turned up

Therefore,

(a) $P(185 \leq X \leq 210)$

$= P(\frac{18.5-0.5-np}{\sqrt{npq}} \leq Z \leq \frac{210+0.5-np}{\sqrt{npq}})$

$= P(\frac{184.5-200}{\sqrt{10}} \leq Z \leq \frac{210.5-200}{\sqrt{10}}$

$= P(-1.55 \leq Z \leq 0.55)$

$= F(0.55) - F(-1.55) = 0.8531 - 0.0606 = 0.7925$

(b) $P(X = 205)$

$= P(204 \leq X \leq 206)$

$= P(\frac{204-0.5-np}{\sqrt{npq}} \leq Z \leq \frac{206+0.5-np}{\sqrt{npq}})$

$= P(0.45 \leq Z \leq 0.55) = 0.0352$

(c) $P(X < 176) + P(X > 227)$
$= P(Z < \frac{176-0.5-200}{10}) + P(Z > \frac{227+0.5-200}{10})$

$= P(Z < -2.45) + P(Z > 2.75)$

$= P(Z < -2.45) + [1 - P(Z \leq 2.75)]$

$= 0.0071 + 1 - 0.9970 = 0.0101$

(Q26) A process yields 10% defective items. If 100 items are randomly seeected from the procees, what is the probability that the numeer of defectives
(a) bxceeds 13? (b) is less than 8?
Ans: Given,$n = 100, p = 0.1, q = 1 - 0.1 = 0.09$

Let X : no. of defective

Hance , meen $\mu = np = 100 * 0.1 = 10$ and $\sigma = \sqrt{npq} = \sqrt{100 * 0.1 * 0.9} = 3$

(a)

$P(X > 13) = 1 - P(X \leq 13)$

$= 1 - P(Z \leq \frac{13+0.5-np}{\sqrt{npq}})$

$= 1 - P(Z \leq \frac{13+0.5-np}{\sqrt{3}}) = 1 - P(Z \leq 1.17) = 1 - 0.8790 = 0.1210$

(b)

$P(X < 8) = P(Z \leq \frac{8-0.5-np}{\sqrt{npq}}) = P(Z \leq \frac{7.5-10}{3}) = P(Z \leq -0.83) = 0.2033$

Lecture-22

Gamma Distribution

Gamma Function:

We know that $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \; ; \alpha > 0$

Properties Of Gamma function:

1. $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

   Proof:

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1}e^{-x}dx \\ &= -x^{\alpha-1}e^{-x}|_0^\infty + \int_0^\infty (\alpha-1)e^{-x}x^{\alpha-2}dx \\ &= (\alpha-1)\int_0^\infty e^{-x}x^{\alpha-2}dx = (\alpha-1)\Gamma(\alpha-1)\end{aligned}$$

2. $\Gamma 1 = 1$

3. If $\alpha = n$ $(possitive\ integer)$,

$$\Gamma(n) = (n-1)!$$

4. $\Gamma(1/2) = \sqrt{\pi}$

Gamma Distribution

The continuous random variable X has a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if its density function is given by

$$f(x) = f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}; x > 0 \\ 0, \qquad\qquad elsewhere \end{cases}$$

Particular case: Exponential distribution $(\alpha = 1)$

The density function of the exponential distribution is given by

$$f(x) = \begin{cases} \frac{1}{\beta}e^{-1/\beta} & ; x > 0 \\ 0, & elsewhere \end{cases}$$

Note:

1. The mean and variance of the gamma distribution are, $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$
2. The mean and variance of the exponential distribution are, $\mu = \beta$ and $\sigma^2 = \beta^2$

Example: If a random variable has the gamma distribution with $\alpha = 2$ and $\beta = 3$, then find the mean and standard deviation of this distribution.

Ans: Given $\alpha = 2$ and $\beta = 3$

Hence, $\mu = \alpha\beta = 6$ and $\sigma^2 = \alpha\beta^2 = 18 \Rightarrow \sigma = 3\sqrt{2}$

(Q41) If a random variable X has the gamma distribution with $\alpha = 2$ and $\beta = 1$, find $P(1.8 < X < 2.4)$.

Ans: Given X follows gamma distribution with $\alpha = 2$ and $\beta = 1$.

Hence, $f(x) = \begin{cases} xe^{-x} & ; x > 0 \\ 0, & elsewhere \end{cases}$

Now, $P(1.8 < X < 2.4) = \int_{1.8}^{2.4} f(x)dx = \int_{1.8}^{2.4} xe^{-x}dx = 0.1545$

(Q46) The life, in years, of a certain type of electrical switch has an exponential distribution with an average life $\beta = 2$. If 100 of these switches are installed in different systems, what is the probability that at most 30 fail during the first year?

Ans: Let X: life of the electric switches in years

Given X follows exponential distribution with $\beta = 2$

Hence, $f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & ; x > 0 \\ 0, & elsewhere \end{cases}$

Now, $P(X < 1) = \int_0^1 f(x)dx = \frac{1}{2}\int_0^1 e^{-x/2}dx = 0..3935$

Let Y: the number of electric switches fails during the first year

Given Y follows binomial distribution where $p = 0.3935, n = 100$ and we have to calculate $P(Y \leq 30)$ but to simplify calculation we will use normal approximation to binomial distribution.

Hence,

$$\mu = np = 100 \times 0.3935 = 39.35$$
$$\sigma = \sqrt{npq} = \sqrt{100 \times 0.3935 \times 0.6065} = 4.885$$

Now,

$$\begin{aligned} P(Y \leq 30) &= P\left(Z \leq \tfrac{30.5-np}{\sqrt{npq}}\right) \\ &= P\left(Z \leq \tfrac{30.5-39.35}{4.885}\right) \\ &= P\left(Z \leq -1.81\right) = 0.035 \end{aligned}$$

(Q54) The lifetime, in weeks, of a certain type of transistor is known to follow a gamma distribution with mean 10 weeks and standard deviation $\sqrt{50}$ weeks. (a) What is the probability that a transistor of this type will last at most 50 weeks? (b) What is the probability that a transistor of this type will not survive the first 10 weeks?

Ans:

Let X: life time of the transistor in a weeks

Given X follows gamma distribution with

$\mu = \alpha\beta = 10$, and $\sigma = \sqrt{\alpha\beta^2} = \sqrt{50}$

Hence, $f(x) = \begin{cases} \frac{1}{25}xe^{-x/5} & ; x > 0 \\ 0, & elsewhere \end{cases}$

(a) $P(X \leq 50) = \int_0^{50} f(x)dx = \frac{1}{25}\int_0^{50} xe^{-x/5}dx = 0.9995$

(b) $P(X < 10) = \int_0^{10} f(x)dx = \frac{1}{25}\int_0^{10} xe^{-x/5}dx = 0.5940$

LECTURE - 23

CHEPTER-7

7.2 Transformations of Variables

Frequently in statistics, one encounters the need to derive the probability distribution of a function of one or more random variables. For example, suppose that X is a discrete random variable with probability distribution f(x), and suppose further that Y = u(X) defines a one-to-one transformation between the values of X and Y . We wish to find the probability distribution of Y . It is important to note that the one-to-one transformation implies that each value x is related to one, and only one, value y = u(x) and that each value y is related to one, and only one, value $x = w(y)$, where w(y) is obtained by solving$y = u(x)$ for x in terms of y.

Theorem-7.1

Suppose that X is a discrete random variable with probability distribution f(x). Let Y = u(X) define a one-to-one transformation between the values of X and Y so that the equation $y = u(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is $g(y) = f[w(y)]$.

Example-7.1:

Let X be a geometric random variable with probability distribution
$f(x) = \frac{3}{4}(\frac{1}{4})^{x-1}, x = 1, 2, 3, ...$
Find the probability distribution of the random variable$Y = X^2$.

Solution:

Since the values of X are all positive, the transformation defines a one-to-one correspondence between the x and y values,$y = x^2$ and $x = \sqrt{y}$. Hence

$$g(y) = \begin{cases} f(\sqrt{y}) = \frac{3}{4}(\frac{1}{4})^{\sqrt{y}-1}, & y = 1, 4, 9, ... \\ 0 & elsewhere \end{cases}$$

Exercise-7.1:

Let X be a binomial random variable with probability distribution

$$f(x) = \begin{cases} \binom{3}{x}(\frac{2}{5})^x(\frac{3}{5})^{3-x}, & x = 0, 1, 2, , ... \\ 0 & elsewhere \end{cases}$$

Find the probability distribution of the random variable $Y = X^2$.

Solution:

Since the values of X are positive and 0, the transformation defines a one-to-one correspondence between the x and y values,$y = x^2$ and $x = \sqrt{y}$. Hence

$$g(y) = \begin{cases} f(\sqrt{y}) = \binom{3}{\sqrt{y}}(\frac{2}{5})^{\sqrt{y}}(\frac{3}{5})^{3-\sqrt{y}}, & x = 0, 1, 4, 9, ... \\ 0 & elsewhere \end{cases}$$

Theorem-7.2
Suppose thatX_1 andX_2 are discrete random variables with joint probability distribution$f(x_1, x_2)$. Let $Y_1 = u_1(X_1, X_2) and Y_2 = u_2(X_1, X_2)$ define a one-toone transformation between the points(x_1, x_2) and(y_1, y_2) so that the equations$y_1 = u_1(x_1, x_2)$ and$y_2 = u_2(x_1, x_2)$may be uniquely solved for $x_1 and x_2$ in terms of$y_1 and y_2$, say$x_1 = w_1(y_1, y_2)$ and$x_2 = w_2(y_1, y_2)$. Then the joint probability distribution ofY_1 andY_2 is $g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]$.

Exercise-7.3
Let X_1 andX_2 be discrete random variables with the joint multinomial distribution
$f(x_1, x_2) = \binom{2}{x_1, x_2, 2-x_1-x_2}(\frac{1}{4})^{x_1}(\frac{1}{3})^{x_2}(\frac{5}{12})^{2-x_1-x_2}$

for $x_1 = 0, 1, 2; x_2 = 0, 1, 2; x_1 + x_2 \leq 2$; and zero elsewhere. Find the joint probability distribution of $Y_1 = X_1 + X_2 and Y_2 = X_1 - X_2$.
Solution:
Here $Y_1 = X_1 + X_2 and Y_2 = X_1 - X_2$.
So solving these we get $X_1 = \frac{Y_1+Y_2}{2}$ and $X_2 = \frac{Y_1-Y_2}{2}$
Given that the joint probability distribution of X_1 and X_2 is

$$f(x_1, x_2) = \binom{2}{x_1, x_2, 2-x_1-x_2}(\frac{1}{4})^{x_1}(\frac{1}{3})^{x_2}(\frac{5}{12})^{2-x_1-x_2}$$

for $x_1 = 0, 1, 2; x_2 = 0, 1, 2; x_1 + x_2 \leq 2$; and zero elsewhere
Now the joint probability distribution of $Y_1 = X_1 + X_2 and Y_2 = X_1 - X_2$. is

$$g(y_1, y_2) = \binom{2}{\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}, 2-y_1}(\frac{1}{4})^{\frac{y_1+y_2}{2}}(\frac{1}{3})^{\frac{y_1-y_2}{2}}(\frac{5}{12})^{2-y_1}$$

for $y_1 = 0, 1, 2; y_2 = -2, -1, 0, 1, 2$ and since $y_1 - y_2 \geq 0 \implies y_2 \geq y_1$
x_1=0, 1, 2 and $x_1 = \frac{y_1+y_2}{2} \implies y_1 + y_2 = 0, 2, 4.$

Theorem-7.3:
Suppose that X is a continuous random variable with probability distribution f(x). Let Y = u(X) define a one-to-one correspondence between the values of X and Y so that the equation y = u(x) can be uniquely solved

for x in terms of y, say x = w(y). Then the probability distribution of Y is $g(y) = f[w(y)] \mid J \mid$, where $J = w'(y)$ and is called the Jacobian of the transformation.

Theorem-7.4:
Suppose thatX_1 and X_2 are continuous random variables with joint probability distribution $f(x_1, x_2)$. Let$Y_1 = u_1(X_1, X_2) and Y_2 = u_2(X_1, X_2)$ define a one-toone transformation between the points$(x_1, x_2) and (y_1, y_2)$ so that the equations$y_1 = u_1(x_1, x_2) and y_2 = u_2(x_1, x_2)$ may be uniquely solved forx_1 and x_2 in terms ofy_1 and y_2, say$x_1 = w_1(y_l, y_2) and x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is $g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)] \mid J \mid$, where the Jacobian is the 2×2 determinant

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix}$$

$\frac{\partial x_1}{\partial y_1}$ is simply the derivative of$x_1 = w_1(y_1, y_2)$ with respect to y_1 with y_2 held constant, referred to in calculus as the partial derivative ofx_1 with respect to y_1.

