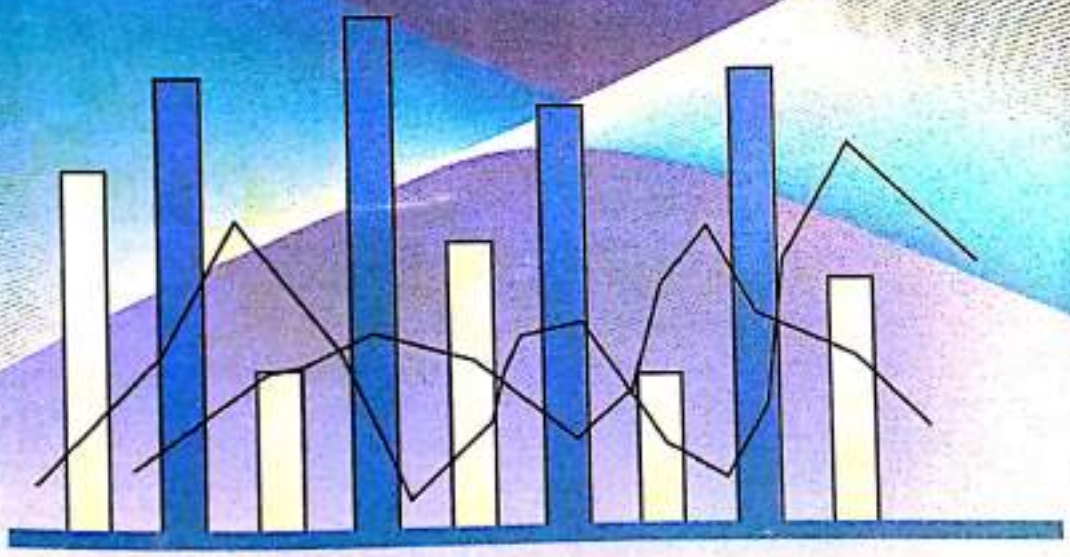
New Edition

A Reference Book of

PROBABILITY AND STATISTICS for ENGINEERS

Bachelor Level of Engineering
with P.U. Exam Question Solution



Written by
D.R. Pandey

Scanned by CamScanner

A Reference Book of

PROBABILITY AND STATISTICS
for
ENGINERS
Bachelor Level of Engineering
With PU Exam Question Solution

D.R. Pandey

KEC Publication and Distribution Pvt. Ltd.
Kathmandu, Nepal
Phone: 01-4168301, 01-4241777

Scanned by CamScanner

A Reference Book of

PROBABILITY AND STATISTICS
for
ENGINERS

Edition	:	First 2075
Copyright	:	Author
Publisher	:	KEC Publication and Distribution Pvt. Ltd. Kathmandu Ph. 014168301, 4241777
Email	:	kecpublication14@gmail.com
Price	:	Rs. 425/-

Scanned by CamScanner

Preface...

This book has been presented as a reference book for the students/reader of Probability and Statistics for Engineers. Especially, it deals the problems of the book **"Probability and Statistic–I"** that is based on the latest syllabus of 4-year Bachelor of Engineering of all programme of Pokhara University. Moreover, the book suggests when and how the process helps to solve the problem(s), if necessary. Also, it contains the required theoretical part that is necessary for final examination as well as for the solving problems. And, the book includes university final exam questions and there solution till 2017 that helps to the student to prepare the final exam.

We found many leaner are facing the trouble about the syllabus and teaching hour(s) distributed by the syllabus as per chapter/unit. So, we include recent syllabus that offered by the Pokhara University.

We gratefully acknowledge the inspiration, encouragement and valuable suggestions received from the well wishers during the preparation of the book. We are extremely thankful to KEC Publication family, Kathmandu, without whose support the book would not have been published.

We do not claim that the book is perfect and free from any mistakes. So, we would like to apologize. Suggestions for the further improvement of the book will be warmly welcomed and thankfully acknowledged.

Author

Scanned by CamScanner

Contents...

Chapter 1 Introduction and Descriptive Statistics 1

Chapter 2 Collection of Data 3

Chapter 3 Diagrammatic and Graphical Representation 10

Chapter 4 Measures of Central Tendency 48

Chapter 5 Measures of Dispersion 78

Chapter 6 Mathematical Preliminaries 106

Chapter 7 Theory of Probability 113

Chapter 8 Random Variable and Probability Distribution 155

Chapter 9 Mathematical Expectation 172

Chapter 10 Theoretical Discrete Distribution 202

Chapter 11 Theoretical Continuous Distribution 237

Chapter 12 Joint Probability Distribution 281

Chapter 13 Sampling and Estimation 313

Chapter 14 Test of Hypothesis 331

Chapter 15 Simple Linear Regression and Correlation 373

Scanned by CamScanner

1

Chapter

Introduction and Descriptive Statistics

Meaning and Definition of Statistics

The term statistics has been derived from the Latin word "status Italian word "statista" or the German word "statistik". All of these words have same meaning, the political state.

The term 'statistics' has been defined in two ways; (i) singular sense and (ii) plural sense.

- In singular sense statistics can be defined as a statistical method. By Croxton and Cowden, "Statistics in the science of collection, presentation, analysis and interpretation of numerical data."

In plural sense Horace Secretes defines statistics as the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated accordingly to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".

Hence, the statistics is the branch of applied mathematics and may be regarded as mathematics applied to observational data.

"Use and Importance of Probability and Statistics [PU 2015 Spring/2016 Fall]

In this modern word of computer and technology the application (importance) of probability and statistics is very well recognized by all the fields. It is used in social as well as physical such as biology, psychology, education, economics, business management, astronomy, medical science, war, industry, planning, agriculture, engineering, etc.

Scanned by CamScanner

Statistics has various engineering applications. Some of them are listed below.

i. It is used to test and account for the variations in materials and goods.

ii. It is used to analyze the signal and noise.

iii. It is used for ranging from quality control and quality assurance to communication theory.

iv. It is used to assess experimental data and control as well as improve chemical process.

v. It is used to check the ability, performance, reliability of any system machine of equipment.

Function of Statistics

Statistics is used in broad area as well as almost every field of human being some important functions are listed below.

i. It simplifies complexity.

ii. It presents the numerical facts indefinite form.

iii. It helps in classification of data according to its nature.

iv. It helps in predicting future trends.

v. It helps to compare as well as make decisions.

vi. It helps in formulating and hypothesis testing.

Limitations of Statistics

Although statistics has wide range of applications, it has some limitations which are listed below.

i. Statistics does not deal with individual.

ii. It does not study about the qualitative phenomenon.

iii. Statistical laws are not exact.

iv. Statistics is liable to be misused.

v. Statistics is only a means

Exam Asked Questions

1. Write short note on the application of probability and statistics in engineering.
[PU 2015 Spring/2016 Fall/2006] [TU 2066/2068/2069/2072]

Solution

See in theoretical part of this chapter.

Scanned by CamScanner

2
Chapter

Collection of Data

Statistics is the collection of information about the concern study, mostly in the form of data. It can be classified under two categories depending upon the sources utilized. They are; (i) primary data and (ii) secondary data

Primary data **[PU 2016 Spring/ Purbanchal 2010]**

The data which are collected by an investigator directly for the purpose of certain study or information, are known as primary data. For example:

i. The census data

ii. Data released in Nepal Rasta Bank bulletin, etc.

The primary data can be collected by the following methods:

i. Direct personal interviews

ii. Indirect oral interviews

iii. Information from correspondents

iv. Mailed questionnaire method

iv. Schedules sent through enumerators

Secondary data

The data which are originally collected by certain people or agency and statistically treated and the information contained in it used again from records, processed and statistically analyzed to extract some information for other purpose is said to be secondary data. For examples; (i) published thesis, (ii) research papers, (iii) project reports, etc.

Scanned by CamScanner

Frequency distribution

Variable (variate): A quantity which can vary from one individual to another is called a variable. For example heights, weights, ages, wages of person, rainfall records, etc.

Frequency: Frequency is the number of times which a variate value occurs. If the variate value x repeats 5 times then the frequency of x is 5.

Frequency distribution

A frequency distribution is simply a table in which the presented and classified data are expressed into different classes, showing also the number of items belonging to each class.

Types of frequency distribution

There are two types of frequency distributions: univariate frequency distribution and bivariate frequency distributions. Univariate frequency distributions may be classified as:

i. Individual frequency distributions or individual series

ii. Discreet frequency distributions or discrete series

iii. Continuous frequency distributions or continuous series

Sturges' rule

i. According to Sturges' rule the number of classes can be determined by the formula:

$$k = 1 + 3.322 \log_{10} N$$

Where, N = Total number of observations

k = Number of class intervals

ii. The width of the class-interval by using Sturges' rule

$$= \frac{\text{Largest value} - \text{Smalleest value}}{K} = \frac{L - S}{K}$$

Types of class intervals

There are three methods of classifying the data according to class intervals:

i. Exclusive method

ii. Inclusive method

iii. Open-end classes

i. **Exclusive method:** When the class intervals are so fixed that the upper limit of one class is the lower limit of the next class is known as the exclusive method of classification. For example,

Temperature (in °C)	10-20	20-30	30-40	40-50
No. of days	5	7	3	2

Note: The upper limit is always excluded in exclusive class interval.

ii. **Inclusive method:** When the both lower and upper limits are included in the class interval such type of class interval is called inclusive class interval. For example,

Temperature (in °C)	10-19	20-29	30-39	40-49
No. of days	5	7	3	2

Scanned by CamScanner

Note:

a. Both limits are included

b. We can change the class of inclusive method into exclusive method by subtracting the correction factor C_f from the lower limit and adding the correction factor C_f on the upper limit.

$$\text{Correction factor } C_f = \frac{\text{Lower limit of second class} - \text{Upper limit of first class}}{2}$$

iii. **Open-end classes:** A frequency distribution is said to be open end classes if the lower limit of the first class or upper limit of the last class is not specified. For example,

Weights (in kg)	below 30	30 - 40	40 - 50	50 - 60	above 60
No. of persons	4	5	7	2	9

Exercise 2

1. Construct the discrete frequency distribution from the following data.

Marks: 50, 67, 71, 50, 67, 79, 50, 67, 71, 50, 80, 80, 79, 71, 50

Solution 1

The discrete frequency distribution of the given data is as follows:

Marks	Tally marks	Frequency
50	IIII	5
67	III	3
71	III	3
79	II	2
80	II	2
		N = 15

2. Construct age distribution of 17 persons taking class width as 10.

Age: 10, 25, 15, 35, 30, 25, 40, 35, 30, 32, 42, 25, 33, 40, 23, 17, 12

Solution 2

Age distribution of 20 students is given as follows:

Age	Tally marks	No. of person
10 - 20	IIII	4
20 - 30	IIII	4
30 - 40	IIII I	6
40 - 50	III	3
		N = 17

3. The birth weights (kilogram) of 30 children were recorded as follows:

2.0	2.1	2.3	3.0	3.1	2.7	2.8	3.5	3.1	3.7
4.0	2.3	3.5	4.2	3.7	3.2	2.7	2.5	2.7	3.8
3.1	3.0	2.6	2.8	2.9	3.5	4.1	3.9	2.8	2.2

Construct frequency distribution using Sturge's formula.

Solution 3

We have,

N = 30

Scanned by CamScanner

No. of class, $k = 1 + 3.322 \log_{10} N = 5.9 \approx 6$

Width of class interval $= \dfrac{\text{Largest} - \text{Smallest}}{k} = \dfrac{4.2 - 2.0}{6} = 0.367 \approx 0.4$

Since, the data varies between 2.0 to 4.2 so,

Birth weights (kg)	Tally marks	No. of children
2.0 - 2.4	卌	5
2.4 - 2.8	卌	5
2.8 - 3.2	卌 III	8
3.2 - 3.6	卌	5
3.6 - 4.0	IIII	4
4.0 - 4.4	III	3
		N = 30

4. The following data refers to the weights of large oil tankers (in thousands of tons):

140	130	142	144	136	153	144	155
190	210	240	230	260	290	305	309
198	297	278	268	265	241	235	236
292	202	205	207	149	180	178	172
170	190	201	180	194	220	224	226
246	292	310	302	282	274	261	194
200	243						

Construct a frequency distribution of equal length.

Solution 4

Given,

$N = 50$

No. of class, $k = 1 + 3.322 \log_{10} N = 6.64 \approx 7$

Width of class interval $= \dfrac{\text{Largest} - \text{Smallest}}{k} = \dfrac{310 - 130}{7} = 25.7 \approx 30$

Since, the data varies between 130 to 310 so,

Weights (000 tons)	Tally marks	No. of tankers
130 - 160	卌 IIII	9
160 - 190	卌	6
190 - 220	卌 卌 I	11
220 - 250	卌 卌	10
250 - 280	卌 I	6
280 - 310	卌 II	7
310 - 340	I	1
		N = 50

5. Represent the following set of marks out of 50 obtained by 40 students.

46	12	14	16	17	21	34	35
40	42	3	7	9	2	17	18
24	3	7	18	4	40	37	36
32	29	26	39	41	5	3	2
36	26	16	17	19	21	24	25

Scanned by CamScanner

Solution 5

Since the frequency distribution varies between 4 and 46. Then,

Marks	Tally marks	No. of students
0 - 5	~~IIII~~ I	6
5 - 10	IIII	4
10 - 15	II	2
15 - 20	~~IIII~~ III	8
20 - 25	IIII	4
25 - 30	IIII	4
30 - 35	II	2
35 - 40	~~IIII~~	5
40 - 45	IIII	4
45 - 50	I	1
		N = 40

6. The minimum temperature (in °C) for Kathmandu for the month August, 2004 s reported by the meteorological department is given below. Construct a frequency distribution table for it.

14	24	25	22	21.2	30	24.3	27
16	17	26	30	31	19.5	26.2	18.3
16	17.2	22.1	24	26.1	23	28	
24	19	23.3	26	27	24	18	

Solution 6

Given,

N = 30

No. of class, $k = 1 + 3.322 \log_{10} 30 = 5.9 \approx 6$

Width of class interval $= \frac{\text{Largest} - \text{Smallest}}{k} = \frac{31 - 14}{6} = 2.83 \approx 3$

Since, the data varies between 14 and 31 inclusive. So,

Temperature (°C)	Tally marks	Frequency
14 - 17	IIII	4
17 - 20	~~IIII~~ I	6
20 - 23	III	3
23 - 26	~~IIII~~ II	7
26 - 29	~~IIII~~ II	7
30 - 32	III	3
		N = 30

7. The following are the monthly rents (in Rs.) of 40 shops. Tabulate the data by grouping in interval of Rs. 8.

36	44	74	72	42	46	54	53
84	31	62	54	53	34	82	76
74	72	71	67	30	41	54	52
59	64	67	69	70	72	82	80
47	39	34	38	47	40	54	56

Solution 7

Since, the data varies between 30 and 84

Scanned by CamScanner

Rents	Tally marks	Frequency
30 - 38	~~IIII~~	5
38 - 46	~~IIII~~ I	6
46 - 54	~~IIII~~ I	6
54 - 62	~~IIII~~	5
62 - 70	~~IIII~~	5
70 - 78	~~IIII~~ III	8
78 - 86	~~IIII~~	5
		N = 40

8. Prepare a frequency distribution taking a class interval c fingers relating to the wages in Rs. of 50 workers:

28	17	46	54	50	82
19	21	55	72	61	64
49	43	39	17	26	29
51	62	67	70	74	78
26	29	32	41	45	58
81	71	61	51	41	31
24	48				

Solution 8

Since, the data varies between 17 and 86

Wages (in Rs)	Tally marks	Frequency
10 - 20	~~IIII~~	5
20 - 30	~~IIII~~ IIII	9
30 - 40	IIII	4
40 - 50	~~IIII~~ II	7
50 - 60	~~IIII~~ IIII	9
60 - 70	~~IIII~~ II	7
70 - 80	~~IIII~~	5
80 - 90	IIII	4
		N = 50

9. Form a frequency distribution from the following data taking 4 as the magnitude of class interval:

3	7	9	8	14	18
13	36	42	4	17	26
14	17	30	28	35	37
20	24	5	8	36	24
36	38	30	32		

Solution 9

Since, the data varies between 3 and 40

Class-interval	Tally marks	Frequency
3 - 7	IIII	4
8 - 11	IIII	4
12 - 15	IIII	4
16 - 19	III	3
19 - 22	I	1
22 - 25	III	3
26 - 29	IIII	4
30 - 33	IIII	4
34 - 37	~~IIII~~ I	6
38 - 41	II	2
42 - 45	I	1
		N = 36

Scanned by CamScanner

of 10 from the following

82	84	86
64	52	51
29	32	52
78	19	21
58	63	67
31	21	18

a by 'Inclusive methods'

18	26	40
36	29	31
37	11	14
34	23	34

10. The weights (in grams) of 40 oranges picked at random from a basket are as follows:
45, 55, 30, 110, 75, 100, 40, 60, 65, 40, 100, 75, 70; 60, 70, 95, 85, 80, 35, 45, 40, 50, 60, 65, 55, 45, 90, 85, 75, 85, 75, 70, 110, 105, 105, 80, 70, 55, 30, 70, 42
Construct an ordinary frequency table.

Solution 10

Given, N = 40

Since, the data varies between 30 and 110

Weights (in grams)	Tally marks	Frequency
30	II	2
35	I	1
40	III	3
42	I	1
45	III	3
50	I	1
55	III	3
60	III	3
65	II	2
70	IIII	5
75	IIII	4
80	II	2
85	III	3
90	I	1
95	I	1
100	II	2
105	I	1
110	II	2
		N = 40

11. The following table gives the weight (to the nearest kg) of 40 students at a college. Construct a frequency distribution with 7 classes.

40, 37, 47, 80, 53, 54, 77, 92, 96, 52, 57, 63, 32, 42, 46, 34, 56, 88, 70, 66, 60, 50, 59, 53, 73, 83, 93, 97, 92, 86, 54, 37, 42, 40, 50, 80, 90, 91, 34, 35

Solution 11

Given, N = 40 No. of class = 7

$$\text{Width of class} = \frac{\text{Largest} - \text{Smallest}}{7} = \frac{97-32}{7} = 9.28 = 10$$

Weights (in kg)	Tally marks	Frequency
30 - 40	IIII I	6
40 - 50	IIII I	6
50 - 60	IIII IIII	10
60 - 70	III	3
70 - 80	III	3
80 - 90	IIII	5
90 - 100	IIII II	7
		N = 40

Example Asked Questions

Short Questions

1. Write short notes in primary data [PU 2016 Spring/Purbanchal 2010]

Solution

See in theoretical part of this chapter.

Scanned by CamScanner

3

Chapter

Diagrammatic and Graphical Representation

There is and old saying that "one picture is worth a thousand words." Here, we study about some diagram and graphs of numerical data.

General rules for constructing diagram

a. **Title:** A suitable, short, clear and self explanatory title
b. Selection of diagram
c. Selection of scale (scale should be mentioned)
d. Proportion between width and height
e. Neatness and cleanness
f. Index
g. Footnote (if necessary)
h. Simplicity

Types of diagrams

There are different types of diagrams in use. Some common used diagrams are given below:

a. One dimensional diagrams (All bar diagrams)
b. Two dimensional diagrams (Squares and circles)
c. Three dimensional diagrams (Cubes)

a. **One dimensional diagrams**

i. Simple bar diagram
ii. Multiple bar diagram
iii. Sub-divided bar diagram
iv. Percentage bar diagram

Scanned by CamScanner

b. **Two dimensional diagrams**

i. Rectangular diagram ii. Percentage diagram

iii. Pie-chart (pie-diagram)

Frequency distribution

i. Histogram ii. Frequency polygon

iii. Frequency curve iv. Ogive (cumulative frequency)

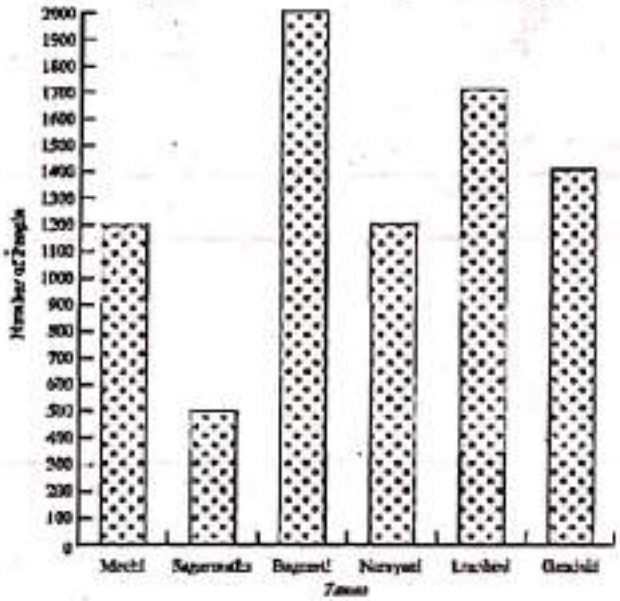
Exercise 3

1. Draw the simple bar diagram of the total number of people in different zones in Nepal in the years 1972.

Zones:	Gandaki	Lumbini	Naryani	Bagmati	Sagarmatha	Mechi
No. of peoples	1400	1700	1200	2000	500	1200

Solution 1

The given data can be represented in the simple bar diagram in the given way:



2. The following table gives the total expenditure of a family in the year 2004, represent this data by a pie-diagram.

Items	Expenditure (Rs)
Food	2600
Clothing	700
Education	2100
House rent	700
Fuel	350
Extra	1050

Solution 2

We have total expenditure of a family Rs. 7500

Let Rs. 7500 = 360° ⇒ Rs. 1

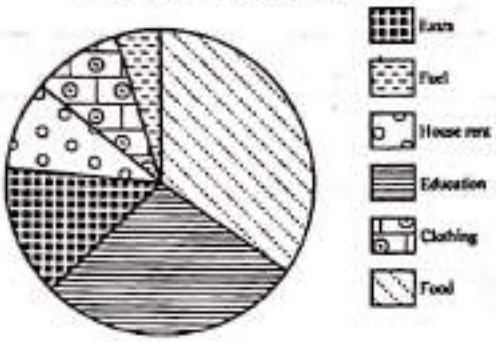
$$= \left(\frac{360}{7500}\right)^{\circ} = (0.048)^{\circ}$$

Scanned by CamScanner

Items	Expenditure (Rs)	Expenditure (In °)	Cum (°)
Food	2600	$2600 \times (0.048)^\circ = 124.8^\circ \approx 125^\circ$	125°
Clothing	700	$700 \times (0.048)^\circ = 33.6^\circ \approx 34^\circ$	159°
Education	2100	$2100 \times (0.048)^\circ = 100.8^\circ \approx 100^\circ$	259°
House rent	700	$700 \times (0.048)^\circ = 33.6^\circ \approx 34^\circ$	293°
Fuel	350	$350 \times (0.048)^\circ = 16.8^\circ \approx 17^\circ$	310°
Extra	1050	$1050 \times (0.048)^\circ = 50.4^\circ \approx 30^\circ$	360°
Total	7500	360°	

The pie-chart of the given data is showing in figure.

Pie-chart of Expenditure



3. Draw a pie-diagram to represent the following population in a certain village A:

Male	Female	Girls	Boys	Total
20	18	15	19	72

Solution 3

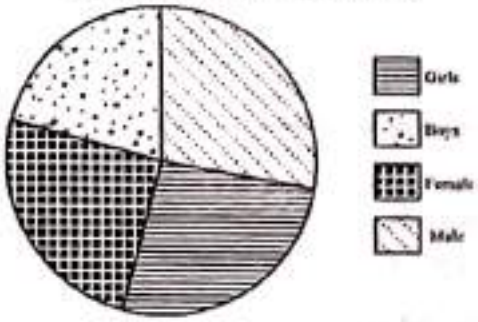
Here, total population is 72

$$\text{Let } 72 = 360^\circ \Rightarrow 1 = \left(\frac{360}{72}\right)^\circ = 5^\circ$$

	Population	Population (In °)	Cum (°)
Male	20	$20 \times (5)^\circ = 100^\circ$	100°
Female	18	$18 \times (5)^\circ = 90^\circ$	190°
Girls	15	$15 \times (5)^\circ = 75^\circ$	265°
Boys	19	$1 \times (5)^\circ = 95^\circ$	360°
Total	72	360°	

The pie-diagram of given data is shown in figure.

Pie-diagram of Population



Scanned by CamScanner

4. Compare number of students of the data by using pie-chart of two colleges A and B.

Program	Colleges	
	A	B
Civil	500	400
Information Technology	300	200
Computer	400	100
Mechanical	300	50
Electrical	300	150

Solution 4

Here, Total no. of students in College A = 1800

$$\text{Let } 1800 = 360^\circ \Rightarrow 1 = \left(\frac{360}{1800}\right)^\circ = (0.2)^\circ$$

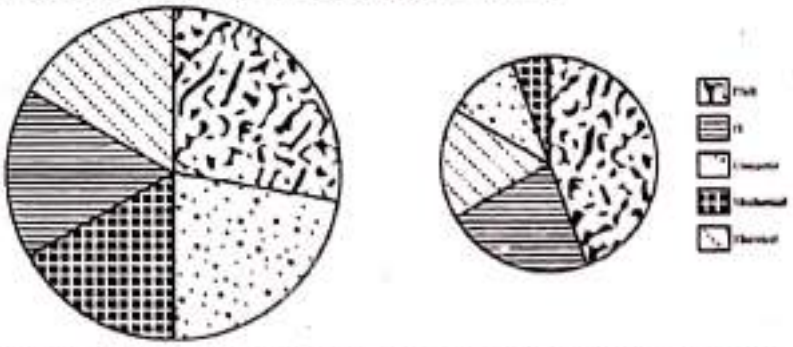
Also, Total no. of students in College B = 900

$$\text{Let } 900 = 360^\circ \Rightarrow 1 = \left(\frac{360}{900}\right)^\circ = (0.4)^\circ .$$

For College A			
Program	No. of students	No. of students (In °)	Cum (°)
Civil	500	$500 \times (0.2)^\circ = 100^\circ$	100°
IT	300	$300 \times (0.2)^\circ = 60^\circ$	160°
Computer	400	$400 \times (0.2)^\circ = 80^\circ$	240°
Mechanical	300	$300 \times (0.2)^\circ = 60^\circ$	300°
Electrical	300	$300 \times (0.2)^\circ = 60^\circ$	360°
Total	1800	360°	
Square root	42.43		
Ratio of circle	1.41		

For College B			
Program	No. of students	No. of students (In °)	Cum (°)
Civil	400	$400 \times (0.2)^\circ = 160^\circ$	160°
IT	200	$200 \times (0.2)^\circ = 80^\circ$	240°
Computer	100	$100 \times (0.2)^\circ = 40^\circ$	280°
Mechanical	50	$50 \times (0.2)^\circ = 20^\circ$	300°
Electrical	150	$150 \times (0.2)^\circ = 60^\circ$	360°
Total	900	360°	
Square root	30		
Ratio of circle	1		

Pie-chart of B.E. students in two colleges A and B



5. The following data relates to the expenditure of families (per month):

Items	Family A	Family B
Food	200	300
Education	50	100
Clothing	50	70
Fuel	50	40
House rent	100	50
Miscellaneous	50	100

Represent these data by angular diagrams:

Scanned by CamScanner

Solution 5

Here, total expenditure of family A is 500

$$\text{Let } 500 = 360^\circ \Rightarrow 1 = \left(\frac{360}{500}\right)^\circ = (0.72)^\circ$$

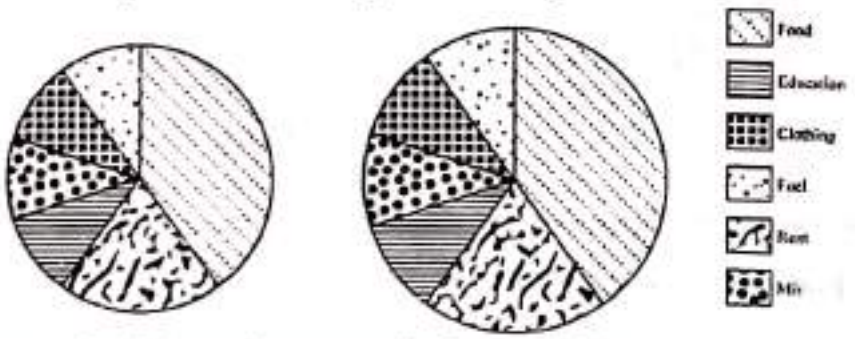
Also, total expenditure of family B is 660

$$\text{Let } 660 = 360^\circ \Rightarrow 1 = \left(\frac{360}{660}\right)^\circ = (0.55)^\circ$$

For College A			
Program	No. of students	No. of students (in °)	Cum (°)
Food	200	$200 \times (0.74)^\circ = 144^\circ$	144°
Education	50	$50 \times (0.74)^\circ = 36^\circ$	180°
Clothing	50	$50 \times (0.74)^\circ = 36^\circ$	216°
Fuel	50	$50 \times (0.74)^\circ = 36^\circ$	252°
Rent	100	$100 \times (0.74)^\circ = 72^\circ$	324°
Mis.	50	$50 \times (0.74)^\circ = 36^\circ$	360°
Total	500	360°	
Square root	22.36		
Ratio of circle	1		

For College B			
Program	No. of students	No. of students (in °)	Cum (°)
Food	300	$300 \times (0.55)^\circ = 165^\circ$	165°
Education	100	$100 \times (0.55)^\circ = 55^\circ$	220°
Clothing	70	$70 \times (0.55)^\circ = 38.5^\circ$	258.5°
Fuel	40	$40 \times (0.55)^\circ = 22^\circ$	280.5°
Rent	50	$50 \times (0.55)^\circ = 27.5^\circ$	305°
Mis.	100	$100 \times (0.55)^\circ = 55^\circ$	360°
Total	660	360°	
Square root	25.69		
Ratio of circle	1.15		

Pie-chart of expenditure of family A and family B



6. Represent the following frequency distribution by histogram.

Scores	Frequency
90 - 99	2
80 - 89	12
70 - 79	14
60 - 69	17
50 - 59	2
40 - 49	5
30 - 39	2

Solution 6

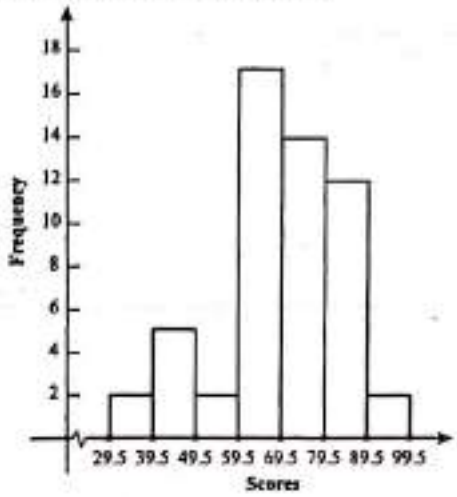
Since, the given frequency distribution is in inclusive, but we have to change into exclusive. For this, correction factor $C_f = \frac{39-40}{2} = 0.5$

Subtracting C_f from beginning value and adding in the last value, we have

Scores	29.5-39.5	39.5 - 49.5	49.5 - 59.5	59.5 - 69.5	69.5 - 79.5	79.5 - 89.5	89.5 - 99.5
Frequency	2	5	2	17	14	12	2

Scanned by CamScanner

The suitable histogram is shown in the figure.

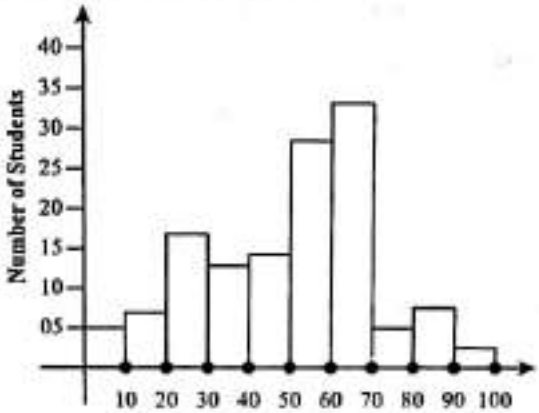


7. Represent the following using histogram:

Marks	No. of students
0 - 10	5
10 - 20	7
20 - 30	16
30 - 40	12
40 - 50	13
50 - 60	27
60 - 70	33
70 - 80	5
80 - 90	7
90 - 100	2

Solution 7

We have given the exclusive frequency distribution with the equal class interval. So, the suitable histogram is in the figure.



8. Given below the pre-tax monthly income (in rupees)of the workers of FEWA hydropower in 2000. [PU 2012 Spring]]

Income more than ('000' Rs.)	7	6	5	4	3	2	1	0
No. of workers	2	8	10	15	25	40	55	60

Draw less than ogive and hence find out,

i. The highest income of the lowest 50% of the worker.

ii. The minimum income earned by the top 5% of the workers.

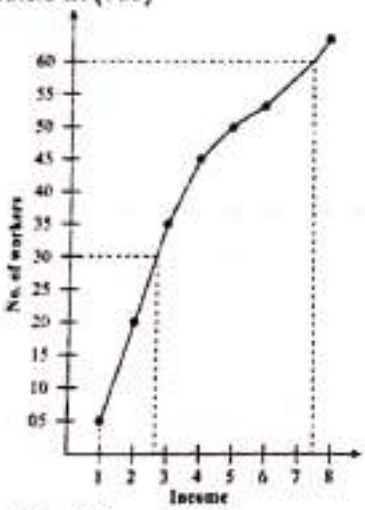
Scanned by CamScanner

Solution 8

Given frequency distribution can be written as

Income (000)	Frequency	Income (000)	Frequency
0 - 1	60 – 55 = 5	Less than 1	5
1 - 2	55 – 40 = 15	Less than 2	20
2 - 3	40 – 25 = 15	Less than 3	35
3. - 4	25 – 15 = 10	Less than 4	45
4 - 5	15 - 10 = 5	Less than 5	50
5 - 6	10 – 8 = 2	Less than 6	52
6 - 7	8 – 2 = 6	Less than 7	58
7 above	2	Less than 8	60

Less than O-give of income in (000)



i. Total number of workers N = 60

50% of workers $= \frac{50}{100} \times 60 = 30$

For this draw a parallel line from 30 to less than of ogive and draw perpendicular from the meeting point of ogive to the x-axis $= 2 + \frac{2}{3}(3 - 1) \times 1000 =$ Rs. 2670

ii. 5% of the workers from the top

$= \frac{5}{100} \times 60 = 3$ [i.e. 60 – 3 = 57] workers

$= 7.5 \times 1000 =$ Rs. 75000 [because it lies between 7 and 8]

9. From the following distribution of 500 students of a college find the marks if only 20% students had failed and also the minimum marks obtained by the top 25% of the students.

Marks	0-20	20-40	40-50	50-60	60-80	80-100
No. of students	50	100	150	90	60	50

Also represent the data by histogram so locate the model.

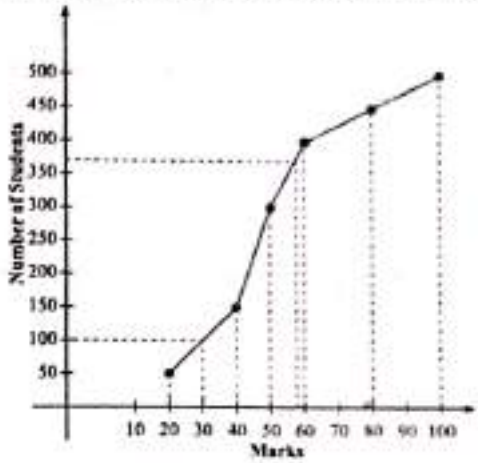
[PU 2017 Fall/ 2013 Spring]

Solution 9

Marks	No. of students	Marks	f
0 - 20	50	Less than 20	50
20 - 40	100	Less than 40	150
40 - 50	150	Less than 50	300
50 - 60	90	Less than 60	390
60 - 80	60	Less than 80	450
80 - 100	50	Less than 100	500

Scanned by CamScanner

Less than ogive of marks of students



i. Since 20% of the students are failed so,

Number of failed students $= 20\%$ of $500 = \frac{20}{100} \times 500 = 100$

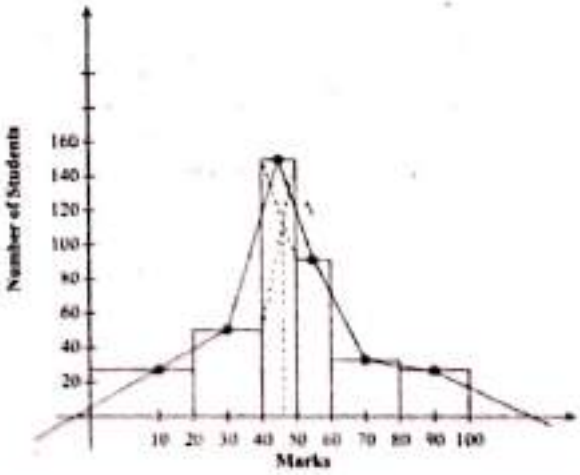
Now, the highest marks obtained by 20% (100) students = 30 marks

ii. The minimum marks of top 25% of the students $= 500 - \frac{25}{100} \times 500 = 375$

i.e. the mark obtained by 375 is $= 50 + \frac{3}{4} \times (60 - 50) = 57.5$

Also for the histogram

Marks	No. of students	Magnitude of class	Adjust frequency
0 - 20	50	20 = 10 × 2	50/2 = 25
20 - 40	100	20 = 10 × 2	100/2 = 50
40 - 50	150	10	150
50 - 60	90	10	90
60 - 80	60	20 = 10 × 2	60/2 = 30
80 - 100	50	20 = 10 × 2	50/2 = 25



Since, highest frequency is 150 which lies in the interval 40 - 50. So, mode lies between 40 to 50. Therefore mode = 45 (approximately).

Scanned by CamScanner

10. Define histogram, the weekly salaries of 84 workers in a factory are given below:

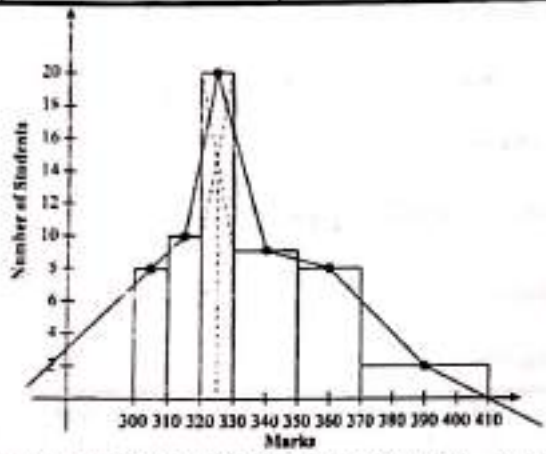
Salaries	300-310	310-320	320-330	330-350	350-370	370-410
No. of workers	8	10	20	18	16	12

Construct histogram, frequency polygon and frequency curve. Also find the mode from histogram.

Solution 10

Histogram: A bar diagram which is suitable for frequency distribution of continuous classes is called histogram.

Salaries	Class-width	Frequency	Adjusted frequency
300 - 310	10	8	8
310 - 320	10	10	10
320 - 330	10	20	20
330 - 350	20	18	18/2 = 9
350 - 370	20	16	16/2 = 8
370 - 410	40	12	12/4 = 3



Since, highest frequency is 20 which lies between 320 - 330. Thus, mode lies in the interval 320 - 330.

∴ Mode = 326 approximately

11. The following are measurements of the breaking strength (in ounces) of sample of 30 linen threads:

32.5	15.2	35.4	21.3	28.4	26.9	24.6	29.3
24.5	31.8	21.2	28.3	27.1	25.0	32.7	29.5
30.2	23.9	23.0	26.4	27.3	33.7	29.4	21.9
29.3	17.3	29.0	36.8	29.2	23.5		

group these measurements into continuous frequency distribution of suitable class interval and construct histogram, frequency, polygon and frequency curve.

Solution 11

Here,

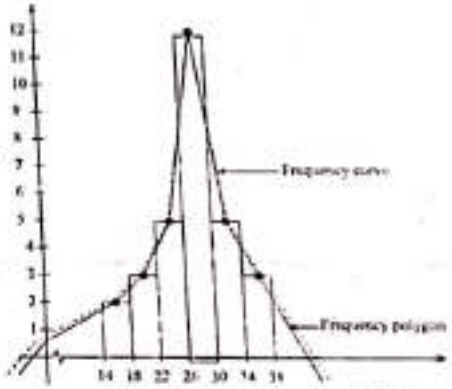
$N = 30$

No. of class intervals $(k) = 1 + 3.322 \log_{10} N = 5.9 \approx 6$

$$\text{Class width} = \frac{\text{Largest value} - \text{Smallest value}}{k} = \frac{35.4 - 15.2}{6} = 3.37 \approx 4$$

Strength (in ounces)	Frequency (f)
14 - 18	2
18 - 22	3
22 - 26	5
26 - 30	12
30 - 34	5
34 - 38	3

Scanned by CamScanner



12. An investigator wants to study the speed of vehicles at Araniko highway and he collected the speed of 30 vehicles and the speed were:

30	35	37	42	45	47	48	50	55	67
70	75	80	90	95	94	45	55	60	71
63	70	65	80	55	40	35	36	65	79

From the above data construct continuous frequency distribution using Struge's rule and construct the cumulative frequency curve and find median speed, speed of first 25% vehicles, speed of first 75% of vehicles and also compute the percentage of vehicles whose speed lies between 40 to 70 km.

Solution 12

Here,

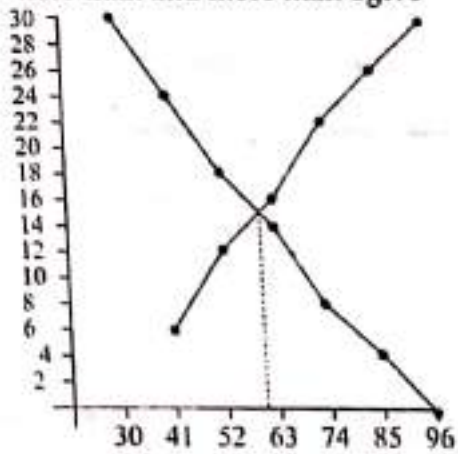
$N = 30$

No. of class intervals (k) = $1 + 3.322 \log_{10}N = 5.9 = 6$

Class width $= \dfrac{\text{Largest value} - \text{Smallest value}}{k} = \dfrac{95-30}{6} = 10.8 = 11$

Speed (km/s)	No. of vehicles	Speed (km/h)	f	speed	f
30 - 41	6	Less than 41	6	More than 30	30
41 - 52	6	Less than 52	12	More than 41	24
52 - 63	4	Less than 63	16	More than 52	18
63 - 74	6	Less than 74	22	More than 63	14
74 - 85	4	Less than 85	26	More than 74	8
85 - 96	4	Less than 96	30	More than 85	4
				More than 96	0

Less than and more than ogive



Here, median M_d = 60 (approximately)

Speed of first 25% vehicles $= \dfrac{25}{100} \times 30 = 7.5$ vehicles = 45 km/hr

Scanned by CamScanner

Speed of first 75% vehicles $= \frac{75}{100} \times 30 = 22.5$ vehicles $= 80$ km/hr

The number of vehicles whose speed is 40 to 70 km/hr is 20.

Percentage $= \frac{20}{30} \times 100\% = 67\%$ (Approximately)

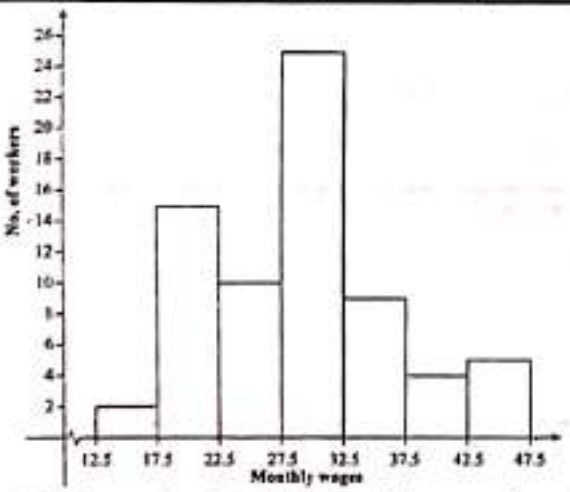
Vehicles has speed between 40 to 70 km/hr.

13. Draw a histogram representing the frequency distribution.

Monthly wages (in Rs)	No. of Workers
15	2
20	15
25	10
30	25
35	9
40	4
45	5

Solution 13

Monthly waves (in Rs)	Class Interval	No. of workers
15	12.5 - 17.5	2
20	17.5 - 22.5	15
25	22.5 - 27.5	10
30	27.5 - 32.5	25
35	32.5 - 37.5	9
40	37.5 - 42.5	4
45	42.5 - 57.5	5



14. The following table shows the distribution of the number of students per teacher in 750 colleges:

Students	1	4	7	10	13	16	19	22	25	28
Frequency	7	46	165	195	189	89	28	19	9	3

Draw the histogram for the data and superimpose on it the frequency polygon.

Solution 14

Do yourself

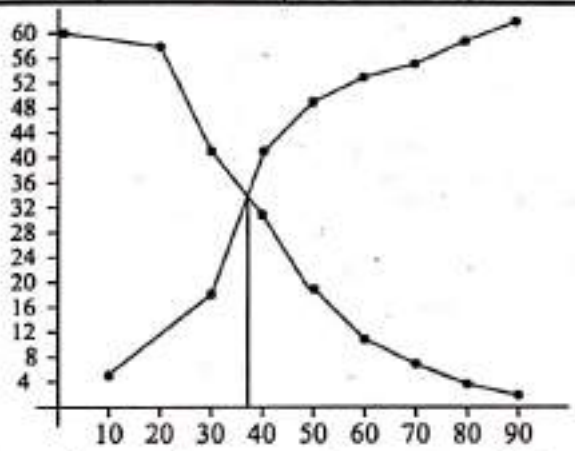
15. Represent the following distribution by an ogive:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students	5	13	12	11	8	4	2	3	2

Scanned by CamScanner

Solution 15

Marks	No. of students	Marks	c.f.	Marks	c.f.
0 - 10	5	Less than 10	5	More than 0	60
10- 20	13	Less than 20	18	More than 10	55
20 - 30	12	Less than 30	30	More than 20	42
30 - 40	11	Less than 40	41	More than 30	30
40 - 50	8	Less than 50	49	More than 40	19
50 - 60	4	Less than 60	53	More than 50	11
60 - 70	2	Less than 70	55	More than 60	7
70 - 80	3	Less than 80	57	More than 70	4
80 - 90	2	Less than 90	60	More than 80	2



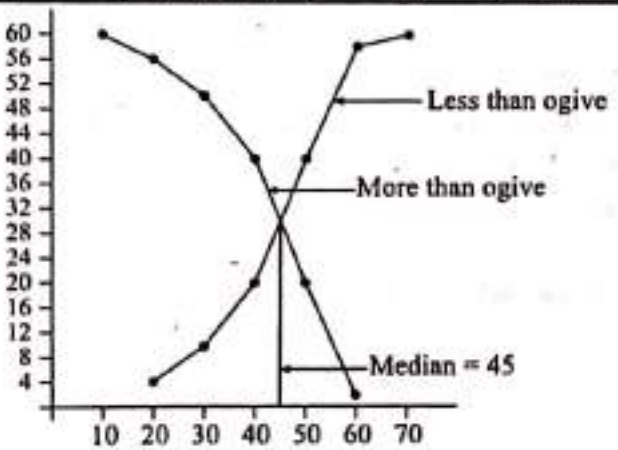
16. Draw less than and more than an ogive curve of the following frequency distribution:

Marks	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	4	6	10	20	18	2

And then find median value.

Solution 16

Marks	No. of students	Marks	c.f.	Marks	c.f.
10- 20	4	Less than 20	4	More than 10	60
20 - 30	6	Less than 30	10	More than 20	56
30 - 40	10	Less than 40	20	More than 30	50
40 - 50	20	Less than 50	40	More than 40	40
50 - 60	18	Less than 60	58	More than 50	20
60 - 70	2	Less than 70	60	More than 60	2



Scanned by CamScanner

17. The frequency distribution of wages in a certain factory is as follows:

Wages (in Rs)	No. of students
250-259	10
260-269	18
170-279	27
280-289	20
290-99	15
300-309	8
310-319	2

Draw an ogive for this distribution.

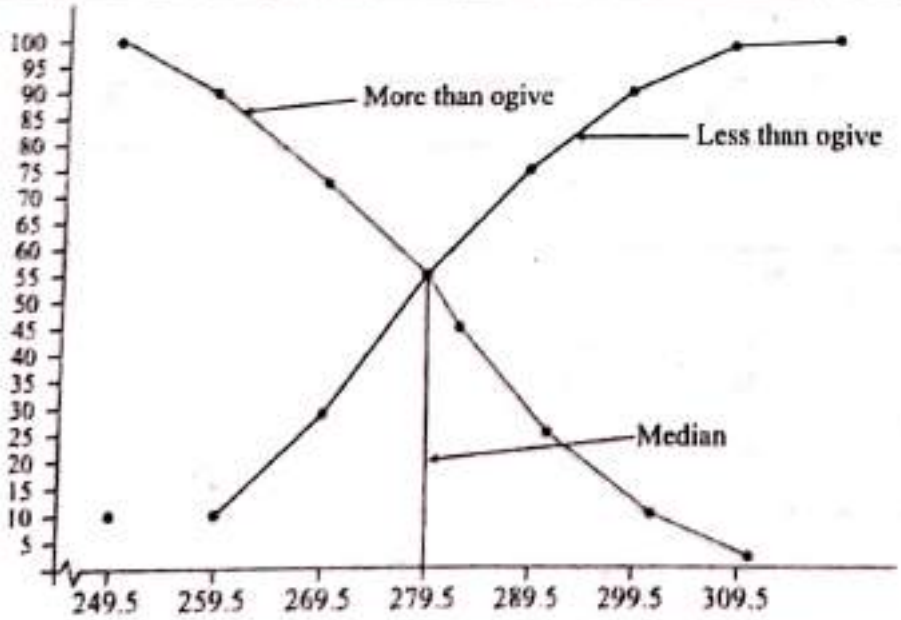
Solution 17

Since the given data is in inclusive. So first we have to change it into exclusive.

For this, correction factor $C_f = \left(\frac{260-259}{2}\right) = 0.5$.

Subtracting from first and adding in last it changes into exclusive.

Wages	No. of students	Wages (in Rs.)	c.f.	Wages (in Rs.)	c.f.
249.5 - 259.5	10	Less than 259.5	10	More than 249.5	100
259.5 - 269.5	18	Less than 269.5	28	More than 259.5	90
269.5 - 279.5	27	Less than 279.5	55	More than 269.5	72
279.5 - 289.5	20	Less than 289.5	75	More than 279.5	45
289.5 - 299.5	15	Less than 299.5	90	More than 289.5	25
299.5 - 309.5	8	Less than 309.5	98	More than 299.5	10
309.5 - 319.5	2	Less than 319.5	100	More than 309.5	2



18. The marks obtained by 50 students in an examination are given below:

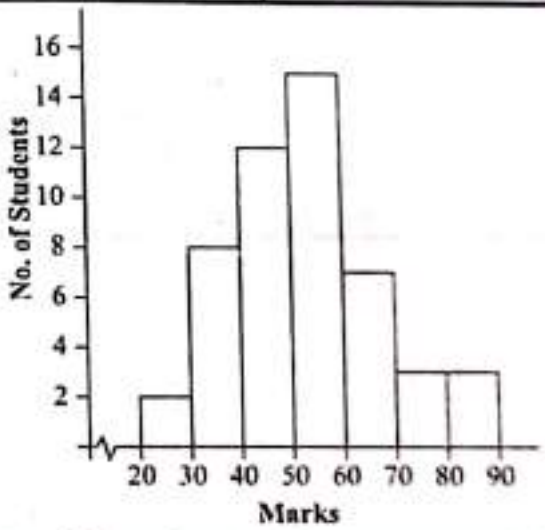
25, 30, 67, 84, 28, 32, 65, 70, 88, 43, 54, 49, 52, 42, 42, 52, 51, 58, 59, 50, 44, 38, 61, 76, 87, 41, 42, 50, 50, 56, 58, 37, 30, 35, 36, 35, 63, 65, 60, 49, 35, 54, 45, 59, 40, 44, 52, 51, 73, 61, 46

Classify the above data taking class-interval 20-30, 30-40, etc. and plot a histogram.

Scanned by CamScanner

Solution 18

Marks	No. of Students (f)
20 - 30	2
30 - 40	8
40 - 50	12
50 - 60	15
60 - 70	7
70 - 80	3
80 - 90	3



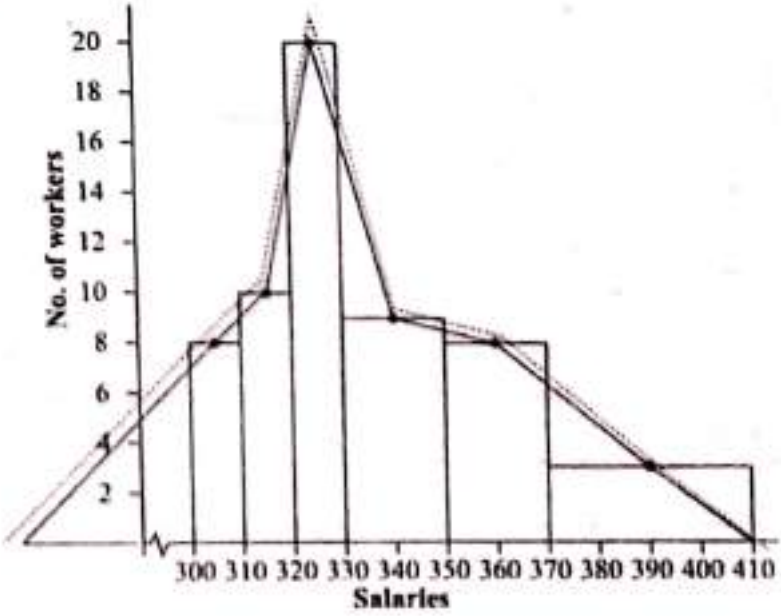
19. The weekly salaries of 90 workers in a factory are given below:

Salaries	300-310	310-320	320-330	330-350	350-370	370-410
No. of workers	8	10	20	18	16	12

Construct histogram, frequency polygon and frequency curve.

Solution 19

Histogram, frequency polygon and frequency curve of weekly salary of workers in a factory.



Scanned by CamScanner

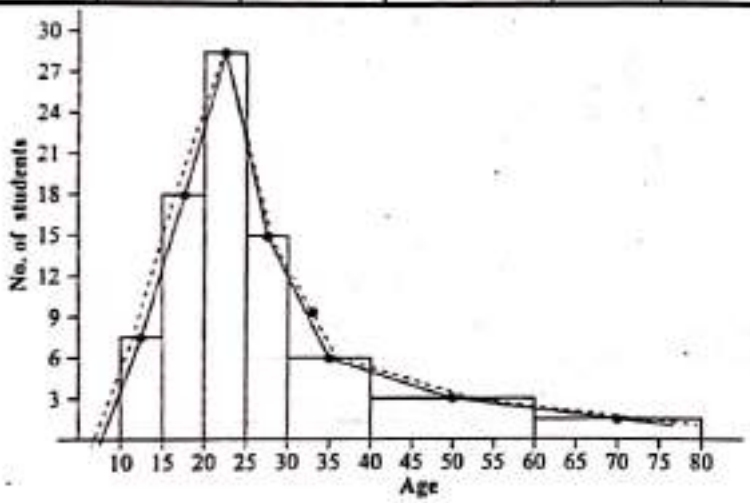
20. The following data represents age workers and no. of workers in a certain survey.

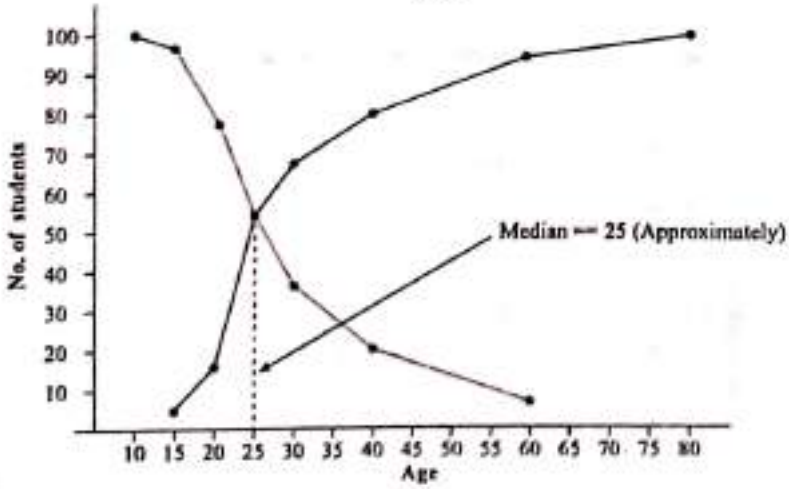
Age	10-15	15-20	20-25	25-30	30-10	40-60	60-80
No. of workers	7	18	28	15	12	12	8

Construct histogram, frequency polygon, frequency curve and ogive and indicate a median in your figure.

Solution 20

Age	No. of workers	Class-width	Adjusted Frequency	Age	cf	Age	cf
10 - 15	7	5	7	Less than 15	7	More than 10	100
15 - 20	18	5	18	Less than 20	25	More than 15	90
20 - 25	28	5	28	Less than 25	53	More than 20	75
25 - 30	15	5	15	Less than 30	68	More than 25	47
30 - 40	12	10 = 5 × 2	12/2 = 6	Less than 40	80	More than 30	32
40 - 60	12	20 = 5 × 4	12/4 = 3	Less than 60	92	More than 40	20
60 - 80	8	20 = 5 × 4	8/4 = 2	Less than 80	100	More than 60	8





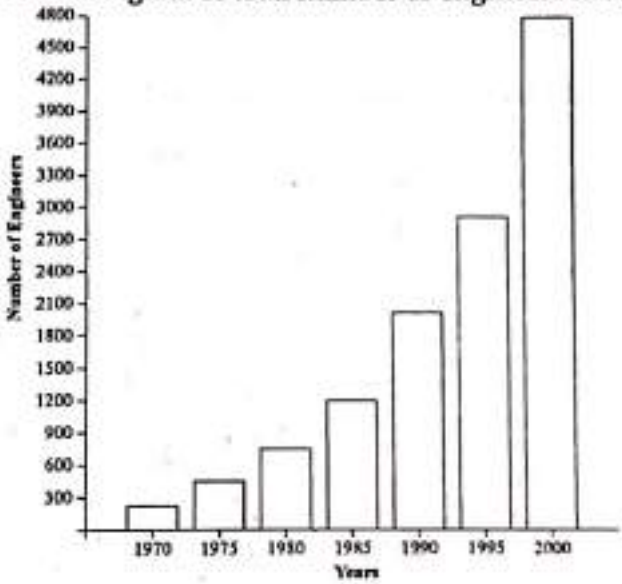
21. Draw the simple bar diagram of the total number of engineers in Nepal in different years.

Years	1970	1975	1980	1985	1990	1995	2000
No. of Engineers	100	350	700	1200	2000	3200	4500

Scanned by CamScanner

Solution 21

Simple bar-diagram of total number of engineers in Nepal

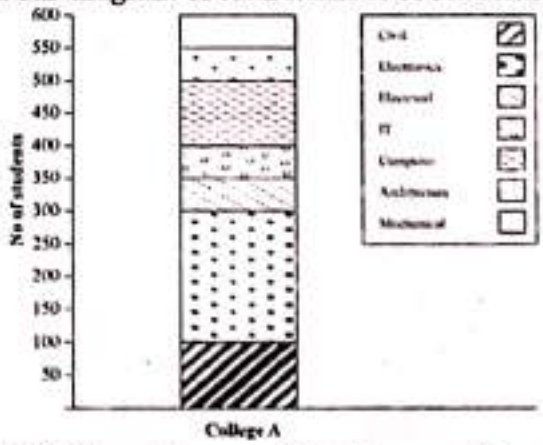


22. Draw the subdivided bar diagram of the total number of students in the college A in the years 2004.

Program	No. of Students
Civil	100
Electronics	200
Electrical	50
Information Tech	50
Computer	100
Architecture	50
Mechanical	50

Solution 22

Sub-divided bar-diagram of total number of students in a college



23. Draw the sub-divided bar diagram of total expenditure of two family A and family B in the years in the years 2003.

Items	Family A	Family B
Food	400	500
Clothing	200	150
Education	300	100
Fuel	200	150
House rent	150	100
Miscellaneous	400	300

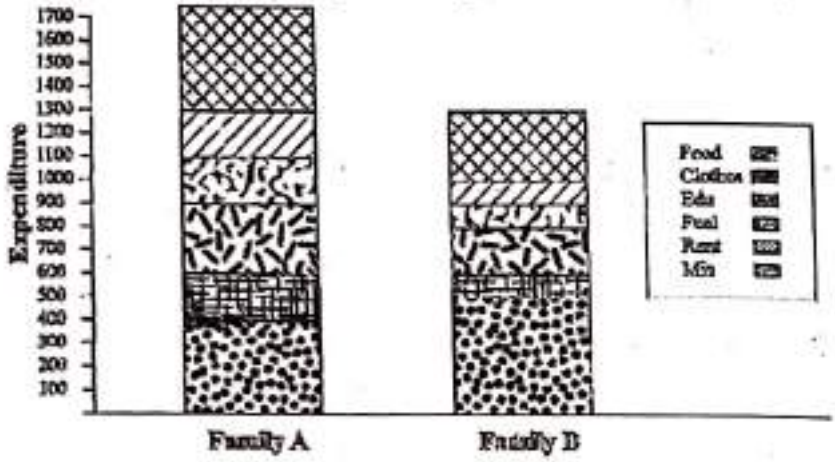
Scanned by CamScanner

Solution 23

We have given the total expenditure of family A is Rs. 1650 and family B is Rs. 1300.

Thus, the ratio of total expenditure of family A and B is $\frac{1650}{1300}$ =33 : 26

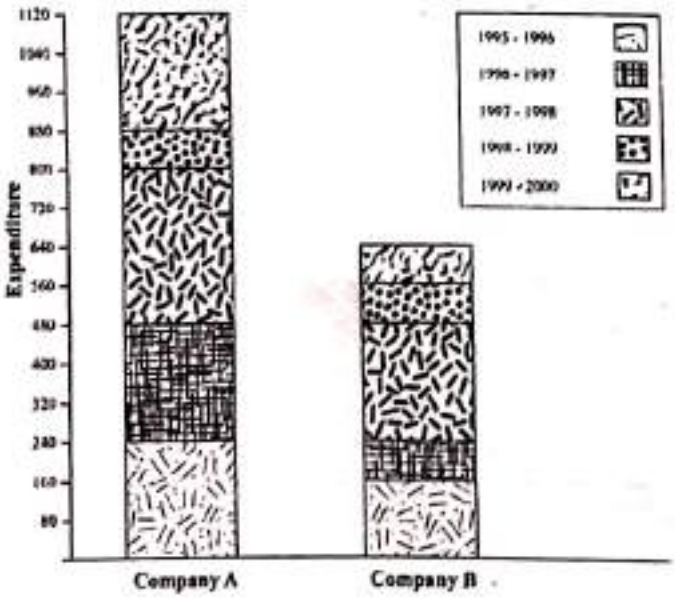
Hence, the breadth of the rectangle will be given in the ratio [33 : 26]



24. The data presented below give the yearly profits (in '000 Rs) of the two companies A and B. Represent this data by sub-divide bar diagrams.

Years	Profits (in '000 RS)	
	Company A	Company B
1995-1996	220	130
1996-1997	240	120
1997-1998	350	240
1998-1999	50	100
1999-2000	230	35

Solution 24



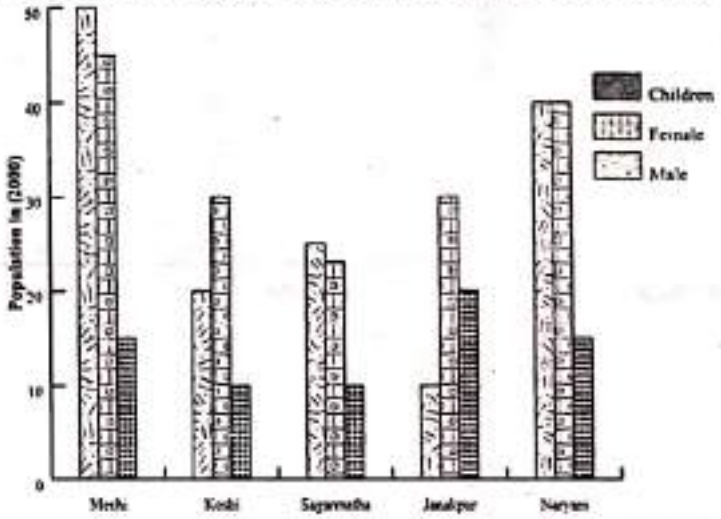
Scanned by CamScanner

25. Draw the multiple bar diagram of the following data.

Zones	Population (2000)		
	Male	Female	Children
Mechi	50	45	15
Koshi	20	30	10
Sagarmatha	25	22	10
Janakpur	40	30	20
Narayani	40	40	15

Solution 25

Multiple Bar-Diagram for Population of Different Zones

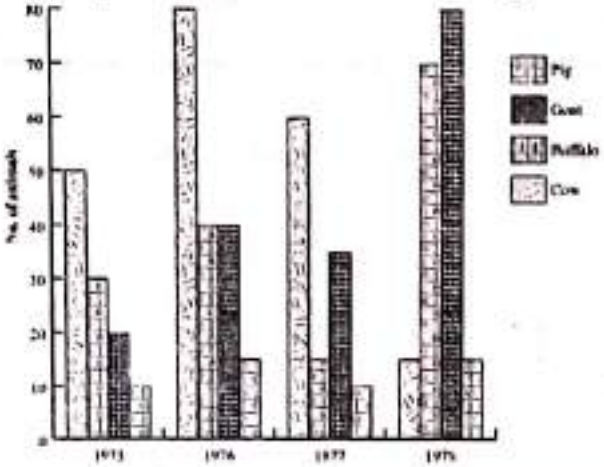


26. Represent the following data related to the production of Animals in different years from Janakpur by multiple bar-diagram.

Number of Years	Cow	Buffalo	Goat	Pig
1975	50	30	20	5
1976	80	40	40	15
1977	60	15	35	10
1978	16	70	60	15

Solution 26

Multiple bar diagram production of animals in Janakpur in different years



27. Represent the following data of the total number of students in a college A by using percentage bar diagram.

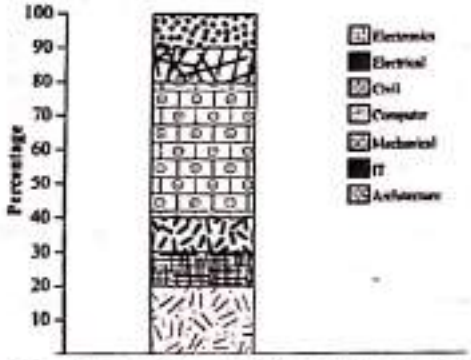
Scanned by CamScanner

Program	No. of Students
Electronics	200
Electrical	100
Civil	200
Computer	400
Mechanical	100
Information Tech	100
Architecture	50

Solution 27

Programme	No. of students	No. of students in %	Cumulative %
Electronics	200	17	17
Electrical	100	9	26
Civil	200	17	43
Computer	400	35	78
Medical	100	9	87
IT	100	9	96
Architecture	50	4	100
Total	1150	100%	

Percentage bar-diagram of total number of students in a college

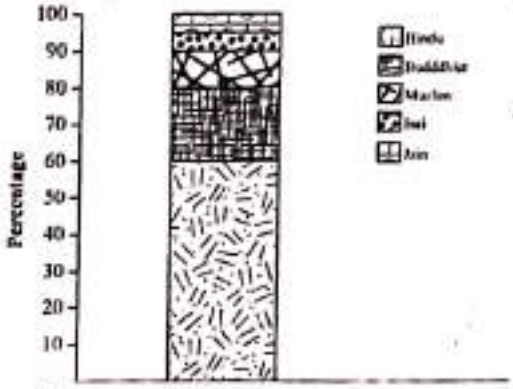


28. Represent the following data of the different types of religious people in Nepal, by percentage bar diagram:

Religious	No. of People (in %)
Hindu	60%
Buddhist	20%
Muslim	10%
Isali	5%
Jain	5%

Solution 28

Percentage bar-diagram of different religious people in Nepal



Scanned by CamScanner

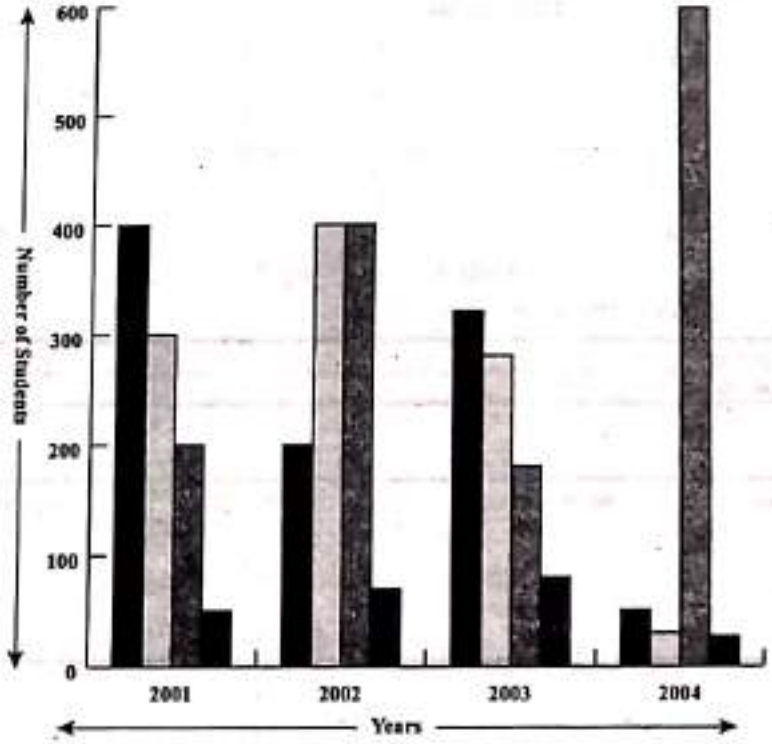
29. The following table give the result of the B.E. examination of a university for past four years:

Years	No. of student passed in			
	1st div	2nd div	3rd div	Failed
2001	400	300	200	50
2002	200	400	400	60
2003	320	280	240	60
2004	50	40	600	30

Represent the above data by suitable diagram.