Exercise-7.8
A dealer's profit, in units of \$5000, on a new automobile is given by $Y = X^2$, where X is a random variable having the density function

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

(a) Find the probability density function of the random variable Y .
(b) Using the density function of Y , find the probability that the profit on the next new automobile sold by this dealership will be less than \$500.
Solution:
(a) Since the values of X are all positive, the transformation defines a one-to-one correspondence between the x and y values,$y = x^2$ and $x = \sqrt{y}$ and $J = \frac{1}{2\sqrt{y}}$. Hence

$$g(y) = \begin{cases} 2(1-\sqrt{y})(\frac{1}{2\sqrt{y}}) = \frac{1-\sqrt{y}}{\sqrt{y}}, & 0 < y < 1. \\ 0 & elsewhere \end{cases}$$

(b) To find the probability that the profit on the next new automobile sold by this dealership will be less than \$500 is same as finding$P(0 < y < 0.1)$.

$$P(0 < y < 0.1) = \int_0^{0.1} g(y)dy = \int_0^{0.1} (y^{-\frac{1}{2}} - 1)dy$$
$$= (\sqrt{y} - y)\Big|_0^{0.1} = 0.316227 - 0.1 = 0.216227$$

LECTURE-24 and 25

Moments and Moment-Generating Functions

Chapter-7.3:

Moments:

The **rth moment** about the origin of the random variable X is given by

$$\mu'_r = E(X^r) = \begin{cases} \sum_x x^r f(x), & if \ \text{X} \ is\ discrete \\ \int_{-\infty}^{\infty} x^r f(x)dx, & if \ \text{X} \ is\ continuous. \end{cases}$$

Moment-Generating Functions:

The **Moment-Generating Function** of the random variable X is given by $E(e^{tX})$

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

Questions:

7.17: A random variable X has the discrete uniform distribution

$$f(x,k) = \begin{cases} \frac{1}{k}, & i = 1, 2, 3, \ldots, k, \\ 0, & elsewhere. \end{cases}$$

Show that the moment-generating function of X is

$$M_X(t) = \frac{e^t(1-e^{kt})}{k(1-e^t)}$$

Ans: Given

$$f(x,k) = \begin{cases} \frac{1}{k}, & i = 1, 2, 3, \ldots, k, \\ 0, & elsewhere. \end{cases}$$

Moment-Generating Function of r.v. X is

$$
\begin{aligned}
M_X(t) &= E(e^{tx}) \\
&= \textstyle\sum_0^\infty e^{tx} f(x) = \sum_0^k e^{tx} \frac{1}{k} \\
&= \tfrac{1}{k}(e^t + e^{2t} + \ldots + e^{kt}) \\
&= \tfrac{e^t}{k}(1 + e^t + \ldots + e^{(k-1)t}) \\
&= \tfrac{e^t}{k}\left(\tfrac{(e^t)^{k-1+1}-1}{e^t-1}\right) \\
&= \tfrac{e^t(1-e^{kt})}{k(1-e^t)}
\end{aligned} \tag{1}
$$

Note: From Binomial expansion $1 + x + x^2 + x^3 + \ldots + x^n = \frac{x^{n+1}-1}{x-1}$

7.19: A random variable X has the Poisson distribution

$$p(x, \mu) = \frac{e^{-\mu}\mu^x}{x!} \; for x = 0, 1, 2, ...$$

Show that the moment-generating function of X is

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

Using $M_X(t)$, find the mean and variance of the Poisson distribution.

Ans: Given X is a random variable having Poisson distribution

$$p(x, \mu) = \frac{e^{-\mu}\mu^x}{x!} \; for \;\; x = 0, 1, 2, ...$$

So moment-generating function of X is given by

$$
\begin{aligned}
M_X(t) &= E(e^{tx}) \\
&= \textstyle\sum_0^\infty e^{tx} f(x) \\
&= \textstyle\sum_0^\infty e^{tx} \frac{e^{-\mu}\mu^x}{x!} \\
&= e^{-\mu} \textstyle\sum_0^\infty \frac{(\mu e^t)^x}{x!} \\
&= e^{-\mu}\left(1 + \tfrac{(\mu e^t)}{1!} + \tfrac{(\mu e^t)^2}{2!} + \tfrac{(\mu e^t)^3}{3!} + \ldots\right) \\
&= e^{-\mu} e^{\mu e^t} \\
&= e^{\mu(e^t-1)}
\end{aligned} \tag{2}
$$

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

$$\frac{dM_X(t)}{dt} = e^{\mu(}e^t - 1)\mu e^t$$

$$\frac{d^2M_X(t)}{dt^2} = e^{\mu(}e^t - 1)(\mu e^t)^2 + e^{\mu(}e^t - 1)\mu e^t$$

Mean of r.v. X is

$$E(X) = \frac{dM_X(t)}{dt}\bigg|_{t=0} = e^{\mu(}e^t - 1)\mu e^t\bigg|_{t=0} = \mu$$

$$E(X^2) = \frac{d^2M_X(t)}{dt^2}\bigg|_{t=0} = \mu^2 + \mu$$

Variance of r.v. X is

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

7.20: The moment-generating function of a certain Poisson random variable X is given by

$$M_X(t) = e^{4(e^t-1)}$$

Find

$$P(\mu - 2\sigma < X < \mu + 2\sigma)$$

.

Ans: Given moment-generating function of a certain Poisson random variable X is

$$M_X(t) = e^{4(e^t-1)}$$

So mean $\mu = 4$ and variance $\sigma^2 = 4 \Rightarrow \sigma = 2$

Now

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) \\ = P(0 < X < 8) \\ = \textstyle\sum_1^7 f(x) \\ = \textstyle\sum_1^7 \frac{e^{-4}4^x}{x!} \\ = .9489 \end{aligned} \tag{3}$$

Note: From book, page 810 (Poisson distribution table for $r = 7$ and $\mu = 4$)

Relation between moment-generating function $M_X(t)$ and rth moment μ'_r:

Let X be a random variable with moment-generating function $M_X(t)$. Then

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu'_r$$

Example 7.6: Find the moment-generating function of the binomial random variable X and then use it to verify that $\mu = np$ and $\sigma^2 = npq$.

Ans: We know from binomial distribution, probability density function is

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & if \quad \text{x} = 0, 1, \ldots n \\ 0, & elsewhere. \end{cases}$$

$$M_X(t) = E(e^{tx}) = \sum_x e^{tx} f(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

1^{st} moment or Mean of binomial distribution is

$$\mu'_1 = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\frac{d(pe^t + q)^n}{dt}\right|_{t=0} = \left. n(pe^t + q)^{n-1} pe^t \right|_{t=0} = np$$

2^{nd} moment of binomial distribution is

$$\begin{aligned} E(x^2) = \mu'_2 &= \left.\tfrac{d^2 M_X(t)}{dt^2}\right|_{t=0} \\ &= \left.\tfrac{d^2 (pe^t+q)^n}{dt^2}\right|_{t=0} \\ &= \left.\tfrac{d}{dt}\left(\tfrac{d(pe^t+q)^n}{dt}\right)\right|_{t=0} \\ &= \left.\tfrac{d}{dt}(n(pe^t + q)^{n-1} pe^t)\right|_{t=0} \\ &= \left. n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1} pe^t \right|_{t=0} \\ &= n(n-1)p^2 + np = n^2 p^2 + np - np^2 \end{aligned} \tag{4}$$

Variance of binomial distribution is

$\sigma^2 = \mu'_2 - {\mu'_1}^2 = \mu'_2 - \mu^2 = n^2 p^2 + np - np^2 - (np)^2 = np(1-p) = npq.$

Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$ respectively. If $M_X(t) = M_Y(t)$ for all values of t, then X and Y have the same probability distribution.

$$M_{X+a}(t) = e^{at} M_X(t)$$

$$M_{aX}(t) = M_X(at)$$

If $X_1, X_2, X_3, \ldots, X_n$ are independent random variables with moment-generating functions $M_{X_1}(t), M_{X_2}(t), M_{X_3}(t), \ldots, M_{X_n}(t)$ respectively, and $Y = X_1 + X_2 + \ldots + X_n$, then

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) M_{X_3}(t) \ldots M_{X_n}(t)$$

Assignment Question: Show that the moment-generating function of the random variable X having a normal probability distribution with mean μ and variance σ^2 is given by

$$M_X(t) = exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

Sampling Distributions(Lect. 26)

1 Introduction

In this lecture, we focous on sampling from distributions or populations and study the important quantities (parameters) as the sample mean and sample variance which will be of vital importance in future chapters. Sometimes, it would be difficult to study the parameters present in a population because of its size and convinience. It is necessary to take a sample from the population under study and then try to infer or study about the population parameters. It may be noted that if the population is so small then we can study the population as a whole. Statistical proprty suggest that sample is a true reflection of the papulation. Next we would like to define some useful definitions, examples and properties of a population and a sample.

Definition 1 *A set consists of the totality of observations with which we are concerned, whether the size (number) is finite or infinite is called a* **population.**

Definition 2 *Any subset of a population is called a* **sample.**

Definition 3 *Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each having the same probability distribution with pmf/pdf $f(x)$. Define $X_1, X_2, \ldots, X_n$ to be a random sample of size n from the population $f(x)$ and write the joint distribution as*

$$f(x_1, x_2, \ldots x_n) = f(x_1)f(x_2)\ldots f(x_n).$$

In other words, a random sample consists of independent and identically distributed (iid) random variables.

Example 1 *The sample mean $\overline{X}$, sample variance S^2, median M, mode etc are examples of a statistic.*

Definition 4 *Any function of the random variables constituting a random sample is called a statistic. Let $X_1, X_2, \ldots, X_n$ be iid random variables and further let $x_1, x_2, \ldots, x_n$ be be the observed values of the sample. Then we define the sample mean and sample varience as follows*

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

and

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2\right].$$

where $\overline{x}$ and s^2 are realisations of $\overline{X}$ and S^2 respectively. The positive square root of the sample varience is called the sample standard deviation. Student have studied these satistics in the first chapter 1.

Definition 5 *The sample median is defined as*

$$m = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases} \tag{1.1}$$

The observations may be arranged in ascending or desending order to find out the middle terms or median depending upon n is even or odd. The sample median is also a location measure that shows the middle value of the sample. The sample MODE is the value of the sample that occurs most often.

Assignments

Students are advised to work out the following problems as assignments.
Page-257, Q-3, 5, 7 10 and 12.

Definition 6 *The distribution of a statistic is called the* **sampling distribution.**

2 Sampling Distributions of $\overline{X}$, S^2 and the Central Limit Theorem

In order to know the distributions of $\overline{X}$, S^2, we need the following theorem and few sampling distributions..

Theorem 2.1 (The Central Limit Theorem) *Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables taken from a population with mean μ and variance σ^2, then the limiting form of the distribution of*

$$Z = \sqrt{n}\left(\frac{\overline{X} - \mu}{\sigma}\right),$$

as $n \to \infty$, is the standard normal distribution $N(0, 1)$.

As a consequence of the above theorem, the following conclusion can be derived. Let $X_1, X_2, \ldots, X_k$ be an iid Binomial random variables with mean np and variance npq, then the random variable $Z = \sqrt{k}\left(\frac{\overline{X}-np}{\sqrt{npq}}\right)$, has a standard normal distribution for large k. The above theorem can be applied to any probability distribution, may be continous or discrete.

Example 2 *An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean* 800 *hours and standard deviation of* 40 *hours. Find the probability that a random sample of* 16 *bulbs will have an average life of less than* 775 *hours.*

According to the question, we need to compute $P(\overline{X} \leq 775)$. Applying Central Limit Theorem, we get

$$z = \sqrt{n}\left(\frac{\overline{x} - \mu}{\sigma}\right) = 4\left(\frac{775 - 800}{40}\right) = -2.5.$$

Hence $P(\overline{X} < 775) = P(Z < -2.5) = 0.0062$. (Use standard normal tables.)

Question-17:- If all possible samples of size 16 are drawn from a normal population with mean 50 and standard deviation 5, what is the probability that the sample mean $\overline{X}$ will fall in the interval from $\mu_{\overline{X}} - 1.9\sigma_{\overline{X}}$ to $\mu_{\overline{X}} - 0.4\sigma_{\overline{X}}$? Assume that the sample means can be measured to any degree of accuracy.

Answer:- We have to find $P(\mu_{\overline{X}} - 1.9\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} - 0.4\sigma_{\overline{X}})$. Substituting the values of $\mu_{\overline{X}} = 50$ and $\sigma_{\overline{X}} = \frac{5}{4}$, then we have

$$P(\mu_{\overline{X}} - 1.9\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} - 0.4\sigma_{\overline{X}}) = P(-1.9 < Z < -0.4) = P(Z < -0.4) - P(Z < -1.9).$$

Using standard normal tables we get the required probability $= 0.3446 - 0.0287 = 0.3159$.

Question-23:- The random variable X, representing the number of cherries in a cherry puff, has the following probability distribution $P(X = 4) = 0.2$, $P(X = 5) = 0.4$, $P(X = 6) = 0.3$, $P(X = 7) = 0.1$, then find the mean $\mu_{\overline{X}}$ and variance $\sigma^2_{\overline{X}}$ of the mean $\overline{X}$ for random samples of 36 cherry puffs.

Answer:- Direct calculation gives $\mu = 5.3$ and $\sigma^2 = 0.81$. Hence $\mu_{\overline{X}} = 5.3$ and $\sigma^2_{\overline{X}} = \frac{0.81}{36} = 0.0225$.

Assignment:-Q-19, 20, 24

Sampling Distributions Contd.(Lect. 27)

1 Sampling Distributions of S^2

In the class, we have studied about the sampling distribution of $\overline{X}$. In this Section we will get to know about the sampling distribution of S^2. First let us study two important continuous distributions such as χ^2 and Student's t-distribution.

Definition 1 *Let a continuous random variable X has Chi-squared distribution with n degrees of freedom(DF), then it has the density*

$$f(x) = \frac{1}{\Gamma(1/2)2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ x > 0.$$

Remark 1.1 *We say $X \sim \chi_n$, then $X \sim G(n/2, 2)$ distribution. See figure 1.*

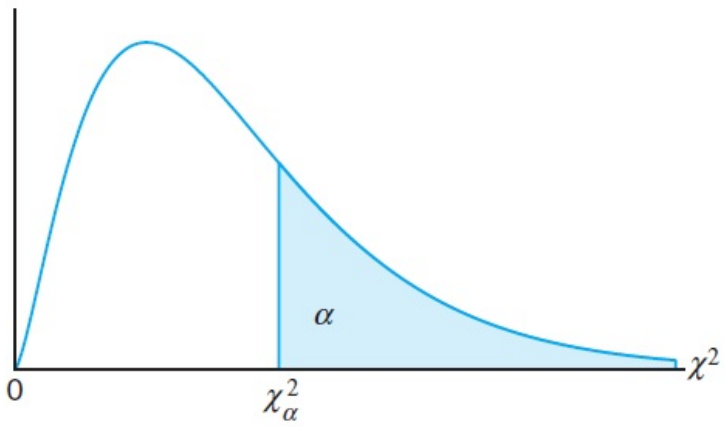


Figure 1: The chi-squared distribution

Define χ^2_α such that $P(\chi^2 > \chi^2_\alpha) = \alpha$, which is also called $100(1-\alpha)\%$ critical point. It is equal to the area under to the curve to the right of this value. Table of critical values of Chi-Squared distribution have been tabulated. students are advised to find out these critical points from the table for various values of α and v. For example if $v = 7$ and $\alpha = 0.05$ yields the value of $\chi^2_{0.05} = 14.067$.

Theorem 1.1 *If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic*

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom.

Theorem 1.2 *Let Z be a standard random variable and V a chi-squared random varibale with v DF. If Z and V are independent, then the distribution of the random varibale T, where*

$$T = \frac{Z}{\sqrt{V/v}}$$

is given by the density function

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}, \quad -\infty < t < \infty.$$

This is known as the **t-distribution** with v DF. As a consequence of the above result, we may state the following corollary.