Solution 29

Multiple bar-diagram of result of a university in different years



30. Draw the suitable diagram for the given data:

Expenditure	Family A	Family B
	Total income (Rs 4000)	Total income (Rs 5000)
Food	1500	3000
Education	500	2000
Clothing	300	500
Fuel	500	500
Saving or deficit	+1200	−1000

Solution 30

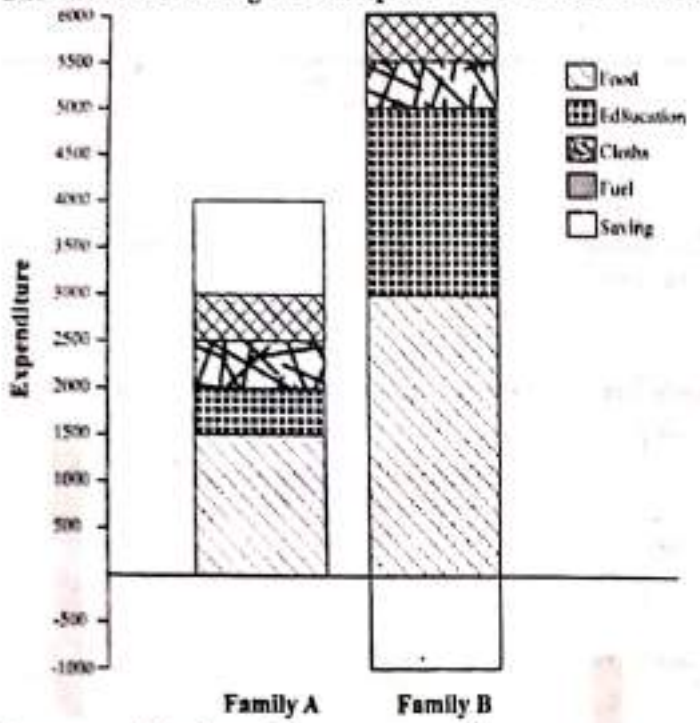
Total income of a family A is 400 and family B is 5000.

The ratio of income of A and B is $\frac{4000}{5000} = 4 : 5$.

Thus, the width of rectangle will be in the ratio of 4 : 5.

Scanned by CamScanner

Sub-divided bar-diagram of expenditure of two families

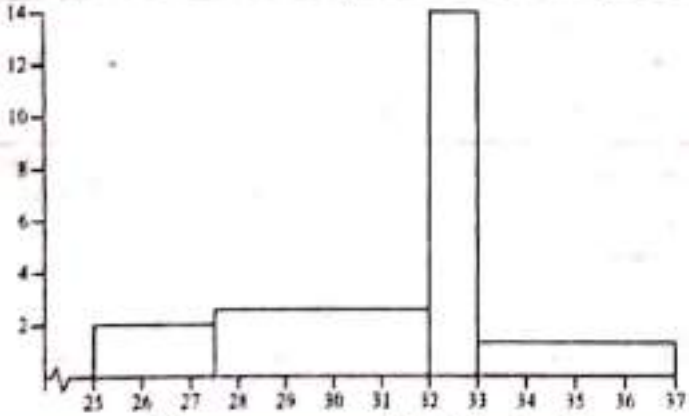


31. Draw a histogram of the given data:

Weights (in kg)	25 - 27.5	27.32	32-33	33-37
No. of students	5	10	14	5

Solution 31

Weights (in kg)	No. of students	Class-width	Adjusted frequency
25 - 27.5	5	2.5	5/2.5 =2
27.5 - 32	10	4.5	10/4.5 = 2.22
32 - 33	14	1	14
33 - 37	5	4	1.25



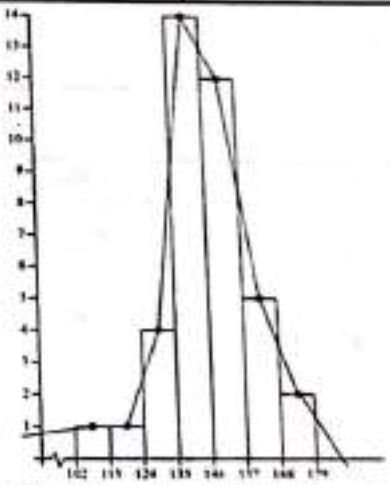
32. Draw a histogram and frequency curve of the following frequency distribution.

Weight (in kg)	No. of Buffalos : (f)
102 - 113	1
113-124	1
124-135	4
135-146	14
146-157	12
157-168	5
168-179	3

Scanned by CamScanner

Solution 32

Weights (in kg)	No. of buffaloes
102 - 113	1
113 - 124	1
124 - 135	4
135 - 146	14
146 - 157	12
157 - 168	5
168 - 179	3



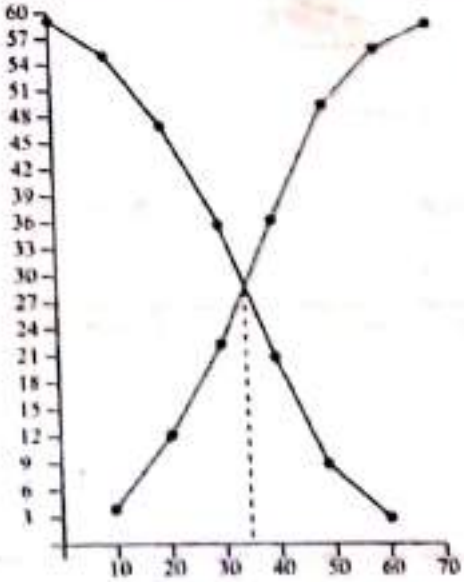
33. Draw the two ogive for the following distribution showing marks of 59 students:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	4	8	11	15	12	6	3

(Assume no negative mark) and then find median value.

Solution 33

Marks	No. of students	Marks	C_l	Marks	C_u
0 - 10	4	Less than 10	4	More than 0	59
10 - 20	8	Less than 20	12	More than 10	55
20 - 30	11	Less than 30	3	More than 20	47
30 - 40	15	Less than 40	38	More than 30	36
40 - 50	12	Less than 50	50	More than 40	21
50 - 60	6	Less than 60	56	More than 50	9
60 - 70	3	Less than 70	59	More than 60	3



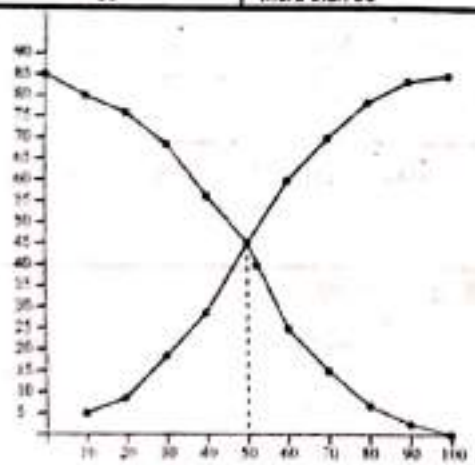
Scanned by CamScanner

34. Draw the two ogive for the following frequency distribution:

Marks	No. of Students
Less than 10	5
Less than 20	9
Less than 30	17
Less than 40	29
Less than 50	45
Less than 60	60
Less than 70	70
Less than 80	78
Less than 90	83
Less than 100	85

Solution 34

Marks	C_l	Marks	C_u
Less than 10	5	More than 0	85
Less than 20	19	More than 10	80
Less than 30	17	More than 20	76
Less than 40	29	More than 30	68
Less than 50	45	More than 40	56
Less than 60	60	More than 50	40
Less than 70	70	More than 60	25
Less than 80	78	More than 70	15
Less than 90	83	More than 80	7
Less than 100	85	More than 90	2



35. Draw a histogram of the following frequency distribution and use it to find the total number of workers in the age group 19-32 years.

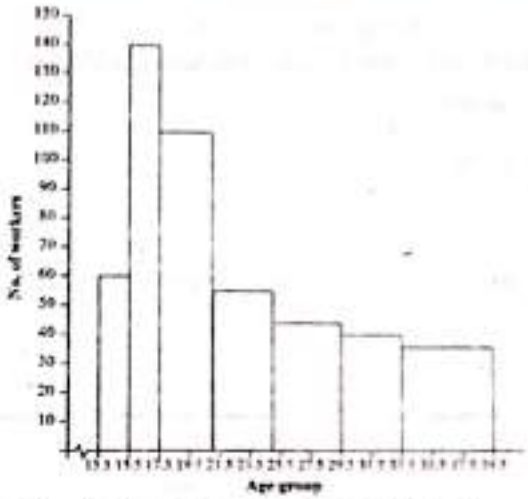
Age group	14-15	16-17	18-20	21-24	25-29	30-34	35-39
No. of workers	60	140	150	110	110	100	90

Solution 35

Since, the given frequency distribution is inclusive. First make it exclusive by finding correction factor $C_l = \frac{16 - 15}{2} = 0.5$

Age group	No. of workers	Class-width	Adjusted frequency
13.5 - 15.5	60	2	60
15.5 - 17.5	140	2	140
17.5 - 20.5	150	3	150 × 2 ÷ 3 = 100
20.5 - 24.5	110	4	110 ÷ 2 ÷ 55
24.5 - 29.5	110	5	110 × 2 ÷ 5 = 44
29.5 - 34.5	100	5	100 × 2 ÷ 5 = 40
34.5 - 39.5	90	5	90 × 2 ÷ 5 = 36

Scanned by CamScanner



The total number of students in the age group (19 - 32)

$$= \frac{1}{2}(17.5 - 20.5) + (20.5 - 24.5) + (24.5 - 29.5) + \frac{1}{2}(29.5 - 34.5)$$

$$= \frac{1}{2} \times 150 + 110 + 110 + \frac{1}{2} \times 100 = 345$$

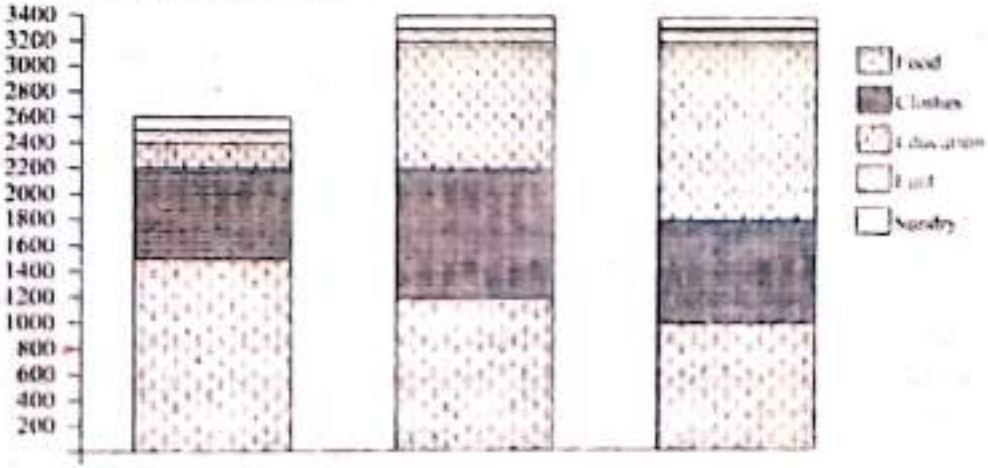
36. Represent the following data by suitable diagram.

Item of Expenditure	Family A	Family B	Family C
	Expenditure (Rs)	Expenditure (Rs)	Expenditure (Rs)
Food	1500	1250	1000
Clothing	750	1000	750
Education	250	1000	1500
Fuel and electricity	25	50	100
Sundry	50	100	50
Saving or deficit	+1000	-200	+750

Solution 36

We have given total expenditure of family A is 2575 and family B is 3400 and family C is 3400.

The ratio of expenditure of family A and B is 2575 : 3400 = 103 : 136 and ratio of expenditure of family A and C is 2575 : 3400 = 103 : 136. The width of diagram is in the ratio of 103 : 136 : 13.



Scanned by CamScanner

38. A business company investigated the delay associated with processing credit car applications, it collected the data into following categories:

Categories	Frequency
No signature	40
Residential address not valid	09
Non-legible handwriting	22
Already a customer	15
Other	08

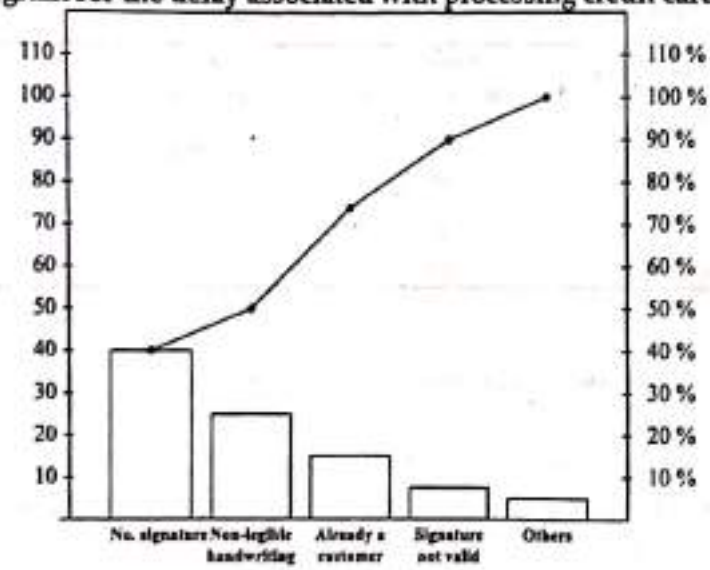
Draw a Pareto-chart and interpret your result.

Solution 38

Analysis Sheet

Categories	Frequency	%	Cumulative percentage
No signature	40	42.5	42.5
Residential signature not valid	90	9.5	52
Non-legible handwriting	22	23.5	75.5
Already a customer	15	16	91.5
Others	8	8.5	100
	N = 94	100	

A pareto diagram for the delay associated with processing credit card applications



From the above chart the key problem being delay with processing credit car applications is, no-signature with the percentage $\frac{40}{94} \times 100\% = 42.55\%$.

Also, the percentage of delay with no signature and non-legible handwritin

$$\left(\frac{40+22}{94}\right) \times 100\% = 66\%$$

39. Out of 110 diesel engines tested, a rework and repair facility found 9 had leak water pumps, 15 had faulty cylinders, 4 had ignition problems, 52 had oil leak and 30 had cracked blocks. Draw a Pareto chart to identify the key problem in th production process.

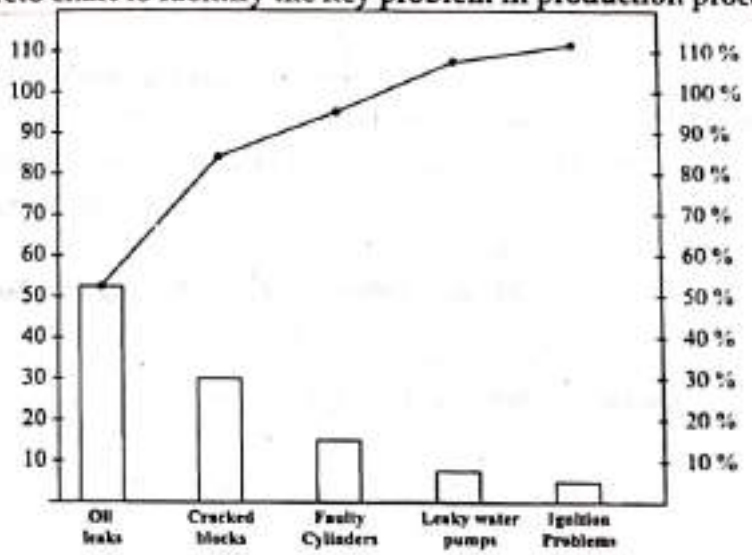
Scanned by CamScanner

Solution 39

Analysis Sheet

Categories	Frequency	%	Cumulative percentage
Leaky water pumps	9	8	8
Faulty cylinders	15	14	22
Ignition problems	4	4	26
Oil leaks	52	47	73
Cracked blocks	30	27	100
	N = 110	100	

Pareto chart to identify the key problem in production process



The key problem in the production process of the diesel engines in oil leaks which contains 47% of the problem. The oil leaks and cracked blocks contains 74.5% problems.

40. The test scores of the students in probability and statistical are listed below. Construct a stem and leaf plot of the score. [PU 2015 Fall]

92	78	73	89	98	89	83	75	83	94
99	69	71	96	67	81	73	88	86	82
63	73	76	82	84	89	92	95	78	87

Solution 40

Table: Steam and Leaf plot for the give scores

Steam	Leaf	Frequency
6	.3	1
6	9, 7	2
7	3, 3, 1, 3	4
7	8, 6, 5, 8	4
8	2, 4, 1, 3, 3, 2	6
8	9, 9, 9, 8, 6, 7	6
9	2, 2, 4	3
9	9, 6, 8, 5	4
		N = 30

Scanned by CamScanner

Exam Asked Questions

Long Questions

1. The marks obtained by 50 students in numerical methods are given as:

65	36	49	84	79	56	28	43	67	36
43	78	37	40	68	72	70	55	62	82
88	50	60	56	57	46	39	57	22	65
59	48	76	74	80	69	51	40	56	45
35	21'	62	52	63	32	86	64	53	34

i. Construct a frequency distribution from the above data. What additional information did your receive from frequency distribution that you didn't receive from the original raw data?

ii. Represent the data by means of cumulative frequency curve.

Identify the median and first quartile from the curve. Also interpret the result of first Quartile.

[PU. 2017 Fall / 2011 Fall / 2010 Spring / Purbanchal 2011 Fall]

Solution 1

We have given total number of data N = 50, L = 88, S = 21

Then by Sturges' rule the no. of classes is given by

$$k = 1 + 3.322 \log_{10} N = 1 + 3.322 \times 10_{10} 50 = 6.64 \approx 7.$$

Also, the width of class interval $= \frac{L-S}{K} = \frac{88-21}{7} = 9.57 \approx 10$

Frequency distribution of marks obtained by students

Marks	20-30	30-40	40-50	50-60	60-70	70-80	80-90	Total
Tally marks	III	卌 II	卌 III	卌 卌 I	卌 卌	卌 I	卌	
Frequency	3	7	8	11	10	6	5	50

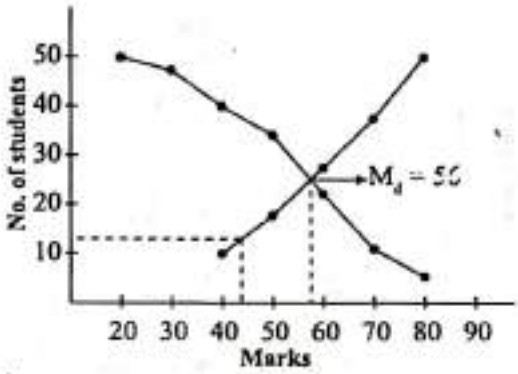
From the frequency distribution table, we see that the no. of students getting 40 to 70 is more higher than getting less than 40 and more than 70 marks. That is the no. of students getting average marks is more higher.

Second Part

Cumulative frequency distribution of marks obtained by students

Marks	No. of students	Marks	No. of students	Marks	No. of students
20-30	3	More than 20	50	Less than 30	3
30-40	7	More than 30	47	Less than 40	10
40-50	8	More than 40	40	Less than 50	18
50-60	11	More than 50	32	Less than 60	29
60-70	10	More than 60	21	Less than 70	39
70-80	6	More than 70	11	Less than 80	45
80-90	5	More than 80	5	Less than 90	50
Total	N = 50				

Scanned by CamScanner



From the above graph, median = 56

∴ First quartile $= \frac{N}{4} = \frac{50}{4} = 12.5$

From the graph first quartile = 43

The maximum marks obtained by first 25% students is 43.

2. The ordered arrays in the accompanying table deal with the life (in hours) of a sample of forty 100-watt light-bulbs produced by manufacturer A.

684	972	697	720	773	821	939
831	977	835	848	852	852	943
859	984	860	868	870	876	946
893	1005	899	905	909	911	954
922	1014	924	926	926	938	971

i. Form the frequency distribution of manufacturer A using the appropriate class widths.

ii. Construct the histogram and frequency polygon from the above frequency distribution. What additional information can you draw from this histogram and polygon? **PU 2010 Spring**

Solution 2

i. Here, given

N = 40, L = 1093 and S = 684

Now, by Sturge's rule the number of class interval is given by

$$k = 1 + 3.322 \log_{10}N = 1 + 3.322 \log_{10}40 = 6.322 = 7$$

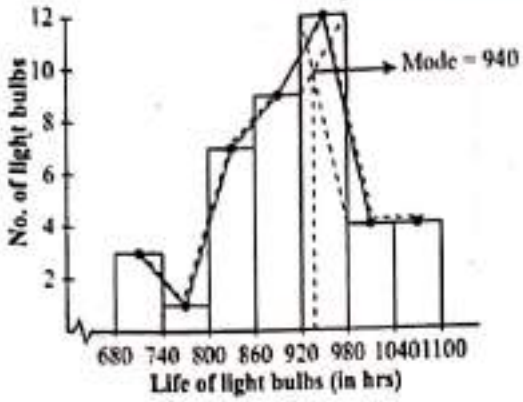
Also, the class width $= \frac{L - S}{K} = \frac{1093 - 684}{7} = 58.42 \approx 60$

Frequency distribution and percentage distribution

Life (in hours)	No. of light bulbs
680 - 740	3
740 - 800	1
800 - 860	7
860 - 920	9
920 - 980	12
980 - 1040	4
1040 - 1100	4
Total	N = 40

ii.

Scanned by CamScanner



Since, the life of light bulbs produced by the manufacturer A has maximum number of bulbs is 920 to 980. Also, mode life is 940 from the histogram.

3. An article in Environmental Concentration and Toxicology report the amount of cadmium in sea scallops observed at a number of different stations in North Atlantic water. The observed values are as follows:

5.1	14.4	14.7	10.8	6.5	5.7	7.7	14.1
3.7	8.9	7.9	7.9	4.5	10.1	5.0	9.6
5.1	11.4	8.0	12.1	7.5	8.5	13.1	6.4
27.	18.9	10.8	13.1	8.4	16.9	2.7	9.6
12.4	5.5	12.7	17.1	9.5	5.5	18	4.5

i) Construct a frequency distribution for the data set using 0.0 - 4.0 and 4.0 -8.0 as the first and second class interval respectively.

ii) Draw the histogram and frequency curve of the above frequency distribution. [PU 2009 Spring]

Solution 3

Please refer solution 2

4. Define classification. Discuss the general rules of classification. The following is a record of a grades gained by 40 examinees at an examination:

30	90	54	57	64	58	69	78	28	44
83	88	17	70	93	33	59	55	20	46
23	91	51	63	18	53	38	69	27	65
73	85	41	87	95	67	75	15	40	

Prepare a continuous frequency distribution table and hence represent the data by means of Histogram. [PU 2003 Fall]

Solution 4

Please refer solution 2

5. The high temperature (degree Fahrenheit) in a random sample of 30 days during summer season in Nepal Terai region is recorded as below.

94	78	88	95	102	89	89	88	99	89	105	78	97	88	82
98	87	94	96	86	87	94	92	102	84	75	69	89	94	84

i) Enumerate the steps you would take to convert this data into a frequency distribution. What class interval size would you use and why?

ii) Construct a frequency distribution and hence represent the data by means of histogram.

Scanned by CamScanner

iii. What can you conclude from the frequency distribution and histogram that you could not infer from the original raw data?

[PU 2007 Fall / 2012 Fall/ Purbanchal 2012 Fall]

Solution 5

i. Here, total no. of data N = 50, largest value = 105, smallest value = 69

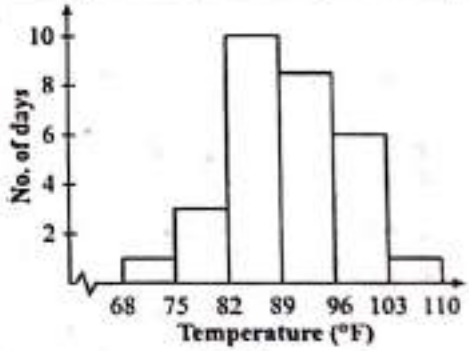
Now, by Sturges' rule

$$k = 1 + 3.322 \log_{10}N = 1 + 3.322 \log_{10}30 = 5.9 \approx 10$$

Also, the width of class interval $= \frac{L - S}{K} = \frac{105 - 69}{6} = 7$

ii. **Frequency distribution of highest temperature**

Temperature (in °F)	68-75	75-82	82-89	89-96	96-103	103-110	Total
Tally Marks	I	III	~~IIII~~ ~~IIII~~	~~IIII~~ ~~IIII~~	~~IIII~~ I	I	
No. of days	1	3	10	9	6	1	N = 30



iii. From the above histogram we find the mode temperature is 87°F. Also, the temperature lies between 82 and 130°F most of the days.

6. The following table shows the length of the bridge

68	84	73	82	68	90	62	88	76	93	73	79	75
73	60	93	71	59	85	75	61	65	88	87	74	62
95	78	63	72	66	78	75	75	94	77	69	74	68
60	96	78	82	61	95	75	60	79	83	71	79	62
89	97	78	85	76	65	71	75	65	80	67	57	88
78	62	76	53	74	86	67	73	81	72	63	76	75
85	77											

i) Construct the grouped frequency distribution having class width 10.

ii) Draw less than ogive and more than ogive in same graph hence locate the median.

iii) By the help of less than ogive find the number of bridge having length less than 65 meter. [PU 2015 Spring]

Solution 6

Here, given total number of data (N) = 80

The smallest value (S) = 53 The largest value (L) = 97

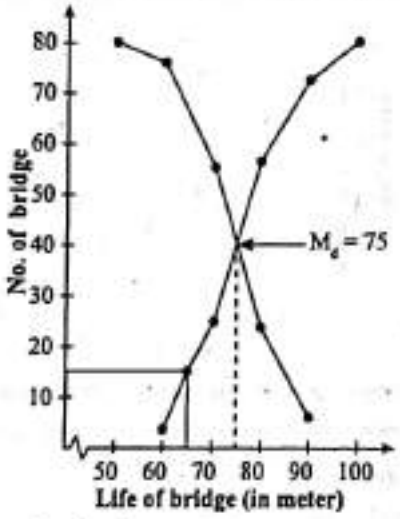
Scanned by CamScanner

i. Frequency distribution of length of bridges

Length of bridge (in meter)	Tally marks	No. of bridges (frequency)
50 - 60	III	3
60 - 70	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ I	21
70 - 80	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III	33
80 - 90	~~IIII~~ ~~IIII~~ ~~IIII~~	15
90 - 100	~~IIII~~ III	8
		N = 80

ii. Less than and more than frequency distribution of length of bridges

Length (in meter)	Frequency	Length (in meter)	Frequency
Less than 60	3	More than 50	80
Less than 70	24	More than 60	77
Less than 80	57	More than 70	56
Less than 90	72	More than 80	23
Less than 100	80	More than 90	8



Hence, the median length of bridges is 75 meter.

iii. For the number of bridge having length less than 65 meter

Draw a vertical line from the x-axis (length of bridge) to less than ogive. Also draw a horizontal line from the point of intersection of vertical line and less than ogive to the y-axis, (no. of bridges) which meets at 14. (From graph)

Hence, no. of bridges having length 65 meter is 14.

$$\therefore \quad \left[(50-60)+\frac{60-70}{2}\right] = [3+\frac{21}{2} = 3+10.5 = 13.5 \approx 14]$$

7. The weight (in lbs) of 40 boys in a class are as follows:

138	172	145	147	150	119	158	152	168
142	157	147	102	144	165	136	164	163
128	135	126	150	146	148	145	125	146
153	138	156	173	140	135	149	140	144
132	154	142	135					

i) Construct a frequency distribution.

ii) Draw less than ogive and find the number of boys whose weight is less than 165 lbs. [PU 2013 Spring]

Scanned by CamScanner

Solution 7

Here, we have, N = 40, L = 173, S = 102

Then, by Sturges' rule, the no. of class intervals is given by

$$k = 1 + 3.322 \log_{10}N = 1 + 3.322 \log_{10}40 = 6.32 \approx 7$$

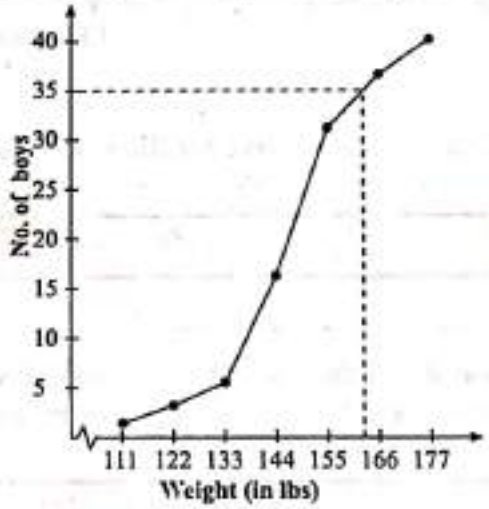
Also, the width of class interval $= \frac{L - S}{K} = \frac{173 - 102}{7} = 10.14 \approx 11$

i. **Frequency distribution of weight (in lbs) of boys**

Weight (in lbs)	100-111	111-122	122-133	133-144	144-155	155-166	166-177	Total
No. of boys	1	1	4	10	15	6	3	40

ii. **Less than frequency distribution of weight of boys**

Weight (in lbs)	Less than 111	Less than 122	Less than 133	Less than 144	Less than 155	Less than 166	Less than 177
No. of boys	1	2	6	16	31	37	40



For the number of boys having weight less than 165 lbs, (draw a vertical line from (x-axis) weight of the boys to the less than ogive. Again, draw vertical line from point of intersection to the no, of boys) which meets at 35.)

Hence, the no. of boys having weight less than 165 lbs is 35.

8. Represent the following data by means of histogram, frequency curve and polygon. [PU 2010 Fall /2005 Fall / Purbanchal 2010]

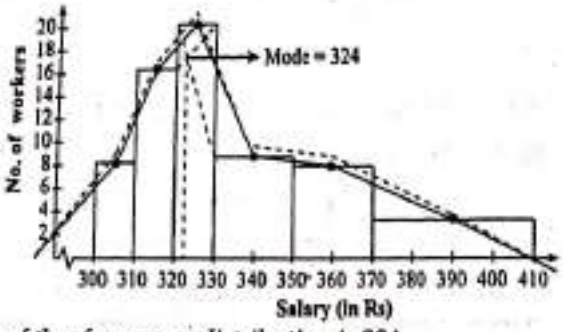
Wages	10-15	15-20	20-25	25-30	30-40	40-60	60-80
No. of workers	7	19	28	15	12	12	8

Solution 8

Frequency distribution for the histogram

Salary (Rs)	No. of workers	Class Interval	Adjusted frequency
300 - 310	8	10	8
310 - 320	16	10	16
320 - 330	20	10	20
330 - 350	18	20	18/2 = 9
350 - 370	16	20	16/2 = 8
370 - 410	12	40	12/4 = 3

Scanned by CamScanner



Hence, mode of the frequency distribution is 324.

9. The weekly salaries of 90 workers in a factory are given below:

Salary (Rs)	300-310	310-320	320-330	330-350	350-370	370-410
No. of workers	8	16	20	18	16	12

i. Represent the data by means of the histogram and frequency polygon.
ii. Draw an ogive for the data and estimate the median age.

[PU 2006 Spring / 2014 Spring]

Solution 9

Please refer Q.No. 8

10. Plot the histogram and frequency curve for the following frequency distribution and locate the mode with the help of histogram. [PU 2011 Spring]

Marks	0-10	10-20	20-30	30-40	40-60
No. of students	5	11	17	15	10

Solution 10

Please refer Q.No. 8

11. Following information shows the daily wages of workers of certain hydropower company. Prepare the suitable ogive that helps to give the answers of the following questions

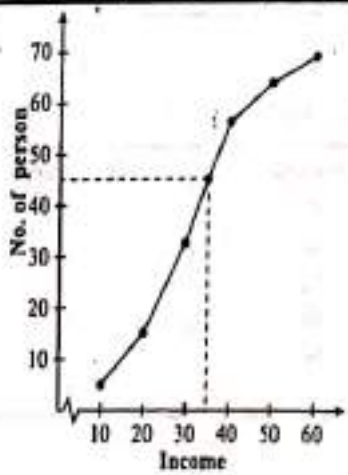
Daily Wages ($ 000)	0-20	20-40	40-60	60-80	80-100
No. of workers	41	51	64	38	7

i. About what wages above the 50% workers earn.
ii. What would be the daily wage limit of middle 30% workers? [PU 2013 Fall]

Solution 11

Less than frequency distribution of income (Rs 000)

Income (Rs 000)	0-10	10-20	20-30	30-40	40-50	50-60
No. of person	5	10	18	23	7	6
Income	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60
No. of person	5	15	33	56	63	69



Scanned by CamScanner

i. For the number of person having income less than 3500

[Draw a vertical line from income 35 to less than ogive and draw horizontal line from point of intersection to the no. of person it meets at 45]

Hence, 45 person have income less than 3500.

ii. From the graph we see that the number of person having income less than 20,000 is 15. Also, the number of person having income less than 50,000 is 63.

Therefore, the number of person having income between 20,000 to 50,000 = 63 – 15 = 48

iii. From the above graph the no. of person having less than 25000 is 24. Therefore, total no. of person having income more than 25000 is (69 – 24) = 45.

12. From the following frequency distribution

Income (Rs 000)	0-10	10-20	20-30	30-40	40-50	50-60
No. of persons	5	10	18	23	7	6

Construct an ogive that will help you the answer to find number of persons having income

i) Less than Rs 35000 ii) between Rs 20000 to 50000

iii) more than Rs 25000 [PU 2016 Spring / 2013 Fall]

Solution 12

Please refer solution 11

13. The following data relate to the expenditure of families (per month): Represent these data by angular (pie) diagram [PU 2009 Fall]

Items	Family A	Family B
Food	2000	3000
Education	500	1000
clothing	500	700
Fuel	500	400
House rent	1000	500
Miscellaneous	500	1000

Solution 13

The total expenditure of family A is 5000.

Therefore, $5000 = 360° \Rightarrow 1 = \frac{360}{5000} = 0.072°$

Similarly, the total expenditure of family B is 66000

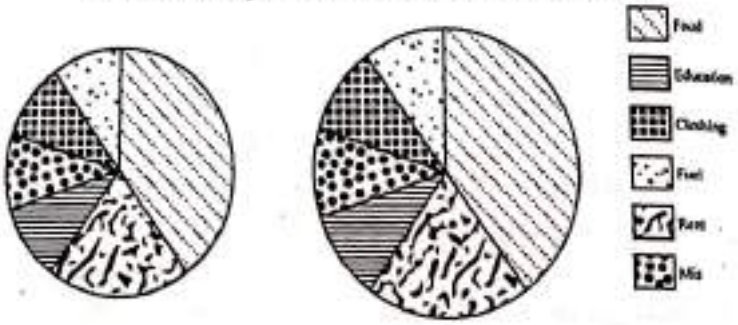
Therefore, $6600 = 360° \Rightarrow 1 = \frac{360}{6600} = 0.072° = (3/55)°$

Calculation for Pie chart

For Family A			For Family B	
Items	Expenditure	Expenditure in degree	Expenditure	Expenditure in degree
Food	2000	144°	3000	3000 × 3/55 = 164°
Education	500	36°	1000	55°
Clothing	500	36°	700	38°
Fuel	500	36°	400	22°
House rent	1000	72°	500	27°
Miscellaneous	500	36°	1000	55°
Total	5000	360°	6600	360°
Square root	70.71		81.24	
Ratio of circle	1		1.15	

Scanned by CamScanner

Pie-chart of expenditure of family A and family B



Solution 13

The test scores of the students in probability and statistics are listed below. Construct a stem and leaf plot of the scores.