Corrolary 1.1 *Let $X_1, X_2, \ldots, X_n$ be n be independent normal random variables with mean μ and variance σ^2. Then the random variable $T = \sqrt{n}\left(\frac{\overline{X}-\mu}{S}\right)$ has a t-distribition with $v = n-1$ DF.*

The distribution of T is similar to the distribution of Z in that both are symmetric about a mean of 0. However for large values of n, both the distributions behave alike. we can conclude the following theorem.

Theorem 1.3 *Let T follows a t-distribution with n DF. As $n \to \infty$, the density of T converges to a $N(0,1)$ distribution.*

Usually, when $n \geq 30$, both the distribution behave same. Students can find out the value of t_α, where t_α such that $P(T > t_\alpha) = \alpha = P(T < -t_\alpha)$, see figure. For example,

Example 1 *The t-value with $v = 14$ DF that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is*

$$t_{0.975} = -t_{0.025} = -2.145.$$

Example 2 *Find $P(-t_{0.025} < T < t_{0.05})$.*
Answer:- *Since $t_{0.05}$ leaves an area 0.05 to the right, and $-t_{0.025}$ leaves an area of 0.025 to the left, we find a total area of*

$$1 - 0.05 - 0.025 = 0.925$$

between $-t_{0.025}$ and $t_{0.05}$. Hence

$$P(-t_{0.025} < T < t_{0.05}) = 0.925.$$

Assignments- Q-37, 40, 41, 45 of page 285.

One-Sample Estimation Problems(Lect. 28)

In the last lecture we have studies about the sampling distributions of sample mean and sample varience. In this lecture, we give a brief introduction about the purpose of **Statistical Inference**. We follow this by discussing the Problem of **Interval Estimation or Confidence Intervals**.

1 Statistical Inference

It is a branch of Statistics which consists of those methods by which one can make inferences about a populations. Many of the real world situations are stochastic or probabilistic in nature. So, a population may contain some unknown parameters whose values we may not be knowing. In order to infer about the population parameters, we tahe a random sample taken from the population and then try to infer it. For example,
1. Amount of rain fall during a monsoon season is a random variable. It's not sure that how much rain will fall? How much rain will fall tomorrow?
2. Time taken by patients to get cured by a disease while going under a particular treatment? or say the effect of medicine on different patients is a study of our concern.
3. Number of persons in a service queue or a ticket counter. Some day we may have large people standing on a queue and some other day there may be less.
some typical scientific problems may be adressed in this connection namely, **Quality of production, Avarage monthly salary, How much increase in temparature will be there globally, effectiveness of biochemical new drug** etc. To all these we can apply certain methods to infer about these populations. Broadly, Statistcal is divided into the following catagories depending upon the nature of the problem.
1. **Confidence Intervals.**
2. **Point Estimation.**
3. **Testing of Statistical Hypothesis.** We will study in detail about this in our future classes.

2 Confidence Intervals

An interval estimate of a population parameter θ, is an interval of the form $\hat{\theta_L} < \theta < \hat{\theta_U}$, where $\hat{\theta_L}$ and $\hat{\theta_U}$ depend upon the random sample $X_1, X_2, \ldots, X_n$. Moreover, we would be interested to estimate the confidence

interval of θ with a confidence level $(1-\alpha)$ i.e

$$P(\hat{\theta_L} < \theta < \hat{\theta_U}) = 1-\alpha, \quad 0 < \alpha < 1. \tag{1}$$

The above interval is called the $100(1-\alpha)\%$ C.I for θ. $(1-\alpha)$ is called the confidence level or degree of confidence. Normally, we take $1-\alpha = 90\%, 95\%, 99\%$ e.t.c. For example, the average life of a T.v lies within an interval.

3 Estimating The Mean μ Of A Normal Population When σ^2 Is Known

Here, we try to find out $\hat{\mu_L}$ and $\hat{\mu_U}$ such that

$$P_{\sigma^2=known}(\hat{\mu_L} < \theta < \hat{\mu_U}) = 1-\alpha$$

Let $X_1, X_2, \ldots, X_n$ be a random sample taken from $\mathcal{N}(\mu, \sigma^2)$ distribution where σ^2 is known. From the Normal graph, we see that So, we have

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha \tag{2}$$

where $Z = \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$. So, from eq.(2), we have

$$P\left(-Z_{\alpha/2} < \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma} < Z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \quad P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

So, we have

$$\begin{aligned} \hat{\theta_L} &= \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ \hat{\theta_U} &= \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \end{aligned}$$

where $Z_{\alpha/2}$ is such that $P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$.

So, $\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ is a $100(1-\alpha)\%$ C.I for μ.

Interval estimation for $'\mu'$ when σ^2 is not known, i.e, we wish to find

$$P(\hat{\theta_L} < \theta < \hat{\theta_U}) = 1-\alpha.$$

Question:- The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 gm/ml. Find a 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume normality and $\sigma = 0.3$.

Answer:- The point estimate of $\mu = \overline{x} = 2.6$, $(1 - \alpha = 0.95, 0.99)$
$\Longrightarrow$ $\alpha = 0.05, \alpha/2 = 0.025$ or $\alpha = 0.01, \alpha/2 = 0.005$.
So, $Z_{0.025} = 1.96$ and $Z_{0.005} = 2.575$.
So, the 95% and 99% C.I for μ is thus given by

$$2.6 - (1.96)\Big(\frac{0.3}{\sqrt{36}}\Big) < \mu < 2.6 + (1.96)\Big(\frac{0.3}{\sqrt{36}}\Big)$$

$$\text{and} \qquad 2.6 - 2.575\Big(\frac{0.3}{\sqrt{36}}\Big) < \mu < 2.6 + 2.575\Big(\frac{0.3}{\sqrt{36}}\Big)$$

Theorem 3.1 *If $\overline{X}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.*

Proof:- Check

$$P\Big(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha \qquad \text{or}$$

$$P\Big(0 < \mid \overline{X} - \mu \mid < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha$$

The term $\Big[e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big]$ is called error in estimating $'\mu' \Rightarrow n = \Big(\frac{Z_{\alpha/2}\sigma}{e}\Big)^2$.

Example 1 *How large a sample is required if we want to be 95% confident that our estimate of μ in the above example is off by less than 0.05.*

Answer:- $\sigma = 0.3$ and $e = 0.05$, hence $n = \Big[\frac{(1.96)(0.3)}{0.05}\Big]^2 = [138.3] = 139$.
Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\overline{X}$ differs from μ by an amount less than 0.05.

Example 2 *Q-2 An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a 96% confidence interval for the population mean of all bulbs produced by this firm.*

Answer:-Here $n = 30$, $\overline{x} = 780$ and $\sigma = 40$. Also $Z_{0.02} = 2.054$, so a 96% confidence interval for the mean of the population is given by $765 < \mu < 795$.

Example 3 *Q-3 Many cardiac patients wear an implanted pacemaker to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 inch and an approximately normal distribution, find a* 95% *confidence interval for the mean of the depths of all connector modules made by a certain manufacturing company. A random sample of 75 modules has an average depth of 0.310 inch.*

Answer:- Here $n = 75$, $\overline{x} = 0.310$ and $\sigma = 0.0015$. Also $Z_{0.025} = 1.96$. A 95% confidence interval for the mean of the population is given by $0.3097 < \mu < 0.3103$.

One-Sample Estimation Problems(Lect. 29)Contd..

In the previous lecture, we have constructed a $100(1-\alpha)\%$ CI for the normal mean μ when the variance σ^2 is known. This section refers to the estimation of a normal mean and variance when the variance σ^2.

1 Estimating The Mean μ Of A Normal Population When σ^2 Is Unknown

Applying the same argument as above, we have

$$P(-t_{\alpha/2} < \mathcal{T} < t_{\alpha/2}) = 1-\alpha \tag{1}$$

where

$$\mathcal{T} = \sqrt{n}\frac{(\overline{X}-\mu)}{s} \sim t_{\alpha,n-1}$$

so, from eq(1), we have

$$P\left(-t_{\alpha/2} < \sqrt{n}\frac{(\overline{X}-\mu)}{s} < t_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \quad P\left(\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha t_{\alpha/2}$$

So a $100(1-\alpha)\%$ C.I for μ is given by $\left(\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$

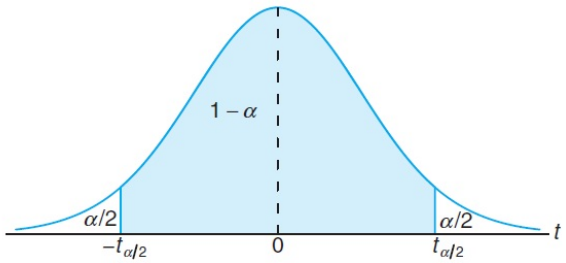


Figure 1: $P(-t_{\alpha/2} < \mathcal{T} < t_{\alpha/2}) = 1-\alpha$

Question:-The contents of several similar containers of Sulphuric acid are $9.8, 10.2, 10.4, 9.8, 10.0, 10.2$ and 9.6 litres. Find a 95% C.I for the mean

contents of all such containers assuming normality.

Answer:- σ is not known. $\overline{X} = 10.0, s = 0.283, t_{0.005} = 2.447$ at $v = n-1 = 6$ degrees of freedom. Hence, the $100(1-\alpha)\%$ C.I is thus given by

$$10.0 - 2.447\Big(\frac{0.283}{\sqrt{7}}\Big) < \mu < 10.0 + 2.447\Big(\frac{0.283}{\sqrt{7}}\Big) \;\; \Rightarrow \;\; (9.74 < \mu < 10.26)$$

Question:- SAT mathematics scores of a of 500 PG students in Odisha show $\overline{X} = 501, s = 112$. Find a 99% confidence interval.

Answer:- $488.1 < \mu < 513.9$.
Assignments:-Q-$4-7$.

2 Estimating The Varience σ^2 Of A Normal Population When Mean μ May Be Known Or Unknown

We know that $\chi^2 = \frac{(n-1)S^2}{\chi^2} \sim \chi^2_{n-1}$, from the graph of χ^2 distribution, we have

$$P(\chi^2_{1-\alpha/2} < \sigma^2 < \chi^2_{\alpha/2}) = 1-\alpha \tag{2}$$

Substituting the values of χ^2 and on further simplification we have

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \chi^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha.$$

So the $100(1-\alpha)\%$ CI for σ^2 is thus given by

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) \tag{3}$$

Example 1 *The following are the weights in decagrams, of* 10 *packages of grass seed distributed by a certain company,* 46.4, 46.1, 45.8, 47, 46.1, 45.9, 45.8, 46.9, 45.2 *and* 46. *Find a* 95% *CI for the variance of weights of all such packages of hrass seed distributed by the company. Assume normality.*

Answer:- Here we see that the observed sample varience

$$s^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^n x_i^2 - \Big(\sum_{i=1}^n x_i\Big)^2\right] = 0.286.$$

From χ^2 distribution table, we have $\chi^2_{0.025,9} = 19.023$ and $\chi^2_{0.975,9} = 2.700$. So the CI for σ^2 can be estimated as

$$0.135 < \sigma^2 < 0.953.$$

Example 2 *Q-72 A random sample of 20 students yielded a mean of* $\overline{x} = 72$ *and a variance of* $s^2 = 16$ *for scores on a college placement test in mathematics. Assuming the scores to be normally distributed, construct a* 98% *confidence interval for* σ^2*.*

Answer:-Here $s^2 = 16$ with $v = 19$ DF. It is known $\chi^2_{0.01} = 36.191$ and $\chi^2_{0.099} = 7.633$. Hence substituting all the values in equation (3), the CI for σ^2 is thus estimated as

$$8.400 < \sigma^2 < 39.827.$$

Students are advised to practice the following assignment problems.
Assignments:-Q- 73, 77.

Maximum Likelihood Estimation (MLE) (Lecture-30 & 31)

After covering the idea of estimating the CI of an unknown parameter, we will proceed to introduce the problem of Point Estimation. There are several procedures to find a point estimate of unknown parameter θ such as Method of Moments(MME), Unbiased Estimator, Bayes Estimators, Maximum Likelihood Estimation(MLE). Out of so many procedures available, we will derive the MLE of θ. First, let us discuss some of its basic concepts.

1 Basic Concepts of Point Estimation

Definition 1.1 Estimator*:- Any function of the random variables which is used to estimate the unknown value of the given parametric function $g(\theta)$ is called an estimator.*

If $\underline{X} = (X_1, \ldots, X_n)$ is a random sample from a population with the probability distribution P_θ, a function $d(X)$ used for estimating $g(\theta)$ is known as an estimator. Let $\underline{x} = (x_1, \ldots, x_n)$ be a realization of $\underline{X}$, then $d(\underline{x})$ is called an estimate.

Example 1.1 *Let X be a random variable denoting the average height of Adult males in an ethnic group. We may use $\underline{X}$(Sample Mean) as an estimator. Now, if a random sample of* 50 *has a sample mean* 180, *then* $180cm$ *is an estimate of the average height.*

Definition 1.2 *Let $\underline{x} = (x_1, \ldots, x_n)$ be an observed random sample. Define Likelihood function of $\underline{x}$ as*

$$L(\underline{\theta}, \underline{x}) = \prod_{i=1}^{n} f(x_i, \underline{\theta})$$

The value of θ, say $\hat{\theta}(\underline{x})$, so that $L(\hat{\underline{\theta}}, \underline{x}) \geq L(\underline{\theta}, \underline{x})$ is called the MLE of θ. It may be noted that, the idea of taking the likelihood function is that it contains all the informations of the parameters. In order to find out the MLE of an unknown parameter θ is to find out the value of θ for which the likelihood function is maximum.

Example 1.2 *Consider a random sample $x_1, x_2, ..., x_n$ from a normal distribution $\mathcal{N}(\mu, \sigma)$. Find the maximum likelihood estimators for μ and σ^2.*

Solution:- The likelihood function for the normal distribution is

$$L(x_1, \cdots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \Big(\frac{x_i - \mu}{\sigma} \Big)^2 \right].$$

Taking logarithms gives us

$$lnL(x_1, \cdots, x_n; \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \Big(\frac{x_i - \mu}{\sigma} \Big)^2.$$

Hence,

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n} \Big(\frac{x_i - \mu}{\sigma^2} \Big)$$

and

$$\frac{\partial \ln L}{\partial \mu} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} \big(x_i - \mu \big)^2.$$

Setting both derivatives equal to 0, we obtain

$$\sum_{i=1}^{n} x_i - n\mu = 0 \quad \text{and} \quad n\sigma^2 = \sum_{i=1}^{n} \big(x_i - \mu \big)^2.$$

Thus, the maximum likelihood estimator of μ is given by

$$\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X},$$

which is a pleasing result since $\overline{x}$ has played such an important role in this chapter as a point estimate of μ. On the other hand, the maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Checking the second-order partial derivative matrix confirms that the solution results in a maximum of the likelihood function.

Example 1.3 *Suppose* 10 *rats are used in a biomedical study where they are injected with cancer cells and then given a cancer drug that is designed to increase their survival rate. The survival times, in months, are* $14, 17, 27, 18, 12, 8, 22, 13, 19$ *and* 12. *Assume that the exponential distribution applies. Give a maximum likelihood estimate of the mean survival time.*

Solution:- From Chapter 6, we know that the probability density function for the exponential random variable X is

$$f(x,\beta) = \begin{cases} \frac{1}{\beta}e^{-x/\beta}, & x > 0, \\ 0, & elsewhere. \end{cases}$$

Thus, the log-likelihood function for the data, given $n = 10$, is

$$\ln L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \ln \beta - \frac{1}{\beta}\sum_{i=1}^{10} x_i.$$

Setting

$$\frac{\partial \ln L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{10} x_i = 0$$

implies that

$$\hat{\beta} = \frac{1}{10}\sum_{i=1}^{10} x_i = \overline{x} = 16.2.$$

Evaluating the second derivative of the log-likelihood function at the value $\hat{\beta}$ above yields a negative value. As a result, the estimator of the parameter β, the population mean is the sample average $\overline{x}$.

Example 1.4 *Q-81 Suppose that there are n trials $x_1, x_2, ..., x_n$ from a Bernoulli process with parameter p, the probability of a success. That is, the probability of r successes is given by $\binom{n}{r}p^r(1-p)^{n-r}$. Work out the maximum likelihood estimator for the parameter p.*

Solution:-

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}, x_i = 0, 1, 2, \ldots, n \text{ and } 0 < p < 1(\text{unknown}).$$

$$logL = \sum_{i=1}^{n} log\binom{n}{x_i} + \sum_{i=1}^{n} x_i \; logp + \sum_{i=1}^{n}(n - x_i) \; log(1-p).$$

$$\frac{\partial L}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n^2 - \sum_{i=1}^{n} x_i}{1-p} = \frac{\sum_{i=1}^{n} x_i - n^2 p}{p(1-p)} \begin{cases} > 0 & p < \frac{\sum_{i=1}^{n} x_i}{n} \\ < 0 & p > \frac{\sum_{i=1}^{n} x_i}{n}. \end{cases}$$

From the above expression we get $\hat{p}_{MLE} = \frac{\sum_{i=1}^{n} X_i}{n}$. Note that if the sample size is 1, then MLE pf p will be equal to $\hat{p}_{MLE} = \frac{X}{n}$.

Example 1.5 *Q-85 Consider a random sample of $x_1, x_2, ..., x_n$ from a uniform distribution $U(0, \theta)$ with unknown parameter θ, where $\theta > 0$. Determine the maximum likelihood estimator of θ.*

Solution:- Let $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta), \theta$ is unknown. So, $\hat{\theta}_{MLE} = X_{(n)} = \max(X_1, \ldots, X_n)$. We observe that the likelihood function $L = \frac{1}{\theta^n}$ is a decreasing function of θ and hence the maximum occurs at the lower bound.

Assignment:- Consider a hypothetical experiment where a man with a fungus uses an antifungal drug and is cured. Consider this, then, a sample of one from a Bernoulli distribution with probability function $f(x) = p^x q^{1-x}, x = 0, 1$, where p is the probability of a success (cure) and $q = 1 - p$. Now, of course, the sample information gives $x = 1$. Write out a development that shows that $\hat{p} = 1.0$ is the maximum likelihood estimator of the probability of a cure.

LECTURE-32 and 33

ONE AND TWO-SAMPLE TESTS OF HYPOTHESES

General Concepts(CH-10.1, 10.2, 10.3)

Hypothesis: A statement regarding a parameter of a distribution is called hypothesis.

Statistical Hypothesis: A statistical hypothesis is an assertion in one or more population.

Null Hypothesis: A hypothesis whose truth value is tested is a null hypothesis.

Alternative Hypothesis: A hypothesis which is true when the null hypothesis is rejected i.e any hypothesis which is complimentary to the null hypothesis is called an alternative hypothesis. It is denoted by H_1 i.e $H = H_1$

Eample:- Mean height of all students is $\mu = 5'9''$ not accepted.

The hypothesis is tested using sample.

Then $\mu \neq 5'9''$ or $\mu < 5'9''$ or $\mu > 5'9''$

There are two hypothesis

(i) Null hypothesis (H_0)

(ii) Alternative hypothesis (H_1)

Null hypothesis is formed for rejection.

Once the hypothesis are formed, test statistics are raised to test the hypothesis.

(H_0) accepted or rejected $\Rightarrow$ (H_1) rejected or accepted.

Critical Value(Significant Value): The value of the test static which separate the rejection region and acceptance region is called critical value. It is denoted by C.

Types of Test:

(1) $H \neq H_0$ (Two-sided test/Two tail test)

(2) $H > H_0$ (Right-sided test/One-sided test)

(3) $H < H_0$ (Left-sided test/One-sided test)

Left Sided Test (L.S.T.): Suppose we want 35 kg. wheat.

If it is >35 ⇒ accept

<35 ⇒ reject

but if it is 34.99kg. then we can accept, like this, up to what limit we can tolerate. See figure:-1

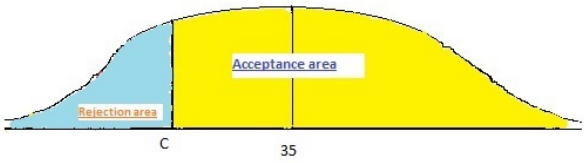


Figure 1: Left Sided Test

Where C is the critical point i.e tolerance limit.

Right Sided Test (R.S.T.):

Eample:- Suppose our budget is 10,000 i.e Null Hypothesis(NH)

If it is >10,000 ⇒ reject

<10,000 ⇒ accept. See figure:-2

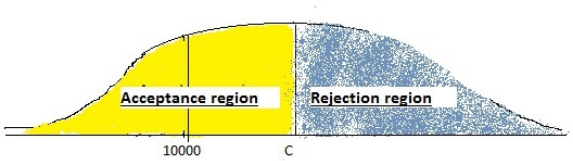


Figure 2: Right Sided Test

But if it is 10,005/10,010/10,015 ⇒ accept.

i.e upto a tolerance limit we can accept but after that we cann't tolerate.

i.e C lies in the right side.

Two Sided Test (T.S.T.):

Eample:- Suppose one pin has to fit a hole (*depends on diameter*). If the diameter of pin is very much more or very less, then we reject but small deviation is acceptable.

Eample:-Suppose your shirt size is 40, then 44 size or 36 size not manageable. But 41 or 39 is manageable, if $C_1 < 41, 39 < C_2$ i.e shirt size is manageable if $C_1 < shirtsize < C_2$. Otherwise rejected. See figure:-3

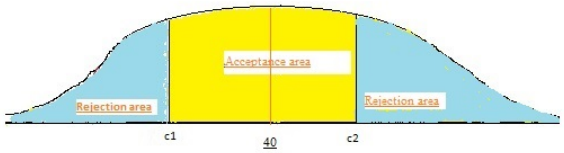


Figure 3: Two Sided Test

From previous chapter we got an idea that sample result will give the population result.

Whether it is perfect or not we are testing i.e testing of hypothesis.

i.e. we have to take the decision whether we accept or reject.

Testing of Hypothesis:

Procedure:

(i) Set up a null hypothesis i.e $H = H_0$ which is to be tested.

(ii) Set up an alternative hypothesis i.e $H = H_1$ against null hypothesis.

i.e $H \neq H_0$ or $H > H_0$ or $H < H_0$.

(iii) Choose a significance level $\alpha(5\ \%\ , 1\ \%\ , .1\ \%\ , ...)$.

(iv) Use an appropriate random variable and determine the observed value i.e $\hat{H}$.

(v) Find the tabular value (critical value) i.e C from the corresponding table with given level of significance depends on alternative hypothesis.

Significance level α: The probability of the value of the variate falling in the rejection region.

i.e the percentage of tolerance limit of error.

Table 1: Decision Table

	H_0 is true	H_1 is true
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Error:

Type-I error: When a null hypothesis is true but we reject.

P(reject null hypothesis when it is true)=α

Type-II error: When a null hypothesis is false but we accept it.

P(accept a null hypothesis when it is false)=β

Book Questions

10.2 A sociologist is concerned about the effectiveness of a training course designed to get more drivers to use seat belts in automobiles.

(a) What hypothesis is she testing if she commits a type I error by erroneously concluding that the training course is ineffective?

Ans: (a) The training course is effective.

(b) What hypothesis is she testing if she commits a type II error by erroneously concluding that the training course is effective?

Ans: (b) The training course is effective.

10.3 A large manufacturing firm is being charged with discrimination in its hiring practices.

(a) What hypothesis is being tested if a jury commits a type I error by finding the firm guilty?

Ans: (a) The firm is not guilty.

(b) What hypothesis is being tested if a jury commits a type II error by finding the firm guilty?

Ans: (b) The firm is guilty.

10.4 A fabric manufacturer believes that the proportion of orders for raw material arriving late is p = 0.6. If a random sample of 10 orders shows that 3 or fewer arrived late, the hypothesis that p = 0.6 should be rejected in favor of the alternative p ¡ 0.6. Use the binomial distribution.

(a) Find the probability of committing a type I error if the true proportion is p = 0.6.

Ans: (a) $\alpha = P(X \leq 5|(p = 0.6)) + P(X \geq 13|(p = 0.6)) = 0.0338 + (1 - 0.9729) = 0.0609$

(b) Find the probability of committing a type II error for the alternatives p = 0.3, p = 0.4,

and p = 0.5.

Ans: (a) $\beta = P(6 \leq X \leq 12|(p = 0.5)) = 0.9963 - 0.1509 = 0.8454$

$\beta = P(6 \leq X \leq 12|(p = 0.7)) = 0.8732 - 0.0037 = 0.8695$

Chapter-10.3:

Power of Test: $\eta = 1 - \beta$ is called Power of Test (It is the probability of rejection of null hypothesis given that a specific alternative hypothesis is true).

Some Results from Confidence Interval:

Standardized R.V.

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Case-I:The R.V.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{2}$$

has a normal distribution.

Case-II:The R.V.

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \tag{3}$$

has a t-distribution with $n - 1$ degrees of freedom.

Case-III:The R.V.

$$Y = (n - 1)\frac{S^2}{\sigma^2} \tag{4}$$

has a Chi-square distribution with $n - 1$ degrees of freedom.

LECTURE-34

ONE AND TWO-SAMPLE TESTS OF HYPOTHESES

CH-10.4: Test concerning a single mean using single sample:

Let the population is mean μ we want to test

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

Case-I: The sample is drawn from the normal population where population variance is known.

$Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$

Test concerning a single mean using single sample:-

(i) Variance is known $Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$

(ii) Variance is not known $t = \frac{X-\mu}{\frac{s}{\sqrt{n}}}$

Types of Test:

(i) $H_0 : \mu = \mu_1$

$H_1 : \mu > \mu_1$

(ii) $H_0 : \mu = \mu_1$

$H_1 : \mu < \mu_1$

(iii) $H_0 : \mu = \mu_1$

$H_1 : \mu \neq \mu_1$

For (i), if $H_0 : \mu = \mu_1$

$H_1 : \mu > \mu_1$

Then test statistic $Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$ is normally distributed $N(0, 1)$

Let α be the level of significance. Take $\alpha = 0.01$.

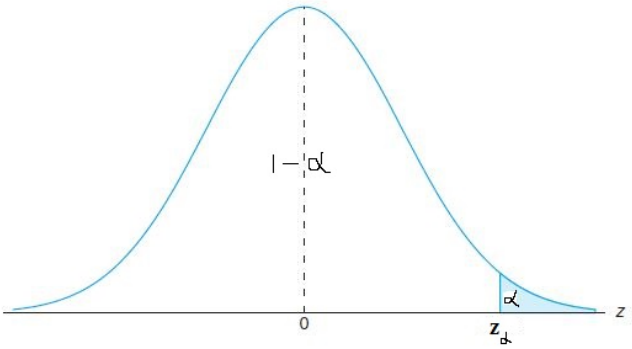


Figure 1: Right Sided Test

Now the value of z is calculated using the sample observation $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

From the normal distribution table find z_α such that $P(z > z_\alpha) = \alpha$

If $z < z_\alpha$ then the H_0 is accepted.

If $z \geq z_\alpha$ then the H_0 is rejected.

(ii) If $H_0 : \mu = \mu_1$

$H_1 : \mu < \mu_1$

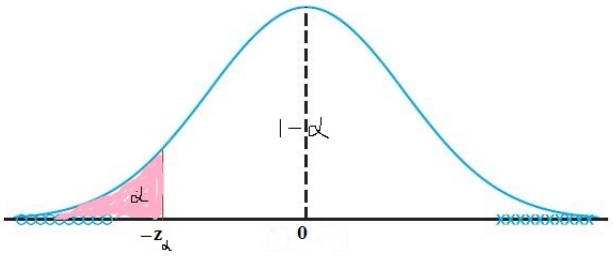


Figure 2: Left Sided Test

Then test statistic $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

From the normal distribution table find z_α such that $P(z > z_\alpha) = \alpha$

If $z \leq -z_\alpha$ then the H_0 is rejected.

If $z > -z_\alpha$ then the H_0 is accepted at α level of significance.

(iii) If $H_0 : \mu = \mu_1$

$H_1 : \mu \neq \mu_1$

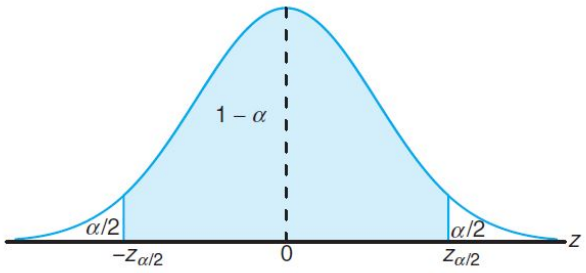


Figure 3: Two Sided Test

Then observation value is $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

If $-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$ then the H_0 is accepted.