92, 78, 73, 89, 98, 89, 83, 75, 83, 94, 99, 69, 71, 96, 67, 81,

73, 88, 86, 22, 63, 73, 76, 82, 84, 89, 92, 95, 78, 87

Find the lowest score of the best 25% students. [PU 2015 Fall]

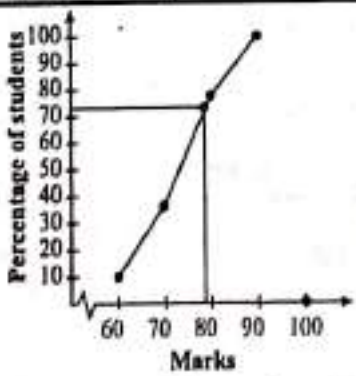
Solution 14

The stem and leaf plot off marks of probability and statistics

Stem	Leaf	Frequency
6	3, 7, 9	3
7	1, 3, 3, 3, 5, 6, 8, 8	8
8	1, 2, 3, 3, 3, 4, 6, 7, 8, 9, 9, 9	12
9	2, 2, 4, 4, 5, 6, 8, 9	7
		N = 30

Now, the relative percentage frequency distribution

Marks	No. of students	Relative % frequency	Cumulative
60-70	3	3/30 × 100% = 10%	10%
70-80	8	8/30 × 100% = 27%	37%
80-90	12	12/30 × 100% = 40%	77%
90-100	7	7/30 × 100% = 23%	100%
	N = 30		



The marks obtained by the top 25% of the students is 79.

Scanned by CamScanner

15. Given below the pre-tax monthly income (in rupees)of the workers of FEWA hydropower in 2000.

Income more than ('000' Rs.)	7	6	5	4	3	2	1	0
No. of workers	2	8	10	15	25	40	55	60

Draw less than ogive and hence find out,

i. The highest income of the lowest 50% of the worker.

ii. The minimum income earned by the top 5% of the workers.[PU 2012 Spring]

Solution 15

See in exercise 3 Q. No. 8

16. From the following distribution of 500 students of a college find the marks if only 20% students had failed and also the minimum marks obtained by the top 25% of the students.

Marks	0-20	20-40	40-50	50-60	60-80	80-100
No. of students	50	100	150	90	60	50

Also represent the data by histogram so locate the model. [PU 2017 Fall/ 2013 spring]

Solution 16

See in exercise Q . No. 9

17. Draw less than and more than ogive and locate the median from the following data. [Purbanchal 2013]

Daily Wages (Rs)	0-20	20-40	40-60	60-80	50-100
No. of workers	41	51	64	38	7

Solution 17

Please refer solution 6

18. Draw the histogram and frequency polygon from the following data[PU 2008 Fall]

Class	10-19	20-29	30-39	40-49	50-59
Frequency	20	40	35	25	30

Solution 18

Since, the given frequency distribution is inclusive so first we have to make it exclusive.

For this correction factor $C_f = \frac{10-9}{2} = \frac{1}{2} = 0.5$

Now, the exclusive frequency distribution is

Milk in liter	Frequency	Length of class interval
4.5 - 9.5	7	5
9.5 - 14.5	10	5
14.5 - 19.5	13	5
19.5 - 24.5	18	5
24.5 - 29.5	8	5
29.5 - 34.5	5	5
34.5 - 39.5	3	5

Since, all the class intervals are equal, so no need to adjust the frequency.

Please draw the figure yourself.

Scanned by CamScanner

19. In a survey it was found that 64 families bought milk in liter. A frequency is given as follows:

Milk in (liter)	5-9	10-14	15-19	20-24	25-29	30-34	35-39
frequency	7	10	13	18	8	5	3

Construct the histogram and frequency polygon of the distribution.

[PU 2007 Spring]

Solution 19

Please refer solution 18

20. For the computer controlled lathe whose performance was below per workers record the following causes and their frequency

Power fluctuation	6
Control not stable	22
Operator error	13
Worn tool not replaced	2
Other	5

Construct a pareto chart and

i. What percentage of the causes are due to an unstable controller?

ii. What percentage of the causes are due to either unstable controller or operator error?

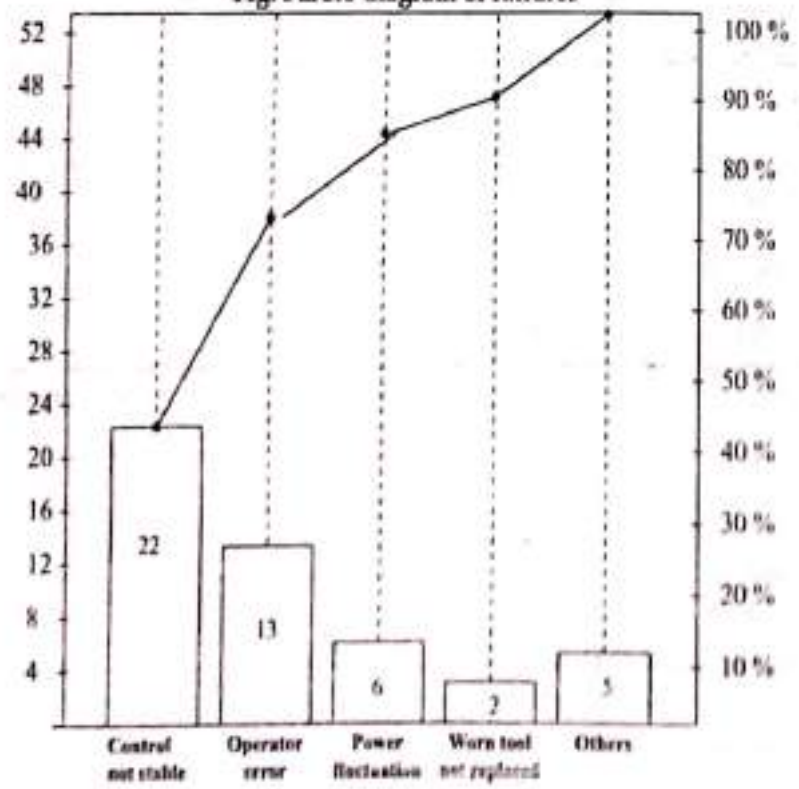
[PU 2016 Fall]

Solution 20

Analysis Sheet of Pareto Chart

Category	Frequency	Percentage	Cumulative %
Control not stable	22	45.8	45.8
Operator error	13	27.1	72.9
Power fluctuation	6	12.5	85.4
Worn tool not replaced	2	4.2	89.6
Others	5	10.4	100
Total	N = 48	100	

Fig: Pareto diagram of failures



Scanned by CamScanner

i. The percentage of causes due to an unstable controller $= \frac{22}{48} \times 100\% = 45.8 = 46\%$

ii. The percentage of causes due to either unstable controller or operator error

$$= \frac{22+13}{48} \times 100\% = \frac{35}{48} \times 100\% = 73\%$$

21. The number of B.E students in two colleges A and B is given below

	No. of students			
College	Civil	Computer	Electronics	Industrial
A	40	60	50	30
B	60	80	50	50

Compare the number of students with the help of angular diagram

[PU 2009 Spring]

Solution 21

Please refer solution 13

Short Questions

22. **Write short note on histogram and frequency polygon.[PU 2009 Fall/2009 Spring]**

Solution 22

A histogram is a bar graph in which the horizontal scale represents classes of data values and the vertical scale represents frequencies. The height of the bars corresponds to the frequency values and the bars are draw adjacent to each other without gaps.

The frequency polygon is also a graphical representation of frequency distribution. For the frequency polygon we find the middle points of every class interval.

Consider the middle points as x -coordinate and then points are plotted as in the discrete frequency distribution. Since, two frequency polygon can be drawn on the graph paper, so two or more frequency distribution can be more effectively compared with the help of frequency polygon.

Scanned by CamScanner

4

Chapte

Measures of Central Tendency

Central Tendency

Central tendency is defined as it is the single value within the range of da which represents a group of individual values in a simple and concise mann and concentrates in the middle of the distribution

Measure of central tendency

The chief measures of central tendencies are

i. Arithmetic mean (Average mean)
   a. Simple arithmetic mean b. Weighted arithmetic mean
ii. Median
iii. Mode
iv. Geometric mean
v. Harmonic mean
vi. Percentile
vii. Decile
viii. Quartiles

Arithmetic Mean (A.M. or $\bar{X}$)

Definition: Sum of observations of a set divided by the number of observations called an arithmet

ic mean. It is denoted by A.M. or $\bar{X}$.

Scanned by CamScanner

Calculation of A.M. $\bar{X}$

	Individual Series	Discrete/Continuous Series
1. Direct method	A.M. = $\bar{X} = \frac{\Sigma x}{n}$	A.M. = $\bar{X} = \frac{\Sigma fx}{n}$
2. Short cut method	A.M. = $\bar{X} = A + \frac{\Sigma d}{n}$	A.M. = $\bar{X} = A + \frac{\Sigma fd}{n}$
Where, A is assumed mean and $d = x - A$		
	A.M. = $\bar{X} = A + \frac{\Sigma d'}{n} \times h$	A.M. = $\bar{X} = A + \frac{\Sigma fd'}{n} \times h$
Where, A is assumed mean, $d' = \frac{X - A}{h}$, h = width of class interval		

Note:

1. If the data are inclusive type, then it is not necessary to adjust data into exclusive type before calculating mean because the mid-points remains same in both cases.
2. Since arithmetic mean depends on all observations. So, we can not find arithmetic mean if the data is open-ended type.
3. Arithmetic mean is the best measures of central tendency because it depends on all observations.
4. Weighted arithmetic mean $\bar{X}_w = \frac{\Sigma Wx}{\Sigma W}$

   Where, W are weights of given data
5. Combined mean or mean of composite series

   Let n_1 and n_2 are the number of data and $\bar{x}_1$, $\bar{x}_2$ are the means of two data.

   Then, the combined mean is denoted and defined by $\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

   Similarly, $\bar{x}_{123} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$ and so on.

Geometric Mean [G.M.]

Geometric mean of n variety values is the nth root of their product. It is denoted by GM or G.

Calculation of G.M.

1. Individual series : $GM = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$
2. Discrete series : $GM = \text{Antilog}\left(\frac{\Sigma f \log x}{N}\right)$ $N = \Sigma f$
3. Continuous series : $GM = \text{Antilog}\left(\frac{\Sigma f \log x}{N}\right)$

   Where, x = mid value of class, $N = \Sigma f$

Scanned by CamScanner

Note: If the variety of values are measured as ratios, proportions or percentage GM gives the better measure of location than other means.

Harmonic Mean [H.M.]

The reciprocal of the arithmetic average of the reciprocal of the given values is called the harmonic mean. It is denoted by H.M. or H.

Calculation of H.M.

1. Individual series : $H.M. = \dfrac{n}{\Sigma\left(f\frac{1}{x}\right)}$, n = no. of data

2. Discrete series : $H.M = \dfrac{N}{\Sigma\left(f\frac{1}{x}\right)}$, $N = \Sigma f$

3. Continuous series : $HM = \dfrac{N}{\Sigma f\left(\frac{1}{x}\right)}$

Note:

1. Harmonic mean is suitable when the data contains the speed ratio, rate and times, etc.
2. $A.M. \geq G.M \geq H.M.$
3. $G.M. = \sqrt{A.M \times H.M.} \Rightarrow (G.M.)^2 = A.M. \times H.M.$

Median (M_d)

It is the most preferable measure of location for asymmetric distribution. Median is the value of the variable that divides the ordered set off values into two equal halves, 50% of values are to the left of the median and 50% are to the left of the median and 50% are to the right of the median. It is denoted by M_d.

Calculation of Median (M_d)

1. Individual series :

   i. Median (M_d) = value of $\left(\dfrac{n+1}{2}\right)^{th}$ item, when n is odd

   ii. Median (M_d) = $\dfrac{\left(\frac{n}{2}\right)^{th} \text{item} + \left(\frac{n}{2}+1\right)^{th} \text{item}}{2}$ when n is odd

2. Discrete series : Median (M_d) = value of $\left(\dfrac{N+1}{2}\right)^{th}$ item, $N = \Sigma f$

3. Continuous series: Median (M_d) = $l + \dfrac{\frac{N}{2} - C.f.}{f} \times i$

   Where, l = lower limit of corresponding median class

Scanned by CamScanner

N = sum of all frequencies (Σf)

f = frequency of corresponding median class

c.f. = cumulative frequency of the class preceding median class

i = width of the class interval

Note:

1. When the class-intervals are unequal, it is not necessary to make the class-interval equal to find the median.
2. The distribution must be exclusive to determine the median.
3. Median is useful when the distribution is open-ended.

Mode

The variate value x having maximum frequency in a distribution is known as mode. It is denoted by M_o.

Calculation of Mode (M_o)

1. Individual series : M_o = maximum times repeated value
2. Discrete series : M_o = value containing maximum frequency
3. Continuous series: The class containing maximum frequency is the modal class

Note:

1. While applying the above formula it is necessary to see that the class intervals are of same size. If they are unequal, they should first be made equal on the assumption that the frequencies are equally distributed through out the class.
2. When mode is ill-defined, the method of grouping also fails its value can be determined by mode = 3 median – 2 mean.

Quartiles

Variate values which divide the distribution into four equal parts and each part equals to 25% are called quartiles. They are denoted by Q_1, Q_2 (or M_d)

Calculation of Quartiles (Q_i)

1. Individual series : Q_i = value of $\left[\frac{i(n+1)}{4}\right]^{th}$ item, i = 1, 2, 3
2. Discrete series : Q_i value of $\left[\frac{i(n+1)}{4}\right]^{th}$ item, i = 1, 2, 3
3. Continuous series : $Q_i = l + \frac{\frac{iN}{4} - cf}{f} \times h$, i = 1, 2, 3

Scanned by CamScanner

Deciles

Variate values which divide the distribution into ten equal parts and each part equal to 10% are called deciles. There are nine deciles; $D_1, D_2, D_3, \ldots, D_9$. [$D_1 < D_2 < D_3 <, \ldots, < D_9$]

Calculation of Deciles (D_i)

1. Individual series : Q_i = value of $\left[\frac{i(n+1)}{10}\right]^{th}$ item, $i = 1, 2, \ldots, 9$
2. Discrete series : Q_i value of $\left[\frac{i(n+1)}{10}\right]^{th}$ item, $i = 1, 2, \ldots, 9$
3. Continuous series : $Q_i = l + \frac{\frac{iN}{10} - cf}{f} \times h$, $i = 1, 2, 3, \ldots, 9$

Percentiles

Variate values which divide the distribution into 100 equal parts and each part equal to 1% are called deciles. There are 99 deciles; $P_1, P_2, P_3, \ldots, P_{99}$.

Calculation of Percentiles (P_i)

1. Individual series : P_i = value of $\left[\frac{i(n+1)}{100}\right]^{th}$ item, $i = 1, 2, \ldots, 99$
2. Discrete series : P_i value of $\left[\frac{i(n+1)}{10}\right]^{th}$ item, $i = 1, 2, \ldots, 99$
3. Continuous series : $Q_i = l + \frac{\frac{iN}{100} - cf}{f} \times h$, $i = 1, 2, 3, \ldots, 99$

Selection of an average

i. A.M. is best when:
   a. Data is quantitative in nature
   b. Distribution is less skewed
   c. Distribution does not have open-end closed

ii. G.M. is best when we have to compute average rates off population growth, rate off interest, ratios, percentage and in measuring relative changes rather absolute changes.

iii. H.M. is best when we have to compute the average of time, distance travelled in a certain time, speed.

iv. Median is best when:
   i. Distribution is open-end class
   ii. Distribution contains unequal classes
   iii. Highly skewed distribution
   iv. Qualitative values are given

v. Mode is best when we have to find the ideal, most common, most likely average.

Scanned by CamScanner

Exercise 4

1a. The following are the monthly salaries of 10 persons:

27, 300, 250, 380, 420, 275, 126, 196, 410, 200

Solution 1a

Arranging the given data in ascending order, we get

126, 196, 200, 250, 275, 275, 300, 380, 410, 420

$$\text{Mean} = \frac{\Sigma X}{n} = \frac{2832}{10} = 283.2$$

Hence, the average salary is Rs.283.2

1b. Rank of beauty contest is given below:

7, 9, 5, 4, 2, 7, 8, 1, 6, 3, 9, 7, 8, 2, 5

Solution 1b

Arranging the given data in ascending order, we get,

1, 2, 2, 3, 4, 5, 5, 6, 7, 7, 7, 8, 8, 9, 9

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{ place item} = 8^{th} \text{ place item} = 6$$

1c. Size of shoes is given below:

$33\frac{1}{2}, 22\frac{1}{2}, 22, 35, 36, 36\frac{1}{2}, 36, 16, 15\frac{1}{2}, 22\frac{1}{2}, 35, 22\frac{1}{2}$

Solution 1c

Arranging the given data in ascending order, we get,

$15\frac{1}{2}, 16, 22, 22\frac{1}{2}, 22\frac{1}{2}, 22\frac{1}{2}, 33\frac{1}{2}, 35, 35, 36, 36, 36\frac{1}{4}$

$\therefore$ Mode = Maximum times repeated number = 22.5

1d. Ratio of husbands and wives is given below:

1.2, 0.96, 0.84, 1.23, 1.1, 1.02, 0.89, 1.13, 1.21, 1.05, 0.99

Solution 1d

Arranging the given data in ascending order, we get

0.84, 0.89, 0.96, 0.99, 1.02, 1.05, 1.1, 1.13, 1.2, 1.21, 1.23

$$\text{Then, G.M.} = \text{anti log}\left[\frac{\Sigma \log x}{n}\right] = \text{anti log}\left[\frac{0.215}{11}\right] = 1.046$$

X	0.84	0.89	0.96	0.99	1.02	1.05	1.1	1.13	1.2	1.21	1.23	Total
log X	−0.08	−0.05	−0.02	−0.004	0.009	0.002	0.04	0.05	0.08	0.08	0.09	0.215

1e. A student run from A to C though B and back to initial stage.

Speed (A to B): 10 mph

Speed (B to C): 9 mph

Speed (C to B): 9 mph

Speed (B to A): 7 mph

Scanned by CamScanner

Solution 1e

We know that when equal distance are covered with unequal speeds, the harmonic mean (H.M.) is the proper average. So, here given,

$x_1 = 10$, $x_2 = 9$, $x_3 = 9$ and $x_4 = 7$

Now, the average speed (H.M.) $= \dfrac{n}{\Sigma\left(\frac{1}{x}\right)} = \dfrac{4}{\frac{1}{10}+\frac{1}{9}\frac{1}{9}+\frac{1}{7}} = \dfrac{4}{0.46} = 8.7$ mph

2. Find mode from following data:

5, 7, 3, 5, 7, 3, 9, 5, 7, 9, 3, 6, 5, 6, 7, 3

Solution 2

Arranging the given data in ascending order, we get,

3, 3, 3, 3, 5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 9, 9

Since mode of individual data is maximum times repeated value, but here 3, 5 and 7 are repeated 4 (same number of) times.

∴ Mode = 3 median – 2 mean

Where, median $= \left(\dfrac{n+1}{2}\right)^{th}$ place item $= \dfrac{5+6}{2} = 5.5$

Also, mean $= \dfrac{\Sigma x}{n} = \dfrac{1258}{16} = 5.625$

∴ Mode = 3 median – 2 mean

$= 3 \times 5.5 - 2 \times 5.625 = 16.5 - 11.25 = 5.625$

3. Calculate an arithmetic mean and median for the following distribution.

x	1	2	3	4-6	7-9	10-12	13-20	21-28	29-36
f	10	5	3	9	6	2	1	10	15

Solution 3

Calculation of mean and median

X	f	mid value (m)	f.m
1	10	1	10
2	5	2	10
3	3	3	9
4-6	9	5	45
7-9	6	8	48
10-12	2	11	22
13-20	1	16.5	16.5
21-28	10	24.5	245
29-36	15	32.5	487.5
	N = 61		Σfm = 893

X	f	c.f.
1	10	10
2	5	15
3	3	18
3.5-6.5	9	27
6.5-9.5	6	33
9.5-12.5	2	35
12.5-20.5	1	36
20.5-28.5	10	46
28.5-36.5	15	61
	N = 61	

Mean $\bar{X} = \dfrac{\Sigma fm}{N} = \dfrac{893}{61} = 14.64$

Scanned by CamScanner

For median:

$$\frac{N}{2} = \frac{61}{2} = 30.5$$

The cumulative frequency just greater than $\frac{N}{2} = 30.5$ is 33. Therefore, median lies in the interval 6.5 to 9.5.

So, the median $M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 6.5 + \frac{30.5 - 27}{6} \times 3 = 6.5 + 1.75 = 8.25$

4. Calculate the arithmetic mean for following data:

Height (in cm):	219	216	213	210	207	204	201	198	195
No. of persons	2	4	6	10	11	7	5	4	1

Solution 4

Calculation of A.M.

Height (in c.m.) (X)	No. of persons (f)	fx
219	2	438
216	4	864
213	6	1278
210	10	2100
207	11	2277
204	7	1428
201	5	1005
198	4	792
195	1	195
	N = 50	Σfx = 10377

$$AM = \frac{\Sigma fx}{N} = \frac{10377}{50} = 207.54$$

Hence, the arithmetic mean of the given data is 207.54 cm.

Alternative method:

- Since, the given data contains large number. Therefore, suppose the assumed mean (A) = 207

Height (in c.m.) (X)	No. of persons (f)	d = X – 207	fd
219	2	12	24
216	4	9	36
213	6	6	36
210	10	3	30
207	11	0	0
204	7	–3	–21
201	5	–6	–30
198	4	–9	–36
195	1	–12	–12
	Σf = N = 50		Σfd = 27

$$A.M. = A + \frac{\Sigma fd}{N} = 207 + \frac{27}{50} = 207.54 \text{ cm}$$

Scanned by CamScanner

5. **Find the average marks of students from the following data:**

Marks	No. of Students	Marks	No. of Students
Above 0	80	Above 60	28
Above 10	77	Above 70	16
Above 20	72	Above 80	10
Above 30	65	Above 90	8
Above 40	55	Above 100	0
Above 50	43		

Solution 5

Calculation of average of the given data

Marks	No. of students
Above - 0	80
Above - 10	77
Above - 20	72
Above - 30	65
Above - 40	55
Above - 50	43
Above - 60	28
Above - 70	16
Above - 80	10
Above - 90	8
Above - 100	0

Marks	f	x-mid value	fx
0 - 1	80 – 77 = 3	5	15
10 - 20	77 – 72 = 5	15	75
20 - 30	72 – 65 = 7	25	175
30 - 40	65 – 55 = 10	35	350
40 - 50	55 – 43 = 12	45	540
50 - 60	43 – 28 = 15	55	825
60 - 70	28 – 16 = 12	65	780
70 - 80	16 – 10 = 6	75	450
80 -90	10 – 8 = 2	85	170
90 - 100	8 – 0 = 8	95	760
Above 100	0		
	N = 80		Σfx = 4140

$$\text{Mean} = \frac{\Sigma fx}{N} = \frac{4140}{80} = 51.75$$

6. **Two hundred persons were interviewed by a public opinion polling agency. The frequency distribution gives the ages of persons interviewed.**

Age groups (Years)	Frequency (f)	Age groups (Years)	Frequency (f)
80 - 89	2	40 - 49	56
70- 79	2	30 - 39	40
60 - 69	6	20 - 29	42
50 - 59	20	10 - 19	32

Calculate the arithmetic mean of the data.

Solution 6

Calculation of arithmetic mean

Age group (years)	Frequency (f)	Mid-value (x)	$d' = \frac{X - 44.5}{10}$	fd'
80 - 89	2	84.5	4	8
70 - 79	2	74.5	3	6
60 - 69	6	64.5	2	12
50 - 59	20	54.5	1	20
40 - 49	56	44.5	0	0
30 - 39	40	34.5	–1	–40
20 - 29	42	24.5	–2	–84
10 - 19	32	14.5	–3	–96
	N = 200			Σfd' = 174

Scanned by CamScanner

$$\text{Arithmetic mean} = A + \frac{\Sigma fd'}{N} \times h = 44.5 + \left(\frac{-174}{200} \times 10\right) = 44.5 - 8.7 = 35.8$$

Hence, the average age is 35.8 years.

7. **Calculate the arithmetic mean from the following data:**

Class Interval	Frequency	Class Interval	Frequency
0-1	8	15-25	11
1-3	8	25-28	10
3-5	10	28-30	9
5-10	12	30-45	8
10-15	18	45-60	6

Solution 7

Calculation of arithmetic mean from the given data

Class-interval	Frequency (f)	Mid-value (x)	fx
0 - 1	8	0.5	4
1 - 3	8	2	16
3 - 5	10	4	40
5 - 10	12	7.5	90
10 - 15	18	12.5	225
15 - 25	11	20	220
25 - 28	10	26.5	265
28 - 30	9	29	261
30 - 45	8	37.5	300
45 - 60	6	52.5	315
	N = Σf = 100		Σfx = 1736

$$\text{A.M.} = \frac{\Sigma fx}{N} = \frac{1736}{100} = 17.36$$

8. **Find the class intervals, if the arithmetic mean of the following distribution is 33 and assumed mean is 35.**

Step deviation (d)	–3	–2	–1	0	1	2
Frequency	5	10	25	30	20	10

Solution 8

Here, we have given,

A.M. = 33, Assumed mean (A) = 35, Class-interval = ?

Step deviation (d)	Frequency (f)	fd	mid value	class interval
– 3	5	– 15	– 3 × 10 + 35 = 5	0 – 10
– 2	10	– 20	– 2 × 10 + 35 = 15	10 – 20
– 1	25	– 25	–1 × 10 + 35 = 25	20 – 30
0	30	0	0 × 10 + 35 = 35	30 – 40
1	20	20	1 × 10 + 35 = 45	40 – 50
2	10	20	2 × 10 + 35 = 55	50 – 60
	N = Σf = 100	Σfd = – 20		

We know that,

$$\text{A.M.} = A + \frac{\Sigma fd}{N} \times h$$

Scanned by CamScanner

or, $33 = 35 + \frac{(-20)}{100} \times h$

or, $-2 = -\frac{2h}{10}$

$\therefore$ $h = 10$

Note: $d = \frac{X - A}{h} \Rightarrow X = dh + A$

9. **The average height of a group of 25 children was calculated to be 78.4 cm. It was later discovered that one value was misread as 69 cm instead of the correct value 96 cm. Calculate the correct mean.**

Solution 9

Here, given, Number of children (n) = 25, Average height $\bar{x}$ = 78.4 cm

But Wrong input = 69 cm Correct value =96 cm

Correct mean = ?

We know that, Correct $(\bar{x}) = \frac{\text{correct } (\Sigma x)}{N} = \frac{25 \times 78.4 - 69 + 96}{25} = 79.48$ cm

10. **A candidate obtains the following percentage in an examination: Math 60; Statistics 75; Basic Electrical 63, Physics 59 and Computer programming 55. Find the weightage mean if weights 2, 1, 5, 5, 3 are allocated to the subjects respectively.**

Solution 10

Here, given data is

Subjects	Obtained percentage (X)	Weight (W)	XW
Maths	60	2	120
Statistics	75	1	75
B. Ele	63	5	315
Physics	59	5	295
C. programming	55	3	165
		ΣW = 16	ΣXW = 970

We know that, weightage mean $(\bar{x}_w) = \frac{\Sigma XW}{\Sigma W} = \frac{970}{16} = 60.625\%$

11. **From the following frequency distribution, calculate the missing frequency.**

No. of tablets	No. of persons cured
4-8	11
8-12	13
12-16	16
16-20	14
20-24	?
24-28	9
28-32	17
32-36	6
36-40	4

The average number of tablets to cure a person is 20.

Solution 11

Scanned by CamScanner

The average number of tablets to cure a person $(\bar{x}) = 20$

Calculation of arithmetic mean from the given data

No. of tablets	Number of persons cured (f)	Mid-value (x)	$d' = \frac{X - 180}{4}$	fd'
4 - 8	11	6	−3	−33
8 - 12	13	10	−2	−26
12 - 16	16	14	−1	−16
16 - 20	14	18	0	0
20 - 24	x	22	1	x
24 - 28	9	26	2	18
28 - 32	17	30	3	51
32 - 36	6	34	4	24
36 - 40	4	38	5	20
	N = 90 + x			Σfd' = 38 + x

We know that,

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h$$

$$20 = 18 + \frac{38 + x}{90 + x} \times 4$$

or, $\frac{38 + x}{90 + x} = \frac{1}{2}$

or, $76 + 2x = 90 + x$

$\therefore$ $x = 14$

Hence, the missing frequency is 14.

12. The mean marks obtained by 300 students in the subject of statistics are 45. The mean of the top 100 of them was found to be 70 and the mean of the last 100 was known to be 20. What is the mean of the remaining 100 students?

Solution 12

Here,

Number of students (n) = 300 Mean marks of statistics $(\bar{x}) = 45$

Mean of top 100 of them $(\bar{x}_1) = 70$ Mean of last 100 of them $(\bar{x}_2) = 20$

Mean of remaining 100 of them $(\bar{x}_3) = ?$

We know that,

$$\text{Combined mean } (\bar{x}) = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \bar{x}_3 n_3}{n_1 + n_2 + n_3}$$

$$\Rightarrow \quad 45 = \frac{70 \times 100 + 20 \times 100 + \bar{x}_3 \times 100}{300}$$

or, $13500 = 9000 + (\bar{x}_3) \times 100$

$\therefore$ $\bar{x}_3 = \frac{4500}{100} = 45$

13. In a certain examination, average grade of all students in class A is 68.4 and that of all students in class B is 71.2. Of the averages of both classes combined is 70, find the ratio of the number of students in class A to the number in class B.

Scanned by CamScanner

Solution 13

Here,

Average grade of class A($\bar{X}_A$) = 68.4

Average grade of class B($\bar{X}_B$) = 71.2

Combined average of both classes ($\bar{X}$) = 70

Ratio of the number of students in class A to B = ?

Let n_A and n_B be the number of students in class A and B respectively. Then,

$$\text{Combined mean} = \frac{n_A \cdot \bar{X}_A + n_B \cdot \bar{X}_B}{n_A + n_B}$$

or, $$70 = \frac{68.4\, n_A + 71.2\, n_B}{n_A + n_B}$$

or, $70\, n_A + 70\, n_B = 68.4\, n_A + 71.2\, n_B$

or, $1.6\, n_A = 1.2\, n_B$

$\therefore$ $$\frac{n_A}{n_B} = \frac{1.2}{1.6} = \frac{3}{4}$$

Hence, the ratio of students is 3 : 4.

14. The mean of marks obtained in an examination by a group of 100 students was found to be 49.96. The mean of the marks obtained in the same examination by another group of 200 students was 52.32. Find the mean of the marks obtained by both the groups of students taken together.

Solution 14

Given,

$\bar{X}_1 = 49.96$, $n_1 = 100$

$\bar{X}_2 = 52.32$, $n_2 = 200$

Combined mean = ?

We know that,

$$\text{Combined mean} = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2}{n_1 + n_2} = \frac{100 \times 49.96 + 200 \times 52.32}{100 + 200} = 51.53\%$$

15. A distribution consists of three components with frequencies 300, 200 and 600 having their means 16, 8 and 4 respectively. Find the mean of the combined distribution.

Solution 15

Given,

$n_1 = 300$ $\bar{X}_1 = 16$

$n_2 = 200$ $\bar{X}_2 = 8$

$n_3 = 600$ $\bar{X}_3 = 4$

Combined mean = ?

We know that,

$$\text{Combined mean} = \frac{\bar{X}_1 n_1 + \bar{X}_2 n_2 + \bar{X}_3 n_3}{n_1 + n_2 + n_3} = \frac{300 \times 16 + 200 \times 8 + 600 \times 4}{300 + 200 + 600} = 8$$

Scanned by CamScanner

17. The mean weight of 150 students in a certain class is 60 kg. The mean weight of boys in the class is 70 kg and that of the girls is 55 kg. Find the number of boys and number of girls in the class.

Solution 17

Given,

$\bar{X} = 60$ kg $\quad n = 150$

$\bar{X}_1 = 70$ kg $\quad \bar{X}_2 = 55$ kg

Let n_1 be the number of boy and n_2 be the number of girls.

Since, total number of students is 150.

$\therefore \quad n_1 + n_2 = 150 \Rightarrow n_2 = (150 - n_1)$

$\therefore$ Combined mean $= \dfrac{\bar{X}_1 n_1 + \bar{X}_2 n_2}{n_1 + n_2}$

$$60 = \frac{n_1 \times 70 + (150 - n_1)\,55}{150}$$

or, $9000 = 70\,n_1 + 8250 - 55\,n_1$

or, $750 = 15\,n_1$

$\therefore \quad n_1 = \dfrac{750}{15} = 50$

Also, $n_2 = 750 - n_1 = 150 - 50 = 100$

Hence, there are 50 boys and 100 girls.

18. **Find the median from the following table:**

x	5	7	9	11	13	15	17	19
f	1	2	7	9	11	8	5	4

Solution 18

Calculation of median from the given data

x	f	cf
5	1	1
7	2	3
9	7	10
11	9	19
13	11	30
15	8	38
17	5	43
19	4	47
		N = 47

Here, $\dfrac{N}{2} = \dfrac{47}{2} = 23.5$

The c.f. is just greater than $\dfrac{N}{2} = 23.5$ is 30.

Hence, the required median value is 13.

Scanned by CamScanner

19. Calculate the mean and median from the following table:

Class Interval	Frequency
6.5 - 7.5	5
7.5 - 8.5	12
8.5 - 9.5	25
9.5 - 10.5	48
10.5 - 11.5	32
11.5 - 12.5	6
12.5 - 13.5	1

Solution 19

Calculation of mean and median from the given data

Class-Interval	Frequency (f)	cf	Mid value (x)	fx
6.5 - 7.5	5	5	7	35
7.5 - 8.5	12	17	8	96
8.5 - 9.5	25	42	9	225
9.5 - 10.5	48	90	10	480
10.5 - 11.5	32	122	11	352
11.5 - 12.5	6	128	12	72
12.5 - 13.5	1	129	13	13
	N = 129			Σfx = 1273

For median

$$\frac{N}{2} = \frac{129}{2} = 64.5$$

The cumulative frequency just greater than $\frac{N}{2} = 64.5$ is 90.

∴ Median lies in the interval 9.5 – 10.5.

$$\therefore \quad \text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i = 9.5 + \frac{64.5 - 42}{48} \times 1 = 9.97$$

$$\text{Mean } (\bar{X}) = \frac{\Sigma fx}{N} = \frac{1273}{129} = 9.87$$

Hence, mean of the given data is 9.87 and median is 9.97.

20. Compute the median from the following data:

Mid value	Frequency	Mid value	Frequency
115	6	165	60
125	25	175	38
135	48	185	22
145	72	195	3
155	116		

Solution 20

Calculation of median from the given data

Mid-value	Class-Interval	Frequency	cf
115	110 - 120	6	6
125	120 - 130	25	31
135	130 - 140	48	79
145	140 - 150	72	151
155	150 - 160	116	267
165	160 - 170	60	327
175	170 - 180	38	365
185	180 - 190	22	387
195	190 - 200	3	390
		N = 390	

Scanned by CamScanner

Here, $\frac{N}{2} = \frac{390}{2} = 195$

The c.f. just greater than $\frac{N}{2} = 195$ is 267.

$\therefore$ Median lies in the interval 150 –160

$\therefore$ $\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i = 150 + \frac{195 - 151}{116} \times 10 = 153.79$

21. Find the mode of the following frequency distribution:

Size	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	3	8	15	23	35	40	32	28	20	45	14	6

Solution 21

Calculation of mode from the given data

Since the given distribution is an irregular distribution. So, we have to find the mode by grouping method, which is as follows:

Size	I	II	III	IV	V	VI
1	3	11	23	26	46	73
2	8					
3	15	38				
4	23		58	98		
5	35	75			107	
6	40		72			100
7	32	60		80		
8	28		48		93	
9	20	65				79
10	45		59	65		
11	14	20				
12	6					

Analysis table

Column	1	2	3	4	5	6	7	8	9	10	11	12
I										1		
II					1	1						
III						1	1					
IV				1	1	1						
V					1	1	1					
VI						1	1	1				
Total				1	3	5	3	1		1		

In this table, the value x = 6 occurs maximum number of times. Hence the mode is 6.

22. Find the mode and median from the following table:

Manes	No. of students
0-10	2
10-20	18
20-30	30
30-40	45
40-50	35
50-60	20
60-70	6
70-80	3

Scanned by CamScanner

Solution 22

Calculation of median and mode

Marks	No. of students	C.F.
0 - 10	2	2
10 - 20	18	20
20 - 30	30	50
30 - 40	45	95
40 - 50	35	130
50 - 60	20	150
60 - 70	6	156
70 - 80	3	159
	N = 159	

For median:

$$\frac{N}{2} = \frac{159}{2} = 79.5$$

$$\therefore \quad M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 30 + \frac{79.5 - 50}{45} \times 10 = 36.56$$

For mode:

Since, maximum frequency is 45 so mode lies in the interval 30 - 40.

$$\therefore \quad M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 30 + \frac{15}{15 + 10} \times 10 = 36$$

where, $\Delta_1 = f_1 - f_0 = 45 - 30 = 15$ and $\Delta_2 = f_1 - f_2 = 45 - 35 = 10$

23. Calculate the mode of the following distribution monthly wages.

Monthly wages (in Rs)	No. of workers
50-70	4
70-90	44
90-110	38
110-130	28
130-150	6
150-170	8
170-190	12
190-210	2
210-230	2

Solution 23

Calculation of mode from the given data

Monthly wages (in Rs.)	Number of Workers (f)
50 - 70	4
70 - 90	44
90 - 110	38
110 - 130	28
130 - 150	6
150 - 170	8
170 - 190	12
190 - 210	2
210 - 230	2

Scanned by CamScanner

Here, we see that maximum frequency is 44.

$\therefore$ Mode lies in the interval 70 - 90

Where, $\Delta_1 = f_1 - f_0 = 44 - 4 = 40$, $\Delta_2 = f_1 - f_2 = 44 - 38 = 4$

$$\text{Mode} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 70 + \frac{40}{40 + 6} \times 20 = 87.39$$

24. An incomplete distribution of families according to their expenditure per week is given below. The median and mode for the distribution are Rs. 25 and Rs. 24 respectively. Calculate the missing frequencies.

Expenditure	0-10	10-20	20-30	30-40	40-50
No. of families	14	?	27	?	15

Solution 24

Here, given median = 25 and mode = 24

Let x and y be the missing frequencies. Then,

Expenditure	No. of families (f)	Mid-value (x)	fx	cf
0 - 10	14	5	70	14
10 - 20	x	15	15x	14 + x
20 - 30	27	25	675	41 + x
30 - 40	y	35	35y	41 + x + y
40 - 50	15	45	675	56 + x + y
	N = 56 + x + y			

Since, Median is 25 so it lies in the interval 20 - 30.

$$\therefore \quad M_d = l + \frac{\frac{N}{2} - cf}{f} \times i$$

$$\text{or,} \quad 25 = 20 + \frac{\frac{56 + x + y}{2} - (14 + x)}{27} \times 10$$

$$\text{or,} \quad 5 = \frac{56 + x + y - 28 - 2x}{2 \times 27} \times 10$$

$$\text{or,} \quad 27 = y - x + 28$$

$$\Rightarrow \quad x - y = 1 \qquad \ldots(1)$$

Also,

$$\text{Mode} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$\text{or,} \quad 24 = 20 + \frac{27 - x}{(27 - x) + (27 - y)} \times 10$$

$$\text{or,} \quad 4 = \frac{27 - x}{54 - x - y} \times 10$$

$$\text{or,} \quad 108 - 2x - 2y = 135 - 5x$$

$$\Rightarrow \quad 3x - 2y = 27 \qquad \ldots(2)$$

Solving equation (i) and (ii), we get x = 25 and y = 24

Scanned by CamScanner

25. Compute the geometric mean of the following data:

x	10	15	18	20	25
f	2	3	5	6	4

Solution 25

Do yourself

26. Find the G.M. and H.M. of the following data:

Variable	75-125	125-175	175-225	225-275	275-325
Frequency	10	20	30	6	4

Solution 26

Calculation of geometric mean from the given data

Variable	Frequency (f)	Mid-value (x)	lnx	f·lnx	1/x	f · 1/x
75 - 125	10	100	4.61	46.1	0.01	0.1
125 - 175	20	150	5.01	10.2	0.0067	0.134
175 - 225	30	200	5.3	159	0.005	0.15
225 - 275	6	250	5.52	33.12	0.004	0.024
275 - 325	4	300	5.7	22.8	0.003	0.012
	N = 70			Σflnx = 361.22		Σf.1/x = 0.42

We know that,

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \ln x}{N}\right] = \text{Antilog}\left[\frac{361.22}{70}\right] = \text{Antilog}\,[5.16] = 174.1$$

Also,

$$\text{H.M.} = \frac{N}{\Sigma f \cdot (1/x)} = \frac{70}{0.42} = 166.67$$

27. Calculate A.M., G.M and H.M. of the following data:

32 35 36 37 39 41 43

Solution 27

Here, given data is 32, 35, 36, 37, 39, 41, 43

We know that,

$$\text{A.M.} = \frac{\Sigma X}{N} = 37.57$$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \ln x}{N}\right] = \text{Antilog}\left[\frac{25.35}{7}\right] = \text{Antilog}\,[3.622] = 37.41$$

$$\text{H.M.} = \frac{n}{\Sigma(1/x)} = \frac{7}{0.188} = 37.23$$

Hence, A.M. > G.M. > H.M.

28. The marks obtained by 25 students in a test are given below:

Marks	11	12	13	14	15
No. of students	5	7	8	3	2

Scanned by CamScanner

Solution 28

Calculation of harmonic mean of the given data

Marks (X)	No. of students (f)	1/x	f. 1/x
11	5	0.091	0.45
12	7	0.083	0.58
13	8	0.077	0.62
14	3	0.071	0.21
15	2	0.067	0.13
	N = 52		Σf. 1/x = 1.99

We know that,

$$\text{H.M.} = \frac{N}{\Sigma f\,(1/x)} = \frac{25}{1.99} = 12.6$$

29. Find at the harmonic mean of the following data:

Marks (out of 150)	No. of students
10	2
20	3
40	6
60	5
120	4

Solution 29

Calculation of Harmonic mean from the given data

Marks (x)	No. of students (f)	1/x	1/x - f
10	2	0.1	0.2
20	3	0.05	0.15
40	6	0.025	0.15
60	5	0.0167	0.083
120	4	0.008	0.033
	N = 20		Σf. 1/x = 0.616

We know that,

$$\text{H.M.} = \frac{N}{\Sigma f\,(1/x)} = \frac{20}{0.616} = 32.5$$

30. Compute the harmonic mean of the following data:

Class	Frequency
0-10	4
10-20	6
20-30	10
30-40	7
40-50	3

Solution 30

Calculation of harmonic mean from the given data

Class	Frequency	Mid-value (x)	f - 1/x
0 - 10	4	5	0.8
10 - 20	6	15	0.4
20 - 30	10	25	0.4
30 - 40	7	35	0.2
40 - 50	3	45	0.67
	N = 30		Σf. 1/x = 1.87

Scanned by CamScanner

We know that,

$$H.M. = \frac{N}{\Sigma f \cdot 1/x} = \frac{30}{1.87} = 16.04$$

31. **Three cities A, B, C are equidistant from each other. A motorist travels from A to B at 30 km/hr, from B to C at 40 km/hr, and from C to A at 50 km/hr. Determine the average speed.**

Solution 31

Let the distance between the cities A, B and C are x km. The some distance are covered with unequal speeds. Here, given $x_1 = 30$, $x_2 = 40$ and $x_3 = 50$. Therefore, the average speed is given by harmonic mean

$$H.M. = \frac{n}{\Sigma\left(\frac{1}{x}\right)} = \frac{3}{\frac{1}{30}+\frac{1}{40}+\frac{1}{50}} = \frac{3}{0.075} = 40 \text{ km/hr}$$

32. **Find the median, quartiles, 7th deciles and 85th percentile from the following data:**

Monthly Rent	No. of families	Monthly Rent	No. of families
20-40	6	120-140	15
40-60	9	140-160	10
60-80	11	160-180	8
80-100	14	180-200	7
100-120	20		

Solution 32

Calculation of median, quartiles, 7th deciles and85th percentile

Monthly rent in (Rs) (X)	No. of families (f)	c.f.
20 - 40	6	6
40 - 60	9	15
60 - 80	11	26
80 - 100	14	40
100 - 120	20	60
120 - 140	15	75
140 - 160	10	85
160 - 180	8	93
180 - 200	7	100
	N = 100	

i. $\frac{N}{2} = \frac{100}{2} = 50$

So, median lies in 100 - 120.

$$M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 100 + \frac{50-40}{20} \times 20 = 110$$

ii. a. $\frac{N}{4} = \frac{100}{4} = 25$

Therefore, Q_1 lies in the interval 60 - 80.

$$\therefore \quad Q_1 = l + \frac{\frac{N}{4} - cf}{f} \times i = 60 + \frac{25-15}{11} \times 20 = 78.18$$

Scanned by CamScanner

b. $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$

Therefore, Q_3 lies in the interval 140 - 160.

$$\therefore \quad Q_3 = l + \frac{\frac{3N}{4} - cf}{f} \times i = 140 + \frac{75 - 75}{10} \times 20 = 140$$

iii. $\frac{7N}{10} = \frac{7 \times 100}{10} = 70$

Therefore, 7th decile D_7 lies in the interval 120 - 140.

$$\therefore \quad D_7 = l + \frac{\frac{7N}{10} - cf}{f} \times i = 120 + \frac{70 - 60}{15} \times 20 = 133.33$$

iv. $\frac{85N}{10} = \frac{85 \times 100}{100} = 85$

Therefore, P_{85} lies in the interval 160 - 180.

$$\therefore \quad P_{85} = l + \frac{\frac{85N}{100} - cf}{f} \times i = 160 + \frac{85 - 85}{18} \times 20 = 160$$

33. An incomplete frequency distribution is given as follow:

Variable	Frequency	Variable	Frequency
10-20	12	50-60	?
20-30	30	60-70	25
30-40	?	70-80	15
40-50	65	Total	229

Given that the median value is 46, determine the missing frequencies.

Solution 33

We have given median (M_d) = 46 and missing frequencies are x and y.

Calculation of median from the given data

Variable	Frequency (f)	Cf
10 - 20	12	12
20 - 30	30	42
30 - 40	x	42 + x
40 - 50	65	107 + x
50 - 60	y	107 + x + y
60 - 70	25	132 + x + y
70- 80	18	150 + x + y
	N = 229	

Here,

$N = 150 + x + y$

$229 = 150 + x + y$

$\Rightarrow \quad x + y = 79 \qquad \ldots(i)$

Since, median is 46. So, it lies in the interval 40 - 50.

$$\therefore \quad M_d = l + \frac{\frac{N}{2} - cf}{f} \times i$$

Scanned by CamScanner

$$46 = 40 + \frac{114.5 - (42 + x)}{65} \times 10$$

or, $\frac{6 \times 65}{10} = 72.5 - x$

or, $39 = 72.5 - x$

$\therefore$ $x = 33.5 = 34$

$\therefore$ $y = 79 - 34 = 45$

Hence, the missing frequencies are 34 and 35.

34. Find the median, lower and upper quartiles for the following distribution:

Marks	No. of Students	Marks	No. of Students
below 10	15	below 50	94
below 20	35	below 60	127
below 30	60	below 70	198
below 40	84	below 80	249

Solution 34

Calculation of median, lower and upper quartiles from the data

Marks	No. of students	Marks (x)	(f)	cf
Below 10	15	Below 10	15	15
Below 20	35	10 - 20	35 – 15 = 20	35
Below 30	60	20 - 30	60 – 35 = 25	60
Below 40	84	30 - 40	84 – 60 = 24	84
Below 50	94	40 - 50	94 – 84 = 10	94
Below 60	127	50 - 60	127 – 94 = 33	127
Below 70	198	60 - 70	198 – 127 = 71	198
Below 80	249	70 - 80	249– 198 = 51	249

i. For median:

$$\frac{N}{2} = \frac{249}{2} = 124.5$$

Therefore, median lies in the interval 50 - 60.

$$\therefore \quad M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 50 + \frac{124.5 - 94}{33} \times 10 = 59.24$$

ii. For lower quartiles

$$\frac{N}{4} = \frac{249}{4} = 62.25$$

Therefore, Q_1 lies in the interval 30 - 40.

$$\therefore \quad Q_1 = l + \frac{\frac{N}{4} - cf}{f} \times i = 30 + \frac{62.25 - 60}{24} \times 10 = 30.94$$

ii. For uper quartiles

$$\frac{3N}{4} = \frac{3 \times 249}{4} = 186.75$$

Therefore, Q_3 lies in the interval 60 - 70.

$$\therefore \quad Q_3 = l + \frac{\frac{3N}{4} - cf}{f} \times i = 60 + \frac{186.75 - 127}{71} \times 10 = 68.42$$

Scanned by CamScanner

35. The median and mode of the following wage distribution are known to be Rs. 33.5 and Rs. 34 respectively. Three frequency values from the table are however, missing. You are required to find out those values.

Wages in Rs	0-10	10-20	20-30	30-40	40-50	50-60	60-70	Total
No. of persons	4	16	?	?	?	6	4	20

Solution 35

Here, given median = 33.5 and mode = 34. Also, let x, y and z be the missing frequencies.

Calculation of median and mode from the given data

Wages in (Rs.)	No. of person (f)	cf
0 - 10	4	4
10 - 20	16	20
20 - 30	x	20 + x
30 - 40	y	20 + x + y
40 - 50	z	20 + x + y + z
50 - 60	6	26 + x + y + z
60 - 70	4	30 + x + y + z
	N = 30	

Here,

$N = 30 + x + y + z$

or, $230 = = 30 + x + y + z$

or, $x + y + z = 200$ ...(i)

Since, median is 33.5 so it lies in the interval 30 - 40.

$\therefore \quad M_d = l + \dfrac{\frac{N}{2} - cf}{f} \times i$

or, $33.5 = 30 + \dfrac{115 - (20 + x)}{y} \times 10$

or, $3.5y = (115 - 20 - x)\,10$

or, $3.5\,y = (95 - x)10$

$\Rightarrow \quad 10x + 3.5y = 950$ ...(ii)

Also, mode is 34, so it lies in the interval 30 - 40.

$\therefore \quad \text{Mode} = l + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} \times i$

or, $34 = 30 + \dfrac{y - x}{2y - x - z} \times 10$

or, $4 = + \dfrac{y - x}{2y - x - z} \times 10$

or, $4y - 2x - 2z = 5y - 5x$

or, $3x - y - 2z = 0$ ...(iii)

Solving equation (i), (ii) and (iii), we get

$x = 60, \quad y = 100 \quad$ and $\quad z = 40$

Scanned by CamScanner

37. Given below is the distribution of 140 candidates obtaining marks X or higher in certain examination call marks are given in whole numbers):

X	10	20	30	40	50	60	70	80	90	100
cf	140	133	118	100	75	45	25	9	2	0

Solution 37

Calculation of mean and median from the given data

Marks X	cf	Class-Interval	f	cf	mid-value x	$d' = \frac{X-55}{10}$	fd'
Above 10	140	10 - 20	140 – 133 = 7	7	15	–4	–28
Above 20	133	20 - 30	133 – 118 = 15	22	25	–3	–45
Above 30	118	30 - 40	118 – 100 = 18	40	35	–2	–36
Above 40	100	40 - 50	100 – 75 = 25	65	45	–1	–25
Above 50	75	50 - 60	75 – 45 = 30	95	55	0	0
Above 60	45	60 - 70	45 – 25 = 20	115	65	1	20
Above 70	25	70 - 80	25 – 9 = 16	131	75	2	32
Above 80	9	80 - 90	9 – 2 = 7	138	85	3	21
Above 90	2	90 - 100	2 – 0 = 2	140	95	4	8
Above 100	0	Above 100	0				
			N = 140				Σfd = –53

$$\text{Mean} = A + \frac{\Sigma fd'}{N} \times h = 55 + \left(\frac{-53}{140}\right) \times 10 = 55 - 3.8 = 51.2$$

$$\frac{N}{2} = \frac{140}{2} = 70$$

∴ Median lies in 50 - 60.

$$M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 50 + \frac{70 - 65}{30} \times 10 = 51.67$$

38. **If A, G and H denote the arithmetic mean, geometric mean and harmonic mean respectively of two positive numbers, prove that:**

i. $A \geq G \geq H$ ii. $G^2 = AH$

Solution 38

i. Let A, G and H be the arithmetic, geometric and harmonic mean of two positive numbers a and b. If A is the A.M. between a and b then,

$$\text{A.M.} = \frac{a+b}{2} \Rightarrow A = \frac{a+b}{2} \quad \text{...(i)}$$

Also, if G is the G.M. between a and b then,

$$\text{G.M.} = \sqrt{ab} \Rightarrow G = \sqrt{ab} \quad \text{...(ii)}$$

Again, H is the H.M. between a and b then,

$$\text{H.M.} = \frac{2ab}{a+b} \Rightarrow H = \frac{2ab}{a+b} \quad \text{...(iii)}$$

$$\text{Now, } A - G = \frac{a+b}{2} - \sqrt{ab} = \frac{a+b-2\sqrt{ab}}{2} = \frac{(\sqrt{a}-\sqrt{b})^2}{2} \geq 0$$

$$\therefore \quad A \geq G \quad \text{...(A)}$$

Scanned by CamScanner

Also,

$$G - H = \sqrt{ab} - \frac{2ab}{a+b} = \frac{(a+b)\sqrt{ab} - 2ab}{a+b}$$

$$= \frac{\sqrt{ab}\,(a+b-2\sqrt{ab})}{a+b} = \frac{\sqrt{ab}\cdot(\sqrt{a}-\sqrt{b})^2}{a+b} \geq 0$$

$G \geq H$ ...(B)

Combining (A) and (B), we get

$A \geq G \geq H$

ii. Also, $A \times H = \frac{a+b}{2} \times \frac{2ab}{a+b} = ab = (\sqrt{ab})^2 = G^2$

$\therefore \quad G^2 = AH$

39. If the mode and mean of a moderately skewed distribution are respectively 16 inches and 20.2 inches, compute the most probable median.

Solution 39

Here, we have given mode = 16

mean = 20.2

median = ?

We know that,

Mode = 3 median − 2 mean

16 = 3 median − 2 × 20.1

$\therefore \quad \text{Median} = \frac{16 + 40.4}{3} = 18.8$

40. A variate takes the values a, ar, ar², ..., ar^{n-1} each with frequency unity. If A, G and H are respectively the arithmetic mean, geometric mean and harmonic mean, show that:

$$A = \frac{A(1-r^n)}{n(1-r)},\ G = ar^{(n-1)/2},\ H = \frac{an(1-r)r^{n-1}}{(1-r^n)}$$

Prove that $G^2 = AH$. Prove that $A > G > H$ unless $n = 1$, when all three means coincide.

Solution 40

Since each variate values has unity frequency so it is an individual series. So,

$$\text{A.M.} = \frac{a + ar + ar^2 + ar^3 + \ldots + ar^{n-1}}{n}$$

$$A = \frac{a(1 + r + r^2 + r^3 + \ldots + r^{n-1})}{n}$$

$$A = \frac{a(1-r^n)}{n(1-r)}$$

Also, $\text{G.M.} = [a \times ar \times ar^2 \times \ldots \times ar^{n-1}]^{1/n}$

$= [a^n (1 \times r \times r^2 \times \ldots \times r^{n-1})]^{1/n}$

$= a[r^{1+2+3+4+\ldots+n-1}]^{1/n}$

Scanned by CamScanner

$$= a\left[r^{\frac{(n-1)(n-1+1)}{2}}\right]^{1/n} = ar^{\left(\frac{n-1}{2}\right)}$$

$$\text{Also, H.M.} = \frac{n}{\left[\frac{1}{a}+\frac{1}{ar}+\frac{1}{ar^2}+\dots\frac{1}{ar^{n-1}}\right]}$$

$$= \frac{n}{\frac{1}{a}\left[1+\frac{1}{r}+\frac{1}{r^2}+\dots\frac{1}{r^{n-1}}\right]} = \frac{an}{1\left[\frac{\left(\frac{1}{r}\right)^n-1}{\frac{1}{r}-1}\right]} = \frac{an(1-r)r^{n-1}}{(1-r^n)}$$

$$\text{Again, } A \times H = \frac{a(1-r^n)}{n(1-r)} \times \frac{an(1-r)r^{n-1}}{(1-r^n)} = a^2r^{n-1} = (ar^{n-1/2})^2 = G^2$$

- Also, we know that, AM > GM > HM for all $a \neq b$

$\therefore$ A > G > H

Exam Asked Questions

Long Questions

1. The test scores of the students in probability and statistics are listed below: **[PU 2015 Fall]**

92	78	73	89	98	89	83	75	83	94
99	69	71	96	67	81	73	88	86	82
63	73	76	82	84	89	92	95	78	87

Construct a stem leaf plot of marks of probability.

Solution 1

The steam and leaf plot of marks of probability

Stem	Leaf	Frequency
6	3, 7, 9	3
7	1, 3, 3, 3, 5, 6, 8, 8	8
8	1, 2, 2, 3, 3, 4, 6, 7, 8, 9, 9, 9	12
9	2, 2, 4, 5, 6, 8, 9	7
		N = 30

Second part

For the lowest score of the best 25% students we have to find the $(100 - 25) = 75^{th}$ percentile i.e. P_{75}.

The above table can be written in the form

Marks	No. of students	cf
60 - 70	3	3
70 - 80	8	11
80 - 90	12	23
90 - 100	7	30
	N = 30	

Scanned by CamScanner

Now, $\frac{75N}{100} = \frac{75 \times 30}{100} = 22.5$

The cumulative frequency just greater than 22.5 is 23.

Therefore, P_{75} lies on the class interval 80 - 90.

$$\text{Therefore, } P_{75} = l + \frac{\frac{75N}{100} - cf}{f} \times i$$

$$= 80 + \frac{22.5 - 11}{12} \times 10 = 89.58$$

Hence, the lowest score of best 25% students is 89.58%.

2. **The median and mean weights of 105 students of a school are 50 kg and 52 kg respectively. Students weighting 48 kg left the school and his place was filled up by a new comer weighing 46 kg. How will the median and mean be affected by this.** **[PU 2011 Fall]**

Solution 2

Here, given total number of students n = 505

Mean weight $\bar{x}$ = 52

Median weight (M_d) = 50

For mean weight

Since the students weighing 48 kg left the school and he is replaced by the student weighing 46 kg. The corrected mean weight of school is

$$\text{Corrected mean} = \frac{\text{Correct } \Sigma x}{n} = \frac{105 \times 52 - 48 + 46}{105} = \frac{5458}{105} = 51.98$$

For median weight

When the student weighing 48 kg is replaced by 46 kg the median remains constant because both weights are less than the median weight (52). Itt does not effect the position of the median.

Since, the mean weight is decreased because the weight of the students is replaced by the less weight but the median remains constant because the number of the students does not change.

3. **From the following distribution of marks of 500 students of a college, find the minimum pass mark if 20% of the students had failed and also find the minimum marks obtained by the top 25% of students.** **[PU 2017 Fall/2013 Spring]**

Marks	0-20	20-40	40-60	50-60	60-80	80-100
No. of students	50	100	150	90	60	50

Solution 3

Since, only 20% of the students had failed the exam. Then 80% of the students had passed the exam. So that the minimum pass mark is given by (100 – 80) = 20th percentile (i.e. P_{20}). Also, the minimum marks obtained by the top 25% of the students is given by (100 – 25) = 75th percentile i.e. P_{75}.

Scanned by CamScanner

Calculation of P_{20} and P_{75}

Marks	No. of students (f)	Cf
0 - 20	50	50
20 - 40	100	150
40 - 50	150	300
50 - 60	90	390
60 - 80	60	450
80 - 100	50	500

For P_{20}

$$\frac{20\,N}{100} = \frac{20 \times 500}{100} = 100$$

The cumulative frequency just greater than 100 is 150. So P_{20} lies on the class interval 20 – 40.

$$\therefore \quad P_{20} = l + \frac{\frac{20\,N}{100} - cf}{f} \times i = 20 + \frac{100 - 50}{100} \times 20 = 30$$

For P_{75}

$$\frac{75\,N}{100} = \frac{75 \times 500}{100} = 375$$

The cumulative frequency just greater than 375 is 390. So, P_{75} lies on the classs interval 50 - 60.

$$\therefore \quad P_{75} = l + \frac{\frac{75\,N}{100} - cf}{f} \times i = 50 + \frac{375 - 300}{100} \times 10 = 58.33 = 58$$

Hence, the minimum pass mark is 30 and the minimum marks obtained by top 25% students is 58.

4. What are the measures of central tendencies? In what circumstances would you consider the median as the most suitable describing the central tendency of a frequency distribution over the mean? 20% of the workers in an engineering firm employing a total of 2000 earn less than Rs 2 per hour, 440 earn from Rs. 2 to 2.24 per hour. 24% earn from Rs. 2.25 to 2.49 per hour, 370 earn from Rs. 2.50 to 2.74 per hour. 12% earn from Rs. 2.75 to 2.99 per hour and rest earn Rs. 3 or more per hour. Calculate the median village.

Solution 4

Central tendency is defined as the single value with in the range of the data which represents a group of individual values in a simple and concise manner and concentrates in the middle off the distribution.

The chief measure of central tendencies are:

i. Arithmetic mean

ii. Geometric mean

iii. Harmonic mean

iv. Median

v. Mode

vi. Quartiles, Deciles, Percentiles, etc.

Scanned by CamScanner

Median is very useful when

i. The given data is open-end classes.

ii. If the classes are unequal.

iii. It is highly skewed distribution.

iv. It is used for qualitative information like honesty, beauty, intelligence, attitude towards disable people, etc.

Short Questions

5. **Measures of central tendencies** **[PU 2010 Spring]**

Solution 5

See in theoretical part of this chapter.

6. **Write short note on partition values.** **[PU 2011 Fall/2011 Spring]**

Variate values which divides the total number off observations into an equal number of parts are called partitions values. There are different types of partition values such as quartiles, deciles, percentiles, etc.

i. **Quartiles:** Variate values which divide the distribution into four equal parts and each equals to 25% are called quartiles. They are denoted by Q_1, Q_2, Q_3.

ii. **Deciles:** Variate values which divides the distribution into 10 equal parts and each equals to 10% are called deciles. They are denoted by $D_1, D_2, D_3, \ldots D_9$.

iii. **Percentiles:** Variate values which divides the distribution into 100 equal parts and each equals to 1% are called percentiles. They are denoted by $P_1, P_2, P_3, \ldots, P_{99}$.

Scanned by CamScanner

5
Chapter

Measures of Dispersion

Dispersion is the spread of data in a distribution. Also, we can say that the measure of dispersion is the scatterness of the items from the central value. It is also called the measures of variation, measures of variability or measure of spread.

Measures of dispersion

The common measures of dispersion are given below:

Absolute Measures	Relative Measures
1. Range	1. Coefficient of range
2. Quartile deviation	2. Coefficient of Q.D.
3. Mean deviation	3. Coefficient of M.D.
4. Standard deviation	4. Coefficient of variation

Range

The difference between two extreme observations is called the range of the distribution. Where range R = L – SA, L = largest item, S = smallest item.

$$\text{Coefficient of range (R)} = \frac{L-S}{L+S}$$

Quartile deviation

The half of the difference between the lower quartile (Q_1) and upper quartile (Q_3) is said to be the quartile deviation. It is denoted by Q.D. Also defined by quartile deviation (Q.D.) = $\frac{Q_3 - Q_1}{2}$

Scanned by CamScanner

Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Mean deviation (M.D.)

Mean deviation is the arithmetic mean of the deviations of a series computed from mean, median or mode; where all deviations are considered positive. It is denoted by M.D.

Measures of mean deviation

	Individual Series	Discrete/Continuous series
M.D. from mean	$\frac{\Sigma\lvert x-\bar{x}\rvert}{n}$	$\frac{\Sigma f\lvert x-\bar{x}\rvert}{N}$
M.D. from median	$\frac{\Sigma\lvert x-M_d\rvert}{n}$	$\frac{\Sigma f\lvert x-M_d\rvert}{N}$
M.D. from mode	$\frac{\Sigma\lvert x-M_o\rvert}{n}$	$\frac{\Sigma f\lvert x-M_o\rvert}{N}$
Coefficient of M.D.	$\frac{\text{M.D. from mean}}{\text{Mean}} = \frac{\text{M.D. from median}}{\text{Median}} = \frac{\text{M.D. from mode}}{\text{Mode}}$	

Standard deviation

The positive square root of the arithmetic mean of the square of the deviations of the observation from their arithmetic mean is called the standard deviation. It is denoted by s.d or Greek letter σ (sigma)

Measures of standard deviation

	Individual Series	Discrete/Continuous series
Simple method	$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$	$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$
Short-cut method	$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}, d = x - A$	$\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$
Step-deviation method	$\sigma = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times h$	$\sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$

Where, $d' = \frac{x-A}{h}$, A - assumed mean, h - width of class

Note:

1. Standard deviation is the best measures of dispersion.
2. The squares of standard deviation is called the variance, it is denoted by σ^2.
3. The combined standard deviation is denoted and defined as,

i. $$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where, $\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$, $d_1 = \bar{x}_1 - \bar{x}_{12}$, $d_2 = \bar{x}_2 - \bar{x}_{12}$

ii. $$\sigma_{123} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + n_3(\sigma_3^2 + d_3^2)}{n_1 + n_2 + n_3}}$$

Scanned by CamScanner

Where, $\bar{x}_{123} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$

$d_1 = \bar{x}_1 - \bar{x}_{12}$, $d_2 = \bar{x}_2 - \bar{x}_{12}$, $d_3 = \bar{x}_3 - \bar{x}_{123}$

4. Coefficient variation $C.V. = \dfrac{\sigma}{\bar{x}} \times 100\%$
5. For comparing the variability of two or more than two sets of data, a distribution having less C.V. means less variable or more uniform or more consistent or more homogeneous or more stable or more equitable.
6. Also the distribution having more CV means more variable or less uniform or less consistent or less homogenous or less stable or less equitable, etc.

Outliers

An observation (or measurement) that is unusually large or small relative to the other values in a data set is called an outlier.

Box- Plots (Box- and Whisker Plot)

A box-and-whisker plot is a graphical representation of the data which consists of a line extending from minimum value to maximum value and a box with the lines drawn at Q_1, Q_2 and Q_3.

Five points summary or (Five number summary)

In a set of observation of the five values:

i. Smallest value of observation
ii. First quartile (Q_1)
iii. Median (Q_2)
iv. Third quartile (Q_3)
v. Maximum value of the observation is called the five point summary or five-value summary.

Exercise 5

1. **Calculate the quartile deviation of 63 students in numerical methods given below:**

Marks	No. of students	Marks	No. of students
0-10	5	50-60	7
10-20	7	60-70	3
20-30	10	70-80	2
30-40	16	80-90	2
40-50	11	90-100	0

Solution 1

Calculation of quartile deviation from the given data:

Marks	No. of students (f)	cf
0 - 10	5	5
10 - 20	7	12
20 - 30	10	22
30 - 40	16	38
40 - 50	11	49
50 - 60	7	56
60 - 70	3	59
70 - 80	2	61
80 - 90	2	63
90 - 100	0	63
	N = 63	

Scanned by CamScanner

Now, $\frac{N}{4} = \frac{63}{4} = 15.75$

Therefore Q_1 lies in the interval 20 - 30.

$$\therefore \quad Q_1 = l + \frac{\frac{N}{4} - cf}{f} \times i = 20 + \frac{15.75 - 12}{10} \times 10 = 23.75$$

Also, $\frac{3N}{4} = 3 \times \frac{63}{4} = 47.25$

So, Q_3 lies in the interval 40 - 50.

$$\therefore \quad Q_3 = l + \frac{\frac{3N}{4} - cf}{f} \times i = 40 + \frac{47.25 - 38}{11} \times 10 = 48.41$$

Now, Quartiles deviation Q.D. $= \frac{Q_3 - Q_1}{2} = \frac{48.41 - 23.75}{2} = 12.33$

Hence, Q.D. marks is 12.33.

2. **Find the mean deviation from mean of the following data:**

Class	0-6	6-12	12-18	18-24	24-30
Frequency	8	10	12	9	5

Solution 2

Calculation of mean-deviation from mean

Class	Frequency (f)	Mid-value (x)	fx	|x – 14|	f|x – 14|
0 - 6	8	3	24	11	88
6 - 12	10	9	90	5	50
12 - 18	12	15	180	1	12
18 - 24	9	21	189	7	63
24 - 30	5	27	135	13	65
	N = 44		Σfx = 618		Σf|x – 14| = 278

We know that,

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{618}{44} = 14$$

$$\text{M.D.} = \frac{\Sigma f|X - \bar{X}|}{N} = \frac{278}{44} = 6.3$$

3. **Following figures give the production of non-fatty dry milk during the twelve months of 2004.**

Months	Production (million Lbs)
January	63.5
February	81.6
March	95.8
April	111.5
May	131.4
June	126.5
July	98.7
August	76.2
September	53.2
October	50.3
November	49.3
December	67.1

Calculate (i) range (ii) mean deviation about mean (iii) variance (iv) coefficient of variation.

Scanned by CamScanner

Solution 3

Calculation of range, M.D. about mean, variance and coefficient of variation from the given data

Months	Production (million lbs)	$\lvert X-\bar{X}\rvert$	$(X-\bar{X})^2$
Jan	83.5	1.92	3.7
Feb	81.6	3.82	14.6
March	95.8	10.38	107.7
April	111.5	26.08	680.2
May	131.5	46.08	2123.4
June	126.5	41.08	1687.6
July	98.5	13.08	171.1
August	76.2	9.22	85
September	53.2	32.22	1038.1
October	50.3	35.12	1233.4
November	49.3	36.12	1304.7
December	67.1	18.32	335.62

i. Here, Largest value (L) = 131.5 Smallest value (S) = 49.3

Range = L – S = 131.5 – 49.5 =82

ii. **For M.D. from mean**

$$\text{Mean } (\bar{X}) = \frac{\Sigma fX}{n} = \frac{1025}{12} = 85.42$$

$$\therefore \quad \text{M.D. from mean} = \frac{\Sigma|X-\bar{X}|}{n} = \frac{273.44}{12} = 22.79$$

iii. $$\text{Variance} = \frac{\Sigma(X-\bar{X})^2}{n} = \frac{8785.12}{12} = 732.1$$

iv. $$\text{Coefficient of variation C.V.} = \frac{Sd}{\text{mean}} \times 100\% = \frac{\sqrt{\text{variance}}}{85.42} \times 100\% = 31.68\%$$

4. **The following table gives the number of branches and number of plants in a patch of field:**

No. of branches (x)	No. of plants (f)
2	16
3	13
4	19
5	12
6	5
7	4

Calculate the mean deviation about median and variance of the number of branches.