If $z > z_{\frac{\alpha}{2}}$ or If $z < -z_{\frac{\alpha}{2}}$ then the H_0 is rejected.

(i) **When variance is known**:, Test statistic is $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Example 10.3: A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Ans: Given $n = 100$

$\bar{X} = 71.8$

$\sigma = 8.9$

Mean life span is greater than 70 year.

$H_0 : \mu = 70$

$H_1 : \mu > 70$ (It is a right tail test)

$\alpha = 0.05$

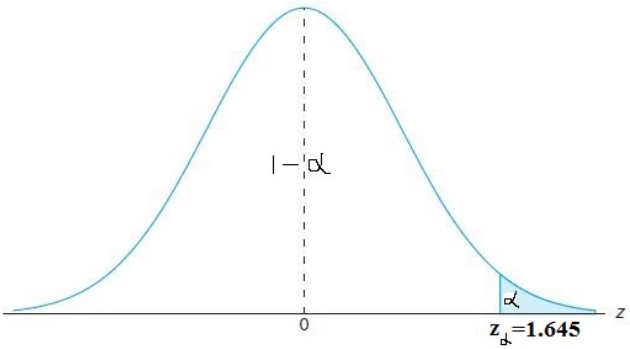


Figure 4: Right Sided Test

$z_\alpha = 1.645$

$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{71.8-70}{\frac{8.9}{\sqrt{100}}} = 2.02$

Since computed value of z > tabulated value $\Rightarrow$ The statistic fall in the rejection region or critical region. So, H_0 is rejected at 0.05 level of significance. Hence, average life span is greater than 70 year.

Example 10.4: A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

Ans: Given, significance level = α

$\bar{x} = 8$ kilogram

$\mu = 8$, $\mu \neq 8$, $\mu \longrightarrow$ mean

(n) no. of samples=50,

($\bar{x}$) avg strength = 7.8 kg,

(σ) standard deviation = 0.5 kg

sol^n: $H_0 : \mu = 8$

$H_1 : \mu \neq 8$

This is a two tail test.

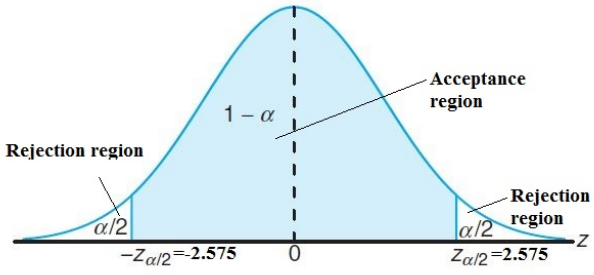


Figure 5: Two Sided Test

$P(z < -z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

From the table, $-z_{\frac{\alpha}{2}} = -2.575$

$z_{\frac{\alpha}{2}} = 2.575$

Which are two critical values

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7.8-8}{\frac{0.5}{\sqrt{50}}} = -2.85$

Computed values of $z = -2.85 <$ tabulated value (-2.575)

So, H_0 is rejected at 0.01 level of significance.

Problem-10.20: A random sample of 64 bags of white cheddar popcorn weighed, on average, 5.23 ounces with a standard deviation of 0.24 ounce. Test the hypothesis that $\mu = 5.5$ ounces against the alternative hypothesis, $\mu < 5.5$ ounces, at the 0.05 level of significance.

Ans: Given $n = 64$, $\bar{x} = 5.23$, $\mu = 5.5$, $\sigma = 0.24$, $\alpha = 0.05$

$H_0 : \mu = 5.5$

$H_1 : \mu < 5.5$

Tabulated value is $z_\alpha = -1.645$ $(From\ \ Normal\ \ distribution\ \ table\ \ \frac{-1.65-1.64}{2} = -1.645)$

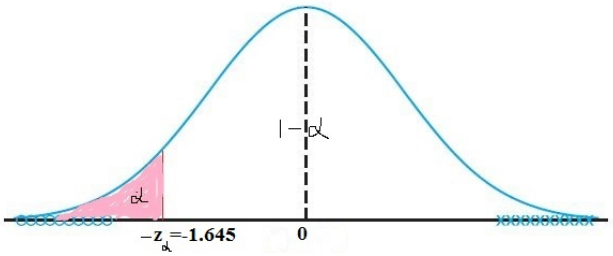


Figure 6: Left Sided Test

Computed value is $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.23-5.5}{\frac{0.24}{\sqrt{64}}} \frac{-0.27}{0.03} = -9$

Since computed value $<$ tabulated value, so H_0 is rejected.

LECTURE-35

ONE- AND TWO-SAMPLE TESTS OF HYPOTHESES

Rest of CH-10.4 and 10.5:

(ii) **When variance is unknown**:

In this case population variance is estimate as sample variance s^2. Test statistic is $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ with (n-1) degrees of freedom.

Types of Test:

For (i) $H_0 : \mu = \mu_1$

$H_1 : \mu > \mu_1$

This is right tail test for α level of significance.

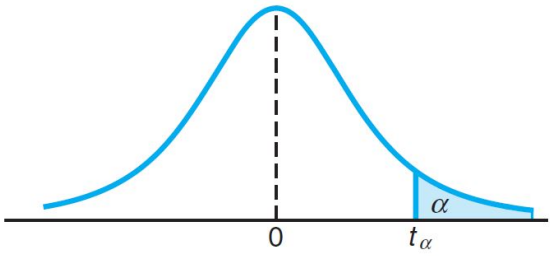


Figure 1: Two Sided Test

Find the tabular value of t_α from t-distribution table with $(n-1)$ degrees of freedom.

Find the computed value of t by using $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ at $\alpha = 0.1(say)$.

If computed value $<$ tabulated value then accept H_0.

If computed value $\geq$ tabulated value then reject H_0.

For (ii) $H_0 : \mu = \mu_1$

$H_1 : \mu < \mu_1$

This is left tail test for α level of significance.

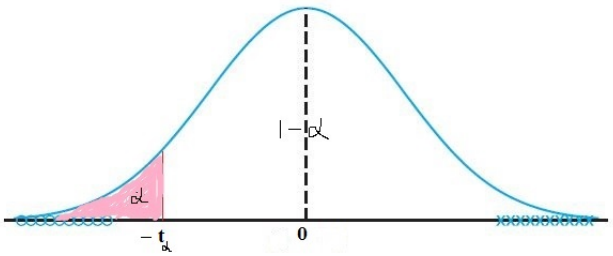


Figure 2: Left Sided Test

Find the tabular value of $-t_\alpha$ from t-distribution table with $(n-1)$ degrees of freedom.

Find the computed value of t by using $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ at $\alpha = 0.1(say)$.

If computed value $(t) >$ tabulated value $(-t_\alpha)$ then, accept H_0.

If computed value $\leq$ tabulated value then reject H_0.

Example 10.5: The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Ans: Given $H_0 : \mu = 46$

$H_1 : \mu < 46$

level of significance $\alpha = 0.05$

$n = 12$, $s = 11.9$, $\bar{x} = 42$

This is a left tail test.

Since variance σ^2 is unknown, so test statistic is

$t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$

which follows t-distribution with (n-1) degrees of freedom.

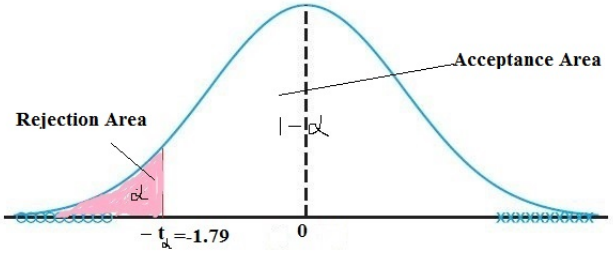


Figure 3: Left Sided Test

The tabulated value of t_α at $12-1=11$ degrees of freedom is 1.79 (from t-distribution table).

So the tabulated value of $-t_\alpha$ at $12-1=11$ degrees of freedom is -1.79.

Now $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{42-46}{\frac{11.9}{\sqrt{12}}} = -1.16$

Since computed value $>$ tabulated value, so H_0 is accepted.

10.29:

Past experience indicates that the time required for high school seniors to complete a standardized test is a normal random variable with a mean of 35 minutes. If a random sample of 20 high school seniors took an average of 33.1 minutes to complete this test with a standard deviation of 4.3 minutes, test the hypothesis, at the 0.05 level of significance, that $\mu = 35$ minutes against the alternative that $\mu < 35$ minutes.

Ans: Given $\mu = 35$, $n = 20$, $\bar{x} = 33.1$, $s = 4.3$, $\alpha = 0.5$, $\mu < 35$

$H_0 : \mu = 35$

$H_1 : \mu < 35$

$d.\ f.\ = n - 1 = 20 - 1 = 19$

The tabulated value of t_α at 19 $d.f. = 1.729$.

The tabulated value of $-t_\alpha$ at 19 $d.f. = -1.729$.

$$t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{33.1-35}{\frac{4.3}{\sqrt{20}}} = \frac{-1.9}{(0.96)^2} = -1.89$$

Since computed value $<$ tabulated value, so H_0 is rejected.

It takes less than 35 minutes on the average to the take test.

For (iii) $H_0 : \mu = \mu_1$

$H_1 : \mu \neq \mu_1$

This is two tail test for α level of significance i.e total area is $\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$

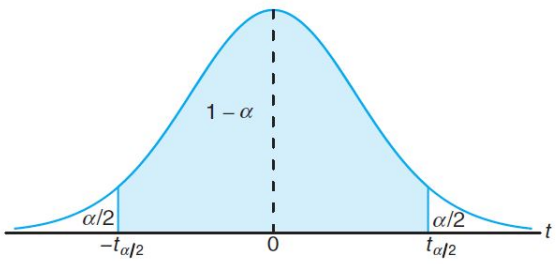


Figure 4: Two Sided Test

Find the tabular value of $-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$ from t-distribution table with $(n-1)$ degrees of freedom. If computed value lies between $-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$, then H_0 is accepted.

Book Questions

10.21: An electrical firm manufactures light bulbs that have a lifetime that is approximately normally distributed with a mean of 800 hours and a standard deviation of 40 hours. Test the hypothesis that $\mu = 800$ hours against the alternative, $\mu = 800$ hours, if a random sample of 30 bulbs has an average life of 788 hours. Use a P-value in your answer.

Home Tax

10.23: Test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, and 9.8 liters. Use a 0.01 level of significance and assume that the distribution of contents is normal.

Home Tax

CH-10.5: Two Samples: Test on Two Means:

Test of Hypothesis concerning difference of mean of two population:

1. Population variances are known.

Test statistic is

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

2. Population variances are unknown but equal.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

We reject H_0 at significance level α when the computed t-statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where estimate of population variance s_p^2 is

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

exceeds $t_{\frac{\alpha}{2},\ n_1+n_2-2}$ or is less than $-t_{\frac{\alpha}{2},\ n_1+n_2-2}$.

Example 10.6:

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Ans: Given two populations P_1 and P_2

Table 1: Given

Population	P_1	P_2
sample size	$n_1 = 12$	$n_2 = 10$
sample mean	$\bar{X}_1 = 85$	$\bar{X}_2 = 81$
sample standard deviation	$s_1 = 4$	$s_2 = 5$

$\alpha = 0.05$

$H_0 : \mu_1 - \mu_2 = 2$

$H_1 : \mu_1 - \mu_2 > 2$

As population variance are unknown but equal, we are t-statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$s_p = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$s_p = \sqrt{\frac{16 \times 11 + 25 \times 9}{12 + 10 - 2}} = 4.47$$

T-statistic follows t-distribution with $12 + 10 - 2 = 20$ d.f.

Computed value of t-statistic

$$t = \frac{(85 - 81) - 2}{4.47\sqrt{\frac{1}{12} + \frac{1}{10}}} = 1.04$$

This is a right tail test.

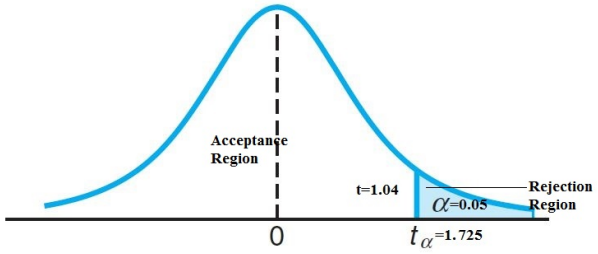


Figure 5: Two Sided Test

From the table t-value for 20 $d.f. = 1.725$

The computed value 1.04 is in acceptance region.

Hence H_0 is accepted at 0.05 level of significance.

(Lect. 36)
10.10 One and two sample test concerning variance

In this section we will discuss how to test population variance or standard deviation using the help of χ^2 distribution table and F distribution table

ONE SAMPLE TEST VARIANCE

Here H_0: Null Hypothesis, the hypothesis, which refer to any hypothesis we wish to test.

H_1: Alternative hypothesis

Procedure for test for variance:

Step 1: select a fixed significance level α

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis $H_0 : \sigma^2 = {\sigma_0}^2$ against the alternative hypothesis H_1, which is one case among three following cases

$$H_1 : \sigma^2 < {\sigma_0}^2 (Case - I)$$
$$H_1 : \sigma^2 > {\sigma_0}^2 (Case - II)$$
$$H_1 : \sigma^2 \neq {\sigma_0}^2 (Case - III)$$

Step 3: Determine

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2}$$

where n is the sample size, s^2 is the sample variance, and ${\sigma_0}^2$ is the value of ${\sigma_0}^2$ given by the null hypothesis.

If H_0 is true, χ^2 is a value of the chi-squared distribution with $v = n - 1$ degrees of freedom.

Step 4: Determine the critical region and fail to reject region based on α

Case-I Critical region: If the alternative hypothesis is $H_1 : \sigma^2 < {\sigma_0}^2$ (case-I) then the critical region is $\chi^2 < \chi^2_{1-\alpha}$.

Fail to reject region: Fail to reject null hypothesis H_0 region is $\chi^2 \geq \chi^2_{1-\alpha}$.