Solution 4

Calculation of M.D from median and variance of the given data

No. of branches (X)	No. of plants (f)	cf	$\lvert X-m_d\rvert$	$f\lvert X-m_d\rvert$	fx	fx^2
2	16	16	2	32	32	64
3	13	29	1	39	13	117
4	19	48	0	76	0	304
5	12	60	1	72	12	360
6	5	65	2	30	10	180
7	4	69	3	28	12	196
	N = 69			$\Sigma f\lvert X-m_d\rvert = 79$		

Scanned by CamScanner

For median

$$\frac{N}{2} = \frac{69}{2} = 35.5$$

The cumulative frequency just greater than $\frac{N}{2} = 48$. Therefore, median is 4.

$$\text{M.D.} = \frac{\Sigma|X - m_d|}{N} = \frac{79}{69} = 1.15$$

For Variance

$$\text{Mean } \overline{X} = \frac{\Sigma fX}{N} = \frac{277}{69} = 4.02$$

$$\text{Variance } (\sigma)^2 = \frac{\Sigma fX^2}{n} - \left(\frac{\Sigma fX}{n}\right)^2 = \frac{1221}{69} - \left(\frac{277}{69}\right)^2 = 17.7 - 16.16 = 1.54$$

5. **Age distribution of hundred life insurance policy holders is as follows:**

Ages as on nearest birthday	Number
17-19.5	9
20-25.5	16
26-35.5	12
36-40.5	26
41-50.5	14
51-55.5	12
56-60.5	6
61-70.5	5

Calculate mean deviation from median age.

Solution 5

To calculate the mean deviation from median the given frequency distribution must be in exclusive. So, correction factor $C_f = \frac{20 - 19.5}{2} = 0.25$

Age	f	cf	Mid-value (X)	$\lvert X - M_d \rvert$	$f\lvert X - M_d \rvert$
16.75 – 19.75	9	9	18.25	20	180
19.75 – 25.75	16	25	22.75	15.5	248
25.75 – 35.75	12	37	30.75	7.5	90
35.75 – 40.75	26	63	38.25	0	0
40.75 – 50.75	14	77	45.75	27.5	105
50.75 – 55.75	12	89	53.25	15	180
55.75 – 60.75	6	95	58.25	20	120
60.75 – 70.75	5	100	65.75	27.5	137.5
	N = 100				$\Sigma f\lvert X - M_d \rvert = 1060.5$

Here, $\frac{N}{2} = \frac{100}{2} = 50$

Cumulative frequency just greater than $\frac{N}{2} = 50$ is 63.

Therefore, median lies in the interval 35.75 – 40.75.

$$\therefore \quad M_d = l + \frac{\frac{N}{2} - cf}{f} \times i = 35.75 + \frac{50 - 37}{26} \times 5 = 38.25$$

$$\text{M.D. from median} = \frac{\Sigma f|X - M_d|}{N} = \frac{1060.5}{100} = 10.6$$

Scanned by CamScanner

6. Calculate the mean and standard deviation of the following distribution:

x	f
2.5-7.5	12
7.5-12.5	28
12.5-17.5	65
17.5-22.5	121
22.5-27.5	175
27.5-32.5	198

x	f
32.5-37.5	176
37.5-42.5	120
42.5-47.5	66
47.5-52.5	27
52.5-57.5	9
57.5-62.5	3

Solution 6

Calculation of mean and s.d. from the given data

X	f	Mid-value (M)	$d' = \frac{M-30}{5}$	fd'	fd'²
2.5 - 7.5	12	5	–5	–60	300
7.5 - 12.5	28	10	–4	–112	448
12.5 - 17.5	65	15	–3	–195	585
17.5 - 22.5	121	00	–2	–242	484
22.5 - 27.5	175	25	–1	–175	175
27.5 - 32.5	198	30	0	0	0
32.5 - 37.5	176	35	1	175	175
37.5 - 42.5	120	40	2	240	480
42.5 - 47.5	66	45	3	198	594
47.5 - 52.5	27	50	4	108	432
52.5 - 57.5	9	55	5	45	225
57.5 - 62.5	3	60	6	18	108
	N = 1000			Σfd' = 1	Σfd'₂ = 4007

We know that,

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h = 30 + \frac{1}{1000} \times 5 = 30.005$$

$$\text{S.D.} = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h = \sqrt{\frac{4007}{1000} - \left(\frac{1}{1000}\right)^2} \times 5 = \sqrt{4.007} \times 5 = 10.009$$

7. A collar manufacture is considering the production of a new type of collar to attract young men. The following statistics of neck circumference are available based upon the measurements of typical group of college students:

Mid value (inches)	No. of students	Mid value (inches)	No. of students
12.5	4	15.0	29
13.0	19	15.5	18
13.5	30	16.0	1
14.0	63	16.5	1
14.5	66		

Compute the mean, standard deviation and variance.

Solution 7

Calculation of Mean, S.D. and Variance

Mid-value (X)	f	d = X – 14.5	fd	fd²
12.5	4	–2	–8	16
13.0	19	–1.5	–28.5	42.75
13.5	30	–1	–30	30
14.0	63	–0.5	–31.5	15.75
14.5	66	0	0	0
15.0	29	0.5	14.5	7.25
15.5	18	1	18	18
16.0	1	1.5	1.5	2.25
16.5	1	2	2	4
	N = 231		Σfd = –62	Σfd² = 136

Scanned by CamScanner

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd}{N} = 14.5 + \left(\frac{-62}{231}\right) = 14.23$$

$$\text{S.D.} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} = \sqrt{\frac{136}{231} - \left(\frac{-63}{231}\right)^2} = \sqrt{0.59 - 0.07} = \sqrt{0.52} = 0.72$$

Variance $= \sigma^2 = (0.72)^2 = 0.52$

8. **Two persons participated in five shooting competition and were able to hit the target correctly out of fifteen shoots as given below:**

Competitor A	Competitor B
6	12
12	15
12	7
10	7
7	4

Find which of the competitors is more consistent in shooting performance. (Observation has taken from survey.)

Solution 8

Calculation of CV from the given data

Competitor A (X)	X^2	Competitor B (Y)	Y^2
6	36	12	144
12	144	15	225
12	144	7	49
10	100	7	49
7	49	4	16
$\Sigma X = 47$	$\Sigma X^2 = 473$	$\Sigma Y = 45$	$\Sigma Y^2 = 483$

$$\text{Mean of A, } \bar{X} = \frac{\Sigma X}{n} = \frac{47}{5} = 9.4$$

$$\text{S.D. of A } (\sigma_A) = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{473}{5} - (9.4)^2} = 2.5$$

$$\text{C.V. of A} = \frac{\text{S.D.}}{\bar{X}} \times 100\% = \frac{2.5}{9.4} \times 100\% = 26.6\%$$

$$\text{Mean of B, } \bar{Y} = \frac{\Sigma Y}{n} = \frac{45}{5} = 9$$

$$\text{S.D. of B } (\sigma_B) = \sqrt{\frac{\Sigma Y^2}{n} - \left(\frac{\Sigma Y}{n}\right)^2} = \sqrt{\frac{483}{5} - (9)^2} = 3.9$$

$$\text{C.V. of B} = \frac{\text{S.D.}}{\bar{X}} \times 100\% = \frac{3.9}{9} \times 100\% = 43.9\%$$

Since, C.V. of competitor A is less than C.V. of B. So, competitor A is more consistent in shooting.

9. **Scores of two golfers for 24 rounds were as follows:**

Golfer A	74	75	78	78	72	77	79
	78	87	76	72	72	77	74
	70	78	79	80	81	74	80
	75	71	73				

Scanned by CamScanner

Golfer B	86	84	80	88	89	85	86
	82	82	79	86	80	82	76
	86	89	87	83	80	88	96
	81	84	87				

Find which golfer may be considered to be a more consistent player?

Solution 9

Calculation of Coefficient of variation from the given data

Golfer A				
X	f	d = X – 75	fd	fd^2
70	1	–5	–5	25
71	1	–4	–4	16
72	3	–3	–9	27
73	1	–2	–2	4
74	3	–1	–3	3
75	2	0	0	0
76	1	1	1	1
77	2	2	4	8
78	4	3	12	36
79	2	4	8	32
80	2	5	10	50
81	2	6	12	72
	n = 24		Σfd = 24	$\Sigma fd^2 = 274$

Golfer B				
Y	f	d = X – 83	fd	fd^2
76	1	–7	–7	49
79	1	–5	–5	25
80	3	–3	–9	27
81	1	–2	–2	4
87	3	–1	–3	3
83	1	0	0	0
84	2	1	2	2
85	1	2	2	4
86	5	3	15	45
87	2	4	8	32
88	2	5	10	50
89	2	6	18	108
	N = 24		Σfd = 29	$\Sigma fd^2 = 349$

$$\bar{X} = A + \frac{\Sigma fd}{N} = 75 + \frac{24}{24} = 76$$

$$\sigma_X = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{274}{24} - \left(\frac{24}{24}\right)^2} = 3.2$$

$$C.V. = \frac{\sigma_X}{\bar{X}} \times 100\% = \frac{3.2}{76} \times 100 = 4.2\%$$

$$\bar{Y} = A + \frac{\Sigma fd}{N} = 83 + \frac{29}{24} = 84.2$$

$$\sigma_Y = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{349}{24} - \left(\frac{29}{24}\right)^2} = 3.62$$

$$C.V. = \frac{\sigma_Y}{\bar{Y}} \times 100\% = \frac{3.61}{84.2} \times 100 = 4.3\%$$

Since, C.V. of golfer A is less than golfer B. So, golfer A is more consistent.

10. **Lives of two models of refrigerators turned in for new models in a recent survey are:**

Life (No. of years)	Number of refrigerators	
	Model A	Model B
0-2	5	2
2-4	16	7
4-6	13	12
6-8	7	19
8-10	5	8
10-12	4	1

Scanned by CamScanner

What is the average life of each model of these refrigerators? Which model has more uniformity? [TU 2015 Spring/2015 Fall/2014 Spring/TU 2056/2065/2068]

Solution 10

Please refer to solution 9

11. An analysis of monthly wages paid to the workers in two firms A and B belongings to the same industry, gave the following results:

	Firm A	Firm B
No. of wage-earners:	986	548
Average monthly wages (Rs)	52.5	47.5
Variance of distribution of wages (Rs)	100	121

i. Which firm, A or B, pays out larger amount as monthly wages?

ii. What are the measures of average monthly wages and the variability in individual wages, of all the workers in the two firms, A and B, taken together?

Solution 11

We have given,

For Firm -A	For Firm B
$n_1 = 986$	$n_2 = 548$
$\overline{X}_1 = 52.5$	$\overline{X}_2 = 47.5$
$\sigma_A^2 = 100$	$\sigma_B^2 = 121$
$\therefore \sigma_A = 10$	$\therefore \sigma_B = 11$

i. Amount of monthly wages in firm A = 52.5 × 986 = 51765

Amount of monthly wages in firm B = 47.5 × 548 = 26030

Hence, firm A pays the larger amount of monthly wages.

ii. $$\text{Combined mean} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2} = \frac{986 \times 52.5 + 548 \times 47.5}{986 + 548} = 50.71$$

$$\text{C.V. of firm A} = \frac{\sigma_A}{\overline{X}_1} \times 100\% = \frac{10}{52.5} \times 100\% = 19\%$$

$$\text{C.V. of firm B} = \frac{\sigma_B}{\overline{X}_2} \times 100\% = \frac{11}{47.5} \times 100\% = 23.16\%$$

Hence, form B has greater variability that firm A.

Also, combined standard deviation is given by the formula

$$\sigma = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}, \quad d_1 = \overline{X}_1 - \overline{X} = 1.79, \quad d_2 = \overline{X}_2 - \overline{X} = 3.2$$

$$= \sqrt{\frac{986(100 + (1.79)^2) + 548(121 + (3.2)^2)}{986 + 548}}$$

$$= \sqrt{\frac{101759.24 + 71954.65}{1534}} = 10.64$$

Scanned by CamScanner

12. **Find the missing information from the following data:**

	Group I	Group II	Group III	Combined
Number	50	?	90	200
S.D.	6	7	?	7.745
Mean	113	?	115	116

Solution 12

Here, we have given

$n_1 = 50, \quad n_2 = ?, \quad n_3 = 90, \quad n = 200$

Since,

$n = n_1 + n_2 + n_3$

$200 = 50 + n_2 + 90$

$\Rightarrow \quad n_2 = 60$

Also,

$\overline{X}_1 = 113, \quad \overline{X}_2 = ?, \quad \overline{X}_3 = 115, \quad \overline{X} = 116$

We also know that,

$$\text{Combined mean } (\overline{X}) = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$$

or, $$116 = \frac{50 \times 113 + 60 \times \overline{X}_2 + 90 \times 115}{200}$$

or, $60\overline{X}_2 = 23200 - 5650 - 10350$

or, $$\overline{X}_2 = \frac{7200}{60} = 120$$

Also given,

$\sigma_1 = 6, \quad \sigma_2 = 7, \quad \sigma_3 = x, \quad \sigma = 7.745$

Combined standard deviation is given by,

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n^2(\sigma_2^2 + d_2^2) + n_3(\sigma_3^2 + d_3^2)}{n_1 + n_2 + n_3}$$

or, $$(7.745)^2 = \frac{50(36 + 9) + 60(49 + 16) + 90\,[x^2 + 1]}{200}$$

or, $90x^2 + 90 = 11997 - 2250 - 3900$

or, $90x^2 = 5757$

or, $x^2 = 63.97$

$\therefore \quad x = 8$

$\therefore \quad \sigma_3 = 8$

13. **A student obtained the mean and standard deviation of 100 observations as 40 and 5 respectively. It was later discovered that he had wrongly copied down an observation as 50 instead of 40. Calculate the correct mean and standard deviation.**

Solution 13

Here, we have given,

Scanned by CamScanner

$n = 100, \quad \overline{X} = 40, \quad S.d. = 5$

Wrong input value = 50, correct input = 40

Correct mean = ? and Correct S.d. = ?

Correct $\Sigma X = 100 \times 40 - 50 + 40 = 3990$

$\therefore$ Correct mean $= \frac{\text{Correct } \Sigma X}{n} = \frac{3990}{100} = 39.9$

Also, correct ΣX^2 = Wrong $\Sigma X^2 - 50^2 + 40^2 = 162500 - 2500 + 1600 = 161600$

$\therefore$ Correct S.d. $= \sqrt{\frac{\text{Correct } \Sigma X^2}{n} - (\text{Correct } \overline{X})^2} = \sqrt{\frac{161600}{100} - (39.9)^2} = 4.9$

14. The scores of two golfers for 10 rounds each are:

A	58	59	60	54	65	66	52	75	69	52
B	84	56	92	65	86	78	44	54	78	68

Which may be regarded as the more consistent player?

Solution 14

From calculation we get,

$\overline{X} = 61 \qquad \overline{Y} = 70.5$

$\sigma_x = 7.3 \qquad \sigma_y = 14.9$

C.V (x) = 11.97% C.V (Y) = 21.13

Since C.V. of golfer A is less than C.V. of golfer B.

Hence, golfer A is more consistent player than player B.

15. The heights and weights of 10 army men are given below. In which characteristics are they more variable?

Heights (in cm)	170	172	168	177	179	179	171	173	178	173
Weights (in kg)	75	74	75	76	77	73	76	75	74	75

Solution 15

From calculation, we get

$\overline{X} = 174 \qquad \overline{Y} = 75$

$\sigma_x = 3.8 \qquad \sigma_y = 1.1$

C.V. (x) = 2.2% C.V (y) = 1.5%

Since C.V. of weights is greater than C.V. of heights so, heights is more variable than the weights.

16. The following are scores of two batsmen A and B in a series of innings:

A	12	115	6	73	7	19	119	36	84	29
B	47	12	16	42	4	51	37	48	43	0

Who is the better run getter and who is more consistent?

Solution 16

From calculation, we get

$\overline{A} = 50 \qquad \overline{B} = 30$

Scanned by CamScanner

$\sigma_A = 41.83$ $\sigma_B = 18.74$

C.V. (A) = 83.66% C.V. (B) = 62.47%

Since C.V. of batsmen A is greater than C.V. of batsmen B.

Hence, batsmen B is more consistent than batsmen A.

17. The average and standard deviation of the heights of two groups of students, numbering 100 and 250 respectively, are given below. What is the average and standard deviation of the heights of all the 350 students taken together?

	Group I	Group II
Average (cm)	150	165
S.D. (cm)	8	10

Solution 17

$$\text{Combined mean } \overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2} = \frac{100 \times 150 + 250 \times 165}{100 + 250} = 160.71$$

Also,

$$\text{Combined S.d } (\sigma) = \sqrt{\frac{n_1(\sigma_1^2 + \sigma_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{100(64 + 114.7) + 250(100 + 18.4)}{100 + 250}} = \sqrt{\frac{17870 + 29600}{350}} = 11.65$$

18. The first of the two groups has 100 items with mean 45 and variance 49. If the combined group has 250 items with mean 51 and variance 130, find the mean and the standard deviation of the second group. [PU 2014 Fall]

Solution 18

Here, we have given

$n_1 = 100$	$\overline{X}_1 = 42$	$\sigma_1^2 = 49$
$n = 250$	$\overline{X} = 51$	$\sigma^2 = 130$
$n_1 = ?$	$\overline{X}_2 = ?$	$\sigma_2^2 = ?$

We have,

$$n = n_1 + n_2 \Rightarrow 250 = 100 + n_2 \Rightarrow n_2 = 150$$

Also, Combined mean

$$(\overline{X}) = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$$

$$\text{or, } 51 = \frac{100 \times 45 + 150 \times \overline{X}_2}{250}$$

$$\text{or, } \overline{X}_2 = \frac{250 \times 51 - 100 \times 45}{150} = 55$$

Combined variance

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$$

Scanned by CamScanner

or, $130 = \frac{100(49 + 100) + 150(\sigma_2^2 + 36)}{250}$

or, $\sigma_2^2 + 36 = \frac{130 \times 250 - 100 \times 149}{150} = 117.33$

$\therefore$ $\sigma_2^2 = 117.33 - 36 = 81.33$

$\therefore$ $\sigma^2 = 9$

19. The means of two population of size 50 and 100 respectively are 54.1 and 50.3 and the standard deviations are 8 and 7. Obtain the mean and standard deviation of the population of size 150 obtained by combining the two population.

Solution 19

Here, we have given

$n_1 = 50$	$n_2 = 100$
$\overline{X}_1 = 54.1$	$\overline{X}_2 = 50.3$
$\sigma_1 = 8$	$\sigma_2 = 7$

Now,

Combined mean $(\overline{X}) = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2} = \frac{50 \times 54.1 + 100 \times 50.3}{50 + 100} = 51.57$

Also,

Combined variance (σ^2) $= \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$

$= \frac{50(64 + 6.4) + 100(49 + 1.6)}{150} = \frac{3520 + 5060}{150} = 57.2$

Combined S.d. $(\sigma) = \sqrt{57.2} = 7.6$

20. A distribution consists of three components with frequencies 200, 250 and 100 having means 25, 10 and 15, standard deviations 3, 4 and 5 respectively. Show that the mean of the combined group is 16 and its standard deviation is 8 approximately.

Solution 20

Here, we have given

$n_1 = 200$	$\overline{X}_1 = 25$	$\sigma_1 = 3$
$n_2 = 250$	$\overline{X}_2 = 10$	$\sigma_2 = 4$
$n_3 = 100$	$\overline{X}_3 = 4$	$\sigma_3 = 5$

Then, combined mean is given by the formula

$$(\overline{X}) = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3} = \frac{200 \times 25 + 250 \times 10 + 100 \times 15}{200 + 250 + 100} = 16.36 = 16$$

Also, the combined standard deviation is

or, $\sigma = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + n_3(\sigma_3^2 + d_3^2)}{n_1 + n_2 + n_3}}$

$= \sqrt{\frac{200(9 + 81) + 250(16 + 36) + 100(25 + 1)}{200 + 250 + 100}} = \sqrt{61} = 7.8 = 8$

Scanned by CamScanner

22. Find the mean and variance of first n-natural numbers.

Solution 22

We know that,

The se of first, n-natural numbers are 1, 2, 3, 4, 5, 6, ..., n

$$\text{Mean } (\bar{X}) = \frac{\Sigma X}{n} = \frac{1+2+3+4+5+\ldots+n}{n} = \frac{n(n+1)}{2} \times \frac{1}{n} = \frac{n+1}{2}$$

Also variance,

$$\sigma^2 = \frac{\Sigma X^2}{n} = \frac{1^2+2^2+3^2+4^2+\ldots+n^2}{n} = \frac{n(n+1)(2n+1)}{6} \times \frac{1}{n} = \frac{(n+1)(2n+1)}{6}$$

23. Following figures give the production of mustard oil, during the twelve months of 1975 in sample.

Months	Production (million LBS)
Jan	96.1
Feb	99.2
Mar	85.2
Apr	87.2
May	90.1
June	90.0

Months	Production (million LBS)
July	76.1
Aug	98.1
Sept	101.0
Oct	79.2
Nov	87.2
Dec	71.3

Calculate

i. range
ii. mean deviation about mean
iii. variance and
iv. Coefficient of variation

Solution 23

Calculation of ...

Months	Production million lbs (X)	d = x – 86	d^2	$\lvert X - \bar{X} \rvert$
Jan.	96.1	10.1	102.01	7.71
Feb.	99.2	13.2	174.24	10.81
Mar.	85.2	–0.8	0.64	3.19
Apr.	87.2	1.2	1.44	1.19
May	90.1	4.1	16.81	1.71
June	90.0	4	16	1.61
July	76.1	–9.9	98.01	12.29
Aug.	98.1	12.1	146.41	9.71
Sept.	101.0	15	225	12.61
Oct.	79.2	–6.8	46.24	9.19
Nov.	87.2	1.2	1.44	1.19
Dec.	71.3	–14.7	216.09	17.09
		$\Sigma d = 27.7$	$\Sigma d^2 = 1044.33$	$\Sigma \lvert X - \bar{X} \rvert = 88.21$

i. Range = Largest – Smallest = 101.0 – 71.3 = 29.7

ii. $\text{Mean } (\bar{X}) = A + \frac{\Sigma d}{n} = 86 + \frac{28.7}{12} = 88.39$

$$\text{M.D.} = \frac{\Sigma \lvert X - \bar{X} \rvert}{n} = \frac{88.21}{12} = 7.35$$

iii. $\text{Variance } (\sigma)^2 = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{1044.33}{12} - \left(\frac{28.7}{12}\right)^2} = 9$

iv. $\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100\% = \frac{6}{88.39} \times 100\% = 10.18\%$

Scanned by CamScanner

24. Direct evidence of Newton's universal law of gravitation as provided from a renowned experiment, by Henry Cavendish (1731 – 1810). In the experiment, masses of objects were determined by weighing and the measured force of attraction was used to calculate the density of the earth. The values of earth's density, in time order by row are:

5.36	5.29	5.58	5.65	5.57	5.53	5.62
5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.24	5.47	5.63	5.34	5.46	5.75	5.68
5.85	5.30					

a. Find mean and standard deviation

b. Find the median Q_1 and Q_3

Solution 24

Do yourself

25. A Michelson (1852 – 1931) made many series of measurement of the speed of light, Using a revolving mirror technique, he obtained

12, 30, 30, 27, 30, 39, 18, 27, 48, 24, 18

For the (difference velocity of light in air) – 299, 700 km/s

a. Find mean, median and mode b. Find variance and standard deviation

Solution 25

Do yourself

26. Civil engineers must monitor flow on river where power is generated. The following are the daily mean follow rate (MGD) on the Namekagon River during the month of May of 47 years.

602.0	517.5	572.5	392.4	505.8	547.5	289.1	497.2
794.8	657.6	904.7	595.5	611.9	482.9	696.6	606.7
986.4	567.7	400.4	634.9	448.4	479.1	415.6	718.5
575.6	743.3	114.6	461.6	644.0	480.8	429.1	626.9
8339.9	889.0	752.6	516.5	817.2	895.8	572.2	563.7
679.3	738.0	618.9	390.8	550.9	425.9	760.6	

i. Obtain the quartiles. ii. Obtain the 90th percentile.

iii. Construct a histogram.

Solution 26

Do yourself

27. J.J Thompson (1856 - 1940) discovered the electron by isolation negatively charged particles for which he could measure the mass-charge ratio. This ratio appeared to be constant over a wide range of experimental conditions and consequently, could be a characteristics of a new particles. His observations, from two different cathode-ray tubes that used air gas, are:

Tube 1	0.57	0.34	0.43	0.32	0.48	0.4	0.4
Tube 2	0.53	0.47	0.47	0.51	0.63	0.61	0.48

Which may be regarded as more consistent tube?

[PU 2014/PU2006 Spring/TU 2062/2068/2069]

Scanned by CamScanner

Solution 27

Calculation of C.V. from the given data

Tube -1		
X	d = X – 0.4	d^2
0.32	–0.08	0.0064
0.34	–0.06	0.0036
0.4	0	0
0.4	0	0
0.43	0.03	0.009
0.48	0.05	0.0025
0.57	0.17	0.0289
	$\Sigma d = 0.11$	$\Sigma d^2 = 0.0504$

Tube -2		
Y	d = Y – 0.5	d^2
0.47	–0.05	0.0025
0.47	–0.05	0.0025
0.48	–0.04	0.0016
0.51	0	0
0.53	0.02	0.004
0.61	0.1	0.01
0.63	0.12	0.0144
	$\Sigma d = 0.1$	$\Sigma d^2 = 0.035$

Tube -1

$$\bar{X} = A + \frac{\Sigma d}{n} = 0.4 + \frac{0.11}{7} = 0.42$$

$$\sigma_X = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{0.0504}{7} - \left(\frac{0.11}{7}\right)^2} = 0.084$$

$$C.V. = \frac{0.084}{0.42} \times 100 = 20\%$$

Tube -2

$$\bar{Y} = A + \frac{\Sigma d}{n} = 0.51 + \frac{0.1}{7} = 0.52$$

$$\sigma_Y = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{0.035}{7} - \left(\frac{0.1}{7}\right)^2} = 0.069$$

$$C.V. = \frac{0.069}{0.52} \times 100 = 13.27\%$$

Hence, tube 2 is more consistent than tube 1.

28. The mean and standard deviation of 20 items is found to be 10 and 2 respectively. In the time checking it was found that one item was 8 was incorrect. Calculate the mean and standard deviation if:

i. The wrong item is omitted and ii. It is replaced by 12.

Solution 28

Here,

$n = 20, \quad \bar{X} = 10, \quad \sigma = 2$

One value is incorrect,

i. If wrong item is omitted then

$n = 19, \quad \Sigma\bar{X} = 20 \times 10 - 8 = 192$

$\therefore$ Correct mean $= \frac{\Sigma X}{n} = \frac{192}{19} = 2016$

For $\Sigma X^2 = 104 \times 20 - 8^2 = 2016$

Scanned by CamScanner

Correction S.d $= \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{2016}{19} - (10.11)^2} = 1.97$

ii. If it is replaced by 12 then,

Correct mean $= \frac{\text{Correct } X}{n} = \frac{20 \times 10 - 8 + 12}{20} = 10.2$

Correct S.d $= \sqrt{\frac{\text{Correct } \Sigma X^2}{n} - \left(\frac{\text{Correct } \Sigma X}{n}\right)^2}$

$= \sqrt{\frac{104 \times 20 - 8^2 + 12^2}{20} - (10.2)^2} = 1.99$

29. From the following sample data, construct the inter-quartile range. Also, give 5-point summary for the data and plot the box and whisker diagram.

171	290	331	152	144	132	119	122
134	221	229	228	243	230	322	334
233	239	235					

Solution 29

Do yourself

30. The cold start engine ignition time of an automobile engine is being investigated by gasoline manufactured. The following times (in seconds) were obtained for a test vehicle.

1.75	1.92	2.62	2.35	3.09	3.15	2.53	1.91

Construct a box plot of the data and analyze it.

Solution 30

Here, given data in ascending order are

1.75, 1.91, 1.92, 2.35, 2.53, 2.62, 3.09, 3.15

Now, $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ place item $= (2.25)^{th}$ place item $= \frac{1.91 + 1.92}{2} = 1.915$

$M_d = \left(\frac{n+1}{2}\right)^{th}$ place item $= (4.5)^{th}$ place item $= \frac{2.35 + 2.53}{2} = 2.44$

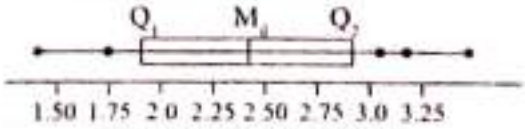
$Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ place item $= (6.75)^{th}$ place item $= \frac{2.62 + 3.09}{2} = 2.855$

Now, inter quartile range $= Q_3 - Q_1 = 2.855 - 1.915 = 0.94$

Lower whisker $= Q_1 - 1.5 \times (IQR) = 1.915 - 1.5 \times 0.94 = 0.505$

Upper whisker $= Q_3 + 1.5 \times (IQR) = 2.855 + 1.5 \times 0.94 = 4.265$

Here, no data lies below 0.505 and above 4.265, so there is no outliers.



Ignition time (in second)

In the above figure, whisker to the left off median is longer than whisker to the right. So, the distribution is negatively skewed.

Scanned by CamScanner

31. The nicotine contents, in milligrams for 40 cigarettes of a certain brand were recorded as follows:

1.08	1.92	2.31	1.79	2.28	1.74	1.47	1.97
0.85	1.24	1.58	2.03	1.70	2.17	2.55	2.11
1.86	1.90	1.68	1.51	1.64	0.72	1.69	1.85
1.82	1.79	2.46	1.68	2.08	1.67	1.37	1.93
1.40	1.64	2.09	1.75	1.63	2.37	1.37	1.69

Construct box and whisker plot for the above data and interpret the result.

Solution 31

Here, given N = 40

No. of class $k = 1 + 3.322 \log N = 6.3$

Width of class interval $= \dfrac{\text{Largest} - \text{Smallest}}{2} = \dfrac{2.55 - 0.72}{2} = 0.29 \approx 0.3$

Nicotine (in milligrams)	No. of cigarettes (f)	C.f.
0.72 – 1.02	2	2
1.02 – 1.32	2	4
1.32 - 1.62	5	9
1.62 - 1.92	18	27
1.92 - 2.22	8	35
2.22 - 2.52	4	39
2.52 - 2.82	1	40
	N = 40	

For Q_1; $\dfrac{N}{4} = \dfrac{40}{4} = 10$

$$\therefore \quad Q_1 = l + \frac{\frac{N}{4} - cf}{f} \times i = 1.62 + \frac{10 - 9}{18} \times 0.3 = 1.636$$

$$M_d = = l + \frac{\frac{N}{2} - cf}{f} \times i = 1.62 + \frac{20 - 9}{18} \times 0.3 = 0.803$$

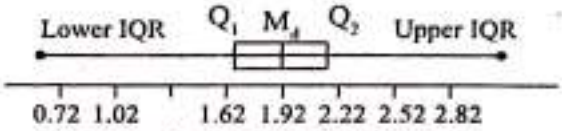
$$Q_3 = l + \frac{\frac{3N}{4} - cf}{f} \times i = 1.92 + \frac{30 - 27}{8} \times 0.3 = 2.03$$

Now, inter quartile range (IQR) $= Q_3 - Q_1 = 2.03 - 1.636 = 0.394$

Lower whisker $= Q_1 - 1.5 \times IQR = 1.636 - 1.5 \times 0.394 = 1.045$

Upper whisker $= Q_3 + 1.5 \times IQR = 2.03 + 1.5 \times 0.394 = 2.621$

Since, there are the numbers which are less than 1.045 and greater than 2.621, so they are outliers.



Nicotine contents (in mg)

In the above figure whisker to the right of median is greater than whisker to the left. So, it is positively skewed.

Scanned by CamScanner

32. **The following are the numbers of minutes that a person had to wait for the bus to work on 15 working days:**

10, 1, 13, 9, 5, 9, 2, 10, 3, 8, 6, 17, 2, 10, 15

Draw a box plot and interpret the result. [TU 2063A/2063M/2069B/PU2014 Spring]

Solution 32

Do yourself

Exam Asked Questions

Long Questions

1. **Over a period of 40 days, the percentage relative humidity in a vegetable storage building was measured. Mean daily values were recorded as below:** [PU 2017Spring]

60	63	64	71	67	73	79	80	83	81
86	90	96	98	98	99	89	80	77	78
71	79	74	84	85	82	90	78	79	79
78	80	82	83	86	81	80	76	66	74

i. Prepare a stem-leaf display for these data. Show the leaves sorted in order of increasing magnitude on each stem.

ii. Draw a box-plot for these data and interpret the data in practical manner.

Solution 1

Arranging the given data in the ascending order

60	63	64	66	67	71	71	71	73	74	74
76	77	78	78	78	79	79	79	79	80	80
80	80	81	81	82	82	83	83	84	85	86
86	89	90	96	98	98	99				

Stem and leaf display

Stem	Leaf	Frequency
6	0, 3, 4, 6, 7	5
7	1, 1, 3, 4, 4, 6, 7, 8, 8, 8, 9, 9, 9, 9	14
8	0, 0, 0, 0, 1, 1, 2, 2, 3, 3, 4, 5, 6, 6, 9	15
9	0, 0, 6, 8, 8, 9	6
		40

ii. **Corresponding frequency distribution**

Vegetable storage	No. of days (f)	Vegetable storage (In%)	f	cf
60 - 69	5	59.5 - 69.5	5	5
70 - 79	14	69.5 - 79.5	14	19
80 - 89	15	79.5 - 89.5	15	34
90 - 99	6	89.5 - 99.5	6	40

$$C_f = \frac{70-69}{2} = 0.5$$

Now, we have,

Scanned by CamScanner

a. Minimum value = 60 b. Maximum value = 99

c. Lower quartile $(Q_1) = \frac{N}{4} = \frac{40}{4} = 10$

Therefore, Q_1 lies in the interval 69.5 – 79.5.

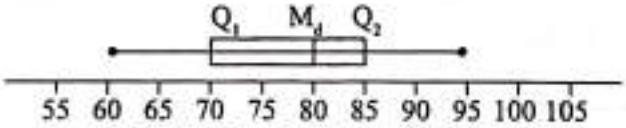
$$\therefore \quad Q_1 = l + \frac{\frac{N}{2} - cf}{f} \times i = 79.5 + \frac{30 - 19}{15} \times 10 = 86.83$$

Also, the interquartile range (IQR) = 86.83 – 70.07 = 16.76

For lower whisker, $Q_1 - 1.5 \times IQR = 70.07 - 1.5 \times 16.76 = 112.97$

For upper whisker, $Q_3 + 1.5 \times IQR = 86.83 + 1.5 \times 16.76 = 112.97$

Since, no data lies below 44.93 and above 111.97, so there is no outlier.



Vegetable storage in percentage

In the above figure, whisker to the left is longer than the whisker to the right, so the distribution is negatively skewed.

2. **The following are the number of minutes that a person had to wait for the bus to work on 15 working days.** **[PU 2010 Spring/TU 2063 A/2063 M/2069 B]**

10, 1, 13, 9, 5, 9, 2, 10, 3, 8, 6, 17, 2, 10, 15

Draw a box plot and interpret the result.

Solution 2

Arranging the given data in ascending order, we get,
1, 2, 2, 3, 5, 6, 8, 9, 9, 10, 10, 10, 13, 15, 17
The minimum value is equal to 1.
The maximum value is equal to 17.

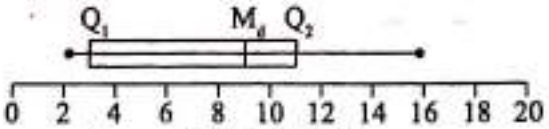
The first quartile (Q_1): The value of $\left(\frac{n+1}{4}\right)^{th}$ item

$$= \left(\frac{15+1}{4}\right)^{th} = 4^{th} \text{ item} = 3$$

The median $(M_2) = \left(\frac{n+1}{2}\right)^{th}$ item $= \left(\frac{15+1}{2}\right)^{th}$ item $= 8^{th}$ item $= 9$

The third quartile $(Q_3) = \left(\frac{3(n+1)}{4}\right)^{th}$ item $= 12^{th}$ item $= 10$

Therefore, five-number summary is 1, 3, 9, 10 and 17.
The inter quartile range (IQR) = $Q_3 - Q_1 = 10 - 3 = 7$
For lower whisker, $Q_1 - 1.5 \times IQR = 3 - 1.5 \times 7 = -7.5$
For upper whisker, $Q_3 + 1.5 \times IQR = 9 + 1.5 \times 7 = 19.5$
Since, there is no data below –7.5 and above 19.5. So, it has no outliers.



Time for Bus

Scanned by CamScanner

3. **A random sample was taken of the thickness off insulation in transformer windings and the following thickness (in millimetes) were recorded as:** **[PU 2016 Fall]**

18	21	22	29	25	31	37	38	41	39	44	48	54	56	57
47	38	35	36	29	37	37	32	42	42	40	48	36	37	37

i. Prepare a stem leaf display for these data.

ii. Prepare a box plot for these data.

Solution 3

Please refer solution 1

4. **The temperature in a chemical reactor was measured every half hour under the same conditions. The results were; (in degree celsius)** **[PU 2017 Spring]**

78.1, 79.2, 78.9, 80.2, 78.3, 78.8, 79.4

Calculate the mean, median, lower quartile, upper quartile, standard deviation and coefficient of variation.

Solution 4

Let X denote the temperature in a chemical reactor. Then,

X (in °C)): 78.1, 78.3, 78.8, 78.9, 79.2, 79.4, 80.2

Here, no of observation, n = 7

i. $\text{Mean} = \frac{\Sigma x}{n} = \frac{78.1 + 78.3 + 78.8 + 78.9 + 79.2 + 79.4 + 80.2}{7} = 78.98°C$

ii. Median = The value of $\left(\frac{n+1}{2}\right)^{th}$ place item

$= \left(\frac{7+1}{2}\right)^{th}$ place item = 4th place item = 78.9

iii. $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ place item $= \left(\frac{7+1}{4}\right)^{th}$ place item = 2th place item = 78.3

iv. $Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ place item = $(3 \times 2)^{th}$ place item = 6th place item = 79.4

v. $\text{S.d.} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{43674.19}{7} - \left(\frac{552.9}{7}\right)^2} = \sqrt{6236.17 - 6237.84} = 1.14$

vi. $\text{Coefficient of variation} = \frac{S.d}{\bar{X}} \times 100\% = \frac{1.14}{78.98} \times 100\% = 1.44\%$

5. **During the 10 weeks of a session the marks scored by two candidates A and B are taking the computer courses are given below**

A	58	59	60	54	65	66	52	75	69	52
B	87	89	78	71	73	84	65	66	56	46

i. Who is better scorer?

ii. Who is more consistent? **[PU 2001]**

Scanned by CamScanner

Solution 5

Calculation of mean, standard deviation and C.V

For candidate A		
X	d = X – 64	d^2
58	–6	36
59	–5	25
60	–4	16
54	–10	100
65	1	1
66	2	4
52	–12	144
75	11	121
69	5	25
52	–12	144
n = 10	$\Sigma d = -30$	$\Sigma d^2 = 616$

For candidate B		
Y	d = Y – 68	d^2
87	19	361
89	21	441
78	10	100
71	3	9
73	5	25
84	16	256
65	–3	9
66	–2	4
56	–12	144
46	–22	484
n = 10	$\Sigma d = -30$	$\Sigma d^2 = 616$

For candidate A

$$\text{Mean } \bar{X} = A + \frac{\Sigma d}{n} = 64 + \frac{(-30)}{10} = 64 - 3 = 61$$

$$\text{S.d. } (\sigma) = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{616}{10} - \left(\frac{-30}{10}\right)^2} = \sqrt{52.6} = 7.25$$

$$\text{C.v} = \frac{\text{S.d}}{\bar{X}} \times 100\% = \frac{7.25}{61} \times 100\% = 11.88\%$$

For candidate B:

$$\text{Mean } \bar{Y} = B + \frac{\Sigma d}{n} = 68 + \frac{35}{10} = 71.5$$

$$\text{S.d. } (\sigma) = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{1833}{10} - \left(\frac{35}{10}\right)^2} = 13.08$$

$$\text{C.v} = \frac{\text{S.d}}{\bar{Y}} \times 100\% = \frac{13.08}{71.5} \times 100\% = 18.29\%$$

i. Since, the average marks of candidate A is less than average marks of candidate B so, candidate B is good (better) scorer than candidate A.

ii. Since c.v. of candidate A is less than c.v. of candidate B, so candidate A is more consistent than candidate B.

6. **The combined mean and variance of salary of 250 workers of Kathmandu and Bhaktapur are Rs. 560 and Rs. 5497 resp. if the mean and variance salary of workers of Bhaktapur are Rs. 500 and Rs. 81 resp. find the variance salary of workers in Kathmandu.** **[PU 2016 / TU 2052/ TU 2067]**

olution 6

Iere, we have given,

Combined mean $\bar{X}_{12} = 560$

Combined variance $\sigma_{12}^2 = 5497$

Mean salary of Bhaktapur $\bar{X}_1 = 500$

Variance salary of Bhaktapur, $\sigma_1^2 = 81$

No. of workers in Bhaktapur $(n_1) = 150$

Variance salary of workers in Kathmandu $(\sigma_2^2) = ?$

Scanned by CamScanner

We know that,

Total number of workers = $(n_1 + n2)$

$250 = 150 + n_2 \Rightarrow n_2 = 100$

Also, the combined mean is given by

$$\bar{X}_{12} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} \Rightarrow 560 = \frac{150 \times 500 + 100 \times \bar{X}_2}{250} \Rightarrow \bar{X}_2 = \frac{560 \times 250 - 150 \times 500}{100} = 650$$

Again, the combined variance is given by,

$$\sigma_1^2 = \frac{n(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2} \quad \text{where } d_1 = |\bar{X}_1 - \bar{X}_{12}| = 60, \; d_2 = |\bar{X}_2 - \bar{X}_{12}| = 90$$

or, $5497 = \dfrac{150(81 + 60^2) + 100(\sigma_2^2 + 90^2)}{250}$

or, $150 \times 3681 + 100\,\sigma_2^2 + 81000 = 5497 \times 250$

$\therefore$ $\sigma_2^2 = \dfrac{1374250 - 1362150}{100} = 121$

7. **If the first group has 100 items with mean 45 and variance 49. If the combined group has 250 items with mean 51 and variance 130. Find the mean and standard deviation of the second group.** **[PU 2014 Fall]**

Solution 7

Please refer book exercise 5 Q.No. 18

8. **The first of two samples has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and variance 13.44. Find the standard deviation of the second group.** **[PU 2005/2007 Spring]**

Solution 8

Please refer solution 6

9. **The life's of two model (A and B) of refrigerators in a recent survey shown below**

Life (No. of Years)	No. of Refrigerators	
	Model A	Model B
0-2	5	2
2-4	16	7
4-6	13	12
6-8	7	19
8-10	5	9
10-12	4	1

i. What is the average life of each model of these refrigerators?

ii. Which model has greater uniformity?

PU 2015 Spring/ 2015 Fall / 2014 Spring /TU 2056/2065/2068

Solution 9

Please refer book exercise 5 Q.No. 10

10. **Calculate the coefficient of variation for the following distribution giving 300 telephones calls according to their duration in second** **[PU 2003]**

Duration (in sec)	0-30	30-60	60-90	90-120	120-150	150-180	180-210
No. of Calls	9	17	43	82	81	44	24

Scanned by CamScanner

Solution 10

Calculation of coefficient of variation of telephone calls

Duration (in second)	No. of calls (f)	Mid-value (x)	$d' \frac{x-105}{30}$	fd'	fd'^2
0 - 30	9	15	–3	–27	81
30 - 60	17	45	–2	–34	68
60 - 90	43	75	–1	–43	43
90 - 120	82	105	0	0	0
120 - 150	81	135	1	81	81
150 - 180	44	165	2	88	176
180 - 210	24	195	3	72	216
	N = 300			$\Sigma fd' = 173$	$\Sigma fd'^2 = 665$

We know that,

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd'}{N} \times h = 105 + \frac{137}{300} \times 30 = 118.7$$

$$\text{S.D } (\sigma) = \sqrt{\frac{\Sigma fd'}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h = \sqrt{\frac{665}{300} - \left(\frac{137}{300}\right)^2} \times 30 = 39.8$$

$\therefore$ Coefficient of variation C.V. $= \frac{S.d}{\bar{X}} \times 100\% = \frac{9.8}{118.7} \times 100\% = 33.5\%$

11. Following information are obtained about two batsman of a cricket team.

	Player A	Player B
No. of Matches	35	40
Average Runs	65	70
Standard Deviation	10	15

Find the combined mean & standard deviation of runs of two players. Which player is more consistent and why?
What would be the effect of combined mean if both player A and B play equal number of matches? [PU 2010 Spring/ 2009 Fall]

Solution 11

Here, we have given

Number of matches A player Plays $n_1 = 35$, $\bar{X}_1 = 65$, $\sigma_1 = 10$

Number of matches B player $n_2 = 40$, $\bar{X}_2 = 70$, $\sigma_2 = 15$

Now, combined mean $\bar{X}_{12} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{35 \times 65 + 40 \times 70}{35 + 40} = 67.67 \approx 68$

Also, the combined standard deviation is given by,

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}} \quad \text{where } d_1 = \bar{X}_1 - \bar{X}_{12}, \; d_2 = \bar{X}_2 - \bar{X}_{12}$$

$$\therefore \quad \sigma_{12} = \sqrt{\frac{35(100 + 9) + 40(225 + 4)}{35 + 40}} = \sqrt{\frac{3815 + 9160}{75}} = 13.15 = 13$$

Hence, the combined mean is 68 and s.d. is 13.

For Second Part

C.V. of player A $= \frac{S.d}{\bar{X}_1} \times 100\% = \frac{10}{65} \times 100\% = 15.38\%$

Scanned by CamScanner

C.V. of player B $= \frac{S.d}{\bar{X}_2} \times 100\% = \frac{15}{70} \times 100\% = 21.43\%$

Since, C.V of player A < C.V. of player B. So, player A is more consistent than player B.

For Third Part

If player A and player B play equal matches (i.e. 35 or 40)

Then $\bar{X}_{12} = \frac{70 + 65}{2} = 67.5$

The combined mean will decrease.

12. A factory produces two types of electric bulbs A and B. In an experiment relating to their life, the following results were obtained. Length of life

Hours	No. of bulb (A)	No. of bulb (B)
500-700	5	4
700-900	11	30
900-1100	26	12
1100-1300	10	8
1300-1500	8	6

Compare the variability of the life of the two types of bulbs using coefficient of variation. [PU2005 Fall / TU 2057]

Solution 12

Calculation of coefficient of variation of length of lifes

For bulb A						For bulb B		
Length of life (in hours)	No. of bulbs	Mid value (x)	$d' = \frac{x-1000}{200}$	f_1d'	$f_1d'^2$	f_2	f_2d'	$f_2d'^2$
500 - 700	5	600	−2	−10	20	4	−8	16
700 - 900	11	800	−1	−11	11	30	−30	30
900 - 1000	26	1000	0	0	0	12	0	0
1100 - 1300	10	1200	1	10	10	8	8	8
1300 - 1500	8	1400	2	16	32	6	12	24
	$N_1 = 60$			$\Sigma f_1d' = 5$	$\Sigma f_1d'^2 = 73$	$N_2 = 60$	$\Sigma f_2d' = -18$	$\Sigma f_2d'^2 = 78$

For bulb A

$$\text{Mean life } (\bar{x}_1) = A + \frac{\Sigma f_1 d'}{N_1} \times h = 1000 + \frac{5}{60} \times 200 = 1016.67$$

$$\text{S.D. } (\sigma_1) = \sqrt{\frac{\Sigma f_1 d'^2}{N_1} - \left(\frac{\Sigma f_1 d'}{N_1}\right)^2} \times h = \sqrt{\frac{73}{60} - \left(\frac{5}{60}\right)^2} \times 200 = 220.33$$

$$\text{C.V.} = \frac{\sigma_1}{\bar{x}_1} \times 100\% = \frac{220.33}{1016.67} \times 100\% = 21.67\%$$

For bulb B

$$\text{Mean life } (\bar{x}_2) = A + \frac{\Sigma f_2 d'}{N_2} \times h = 1000 + \frac{(-18)}{60} \times 200 = 940$$

$$\text{S.D. } (\sigma_2) = \sqrt{\frac{\Sigma f_2 d'^2}{N_2} - \left(\frac{\Sigma f_2 d'}{N_2}\right)^2} \times h = \sqrt{\frac{78}{60} - \left(\frac{(-18)}{60}\right)^2} \times 200 = 220$$

Scanned by CamScanner

$$C.V. = \frac{\sigma_2}{\bar{x}_2} \times 100\% = \frac{220}{940} \times 100\% = 23.40$$

Since, C.V. of life of bulb A is less than C.V. of life of bulb B. So, bulb B is more variable than bulb B.

13. An analysis of monthly wages paid to the workers in two firms A and B belonging to the same industry gives the following results:

	Firm A	Firm B
Number of worker:	500	600
Average monthly wages in $:	186	175
Variance of distribution of wage:	81	100

i. Which firm, A or B, has a large wage bill?
ii. In which firm, A or B is there grater variability in wage?
iii. Calculate the average monthly wages and the variance of the distribution of wages, of all the workers in the firm A and B taken together.

[PU 2009 Spring / 2016 Spring]

Solution 13

Please refer solution 11

14. For a group of 200 candidates, the mean and standard deviation were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the correct figures. **[PU 2010 Fall]**

Solution 14

Here given, $n = 100, \bar{x} = 40, \sigma = 10$
Wrong values = 30 and 70
Correct values = 3 and 27

$$\text{Now, correct mean} = \frac{\text{Correct } \Sigma x}{n} = \frac{n\bar{x} - 30 - 70 + 3 + 27}{100} = \frac{4000 - 100 + 30}{100} = 39.3$$

$$\text{Since, } (S.d)^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 \Rightarrow 100 = \frac{\Sigma x^2}{100} - (40)^2 \Rightarrow \Sigma x^2 = 170000$$

Now, correct $\Sigma x^2 = 170000 - 30^2 - 70^2 + 3^2 + 27^2 = 164932$

$$\text{Correct S.d.} = \sqrt{\frac{\text{correct } \Sigma x^2}{n} - \left(\frac{\text{correct } \Sigma x}{n}\right)^2} = \sqrt{\frac{164932}{100} - (39.3)^2} = 10.24$$

15. J.J Thomson (1856-1940) discovered the electron by isolation of negatively charged particles for which he could measure the mass – charge ratio. This ratio appeared to be constant over a wide range of experimental conditions and consequently, could be a characteristic of a new particle. His observations, from two different cathode-ray tubes that used air as gas, are

Tube1	0.57	0.34	0.43	0.32	0.48	0.4	0.4
Tube2	0.53	0.47	0.47	0.51	0.63	0.61	0.48

Which may be regarded as more consistent tube?

[PU 2014 /PU 2006 Spring \ TU 2062/2068/2069]

Solution 15

Please refer exercise-5 Q.No. 27

Scanned by CamScanner

16. **The mean of 100 items was 40 and standard deviation 10. It was later discovered that two items were wrongly read as 30 and 70 instead of 3 and 27. Find the correct mean and correct standard deviation.** **[PU 2011 Spring]**

Solution 16

Please refer solution 14

17. **The average electricity consumption and hence represent the data set of households are as follows:**

	Set-I	Set-II
Mean	65 units	72 Units
Standard Deviation	13 units	18 units
Number Of House Holds	50	40

i. Which set is more variable?

ii. Calculate the coefficient of variation of the combined set? [PU 2012 Fall]

Solution 17

Please refer solution 11

Short Questions

18. **Write short notes on standard deviation.** **[PU 2003 Fall]**

Solution 18

See theoretical part of this chapter

19. **Write short notes on coefficient of variation.** **[PU 2010/2013 Fall]**

Solution 19

See theoretical part of this chapter

20. **Write short note on Box and whisker plot.** **[PU 2015 Fall]**

Solution 20

See theoretical part of this chapter

Scanned by CamScanner

6

Chapter

Mathematical Preliminaries

The fundamental counting rules

The sum rule: Suppose a work W_1 can be done in m_1 ways and other work W can be done in m_2 ways and both cannot occur simultaneously. Then, at least one of them (W_1 or W_2) can be done in $(m_1 + m_2)$ ways

The product rule: Suppose W_1 and W_2 are two independent works. Let W_1 can b done in n_1 ways and W_2 can be done in n_2 ways. Then both W_1 and W_2 can b done in $n_1 \times n_2$ ways.

Factorial notations

i. $n! = n(n-1)(n-2)(n-3) \ldots 3 \cdot 2 \cdot 1$

ii. $5! = 5 \times 4 \times 3 \times 2 \times 1$

iii $n! = (n-1)!$

iv. $0! = 1$

Permutation

The arrangements of set of objects in same order is called permutation.

i. The permutation of a set of n objects taken r at a time with out repetition denoted and defined by

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

ii. The permutation of a set of n objects taken r at a time with repetition denoted and defined by

$$P(n, r) = {}^nP_r = n^r$$

Scanned by CamScanner

Combination

Any selection off one or more things with out any order is called the combination. The combination of n objects taken r at a time is denoted and defined by

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

Exercise 6

1. **A electronic controlling mechanism required five identical memory chips. In how many ways can this mechanism be assembled by placing the 5 chips in the 5 positions within the controller?**

Solution 1

We have total 5 identical memory chips (n) = 5

We have to assemble them in 5 different positions (r) = 5

The total number of way can this mechanism be assembled by placing the 5 chips in the 5 positions in P(5, 5) ways

$$P(5, 5) = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{120}{1} \text{ ways}$$

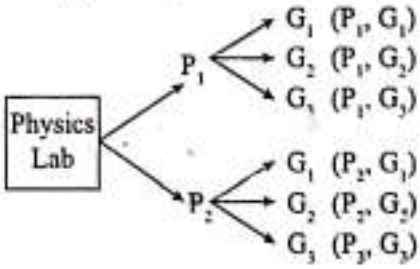
2. **Two professors and three graduate assistants are responsible for the supervision of a physics lab and one professor and one graduate assistant has to be present at all times. Draw a tree diagram showing the points of the corresponding sample space.**

Solution 2

We have two professors say P_1 and P_2.

Also three graduate assistants G_1, G_2 and G_3 (say).

Since, one professor and one graduate assistant has to be present at all times. Then, the sample spaces is given as follows:



$\therefore$ $S = \{(P_1, G_1), (P_1, G_2), (P_1, G_3), (P_2, G_1), (P_2, G_2), (P_2, G_3)\}$

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

Solution 3

The director has to choose 2 chemist among 7. He can do it in C(7, 2) ways

$$= \frac{7!}{(7-2)!\,2!} = \frac{7 \times 6 \times 5!}{5! \times 2} = 21 \text{ ways}$$

Also, the director can select 3 physicist among 9 in (9, 3) ways

Scanned by CamScanner

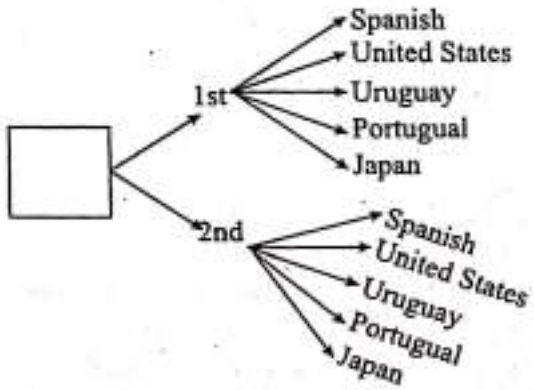
$$= \frac{9!}{(9-3)!\,3!} = \frac{9 \times 8 \times 7 \times 6!}{6! \times 3 \times 2 \times 1} = 3 \times 4 \times 7 = 84 \text{ ways}$$

Hence, the director can choose 2 chemists and 3 physicist among 7 and 9 in (21 × 84) ways = 1764 ways

4. **If the five finalists in an international volleyball tournament are Spanish, United State, Uruguay, Portugal and Japan, draw a tree diagram which shows the various possible first and second place finishers.**

Solution 4

There are two positions 1st position and 2nd position. For the two positions there are fine teams. Spanish, United States, Uruguay, Portugal and Japan. The tree diagram is given as follows:



5. **A carton of 12 rechargeable batteries contains one that is defective. In how many way can an inspector for choose three of the batteries and**

a. get the one that is defective. b. Not get the one that is defective?

Solution 5

Here, total number of batteries is 12 among them 1 is defective.

That is number of non-defective batteries = 11

Number of defective batteries = 1

a. The inspector choose three (3) of them. To get one is defective from the chosen 3 batteries. He have to choose 1 defective out of 1 and 2 non-defective from 11 non-defective,

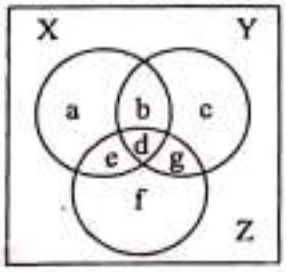
i.e. He can be do this in = C(1, 1) × C(11, 2) = 1 × 55 = 55 ways

b. Among 3 if he cannot get one defective means

Total ways – only one defective

= C(12, 3) – C(11, 2) = 220 – 55 = 165 ways

6. **X, Y and Z are three sets. Write down in set form for each of the letter as event.**



Scanned by CamScanner

Solution 6

$\{a\} = X \cap \overline{Y} \cap \overline{Z}$ $\{b\} = X \cap Y \cap \overline{Z}$ $\{c\} = Y \cap \overline{X} \cap \overline{Z}$

$\{d\} = X \cap Y \cap Z$ $\{e\} = X \cap Z \cap \overline{Y}$ $\{f\} = Z \cap \overline{X} \cap \overline{Y}$

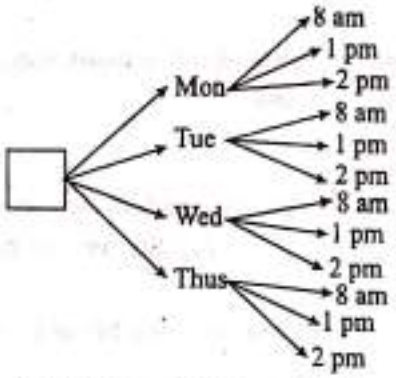
$\{g\} = Y \cap Z \cap \overline{X}$

7. **A building inspector has to check the wiring in a new apartment building either on Monday, Tuesday, Wednesday or Thursday and at 8 a.m., 1 p.m. or 2 p.m. Draw a tree diagram which shows the various ways in which the inspector can schedule the inspection of the wiring of the new apartment building.**

Solution 7

We have four days → Monday, Tuesday, Wednesday, Thursday. He can schedule in three different times → 8 a.m., 1 p.m. or 2 p.m.

The schedule in tree-diagram is given as follows:



8. **How many numbers of 4 different digits can be formed with digits 1, 2, 3, 4?**

Solution 8

Here, we have four different numbers (n) = 4

Also, we have to make 4-digit numbers (r) = 4

So, it can be done in P(4, 4) ways $= \frac{4!}{(4-4)!} = \frac{4!}{0!} = 24$

Hence, we can make 24 different numbers

9. **A man is going from A to C through B. He has 4 options to reach B and 3 option to C from B. How many possibilities are there in reaching C from A.**

Solution 9

We have given a man is going from A to C through B. It means, he has to go from A to B and then B to C. He can go from A to B in 4 different ways and B to C in 3 different ways.

∴ He can go A to C in $4 \times 3 = 12$ different ways

10. **How many three letters words can we make using the letters "Number' allowing no repetition? Also, if repetition is allowed how many words can be formed? How many three words are there with one or more repeated letters?**

Solution 10

i. We have given words is NUMBER which has 6 different letters. The 6 different letters can be arranged in 3 places in P(6,3) ways

Scanned by CamScanner

$$= \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \times 5 \times 4 \times 3!}{3!} = 120$$

ii. If the repetition is allowed then 6 different letters can be arranged in 3 places in 6^3 = 216 words.

iii. Total numbers of 3 letters words with one or more repeated letters

= one repeated + two repeated

$= 6 \times 6 \times 5 + 6 \times 6 \times 6 = 180 + 216 = 432$ words

11. From 4 engineers, 2 physicists and 3 economists, find the number of committees that can be formed consisting of 2 engineers, 1 physicist and 1 economist.

Solution 11

The total number of committees consisting 2 engineers 1 physicist and 1 economist out of 4 engineers, 2 physicist and 3 economists.

$= C(4, 2) \times C(2, 1) \times C(3, 1) \quad 6 \times 2 \times 3 = 36$ committees

12. There are six doors in a hostel. In how many ways can a student enter the hostel and leave by a different door.

Solution 12

There are 6 doors in a hostel. So a student can enter the hostel in 6 different ways. But he has to leave the hostel in the different doors. So, he has 5 different ways to leave the hostel.

Hence, a student can enter and leave the hostel by a different door in (6×5) ways

= 30 ways

Solution 13

The 3 person can sit in 7 vacant chairs in P(7, 3) ways

$$= \frac{7!}{(7-3)!} = \frac{7 \times 6 \times 5 \times 4!}{4!} = 210 \text{ ways}$$

14. How many different words can be formed from the letters of the word 'BUTWALY' when:

a. all the letters are taken,

b. words begin with B

c. the letters B and Y respectively occupy the first and last places

d. the vowels are always together

e. the letters U, T, W are never together

Solution 14

In the word BUTWALY there are 7 different letters.

a. If all the letters are taken then the total number of words can be made with 7 different letters and 7 positions in P(7, 7) ways = 7! = 5040 words

b. For the words begin with B, there is only 1 choice for the beginning letter and the remaining 6 words with 6 places in P(6, 6) ways.

∴ Total number of words = 1 × P(6, 6) = 6! = 720 words

Scanned by CamScanner

c. For the words begin with B and end with Y there are 1/1 choices for the beginning and end letters. Also, the remaining 5 letters can be arranged in P(5,5) ways.

Total number of words = 1 × P(5, 5) × 1 = 120 words

d. The word having vowels always together. Suppose, the two vowels U and A as one (single word). Then, there are six different letters they can be arranged in 6 places in P(6,6) ways.

Also, the two vowels can be interchanged theirs position in 2! ways.

∴ Therefore total number of words = 2! × 6! = 2 × 720 = 1440

e. Suppose the word U, T, W are always together so suppose them as one, we have 5 different letters which can be arranged in 5 places in P(5,5) ways. Also, the letters U, T, W can be interchange their position in 3! ways.

Therefore, total number of words = 3! × P(5, 5) = 3! × 5! = 720

None, U, T, W never come together = Total – U, T, W always come together

= 5040 – 720 = 4320

15. A shipment of 50 transistors contains 40 that are perfect and 10 defectives. In how many ways can 5 of these transistors be selected so that 3 are perfect and 2 are defectives?

Solution 15

We have total 50 transistors among them.

No. of perfect transistors = 40

No. of defective transistors = 10

The selection of 5, among them 3 are perfect and 2 are defective. This can be done in the following ways. 3 perfect transistors can be selected from 40 in C(40, 3) ways.

Also, 2 defective transistors can be selected from 10 in C(10, 2) ways

Total number of ways of 3 perfect and 2 defective transistors can be selected is

C(40, 3) × C(10, 2) ways = 9880 × 45 = 444600 ways

16. How many different ways can 3 red, 4 yellow and 2 blue balls be arranged in string of Christmas tree light with nine sockets?

Solution 16

There are total nine sockets in a string of Christmas tree.

∴ n = 3 + 4 + 2 = 9

Then, we know that 3 red, 4 yellow and 2 blue balls can be arranged in a string of Christmas tree tight with nine sockets = $\frac{96}{3!\,4!\,2!}$ = 1260 ways

17. A box contains two different white balls, 3 different black balls and 4 different red balls. In how many ways can 3 balls be drawn from the box if at least one black ball is to be included in the draw?

Solution 17

The selection of 3 balls out of 2 white, 3 black and 4 red balls including at least one black ball can be done in the following way:

Scanned by CamScanner

Black (3)	White (2)	Red (4)	Total no. of ways	Total no. of way
1	2	0	C(3, 1) × C(2, 2) × C(4, 0)	= 3 × 1 × 1 = 3
1	0	2	C(3, 1) × C(2, 0) × C(4, 2)	= 3 × 1 × 6 = 18
1	1	1	C(3, 1) × C(2, 1) × C(4, 1)	= 3 × 2 × 4 = 24
2	1	0	C(3, 2) × C(2, 1) × C(4, 0)	= 3 × 2 × 1 = 6
2	0	1	C(3, 2) × C(2, 0) × C(4, 1)	= 3 × 1 × 4 = 12
3	0	0	C(3, 3) × C(2, 0) × C(4, 0)	= 1 × 1 × 1 = 1
			Total no. of ways	64 ways

18. From 10 persons in how many ways can a selection of 4 be made:

a. When one particular person is always included?

b. When 2 particular persons are always excluded?

Solution 18

There are 10 persons and we have to select 4 of them.

a. If one particular person is always included then we have to select 3 persons out of 9 in C(9, 3) ways = 84 ways

b. If two particular persons are always excluded then we have to select 4 out of 8 in C(8, 4) ways = 70 ways

19. From 6 gentlemen and 4 ladies, a committee of 5 is to be formed. In how many ways can this be done so as to include at least one lady?

Solution 19

The committee of 5 is to be formed from 6 gentlemen and 4 ladies containing at least one lady, which can be done in the following ways:

Ladies (4)	Gentlemen (6)	No. of ways
1	4	C(4, 1) × C(6, 4) = 4 × 15 = 60
2	3	C(4, 2) × C(6, 3) = 6 × 20 = 120
3	2	C(4, 3) × C(6, 2) = 4 × 15 = 60
4	1	C(4, 4) × C(6, 1) = 1 × 6 = 6
Total no. of ways		246

Hence, we can do it in 246 ways.

20. In how many ways can 10 balls selected out of 20 balls.

Solution 20

We can select 10 balls out of 20 balls in C(20, 10) ways = 184756 ways

21. A bag contains 5 black balls and 3 white balls, in how many ways can:

i. 3 balls be drawn?

ii. 3 black balls be drawn?

iii. 2 black and 2 white balls be drawn?

iv. 2 black or 2 white balls be drawn?

Solution 21

There are 5 black balls and 3 white balls in a bag.

i. If we have to select 3 balls, then we can do it in C(8, 3) ways = 56 ways

ii. If we have to select 3 black balls, then we have to select 3 black balls out of 5 black balls in C(5, 3) ways = 10 ways

iii. If we have to select 2 black and 2 white balls then we can do it in the following ways:

2 black balls can be drawn from 5 in C(5, 2) = 10 ways

2 white balls can be drawn from 3 in C(3, 2) = 3 ways

∴ Total number of ways = 10 × 3 = 30 ways

iv. 2 black or 2 white balls can be drawn in 10 + 3 = 13 ways

Scanned by CamScanner

7

Chapter

Theory of Probability

Basic Terminology in Probability

Random experiment: An experiment is said to be random experiment if outcomes of experiment cannot be predicted with certainty but may be any one of the various possible outcomes.

Sample space: A set of all possible outcomes for a random experiment is called sample space and it is denoted by the symbol S.

Trial and event: A single performance which is obtained from a random experiment is called a trial. The subset of sample space S, which contains specific collection of sample points is called event.

Types of events

i. **Simple and compound events:** An event is said to be simple event if it corresponds to one and only one possible outcomes. Also an event is said to be compound events if it consists of more than one outcomes.

ii. **Mutually exclusive events:** Two or more events are said to be mutually exclusive events if the happening of any one of them excludes the happening of all other in the same experiment.

iii. **Equally likely events:** Two or more events are said to be equally likely if all the events have equal chance of occurrence.

iv. **Independent events:** Events are said to be independent of each other, if the occurrence of any one of them does not affect and is not affected by the occurrence of non-occurrence of the other.

v. **Dependent events:** Events are said to be dependent if the occurrence of one event affects and is affected by the occurrence or non-occurrence of the other.

Scanned by CamScanner

Basis Concept of Probability

The numerical evaluation of chance factor of an event will occur or not is known as probability. If we denote the event is E then the probability that the event E will occur is denoted by P(E).

Glossary of probability terms

Let A and B are two events then

$P(A \text{ or } B) = P(A \cup B)$	:	Probability that at least one of the events A and B will occur
$P(A \text{ and } B) = P(A \cap B)$	:	Probability that both events A and B will occur
$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B})$	:	Probability that neither A nor B will occur
$P(A \cap \bar{B})$	:	Probability that only A will occur
$P(\bar{A} \cap B)$	:	Probability that only B will occur
$P[(A \cap \bar{B}) \cup (\bar{A} \cap B)]$	:	Exactly one of the events A or B will occur

Various Approach to Probability

Mathematical or a priori or classical approach to probability

If there are 'n' equally likely, mutually exclusive and exhaustive outcomes of an experiment and m of them are favourable to the occurrence of an event E, then the probability of occurrence of an event E is denoted by P(E) and defined by

$$P(E) = \frac{\text{Favourable number of cases}}{\text{Total number of cases}} = \frac{m}{n}$$

Remarks

i. Probability of any event is a number lying between 0 and 1 inclusive i.e. $0 \le P(E) \le 1$.

ii. Probability of non-occurrence off an event E is denoted and defined by $P(\bar{E}) = 1 - P(E)$.

iii. $P(E) + P(\bar{E}) = p + q = 1$

iv. If $P(E) = 1$, E is a sure event and if $P(E) = 0$ then E is an impossible event.

Statistical or Empirical Approach

If an event 'E' occur m times in n repetitions under the similar homogeneous conditions of a random experiment, then the value approached by the relative frequency $\frac{m}{n}$ when n becomes infinitely large is said to be the probability of the event E.

i.e. $P(E) = \lim_{n \to \infty} \frac{m}{n}$, provided that the limit exists

Modern Approach of Probability

Given a sample space S of a random experiment, the probability of the occurrence of any event E is defined as a set function P(E) satisfying the following axioms

i. Axiom of non-negativity : i.e. $P(E) \ge 0$

Scanned by CamScanner

ii. Axiom of certainty : i.e. $P(S) = 1$

iii. Axiom of additive : If $E_1, E_2, E_3, \ldots, E_n$ is any finite or infinite sequence of disjoint events of S then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Theorems on Probability

Theorem 1: If ϕ is the impossible event and S is sure event that $P(\phi)$ 0 and $P(S) = 1$.

Theorem 2 : The probability of the complementary event $\bar{A}$ of A is given by

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

Theorem 3 : For any two events A and B defined in the sample space S.

i. $P(\bar{A} \cap B) = P(\text{only } B) = P(B) - P(A \cap B)$

ii. $P(A \cap \bar{B}) = P(\text{only } A) = P(A) - P(A \cap B)$

Theorem 4: If $B \subset A$ then $P(A \cap \bar{B}) = P(A) - P(B)$

Additional Theorem of Probability [Total Probability] [PU 2007 Fall/2009 Spring]

Statement: If A and B are any two events defined in sample space S and are not disjoint then $P(A \cup B) = P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$

Proof: Let S be the sample space and n(S) be the number of elements in the sample space. Then by the definition of probability we have

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} \qquad \ldots(i)$$

Since, A and B are only two events defined in S and are not disjoint. So, by the definition: Probabilities of the events A, B and $(A \cap B)$ are non-zero and given by

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

Now, from equation (i), we have,

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A) + n(B) - n(A \cap B)}{n(S)}$$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

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

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

Remarks

i. The probability of occurrence of at least one of the event A or B is denoted and defined by $P(A \text{ or } B) = P(\cup B) = P(A) + P(B) - P(A \cap B)$

ii. The probability of occurrence of at least one of the events A or B or C is denoted and defined by

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

iii. If the events A and B are mutually exclusive then:

a. $P(A \cup B) = P(A) + P(B)$ $\quad [\because P(A \cap B) = 0]$

b. $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

Scanned by CamScanner

Conditional Probability

If A and B are any two events in a sample space S and $P(B) \neq 0$, then the conditional probability of A given that B has already occurred is denoted by $P(A/B)$ and is defined by $P(A/B) = \frac{P(A \cap B)}{P(A)}$, $P(A) \neq 0$

Where, $P(A \cap B)$ is the joint probability of A and B.