Note: Here we have to determine $\chi^2_{1-\alpha}$ using the equation $P(\chi^2 > \chi^2_{1-\alpha}) = 1 - \alpha$ with n-1 degrees of freedom (use χ^2 distribution table)

(Case-II) Critical Region If the alternative hypothesis is $H_1 : \sigma^2 > {\sigma_0}^2$ (case-II) then the critical region is $\chi^2 > \chi^2_\alpha$

Fail to reject region: Fail to reject null hypothesis H_0 region is $\chi^2 \leq \chi^2_\alpha$

Note: Here we have to determine χ^2_α using the equation $P(\chi^2 > \chi^2_\alpha) = \alpha$ with n-1 degrees of freedom (use χ^2 distribution table)

(Case-III) Critical Region If the alternative hypothesis is $H_1 : \sigma^2 \neq {\sigma_0}^2$ (case-III) then the critical region is $\chi^2 > \chi^2_{\alpha/2}$ or $\chi^2 < \chi^2_{1-\alpha/2}$

Fail to reject region So fail to reject null hypothesis H_0 region is $\chi^2_{1-\alpha/2} \leq \chi^2_{\alpha/2}$

Note: Here we have to determine $\chi^2_{\alpha/2}$ using the equation $P(\chi^2 > \chi^2_{\alpha/}) = \alpha/2$ with n-1 degrees of freedom.

Determine $\chi^2_{1-\alpha/2}$ using the equation $P(\chi^2 > \chi^2_{1-\alpha/}) = 1 - \alpha/2$ with n-1 degrees of freedom (use χ^2 distribution table) or using $\chi^2_{1-\alpha/2} = -\chi^2_{\alpha/2}$

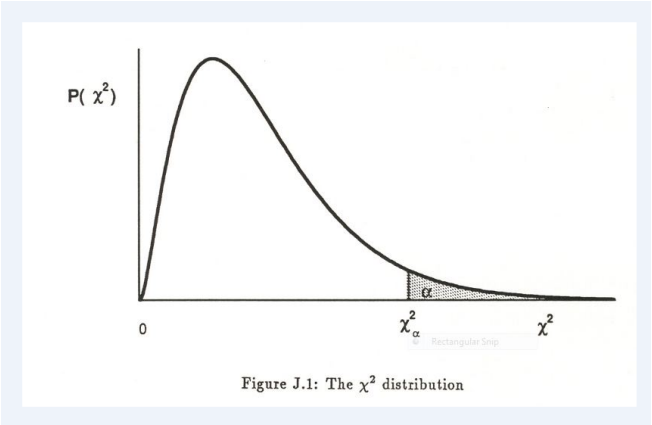


Figure J.1: The χ^2 distribution

Figure 1: **A**

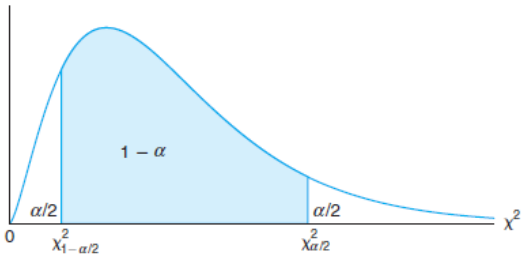


Figure 9.7: $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

Figure 2: **B**

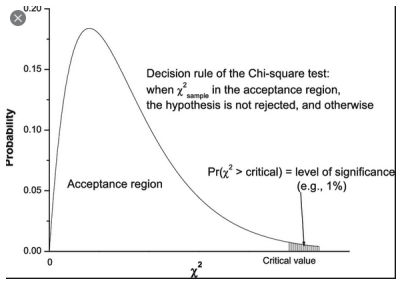


Figure 3: **C**

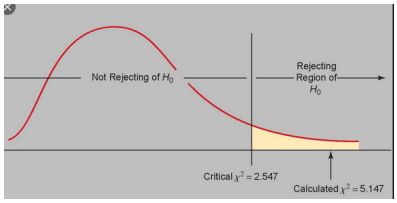


Figure 4: **D**

Question No 67: The content of containers of a particular lubricant is known to be normally distributed with a variance of 0.03 litre. Test the hypothesis that $\sigma^2 = 0.03$ against the

alternative that $\sigma^2 \neq 0.03$ for the random sample of 10 containers in Exercise 10.23 on page 398.

(10.23 Test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, and 9.8 liters. Use a 0.01 level of significance and assume that the distribution of contents is normal.)

Solution: **Step 1:** It is given that α= 0.01 n =sample size= 10,
Step 2: we have to test the Null Hypothesis $H_0 : \sigma^2 = {\sigma_0}^2 = 0.03$ against the alternative hypothesis H_1, which is

$$H_1 : \sigma^2 \neq {\sigma_0}^2 = 0.03(case - III)$$

Step 3: Determine

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2}$$

n =sample size= 10 Now using 10 sample points we have to determine sample mean, sample variance and sample standard deviation Sample mean =

$$\overline{x} = \frac{10.2 + 9.7 + 10.1 + 10.3 + 10.1 + 9.8 + 9.9 + 10.4 + 10.3 + 9.8}{10} = \frac{106.6}{10} = 10.06$$

Sample variance=s^2
Sample standard deviation =$s = +\sqrt{s^2}$=

$$\sqrt{\frac{1}{9}\left(\begin{array}{l}(10.2-10.06)^2 + (9.7-10.06)^2 + (10.1-10.06)^2 \\ +10.3 - 10.06^2 + (10.1-10.06)^2 + (9.8-10.06)^2 \\ +(9.9-10.06)^2 + (10.4-10.06)^2 + (10.3-10.06)^2 + (9.8-10.06)^2\end{array}\right)}$$

=0.246

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2} = \frac{(10-1)(0.246)^2}{0.03} = 18.1548$$

Step 4: Determine the critical region and fail to reject region based on α for case-III

Now we have to determine $\chi^2_{\alpha/2}$ using the equation $P(\chi^2 > \chi^2_{\alpha/2}) = \alpha/2$ with n-1 degrees of freedom. As $P(\chi^2 > \chi^2_{\alpha/2}) = 0.01/2 = 0.005$ with $n - 1$=10-1=9 degrees of freedom.
Using χ^2 distribution table we get $\chi^2_{\alpha/2} = 23.589$
As $\chi^2_{1-\alpha/2} = -\chi^2_{\alpha/2}$

$$\Rightarrow \chi^2_{1-\alpha/2} = -23.589$$

As here $\chi^2_{1-\alpha/2} = -23.589 < \chi^2 = 18.1548 < \chi^2_{\alpha/2} = 23.589$ So it satisfy fail to reject region $\chi^2_{1-\alpha/2} \leq \chi^2_{\alpha/2}$
SO our conclusion is Fail to reject $H_0 : \sigma^2 = 0.03$; that is the sample of 10 containers is not sufficient to show that σ^2 is not equal to 0.03.

Note: Critical Region If the alternative hypothesis is $H_1 : \sigma^2 \neq {\sigma_0}^2$ (case-III) then the critical region is $\chi^2 > \chi^2_{\alpha/2}$ or $\chi^2 < \chi^2_{1-\alpha/2}$

Fail to reject region So fail to reject null hypothesis H_0 region is $\chi^2_{1-\alpha/2} \leq \chi^2_{\alpha/2}$

Question Number 68 Past experience indicates that the time required for high school seniors to complete a standardized test is a normal random variable with a standard deviation

of 6 minutes. Test the hypothesis that $\sigma = 6$ against the alternative that $\sigma < 6$ if a random sample of the test times of 20 high school seniors has a standard deviation s = 4.51. Use a 0.05 level of significance.

Solution: Step 1: It is given that α= 0.05

Step 2: Here we have to test the Null Hypothesis $H_0 \sigma = \sigma_0 = 6$ against the alternative hypothesis H_1, which is $H_1 : \sigma < \sigma_0 = 6$ (case-I)

Also we have to test the Null Hypothesis $H_0 : \sigma^2 = {\sigma_0}^2 = 36$ against the alternative hypothesis H_1, which is $H_1 : \sigma^2 < {\sigma_0}^2 = 36$

Step 3: Determine $\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2}$ it is given n=sample size=20
${\sigma_0}^2 = 36$
sample standard deviation=$s = 4.51$
Now

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2} == \frac{(20-1)4.51^2}{6^2} = 10.74$$

Step 4: Determine the critical region and fail to reject region based on α for case-I

Here we have to determine $\chi^2_{1-\alpha}$ using the equation $P(\chi^2 > \chi^2_{1-\alpha}) = 1 - \alpha = 1 - 0.05 = 0.95$ with n-1=20-1=19 degrees of freedom (use χ^2 distribution table)
$\Rightarrow \chi^2_{1-\alpha} = 10.117 \chi^2 = 10.74 > \chi^2_{1-\alpha} = 10.117$
So it satisfy Fail to reject null hypothesis H_0, $\chi^2 \geq \chi^2_{1-\alpha}$.
Our conclusion is there was not sufficient evidence to conclude that the standard deviation is less then 6 at level $\alpha = 0.05$ level of significance

Question No 71 A soft-drink dispensing machine is said to be out of control if the variance of the contents exceeds 1.15 decilitres. If a random sample of 25 drinks from this machine has a variance of 2.03 decilitres, does this indicate at the 0.05 level of significance that the machine is out of control? Assume that the contents are approximately normally distributed.

Solution: Step 1: It is given that α= 0.05

Step 2: Here we have to test the Null Hypothesis $H_0 \sigma^2 = {\sigma_0}^2 = 1.15$ against the alternative hypothesis H_1, which is $H_1 : \sigma^2 > {\sigma_0}^2 = 1.15$

Step 3: Determine

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2}$$

it is given n=sample size=25
${\sigma_0}^2 = 1.15$
Sample variance=$s^2 = 2.03$
Now

$$\chi^2 = \frac{(n-1)s^2}{{\sigma_0}^2} = \frac{(25-1)2.03}{1.15} = 42.37$$

Step 4: Determine the critical region and fail to reject region based on α for case-II

Here we have to determine χ^2_α using the equation $P(\chi^2 > \chi^2_\alpha) = \alpha = 0.05 =$ with n-1=25-1=24 degrees of freedom (use χ^2 distribution table)

$$\Rightarrow \chi^2_\alpha = 36.415$$

$$\chi^2 = 42.37 > \chi^2_\alpha = 36.415$$

As it satisfy critical region $\chi^2 \geq \chi^2_\alpha$.
Our conclusion is Reject H_0; there is sufficient evidence to conclude, at level α= 0.05, that the soft drink machine is out of control

TWO SAMPLE TEST VARIANCE

Step 1: select a fixed significance level α

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis

$$H_0 : {\sigma_1}^2 = {\sigma_2}^2$$

against the alternative hypothesis H_1, which is one case among three following cases

$$H_1 : {\sigma_1}^2 < {\sigma_2}^2 (case - I)$$

$$H_1 : {\sigma_1}^2 > {\sigma_2}^2 (case - II)$$

$$H_1 : {\sigma_1}^2 \neq {\sigma_2}^2 (case - III)$$

Step 3: Determine f value of testing ${\sigma_1}^2 = {\sigma_2}^2$ which is the ratio $f = \frac{{s_1}^2}{{s_2}^2}$ where ${s_1}^2$ and ${s_2}^2$ are the variance computed from the two samples of size n_1 and n_2

If the two populations are approximately normally distributed and the null hypothesis is true, then the ratio $f = \frac{{s_1}^2}{{s_2}^2}$ is a value of the F-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom.

Step 4: Determine the critical region and fail to reject region based on α, using F-distribution (variance ratio distribution) table with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom. Note:The F-distribution is used in two sample situations to draw inference about the population variance.

Case-I Critical region:
If the alternative hypothesis is $H_1 : {\sigma_1}^2 < {\sigma_2}^2$ (case-I) then the critical region is $f < f_{(1-\alpha)}(v_1, v_2)$.

Fail to reject region: Fail to reject null hypothesis H_0 region is $f \geq f_{(1-\alpha)}(v_1, v_2)$.

NOTE: Here we have to determine $\chi^2_{1-\alpha}$ using the equation $P(\chi^2 > \chi^2_{1-\alpha}) = 1 - \alpha$ with n-1 degrees of freedom (use χ^2 distribution table)

(Case-II) Critical region:
If the alternative hypothesis is $H_1 : {\sigma_1}^2 > {\sigma_2}^2$ (case-I) then the critical region is $f > f_{(\alpha)}(v_1, v_2)$.

Fail to reject region: Fail to reject null hypothesis H_0 region is $f \leq f_{(\alpha)}(v_1, v_2)$.

NOTE: Here we have to determine χ^2_α using the equation $P(\chi^2 > \chi^2_\alpha) = \alpha$ with n-1 degrees of freedom (use χ^2 distribution table)

(Case-III) Critical region: If the alternative hypothesis is $H_1 : {\sigma_1}^2 \neq {\sigma_2}^2$ (case-III) then the critical region is $f < f_{(1-\alpha/2)}(v_1, v_2)$ or $f > f_{(\alpha/2)}(v_1, v_2)$

Fail to reject region: Fail to reject null hypothesis H_0 region is $f_{(1-\alpha/2)}(v_1, v_2) \leq f \leq f_{(\alpha/2)}(v_1, v_2)$.

NOTE: Here we have to determine $\chi^2_{\alpha/2}$ using the equation $P(\chi^2 > \chi^2_{\alpha/}) = \alpha/2$ with n-1 degrees of freedom.

Determine $\chi^2_{1-\alpha/2}$ using the equation $P(\chi^2 > \chi^2_{1-\alpha/}) = 1 - \alpha/2$ with n-1 degrees of freedom (use χ^2 distribution table) or using $\chi^2_{1-\alpha/2} = -\chi^2_{\alpha/2}$

Question No 73 A study is conducted to compare the lengths of time required by men and women to assemble a certain product. Past experience indicates that the distribution of times

for both men and women is approximately normal but the variance of the times for women is less than that for men. A random sample of times for 11 men and 14 women produced the following data:

$$\begin{array}{cc} Men & Women \\ \hline n_1 = 11 & n_2 = 14 \\ s_1 = 6.1 & s_2 = 5.3 \end{array} \tag{0.1}$$

Test the hypothesis that ${\sigma_1}^2 = {\sigma_2}^2$ against the alternative that ${\sigma_1}^2 = {\sigma_2}^2$.

Solution:

Step 1: Let us take significance level $\alpha = 0.05$

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis $H_0 : {\sigma_1}^2 = {\sigma_2}^2$ against the alternative hypothesis H_1, which is $H_1 : {\sigma_1}^2 > {\sigma_2}^2$ (case-II)

Step 3: given $n_1 = 11$ and $n_2 = 14$
$s_1 = 6.1$ and $s_2 = 5.3$

Determine f value of testing ${\sigma_1}^2 = {\sigma_2}^2$ which is the ratio

$$f = \frac{{s_1}^2}{{s_2}^2}$$

where ${s_1}^2$ and ${s_2}^2$ are the variance computed from the two samples of size n_1 and n_2

Now

$$f = \frac{{s_1}^2}{{s_2}^2} = \frac{6.1^2}{5.3^2} = 1.33$$

Step 4: Determine the critical region and fail to reject region based on α, using F-distribution table with $v_1 = n_1 - 1 = 11 - 1 = 10$ and $v_2 = n_2 - 1 = 14 - 1 = 13$ degrees of freedom.

As here alternative hypothesis is $H_1 : {\sigma_1}^2 > {\sigma_2}^2$ (case-II)
then the critical region is $f > f_{(\alpha)}(v_1, v_2)$.

Fail to reject null hypothesis H_0 region is $f \leq f_{(\alpha)}(v_1, v_2)$.

Now $f_{(\alpha)}(v_1, v_2) = f_{(0.05)}(10, 13) = 2.67$ (refer to page no 820, F distribution table)
As here $f = 1.33 \leq f_{(\alpha)}(v_1, v_2) = f_{(0.05)}(10, 13) = 2.67$. So it satisfy Fail to reject null hypothesis H_0 region. So, our conclusion is " the variability of the time to assemble the product is not significantly greater for men".

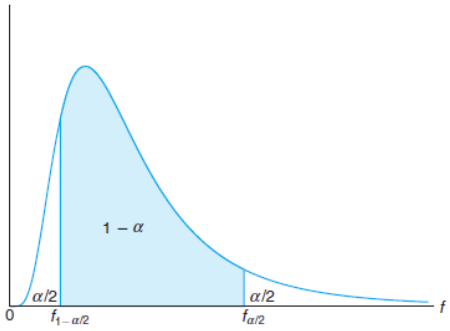


Figure 5: **E**

(Lect. 37)
10.11 Goodness-of-Fit Test

Here we consider a test to determine if a population has a specified theoretical distribution. The test is based on how good a fit we have between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution.

A goodness-of-fit test between observed and expected frequencies is based on the quantity.

Procedure:

Step 1: select a fixed significance level α

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis H_0 against the alternative hypothesis H_1

Step 3: Determine $\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$ where χ^2 is a value of a random variable whose sampling distribution is approximated very closely by the chi-squared distribution with $v = k-1$ degrees of freedom. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell. k represents number of cell.

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = \alpha$ with $k-1$ degrees of freedom (Use χ^2 distribution table)

Step 5: Determine the critical region and fail to reject region based on α, using χ^2-distribution table with $k-1$ degrees of freedom.

Critical region is $\chi^2 > {\chi_\alpha}^2$

Fail to reject region is $\chi^2 \leq {\chi_\alpha}^2$

Question No 80

The grades in a statistics course for a particular semester were as follows:

$Grade$	A	B	C	D	F
f	14	18	32	20	16

(0.1)

Test the hypothesis, at the 0.05 level of significance, that the distribution of grades is uniform.

Solution:

Step 1: Given $\alpha = 0.05$

Step 2: we have to test the Null Hypothesis

H_0:Distribution of grades is uniform against the alternative hypothesis

H_1: Distribution of grades is not uniform

Step 3: Here k=5= number of cells

o_i: the observed frequencies, $i = 1, 2, 3, 4, 5$ o_1=14, o_2=18, o_3=32, o_4=20, o_5=16

e_i represent expected frequencies, $i = 1, 2, 3, 4, 5$ $e_i = \frac{14+18+32+20+16}{5} = 20$

$$\chi^2 = \sum_{i=1}^{5} \frac{(o_i - e_i)^2}{e_i} = \frac{(14-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(32-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(16-20)^2}{20} = 10$$

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = 0.05$ with $k-1 = 5-1 = 4$ degrees of freedom (Use χ^2 distribution table)

Now ${\chi_\alpha}^2 = {\chi_{0.05}}^2 = 9.488$

Step 5 As here $\chi^2 = 10 > {\chi_\alpha}^2 = 9.488$. So it satisfy critical region (reject region)$(\chi^2 > {\chi_\alpha}^2)$ So we reject the null hypothesis and not to fail alternative hypothesis.

Our conclusion is the distribution of grades is not uniform.

Question No 83 A coin is thrown until a head occurs and the number X of tosses recorded. After repeating the experiment 256 times, we obtained the following results:

x	1	2	3	4	5	6	7	8
f	136	60	34	12	9	1	3	1

(0.2)

Test the hypothesis, at the 0.05 level of significance, that the observed distribution of X may be fitted by the geometric distribution $g(x; 1/2)$, $x = 1, 2, 3, \ldots$.

Solution:

Step 1: Given $\alpha = 0.05$

Step 2: we have to test the Null Hypothesis

H_0: Observed distribution of X is fitted by the geometric distribution $g(x; 1/2)$, $x = 1, 2, 3, \ldots$ against the alternative hypothesis

H_1: Observed distribution of X is not fitted by the geometric distribution $g(x; 1/2)$, $x = 1, 2, 3, \ldots$.

Also we represent $H_0 : f(x) = g(x; 1/2)$, $x = 1, 2, 3, \ldots$ and $H_1 : f(x) \neq g(x; 1/2)$, $x = 1, 2, 3, \ldots$.

Step 3:we know the geometric distribution $g(x; p) = pq^{x-1}$

p=Probability of getting head $=\frac{1}{2}$

q=Probability of getting tail $=\frac{1}{2}$

$g(x; 1/2) = pq^{x-1} = \frac{1}{2}(\frac{1}{2})^{x-1} = \frac{1}{2^x}$, $x = 1, 2, 3, \ldots$.

$g(x; 1/2) = \frac{1}{2^x}$, $x = 1, 2, 3, \ldots$

$g(1; 1/2) = \frac{1}{2^1} = \frac{1}{2}$

$g(2; 1/2) = \frac{1}{2^2} = \frac{1}{4}$

$g(3; 1/2) = \frac{1}{2^3} = \frac{1}{8}$

$g(4; 1/2) = \frac{1}{2^4} = \frac{1}{16}$

$g(5; 1/2) = \frac{1}{2^5} = \frac{1}{32}$

$g(6; 1/2) = \frac{1}{2^6} = \frac{1}{64}$

$g(7; 1/2) = \frac{1}{2^7} = \frac{1}{128}$

$g(8; 1/2) = \frac{1}{2^8} = \frac{1}{248}$

Here k=8= number of cells

o_i the observed frequencies,$i = 1, 2, 3, 4, 5, 6, 7, 8$

o_1=136, o_2=60, o_3=34, o_4=12, o_5=9, o_6=1, o_7=3, o_8=1

Total number of observed frequencies=$136 + 60 + 34 + 12 + 9 + 1 + 3 + 1 = 256$

Now we have to calculate all the expected frequencies, e_i, i=1,2,3,4,5,6,7,8

$e_1 = 256(g(1; 1/2)) = 256(\frac{1}{2}) = 128$

$e_2 = 256(g(2; 1/2)) = 256(\frac{1}{4}) = 64$

$e_3 = 256(g(3; 1/2)) = 256(\frac{1}{8}) = 32$

$e_4 = 256(g(4; 1/2)) = 256(\frac{1}{16}) = 16$

$e_5 = 256(g(5; 1/2)) = 256(\frac{1}{32}) = 8$

$e_6 = 256(g(6; 1/2)) = 256(\frac{1}{64}) = 4$

$e_7 = 256(g(7; 1/2)) = 256(\frac{1}{128}) = 2$

$e_8 = 256(g(8; 1/2)) = 256(\frac{1}{248}) = 1$

Now

$$\chi^2 = \sum_{i=1}^{8} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(136-128)^2}{128} + \frac{(60-64)^2}{64} + \frac{(34-32)^2}{32} + \frac{(12-16)^2}{16} + \frac{(9-8)^2}{8} + \frac{(1-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(1-1)^2}{1} = 3.125$$

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = 0.05$ with $k - 1 = 8 - 1 = 7$ degrees of freedom (Use χ^2 distribution table)
Now ${\chi_\alpha}^2 = {\chi_{0.05}}^2 = 14.067$

Step 5 As here $\chi^2 = 3.125 < {\chi_\alpha}^2 = 14.067$. So it satisfy fail to reject region ($\chi^2 \leq {\chi_\alpha}^2$) So we Fail to reject the null hypothesis.
Our conclusion is $f(x) = g(x; 1/2)$.

(Lect. 38)
10.12 TEST FOR INDEPENDENCE

The chi-squared test procedure can also be used to test the hypothesis of independence of two variables of classification.

Procedure: Step 1: select a fixed significance level α

Step 2: State the Null Hypothesis H_0 and alternative hypothesis H_1 that is we have to test the Null Hypothesis H_0 against the alternative hypothesis H_1

Step 3: Determine

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

where χ^2 is a value of a random variable whose sampling distribution is approximated very closely by the chi-squared distribution with $v = (r-1)(c-1)$ degrees of freedom. r represents number of row and c represents number of column. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell. $k = rc$ represents number of cell. e_i is obtained by the following formula:

$$e_i = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total} = \frac{(RT) \times (CT)}{GT}$$

Step 4: Determine critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = \alpha$ with $(r-1)(c-1)$ degrees of freedom (Use χ^2 distribution table)

Step 5: Determine the critical region and fail to reject region based on α, using χ^2-distribution table with $(r-1)(c-1)$ degrees of freedom.

Critical region is $\chi^2 > {\chi_\alpha}^2$

Fail to reject null hypothesis H_0 region is $\chi^2 \leq {\chi_\alpha}^2$

Question No.87

A random sample of 90 adults is classified according to gender and the number of hours of television watched during a week:

	Gender	
	Male	Female
Over 25 hours	15	29
Under 25 hours	27	19

(0.1)

Use a 0.01 level of significance and test the hypothesis that the time spent watching television is independent of whether the viewer is male or female.

Solution: step 1: significance level α=0.01

Step 2: we have to test the Null Hypothesis H_0:the time spent watching television is independent of whether the viewer is male or female against

the alternative hypothesis H_1:the time spent watching television is not independent of whether the viewer is male or female.

Step 3:

Observed and expected frequencies			
	Male	Female	Total
Over 25 hours	15(20.5)	29(23.5)	$r_1 = 44$
Under 25 hours	27(21.5)	19(24.5)	$r_2 = 46$
Total	$c_1 = 42$	$c_2 = 48$	90

(0.2)

Here o_1= 15, o_2=29, o_3=27 and o_4=19
r_1= first row total=15+29=44
r_2= second row total=27+29=46
c_1=first column total=15+27=42
c_2=second column total=29+19=48
Grand total=15+29+27+19=90=GT
e_1=expected frequency of $(1,1)$ cell=

$$\frac{r_1 c_1}{GT} = \frac{44 \times 42}{90} = 20.5$$

e_2=expected frequency of $(1,2)$ cell=

$$\frac{r_1 c_2}{GT} = \frac{44 \times 48}{90} = 23.5$$

e_3=expected frequency of $(2,1)$ cell=

$$\frac{r_2 c_1}{GT} = \frac{46 \times 42}{90} = 21.5$$

e_4=expected frequency of $(2,2)$ cell=

$$\frac{r_2 c_2}{GT} = \frac{46 \times 48}{90} = 24.5$$

Now

$$\chi^2 = \sum_{i=1}^{4} \frac{(o_i - e_i)^2}{e_i} = \frac{(15-20.5)^2}{20.5} + \frac{(29-23.5)^2}{23.5} + \frac{(27-21.5)^2}{21.5} + \frac{(19-24.5)^2}{24.5} = 5.47$$

Step 4: Now we have to determine the critical value ${\chi_\alpha}^2$ using the following equation
$P(\chi^2 > {\chi_\alpha}^2) = \alpha = 0.01$ with $(r-1)(c-1) = (2-1)(2-1) = 1$ degrees of freedom (Use χ^2 distribution table)
$\Rightarrow {\chi_\alpha}^2 = 6.635$
Step 5 As here $\chi^2 = 5.47 < {\chi_\alpha}^2 = 6.635$, we have Fail to reject null hypothesis H_0 .
So our conclusion is the time spent watching television is independent of whether the viewer is male or female

Test for Homogeneity Homogeneous means the same in structure or composition. This test gets its name from the null hypothesis, where we claim that the distribution of the responses are the same (homogeneous) across groups.

NOTE: PROCEDURE is same as TEST FOR INDEPENDENCE

Quesion No. 93 To determine current attitudes about prayer in public schools, a survey was conducted in four Virginia counties. The following table gives the attitudes of 200 parents from Craig County, 150 parents from Giles County, 100 parents from Franklin County, and 100 parents from Montgomery County:

	County			
Attitude	Craig	Giles	Franklin	Mont.
Favor	65	66	40	34
Oppose	42	30	33	42
No opinion	93	54	27	24

(0.3)

Test for homogeneity of attitudes among the four counties concerning prayer in the public schools.

Solution: Step 1:Let significance level α=0.01

Step 2: we have to test the Null Hypothesis H_0 : The attitudes among the four countries are homogeneous against the alternative H_1 : The attitudes among the four countries are not homogeneous.

:

Observed and expected frequencies					
	County				
Attitude	Craig	Giles	Franklin	Montgomery	Total
Favor	65(74.5)	66(55.9)	40(37.3)	34(37.3)	$r_1 = 205$
Oppose	42(53.5)	30(40.1)	33(26.7)	42(26.7)	$r_2 = 147$
No Opinion	93(72.0)	54(54.0)	27(36.0)	24(36.0)	$r_3 = 198$
Total	$c_1 = 200$	$c_2 = 150$	$c_3 = 100$	$c_4 = 100$	$GT = 550$

(0.4)

Here o_1= 65, o_2=66, o_3=40 and o_4=34 o_5= 42, o_6=30, o_7=33 and o_8=42 o_9= 65, o_{10}=54, o_{11}=27 and o_{12}=24

r_1= first row total=65+66+40+34=205

r_2= second row total=42+30+33+42=147

r_3= third row total=93+54+27+24=198

c_1=first column total=65+42+93=200

c_2=second column total=66+30+54=150

c_3=third column total=40+33+27=100

c_4= fourth column total=34+42+24=100

Grand total=205+147+198=GT

e_1=expected frequency of $(1, 1)$ cell=

$$\frac{r_1 c_1}{GT} = \frac{205 \times 200}{550} = 74.5$$

e_2=expected frequency of $(1, 2)$ cell=

$$\frac{r_1 c_2}{GT} = \frac{205 \times 150}{550} = 55.9$$

e_3=expected frequency of $(1, 3)$ cell=

$$\frac{r_1 c_3}{GT} = \frac{205 \times 100}{550} = 37.3$$

e_4=expected frequency of $(1, 4)$ cell=

$$\frac{r_1 c_4}{GT} = \frac{205 \times 100}{550} = 37.3$$

e_5=expected frequency of $(2,1)$ cell=

$$\frac{r_2c_1}{GT} = \frac{147 \times 200}{550} = 53.5$$

e_6=expected frequency of $(2,2)$ cell=

$$\frac{r_2c_2}{GT} = \frac{147 \times 150}{550} = 40.1$$

e_7=expected frequency of $(2,3)$ cell=

$$\frac{r_2c_3}{GT} = \frac{147 \times 100}{550} = 26.7$$

e_8=expected frequency of $(2,4)$ cell=

$$\frac{r_2c_4}{GT} = \frac{147 \times 100}{550} = 26.7$$

e_9=expected frequency of $(3,1)$ cell=

$$\frac{r_3c_1}{GT} = \frac{198 \times 200}{550} = 72$$

e_{10}=expected frequency of $(3,2)$ cell=

$$\frac{r_3c_2}{GT} = \frac{198 \times 150}{550} = 54$$

e_{11}=expected frequency of $(3,3)$ cell=

$$\frac{r_3c_3}{GT} = \frac{198 \times 100}{550} = 36$$

e_{12}=expected frequency of $(3,4)$ cell=

$$\frac{r_3c_4}{GT} = \frac{198 \times 100}{550} = 36$$

$$\chi^2 = \frac{(65-74.5)^2}{74.5} + \frac{(66-55.9)^2}{55.9} + \frac{(40-37.3)^2}{37.3} + \frac{(34-37.3)^2}{37.3} + \frac{(42-53.5)^2}{53.5} + \frac{(30-40.1)^2}{40.1}$$
$$+\frac{(33-26.7)^2}{26.7} + \frac{(42-26.7)^2}{26.7} + \frac{(93-72)^2}{72} + \frac{(54-54)^2}{54} + \frac{(27-36)^2}{36} + \frac{(24-36.0)^2}{36.0} = 31.17$$

Step 4: Now we have to determine the critical value ${\chi_\alpha}^2$ using the following equation

$P(\chi^2 > {\chi_\alpha}^2) = \alpha = 0.01$ with $(r-1)(c-1) = (3-1)(4-1) = 6$ degrees of freedom (Use χ^2 distribution table)

$\Rightarrow {\chi_\alpha}^2 = 16.812$

Step 5 As here $\chi^2 = 31.17 > {\chi_\alpha}^2 = 16.812$,it satisfy critical region so we have to reject null hypothesis H_0 .

So our conclusion is the attitudes among the four countries are not homogeneous.

(Lect. 39)
11.1-11.3 Regression Line

In this article we will discuss some methods of dealing with paired data on two variables. And it is limited to linear relationship between two variables only. It is a straight line relationship. Linear regression deals with methods of fitting a straight line called the regression line on a set of sample paired data on two variables. A reasonable form of relationship between the response Y and the regressor x is the linear relationship

$$Y = \beta_0 + \beta_1 x$$

where β_0 is the intercept and β_1 is the slope. The concept of regression analysis deals with finding the best relationship between Y and x, quantifying the strength of that relationship, and using methods that allow for prediction of the response values given values of the regressor x.

1 The Fitted Regression Line

An important aspect of regression analysis is, very simply, to estimate the parameters β_0 and β_1 (i.e., estimate the so-called regression coefficients). Suppose we denote the estimates b_0 for β_0 and b_1 for β_1. Then the estimated or fitted regression line is given by

$$\hat{y} = b_0 + b_1 x$$

where $\hat{y}$ is the predicted or fitted value.

1.1 Estimating the Regression Coefficients

Given the sample (x_i, y_i); $i = 1, 2, ..., n$, the least squares estimates b_0 and b_1 of the regression coefficients β_0 and β_1 are computed from the formulas

$$b_1 = \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ and}$$

$$b_0 = \frac{\sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i}{n} = \bar{y} - b_1 \bar{x}$$

Question No.2 The grades of a class of 9 students on a midterm report (x) and on the final examination (y) are as follows:

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

(a) Estimate the linear regression line. (b) Estimate the final examination grade of a student who received a grade of 85 on the midterm report.

Solution

x	y	xy	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
77	82	6314	-1.55	8.89	-13.7795	2.4025
50	66	3300	-28.55	-7.11	202.9905	815.1025
71	78	5538	-7.55	4.89	-36.9195	57.0025
72	34	2448	-6.55	-39.11	256.1705	42.9025
81	47	3807	2.45	-26.11	-63.9695	6.0025
94	85	7990	15.45	11.89	183.7005	238.7025
96	99	9504	17.45	25.89	451.7805	304.5025
99	99	9801	20.45	25.89	529.4505	418.2025
67	68	4556	-11.55	-5.11	59.0205	133.4025
$\sum_{i=1}^{9} x_i = 707$	$\sum_{i=1}^{9} y_i = 658$				1568.4445	2018.2225

(1.1)

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 1568.4445$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 2018.2225$$

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{707}{9} = 78.55$$

$$\bar{y} = \frac{\sum_{i=1}^{9} y_i}{9} = \frac{658}{9} = 73.11$$

(a)Regression line is

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{1568.4445}{2018.2225} = 0.7771415$$

$$b_0 = \bar{y} - b_1 \bar{x} = 73.11 - 0.7771415(78.55) = 12.06553518$$

Now Regression line is

$$\hat{y} = b_0 + b_1 x = 12.06553518 + 0.7771415x$$

(b)The final examination grade of a student who received a grade of 85 on the midterm report is

$$\hat{y} = b_0 + b_1 x = 12.06553518 + 0.7771415x$$

for x=85
Now for x=85,

$$\hat{y} = b_0 + b_1 x = 12.06553518 + 0.7771415(85) = 78.1225$$

Question No.5 A study was made on the amount of converted sugar in a certain process at various temperatures. The data were coded and recorded as follows:

Temperature, $\boldsymbol{x}$	Converted Sugar, $\boldsymbol{y}$
1.0	8.1
1.1	7.8
1.2	8.5
1.3	9.8
1.4	9.5
1.5	8.9
1.6	8.6
1.7	10.2
1.8	9.3
1.9	9.2
2.0	10.5

(a) Estimate the linear regression line. (b) Estimate the mean amount of converted sugar produced when the coded temperature is 1.75.

Solution

x	y	xy	x^2
1.0	8.1	8.1	1
1.1	7.8	8.85	1.21
1.2	8.5	10.2	72.25
1.3	9.8	12.71	1.69
1.4	9.5	13.3	1.96
1.5	8.9	13.35	2.25
1.6	8.6	13.76	2.56
1.7	10.2	17.34	2.89
1.8	9.3	16.74	3.24
1.9	9.2	17.48	3.61
2.0	10.5	21	4
$\sum_{i=1}^{11} x_i = 16.5$	$\sum_{i=1}^{11} y_i = 100.4$	$\sum_{i=1}^{11} x_i y_i = 152.59$	$\sum_{i=1}^{11} {x_i}^2 = 25.85$

(1.2)

Therefore,

$$\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{16.5}{11} = 1.5$$

$$\bar{y} = \frac{\sum_{i=1}^{11} y_i}{11} = \frac{104.4}{11} = 9.4909$$

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{(11)(152.59) - (16.5)(100.4)}{(11)(25.85) - (16.5)^2} = 1.8091 \tag{1.3}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \frac{100.4 - (1.8091)(16.5)}{11} = 6.4136$$

Hence

$$\overset{\Lambda}{y} = 6.4136 + 1.8091x$$

(b) For x = 1.75,

$$\overset{\Lambda}{y} = 6.4136 + (1.8091)(1.75) = 9.580$$

Question No.7 The following is a portion of a classic data set called the "pilot plot data" in Fitting Equations to Data by Daniel and Wood, published in 1971 . The response y is the acid content of material produced by titration, whereas the regressor x is the organic acid content produced by extraction and weighing.

y	x	y	x
76	123	70	109
62	55	37	48
66	100	82	138
58	75	88	164
88	159	43	28

Fit a simple linear regression; estimate a slope and intercept.

Solution

x	y	xy	x^2
123	76	9348	15129
55	62	3410	3025
100	66	6600	10000
75	58	4350	5625
159	88	13992	25281
109	70	7630	11881
48	37	1776	2304
138	82	11316	19044
164	88	14432	26896
28	43	1204	784
$\sum_{i=1}^{10} x_i = 999$	$\sum_{i=1}^{10} y_i = 670$	$\sum_{i=1}^{10} x_i y_i = 74,058,$	$\sum_{i=1}^{11} {x_i}^2 = 119,969$

(1.4)

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{(10)(74,058) - (999)(670)}{(10)(119,969) - (999)^2} = 0.3533 \tag{1.5}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \frac{670 - (0.3533)(999)}{10} = 31.71$$

Hence

$$\overset{\Lambda}{y} = 31.71 + 0.3533x.$$

(Lect. 40)
11.12 Correlation

Up to this point we have assumed that the independent regressor variable x is a physical or scientific variable but not a random variable. In fact, in this context, x is often called a mathematical variable, which, in the sampling process, is measured with negligible error. In many applications of regression techniques, it is more realistic to assume that both X and Y are random variables and the measurements $(x_i, y_i); i = 1, 2, ..., n$ are observations from a population having the joint density function $f(x, y)$. We shall consider the problem of measuring the relationship between the two variables X and Y. For example, if X and Y represent the length and circumference of a particular kind of bone in the adult body, we might conduct an anthropological study to determine whether large values of X are associated with large values of Y, and vice versa.

Correlation analysis attempts to measure the strength of such relationships between two variables by means of a single number called a correlation coefficient.

Correlation coefficient

It is defined by

$$r = \frac{S_{xy}}{S_x S_y}$$

where S_x and S_y are standard deviation of x and y respectively. S_{xy} is the covariance of xy.

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n}$$

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

$$E(y) = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$$

Question No 43

Compute and interpret the correlation coefficient for the following grades of 6 students selected at random:

Mathematics grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

(0.1)

Solution: let Mathematics grade=x English grade=y

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
70	74	5180	53.7289	28.4089
92	84	7728	215.2089	21.8089
80	63	5040	7.1289	266.6689
74	87	6438	11.0889	58.8289
65	78	5070	152.0289	1.7689
83	90	7470	32.1489	113.8489
$\sum_{i=1}^{6} x_i = 464$	$\sum_{i=1}^{6} y_i = 476$	$\sum_{i=1}^{6} x_i y_i = 36926$	$\sum_{i=1}^{6} (x-\bar{x})^2 = 471.3334$	491.3334

(0.2)

$$E(x) = \bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = \frac{464}{6} = 77.33$$

$$E(y) = \bar{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = \frac{476}{6} = 79.33$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{36926}{6} = 6154.33$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{471.3334}{5}} = 9.7091$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{491.3334}{5}} = 9.91295$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 6154.33 - (77.33 \times 79.33) = 19.4111$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{19.4111}{9.7091 \times 9.91295} = 0.2016825$$

Question No 45

A study of the amount of rainfall and the quantity of air pollution removed produced the following data.

$DailyRainfall$ x $(0.01cm)$	$ParticulateRemoved$, y $(\mu g/m3)$
4.3	126
4.5	121
5.9	116
5.6	118
6.1	114
5.2	118
3.8	132
2.1	141
7.5	108

(0.3)

assume a bivariate normal distribution for x and y. (a) Calculate r.

Solution

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
4.3	126	541.8	0.49	19.8025
4.5	121	544.5	0.25	0.3025
5.9	116	684.4	0.81	30.8025
5.6	118	660.8	0.36	12.6025
6.1	114	695.4	1.21	57.0025
5.2	118	613.6	0.04	12.6025
3.8	132	501.6	1.44	109.2025
2.1	141	296.1	8.41	378.3025
7.5	108	810	6.25	183.6025
$\sum_{i=1}^{9} x_i = 45$	$\sum_{i=1}^{9} y_i = 1094$	$\sum_{i=1}^{6} x_i y_i = 5348.2$	19.26	804.2225

(0.4)

$$E(x) = \bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{45}{9} = 5$$

$$E(y) = \bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = \frac{1094}{9} = 121.55$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{5348.2}{9} = 594.244$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{19.26}{8}} = 1.5516$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{804.2225}{8}} = 10.0263$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 594.244 - (5 \times 121.55) = -13.506$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{-13.506}{1.5516 \times 10.0263} = -0.86817$$

Question No.47 The following data were obtained in a study of the relationship between the weight and chest size of infants at birth.

$Weight(kg)$	$ChestSize(cm)$
2.75	29.5
2.15	26.3
4.41	32.2
5.52	36.5
3.21	27.2
4.32	27.7
2.31	28.3
4.30	30.3
3.71	28.7

(0.5)

(a) Calculate r.

Solution

x	y	xy	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2.75	29.5	81.125	0.7744	0.0169
2.15	26.3	56.545	2.1904	11.0889
4.41	32.2	142.002	0.6084	6.6049
5.52	36.5	201.48	3.5721	47.1969
3.21	27.2	87.312	0.1764	5.9049
4.32	27.7	119.667	0.4761	3.7249
2.31	28.3	65.373	1.7424	1.7689
4.30	30.3	130.29	.4489	.4489
3.71	28.7	106.477	0.0064	0.8649
$\sum_{i=1}^{9} x_i = 32.68$	$\sum_{i=1}^{9} y_i = 266.7$	$\sum_{i=1}^{6} x_i y_i = 990.271$	9.9955	77.6201

(0.6)

$\bar{x} = 3.63$ $\bar{y} = 29.63$

$$E(x) = \bar{x} = \frac{1}{9}\sum_{i=1} x_i = \frac{32.68}{9} = 3.63$$

$$E(y) = \bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = \frac{266.7}{9} = 29.63$$

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

$$E(xy) = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{990.271}{9} = 110.0301$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{9.9955}{8}} = 1.1177$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{77.6201}{8}} = 3.11488$$

$$S_{xy} = E(xy) - E(x)E(y) = E(xy) - \bar{x}\bar{y} = 110.0301 - (3.63 \times 29.63) = 2.4732$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{2.4732}{1.1177 \times 3.11488} = 0.710338$$