Similarly, the conditional probability of B given that A has already occurred is denoted and defined by, $P(B/A) = \frac{P(A \cap B)}{P(A)}$, $P(A) \neq 0$

Note:

i. Conditional probability is used if two or more events are dependent.

ii. $P(A/B) \geq 0$ for all events A

iii. $P(B/B) = 1$

Product rule of Probability [Theory of Compound Probability]

Statement: For any two events A and B

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

Where, $P(A/B)$ and $P(B/A)$ are conditional probabilities of A and B when B and A has already occurred.

Proof: Let S be the sample space $n(A)$, $n(B)$, $n(A \cap B)$ and $n(S)$ denote the number of elements in A, B $A \cap B$ and S respectively. Then, by the definition of probability, we have,

$$P(A) = \frac{n(A)}{n(S)},\; P(B) = \frac{n(B)}{n(S)} \text{ and } P(A \cap B) = \frac{n(A \cap B)}{n(S)} \quad \text{...(i)}$$

Also, by the definition of conditional probability we have,

$$P(B/A) = \frac{P(A \cap B)}{n(A)} \quad \text{...(ii)}$$

Now, equation (i) reduces to

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

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

Hence, the theorem.

Note: If two events A and B are said to be independent if $P(A \cap B) = P(A) \times P(B)$

Baye's Theorem (Rule for the Inverse Probability)

Statement: If $E_1, E_2, E_3, \ldots, E_n$ be mutually disjoint events of sample space S with $P(E_i) \neq 0$, $i = 1, 2, \ldots, n$ then for any arbitrary event A which is subset of S such that $P(A) > 0$, we have

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

Scanned by CamScanner

Proof: Since A is subset of S and $S = E_1 \cup, E_2 \cup, E3 \cup, \ldots, \cup E_n$. Then, we have,

$$A = A \cap S$$
$$= A \cap [E_1 U, E_2 U, E_3 U, \ldots, U E_n]$$
$$= (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \ldots \cup (A \cap E_n)$$

Since, $E_1, E_2, E_3, \ldots, E_n$ are mutually disjoint events. So, $A \cap E_1, A \cap E_2, A \cap E_3, \ldots, A \cap E_n$ are also mutually disjoint.

$$\therefore \quad P(A) = P[(A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \cup \ldots \cup (A \cap E_n)]$$
$$= P(A \cap E_1) + P(A \cap E_2) + P(A \cap E_3) + \ldots + P(A \cap E_n)$$
$$= \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i) P(A/E_i) \qquad [\because P(A \cap B) = P(B) P(A/B)]$$

Also, for any particular compound event $(A \cap E_i)$

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

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

Exercise 7.1

1.i. Discuss the meaning of probability with suitable examples.

Solution 1

i. There are a number of events in our day to day life about which one is sure to happen and some may or may not happen. But one is always curious to know what chance is there for a happening or event to occur. For instance, one may be interested to estimate whether it will rain today or not, one would live to evaluate his chance of wining for head in a definite number of tosses. Thus, the numerical evaluation of chance factor of an event is known as probability.

1.ii. Explain the mathematical and statistical definition of probabilities. Write down the limitations of these probabilities.

ii. **Mathematical (or Classical) Definition of Probability**

If there are n equally likely, mutually exclusive and exhaustive cases, of an experiment and m of them are favorable to the occurrence of an event either the probability of occurrence or happening of an event E denoted by P(E) or p is defined as

$$P(E) = \frac{\text{Favourable number of cases to E}}{\text{Exhaustive (or total) number of cases}} = \frac{M}{n}$$

This is also known as a priori probability.

Limitations:

In the following case mathematical definition can not be used.

i. The exhaustive (or total) number of outcomes 'n' is indefinitely large, we cannot use this definition.

ii. If each and every outcomes is not equally likely.

Scanned by CamScanner

iii. If the outcomes are not independent.

iv. It is not possible to enumerate all possible outcomes

Statistical Definition of Probability

If an event E occurs m times in n repetitions under the similar and homogeneous conditions of a random experiment, then the value approached by the relative frequency $\frac{m}{n}$ when n becomes infinitely large is said to be the probability of the event E.

Symbolically, $P(E) = \lim_{n \to \infty} \frac{m}{n}$ provided that the limit exists.

- **Limitations**

i. The experimental condition may not remain identical and homogeneous when a random experiment is repeated a large number of times.

ii. The relative frequency $\frac{m}{n}$ may not attain a unique value when $n \to \infty$.

2. Write down the differences with examples.

Solution 2

i. Sample space and sample point:

Sample Space	Sample Point
i. A set of all possible outcomes for a random experiment is called space.	ii. Each element of sample space is called the sample point.
ii. Sample space is a set.	ii. Sample point is a single element of sample space.
iii. In throwing a dice, sample space S = {1, 2, 3, 4, 5, 6}	iii. In a throwing a dice occurring 1 or 2 or 3 or 4 or 5 or 6 is sampling.

ii. Dependent events and Independent events

Dependent Events	Independent Events
i. If the occurrence of one event affects the occurrence or non-occurrence of the other is called dependent events.	i. If the occurrence of one event does not affect the occurrence or non-occurrence of the other is called independent events.
ii. It is affected by the other event.	ii. It is not affected by the other event.

iii. Mutually Exclusive Events and Disjoint Events

Mutually exclusive events	Disjoint events
i. Two or more events are said to be mutually exclusive events if the happening of any one of them excludes the happening of all others.	i. If there is no elements common among the elements then such events are called disjoint events.
ii. In a tossing of a coin the occurrence of head excludes the occurrence of tail.	ii. In a tossing of a coin the head and tail are also disjoint events.

3. Write down the modern approach to probability.

Solution 3

Modern Approach of Probability

Let S be a sample space of random experiment and E be any event in S then P is probability function which is defined in S and the function P(E) satisfying the axioms:

Scanned by CamScanner

I : P(E) is defined, real and non-negative i.e. $P(E) \geq 0$

II : $P(S) = 1$, the total probability

III : If $A_1, A_2, \ldots, A_n$ is any finite or infinite sequences of disjoint events of S, then

$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

4. State and prove the additional and multiplication theorem of probability.

Solution 4

Please refer theoretical part of this chapter

5. State and prove the Baye's theorem.

Solution 5

Please refer theoretical part of this chapter

6. What is the probability of drawing a queen from a pack of card?

Solution 6

Let S denote the pack of card and Q denote the queen.

There are 52 card in a deck of card i.e. $N(S) = 52$

There are 4 queens in a pack of card i.e. $n(Q) = 4$

$\therefore \quad P(\text{Queen}) = \frac{n(Q)}{n(S)} = \frac{4}{52} = \frac{1}{13} = 0.769$

7. There are 5 white and 3 black balls in an urn. What is the probability of getting a black ball in drawing a single ball from the bag?

Solution 7

Let W denote the white and B denote the black balls respectively.

There are 5 white balls i.e. $n(W) = 5$ and

There are 3 black balls i.e. $n(B) = 3$

Total balls in the bag i.e. $n(S) = 5 + 3 = 8$

$P(\text{Black ball}) = P(B) = \frac{n(B)}{n(S)} = \frac{3}{8} = 0.375$

8. Three unbiased coins are tosses simultaneously. Find the probability of getting.

a. exactly all tails

b. at least one tail

c. only one tail

d. at most one tails

e. at least one head

Solution 8

Let S denote the sample space. H denotes head and T denote tail of a coin.

Since, three unbiased coins are tossed then

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

$n(S) = 2 \times 2 \times 2 = 8$

a. $P(\text{Exactly all tails}) = \frac{n(\text{All tails})}{n(S)} = \frac{1}{8} = 0.125$

b. $P(\text{at least one tails}) = \frac{n(\text{at least one tail})}{n(S)} = \frac{7}{8} = 0.875$

Scanned by CamScanner

c. $P(\text{only one tail}) = \frac{n(\text{only one tail})}{n(S)} = \frac{3}{8} = 0.375$

d. $P(\text{at most one tail}) = \frac{n(\text{at most one tail})}{n(S)} = \frac{4}{8} = \frac{1}{2} = 0.5$

e. $P(\text{at least one head}) = \frac{n(\text{at least one head})}{n(S)} = \frac{7}{8} = 0.875$

9. In a single throw of two dices, find the probability that the sum of two faces is

a. 7 or 11 b. 4 exactly c. less then 5

Solution 9

Let S denote the sample space. Since, two dices are thrown at them.

S= {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

$\therefore$ $n(S) = 6 \times 6 = 36$

a. Sum of two faces is 7 or 11.

i.e. $P(7 \text{ or } 11) = P(7) + P(11) - P(7 \text{ and } 11)$

$$= \frac{n(7)}{n(S)} + \frac{n(11)}{n(S)} - 0 = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9} = 0.222$$

b. $P(\text{Sum exactly } 4) = \frac{n(4)}{n(S)} = \frac{3}{36} = \frac{1}{12} = 0.086$

c. $P(\text{Sum less than } 5) = \frac{n(\text{sum less than } 5)}{n(S)} = \frac{6}{36} = \frac{1}{6} = 0.166$

10. An urn contains four white, eight black, six red and two green balls. If there balls are drawn at random. Determine the probability that

a. all white balls b. all red balls

c. 1 white, 1 black and 1 green ball

Solution 10

Let S denote the sample space, W, B, R, G denote the white, black, red and green balls respectively. Then, n(W) = 4, n(B) = 8, n(R) = 6, n(G) = 2

a. $P(\text{All white balls}) = P(3W) = \frac{n(3W)}{n(S)} = \frac{{}^4C_3}{{}^{20}C_3} = \frac{4}{1140} = \frac{1}{285} = 0.0035$

b. $P(3R) = \frac{n(3R)}{n(S)} = \frac{{}^6C_3}{{}^{20}C_3} = \frac{20}{1140} = \frac{1}{57} = 0.0175$

c. $P(1W, 1B, 1G) = \frac{n(1W, 1B, 1G)}{n(S)} = \frac{{}^4C_1 \times {}^8C_1 \times {}^2C_1}{{}^{20}C_3} = \frac{64}{1140} = 161285 = 0.0561$

11. Two balls are drawn in succession from a bag containing 6 white and 2 black balls. What is the probability that both balls are of same color.

Solution 11

Let S denote the sample space, W and B denote the white and black balls respectively. Then,

n(W) = 6 and n(B) = 2

Scanned by CamScanner

P(both same colour) $= P(2W) + P(2B)$

$$= \frac{n(2W)}{n(S)} + \frac{n(2B)}{n(S)} = \frac{{}^6C_2}{{}^8C_2} + \frac{{}^2C_2}{{}^8C_2} = \frac{15}{28} ++ \frac{1}{28} = \frac{4}{7} = 0.571$$

12. A bag contains 3 red, 4 black and 2 white balls. What is the probability of drawing a red and white ball if two balls are drawn one after other with replacement?

Solution 12

Let S denote the sample space and R, B, W denote the red, black and white balls respectively. Then,

$n(R) = 3, n(B) = 4, n(W) = 2, P(R \text{ and } W) = ?$

We know that,

$$P(R \text{ and } W) = P(R) \times P(W) = \frac{n(R)}{n(S)} \times \frac{n(W)}{n(S)} = \frac{3}{9} \times \frac{2}{9} = \frac{2}{27} = 0.074$$

13. Four cards are drawn from a well shuffled pack of 52 cards. Find the probability that

a. all are diamond b. there is one card of each suit and

c. there are two spades and two hearts.

Solution 13

Let S denote the sample spaces and D, S, H denote the diamond, space and hearts respectively.

a. $P(\text{All diamond}) = \frac{n(\text{All diamond})}{n(S)} = \frac{{}^{13}C_4}{{}^{52}C_4} = \frac{715}{270725} = 0.0026$

b. $P(\text{One card of each}) = \frac{n(\text{one card of each})}{n(S)} = \frac{{}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C1}{{}^{52}C_4} = \frac{28561}{270725} = 0.105$

c. $P(2S \text{ and } 2H) = P(2S) \times P(2H) = \frac{n(2S)}{n(S)} \times \frac{n(2H)}{n(S)} = \frac{{}^{13}C_2}{{}^{52}C_4} \times \frac{{}^{13}C_2}{{}^{52}C_4} = \frac{78 \times 78}{270725} = 0.0224$

14. The following is the frequency distribution of annual sales commission from a survey of 300 media sale persons.

Annual Commission (Rs)	frequency
0 – 4999	15
5000 – 9999	25
10000 – 14999	35
15000 – 19999	125
20000 – 24999	70
25000 or above	30

Based on this information, what is the probability that media sale persons make a commission.

a. between 4999 and 1000? (exclude) b. less than 15000?

c. 2000 or more? d. between 14999 and 20000 (exclude)

Solution 14

a. $P(\text{Between 4999 and 10,000}) = \frac{n(\text{Between 4999 and 10,000})}{n(f)} = \frac{25}{300} = 0.0833$

Scanned by CamScanner

b. $P(<15000) = \frac{n(<15,000)}{n(s)} = \frac{15+25+35}{300} = 0.25$

c. $P(2000 \text{ or more}) = \frac{n(2000 \text{ or more})}{n(s)} = \frac{70+30}{300} = 0.333$

d. $P(\text{Between } 14999 \text{ and } 20000) = \frac{n(\text{Between } 14999 \text{ and } 20000)}{n(S)} = \frac{125}{300} = 0.417$

15. Let A and B be two possible outcomes of an experiment and suppose.

$P(A) = 0.4$, $P(A \cup B) = 0.7$ and $P(B) = x$

a. For what choice of x, A and B are mutually exclusive

b. For what choice of x, A and B are independent.

Solution 15

We have given $P(A) = 0.4$, $P(A \cup B) = 0.7$, $P(B) = x$

a. Since A and B are mutually exclusive then

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

or, $0.7 = 0.4 + x$

or, $x = 0.7 - 0.4 = 0.3$

b. Since A and B are independent events then

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

ot, $0.7 = 0.4 + x - 0.4x$

or, $0.7 - 0.4 = (1 - 0.4)x$

$\therefore$ $x = \frac{0.3}{0.6} = 0.5$

16. A market research organization claims that among 100 house wives interviewed 62 regularly but product X, 53 regularly by product Y, 18 regularly buys both of these products and 13 do not buy these products at all. Should the results of this survey be questioned?

Solution 16

Here, given $n(S) = 100$, $n(X) = 62$, $n(Y) = 53$, $n(X \cap Y) = 18$ and $n(\overline{X \cup Y}) = 13$

We know that,

$n(U) = n(X) + n(Y) - N(X \cap Y) + n(\overline{X \cup Y})$

or, $100 = 62 + 53 - 18 + 13$

or, $100 = 110$

which is impossible.

17. A construction company is bidding for two contracts 'A' and 'B'. The probability that the company will get construct A is $\frac{3}{5}$, the probability that the company will get construct B is $\frac{1}{3}$ and the probability that the company will get both the construct is $\frac{1}{8}$. What is the probability that the company will get construct A or B?

Scanned by CamScanner

Solution 17

We have given,

$$p(A) = \frac{3}{5}, \quad P(B) = \frac{1}{3}, \quad P(A \text{ and } B) = \frac{1}{8} \qquad P(A \text{ or } B) = ?$$

We know that,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = \frac{3}{5} + \frac{1}{3} - \frac{1}{8} = \frac{97}{120} = 0.81$$

18. There are three radio stations, A, B and C which can receive in a city of 2000 families. The following information is available on the basis of a surrey.

1200 families listen to radio station A.

1100 families listen to radio station B.

800 families listen to radio station C.

865 families listen to radio station A and B.

450 families listen to radio station A and C.

400 families listen to radio station B and C.

100 families listen to radio station A, B and C.

Find the probability who listen at least one radio stations.

Solution 18

Here, given $N(S) = 2000$, $n(A) = 1200$, $n(B) = 1100$, $n(C) = 800$

$n(A \cap B) = 865$, $n(B \cap C) = 400$, $n(A \cap C) = 450$ and $n(A \cap B \cap C) = 100$ and $P(A \cup B \cup C) = ?$

We know that,

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

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} + \frac{n(C)}{n(S)} - \frac{n(A \cap B)}{n(S)} - \frac{n(B \cap C)}{n(S)} - \frac{n(A \cap C)}{n(S)} + \frac{n(A \cap B \cap C)}{n(S)}$$

$$= \frac{1200}{2000} + \frac{1100}{2000} + \frac{800}{2000} - \frac{865}{2000} - \frac{400}{2000} - \frac{450}{2000} + \frac{100}{2000} = \frac{1485}{2000} = 0.7425$$

19. The odds against a certain event 5 to 2 and odds in favor of another independent event are 6 to 5. Find the chance that at least one of the events will happen.

Solution 19

Here, we have,

$$P(A) = \frac{2}{5+2} = \frac{2}{7}, \qquad P(B) = \frac{6}{6+5} = \frac{6}{11}, \qquad P(A \cup B) = ?$$

Since A and B are two independent events. Then,

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

$$= \frac{2}{7} + \frac{6}{11} - \frac{2}{7} \times \frac{6}{11} = \frac{22+42}{77} - \frac{12}{77} = \frac{52}{77} = 0.675$$

20. A box contains six red, four white and five black balls. A person draws four balls from the box at random. Find the probability that among the balls drawn there is at least one ball of each color.

Solution 20

Scanned by CamScanner

Let, R, W, B denote the red, white and black balls respectively.

n(R) = 6, n(W) = 4 and n(B) 5, P(R or W or B) = ?

We know that,

$$P(\text{at least of one ball of each colour} = \frac{n(\text{at least one ball of each colour})}{n(S)}$$

$$\frac{720}{^{15}C_4} = \frac{720}{1365} = 0.527$$

Where, at least one ball of same colour is given by the following table.

Red (6)	White (4)	Black (5)	Total no. of ways
1	1	2	$^6C_1 \times {}^4C_1 \times {}^5C_2 = 6 \times 4 \times 10 = 240$
1	2	1	$^6C_1 \times {}^4C_2 \times {}^5C_1 = 6 \times 6 \times 5 = 180$
2	1	1	$^6C_2 \times {}^4C_1 \times {}^5C_1 = 15 \times 4 \times 5 = 300$
		Total	720

21. The chances of winning of two race horses are $\frac{1}{3}$ and $\frac{1}{6}$ respectively. What are the probabilities that at least one will win when the horses are running.

a. in different races and

b. in same race?

Solution 21

Let the two horses be A and B. Then,

$P(A) = \frac{1}{3}$, $P(B) = \frac{1}{6}$, $P(A \text{ or } B) = ?$

a. If the horse are running in two different races. Then, the events will independent (events). Thus,

$$P(A \text{ or } B) = P(A) + P(B) - P(A) \times P(B) = \frac{1}{3} + \frac{1}{6} - \frac{1}{3} \times \frac{1}{6} = \frac{9-1}{18} = \frac{8}{18} = 0.44$$

b. If the horses are running in the same race then A and B are two mutua exclusive (disjoint) events. Then,

$$P(A \text{ or } B) = P(A) + P(B) = \frac{1}{3} + \frac{1}{6} = \frac{2+1}{6} = \frac{1}{2} = 0.5$$

22. A card is drawn from a bag containing 10 cards numbered 1 through 10. Let denotes the event that the number on card drawn is less then 7. B is the event that the number on the card drawn is a multiplier of 2. Find the probabilities the events.

a. A b. B c. $A \cap B$ d. $A \cup B$

e. $\bar{A} \cap \bar{B}$ f. A/B g. B/A

Solution 22

Here, we have given,

n(S) = 10, n(A) = 6, (1, 2, 3, 4, 5, 6 < 7), n(B) = 5, (2, 4, 6, 8, 10)

a. $P(A) = \frac{n(A)}{n(S)} = \frac{6}{10} = 0.6$

Scanned by CamScanner

b. $P(B) = \frac{n(B)}{n(S)} = \frac{5}{10} = 0.5$

c. $P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{3}{10} = 0.3$

d. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.5 - 0.3 = 0.8$

e. $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$

f. $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.5} = 0.6$

g. $P(B/A) = \frac{0.3}{0.6} = 0.5$

23. Three newspaper A, B and C are published in a certain city. It is estimated from a survey that 20% read A, 16% read B, 14% read C, 8% read A and B, 5% read A and C, 4% read B and C and 2% read all the three newspaper. What are the probabilities that a normally chosen person

a. does not read any paper
b. does not read C
c. read A but not B
d. read only one of these papers.

Solution 23

We have given,

$P(A) = 20\% = \frac{20}{100} = 0.2$ $\quad P(B) = 16\% = \frac{16}{100} = 0.16$

$P(C) = 14\% = \frac{14}{100} = 0.14$ $\quad P(A \cap B) = \frac{8}{100} = 0.08$

$P(A \cap C) = 5\% = 0.05$ $\quad P(B \cap C) = 4\% = 0.04$

$P(A \cap B \cap C) = 2\% = 0.02$

a. P(does not read any newspaper)

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

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

$= 1 - (0.2 + 0.16 + 0.14 - 0.08 - 0.04 - 0.05 + 0.02) = 1 - 0.35 = 0.65$

b. P(does not read) $= P(\bar{C}) = 1 - P(C) = 1 - 0.14 = 0.86$

c. P(Read A but not B) $= P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.2 - 0.08 = 0.12$

d. P(Only one of these papers) $= \frac{n(\text{only one of thses papers})}{n(S)}$

$= \frac{n_0(A) + n_0(B) + n_0(C)}{100} = \frac{9 + 6 + 7}{100} = \frac{22}{100} = 0.22\%$

24. An urn contains 12 balls of which 4 are white, five balls are drawn at random and laid a side, their color unnoted. Another ball is drawn, what is the probability that it is white.

Solution 24

Let us define the events as follows:

E_0 : selection of 0 white and 5 other balls

E_1 : selection of 1 white and 4 other balls

Scanned by CamScanner

E_2 : selection of 2 white and 3 other balls

E_3 : selection of 3 white and 2 other balls

E_4 : selection of 4 white and 1 other ball

Then,

$$P(E_0) = \frac{{}^4C_0 \times {}^8C_5}{{}^{12}C_5} = \frac{7}{99}, \qquad P(E_1) = \frac{{}^4C_1 \times {}^8C_4}{{}^{12}C_5} = \frac{35}{99}$$

$$P(E_2) = \frac{{}^4C_2 \times {}^8C_3}{{}^{12}C_5} = \frac{42}{99}, \qquad P(E_3) = \frac{{}^4C_3 \times {}^8C_2}{{}^{12}C_5} = \frac{14}{99}$$

and $P(E_4) = \frac{{}^4C_4 \times {}^8C_1}{{}^{12}C_5} = \frac{1}{99}$

Let A denote the event selecting a white ball after 5 ball drawn. Then,

$$P(A/E_0) = \frac{4}{7}, P(A/E_1) = \frac{3}{7}, P(A/E_2) = \frac{2}{7}, P(A/E_3) = \frac{1}{7}, P(A/E_0) = 0$$

Then, the probability of draw a white ball,

$$P(A) = P(E_0) \cdot P(A/E_0) + P(E_1) \cdot P(A/E_1) + P(E_2)\, P(A/E_2) + P(E_3) \cdot P(A/E_3) + P(E_4) \cdot P(A/E_4)$$

$$= \frac{7}{99} \times \frac{4}{7} + \frac{35}{99} \times \frac{3}{7} + \frac{42}{99} \times \frac{2}{7} + \frac{14}{99} \times \frac{1}{7} + \frac{1}{99} \times \frac{0}{7} = \frac{28 + 105 + 84 + 14}{99 \times 7} = \frac{231}{693} = \frac{1}{3}$$

25. Suppose a wildcatter assigns a prior probability of 0.6 for gas. A seismic survey might be used that augurs favorable in 85% of applications where gas was known to be present and unfavorably in 90% of sites where no gas was found. Find the posterior probability of gas.

a. Given a favorable seismic result.

b. Given an unfavorable seismic result.

Solution 25

Do yourself

26. Of 100 patient in a hospital with a certain disease 10 are choose to under go a drug treatment that increase the percentage care rate from 50% in the untreated to 75%. If doctor latter encounters cured patient, what is the probability that he received the drug treatment.

Solution 26

Do yourself

27. The contents of urns I, II and III as follows:

1 white, 2 black and 3 red balls.

2 white, 1 black and 1 red balls.

4 white, 5 black and 3 red balls.

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they came from urn II.

Solution 27

Let E_1, E_2, E_3 denote the events that the urn I, II and III is chosen respectively.

Scanned by CamScanner

Then $P(E_1) = \frac{1}{3}$, $P(E_2) = \frac{1}{3}$, $P(E_3) = \frac{1}{3}$

Also, let A be the events that the two balls taken from the selected urn are white and red. Then,

$$P(A/E_1) = \frac{{}^3C_1 \times {}^3C_1}{{}^6C_2} = \frac{3}{15} = \frac{1}{5} \qquad P(A/E_2) = \frac{{}^2C_1 \times {}^1C_1}{{}^4C_2} = \frac{2}{6} = \frac{1}{3}$$

and $P(A/E_3) = \frac{{}^4C_1 \times {}^3C_1}{{}^{12}C_2} = \frac{4 \times 3}{66} = \frac{2}{11}$

The probability of the two balls (red and white) selected from the II urn i.e. $P(E_2/A)$ is given by the Baye's law as follows:

$$P(E_2/A) = \frac{P(E_2) \cdot P(A/E_2)}{P(E_1) \cdot P(A/E_1) + P(E_2) \cdot P(A/E_2) + P(E_3) \cdot P(A/E_3)}$$

$$= \frac{\frac{1}{3} \times \frac{1}{3}}{\frac{1}{3} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{11}} = \frac{1}{9} \times \frac{495}{118} = \frac{55}{118} = 0.466$$

28. **In Pokhara city, 25% of all cars emitting amount of pollutants. If the probability is 0.99 that a car emitting excessive amounts of pollutants will fail the city vehicular test and the probability is 0.17 that a car not emitting excessive amounts of pollutants will nevertheless fail the test, what is the probability that car which fail the test actually emits excessive amount of pollutants.**

Solution 28

Do yourself

29. **The probabilities of the three persons A, B and C becoming V.C of T.U are $\frac{3}{9}, \frac{2}{7}$ and $\frac{3}{7}$ respectively.**

a. What is the probability that the grade increment system will be introduced?

b. If the grade increment system has been introduced. What is the probability that the V.C appointed is Mr. B?

Solution 29

Let E_1, E_2 and E_3 denote the events that the three persons A, B and C becoming V.C. of T.U. Then, we have,

$$P(E_1) = \frac{3}{9}, P(E_2) = \frac{4}{9} \text{ and } P(E_3) = \frac{2}{9} \text{ respectively}$$

Let A denote the event that 'grade increment system' will be introduced then the probability of A, B, C will be introduced. The grade increment system after being V.C. is given by,

$$P(A/E_1) = \frac{1}{7}, \qquad P(A/E_2) = \frac{2}{7} \qquad \text{and } P(A/E_3) = \frac{3}{7}$$

a. Then, the probability that grade increment system will be introduced is given by

$$P(A) = P(E_1) \cdot P(A/E_1) + P(E_2) \cdot P(A/E_2) + P(E_3) \cdot P(A/E_3)$$

$$= \frac{3}{9} \times \frac{1}{7} + \frac{4}{9} \times \frac{2}{7} + \frac{2}{9} \times \frac{3}{7} = \frac{3+8+6}{63} = \frac{17}{63}$$

Scanned by CamScanner

b. The probability that B being V.C. if the grade increment system has been introduced is given by the Baye's theorem

i.e. $$P(E_2/A) = \frac{P(E_2) \cdot P(A/E_2)}{P(A)} = \frac{\frac{4}{9} \times \frac{2}{7}}{\frac{17}{63}} = \frac{8}{17}$$

Additional Exercise 7.2

1. A committee of 4 members is selected at random from 4 men and 3 women. Find the probability that 3 men and 1 woman are selected.

Solution 1

$$P(3 \text{ men and } 1 \text{ women}) = \frac{n(3 \text{ men and } 1 \text{ women})}{n(S)} = \frac{{}^4C_3 \times {}^3C_1}{{}^7C_4} = \frac{4 \times 3}{35} = 0.34$$

2. If an experiment has three possible and mutually exclusive out-comes A, B and C, check in each case whether the assignment of probabilities is permissible.

a. $P(A) = 1/3$, $P(B) = 1/4$ and $P(C) = -1/5$

b. $P(A) = 1/5$, $P(B) = -1/2$ and $P(C) = 2$

c. $P(A) = 1/3$, $P(B) = 1/3$ and $P(C) = 1/3$

d. $P(A) = 0.57$, $P(B) = 0.2$ and $P(C) = 0.3$

Solution 2

Since A, B and C are three possible and mutually exclusive outcomes then:

a. It cannot be permissible because probability cannot be negative.

b. It cannot be permissible because it must be positive as well as less than 1.

c. It is possible.

d. It is also impossible because total probability is less than 1.

3. A car rental agency has 18 compact cars and 12 intermediate-size cars. If 4 of the cars are randomly selected for a safety check, what is the probability of getting two of each kind?

Solution 3

Let C and I denote the compact and intermediate cars.

$$P(2C \text{ and } 2I) = \frac{n(2c \text{ and } 2I)}{n(S)} = \frac{{}^{18}C_2 \times {}^{12}C_2}{{}^{30}C_4} = \frac{153 \times 66}{27405} = 0.37$$

4. In a group of 160 graduate engineering students, 92 are enrolled in advanced courses in statistics, 63 are enrolled in a course in operational research, and 40 are enrolled in both. How many of these students are not enrolled in either course?

Solution 4

Let S denote the sample space A and O denote the engineers enrolled in advanced course in statistics and operational research respectively. Also, $n(S) = 160$, $n(A) = 92$, $n(O) = 63$ and $N(A \cap O) = 40$. Then,

$$A(\overline{A \cup O}) = ?$$

We know that,

Scanned by CamScanner

$n(S) = N(A) + n(O) - n(A \cap O) + n(\overline{A \cup O})$

$160 = 92 + 63 - 40 + n(\overline{A \cup O})$

$\therefore \quad n(\overline{A \cup O}) = 160 - 115 = 45$

5. **If the probability that a communication system will have high fidelity is 0.81 and the probability that it will have high fidelity and high selectivity is 0.81, what is the probability that a system with high fidelity will also have high selectivity?**

Solution 5

Let F and S denote the event that a communication system will have high fidelity and high selectivity respectively. Also, $P(F) = 0.81$ and $P(F \cap S) = 0.18$. The probability that a system with high fidelity will also have high selectivity is written as $P(S/F)$ and obtained by the formula.

$$P(S/F) = \frac{P(F \cap S)}{P(F)} = \frac{0.18}{0.81} = \frac{2}{9}$$

6. **Two cards are drawn at random from an ordinary deck of 52 playing cards. What is the probability of getting two aces if**
   a. the first card is replaced before the second card is drawn?
   b. the first card is not replaced before the second card is drawn?

Solution 6

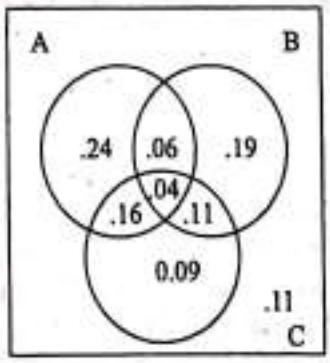
a. If the first card is replaced before the second card is drawn then,

$$P(A \text{ and } A) = P(A) \times P(A) = \frac{4}{52} \times \frac{4}{52} = \frac{1}{169}$$

b. If the first card is not replaced before the second card is drawn then:

$$P(A \text{ and } A) = P(A) \times P(A) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$$

7. Given in probability features.



Solution 7

a. $$P(A/B) = \frac{P(A \cap B)}{n(B)} = \frac{0.06 + 0.04}{0.06 + 0.04 + 0.11 + 0.19} = \frac{0.1}{0.4} = 0.25$$

b. $$P(B/C') = \frac{P(B \cap C')}{P(C')} = \frac{0.06 + 0.19}{0.24 + 0.06 + 0.19 + 0.11} = \frac{0.25}{0.6} = 0.42$$

c. $$P(A \cap B/C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{0.04}{0.04 + 0.16 + 0.11 + 0.09} = \frac{0.04}{0.4} = 0.1$$

Scanned by CamScanner

d. $P(A \cap B \cap C / B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{0.04}{0.04 + 0.11} = \frac{0.04}{0.15} = 0.27$

8. Given P (A) = 0.5, P (B) = 0.3 and P (A∩B) = 0.15, then Verify that:

a. $P(A / B) = P(A / B')$ b. $P(B / A) = P(B / A')$

Solution 8

Here, given, $P(A) = 0.5$, $P(B) = 0.3$ and $P(A \cap B) = 0.15$

a. $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.15}{0.3} = 0.5$

$P(A/B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{0.5 - 0.15}{1 - 0.3} = \frac{0.35}{0.7} = 0.5$

Hence, $P(A/B) = P(A/B')$

b. $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.15}{0.5} = 0.3$

$P(B/A') = \frac{P(B \cap A')}{P(A')} = \frac{P(B) - P(A \cap B)}{1 - P(A)} = \frac{0.3 - 0.15}{1 - 0.5} = \frac{0.15}{0.5} = 0.3$

Hence, $P(B/A) = P(B/A')$

9. If the odds are 5 to 3 that event M will not occurs, 2 to 1 that event N will occurs, and 4 to 1 that they will not both occurs, are the two events M and N independent?

Solution 9

Let M denote the odds 5 to 3 will not occur i.e. $P(M) = \frac{3}{5+3}$

Let N denote the odds 2 to 1 will occur i.e. $P(N) = \frac{2}{2+1}$

Then, $P(M \cap N) = \frac{1}{4+10}$

If M and N are two independent events then we must have,

$P(M \cap N) = P(M) \times P(N)$

or, $\frac{1}{5} = \frac{3}{8} \times \frac{2}{3}$

or, $\frac{1}{5} \neq \frac{1}{4}$

Hence, the relation is unequal. Thus, M and N are two dependent events.

10. If from a lottery of 50 tickets marked 1, 2, 3, ..., 50, 4 tickets are drawn. What is the chance that those marked 1 and 2 are among them?

Solution 10

We have given the lottery ticket marked 1, 2, 3, 4, ...50. Among them 4 tickets are drawn. We are going to find the probability of having marked 1, 2 tickets among the 4 selected tickets. Which is given as follows:

$$P(E) = \frac{n(E)}{n(S)} = \frac{{}^2C_2 \times {}^{48}C_2}{{}^{15}C_4} = \frac{1 \times 1128}{1365} = 0.826$$

Scanned by CamScanner

11. **A bag contains 10 balls, two of which are red, three blue and five black. Three balls are drawn at random from the bag, that is every ball has an equal chance of being included in the three. What is the probability that**

i. the three balls are of different colors

ii. only two balls are of same color and

iii. the balls are of the same color?

Solution 11

Let R, B, B' denote the red, blue and black balls respectively.

$n(R) = 2, \quad n(B) = 3, \quad n(B') = 5$

i. $P(1R, 1B \text{ and } 1B') = \dfrac{n(1R, 1B \text{ and } 1B')}{n(S)} = \dfrac{{}^2C_1 \times {}^3C_1 \times {}^5C_1}{{}^{10}C_3} = \dfrac{2 \times 3 \times 5}{120} = 0.25$

ii. For the probability of having only two same colour is given as follows:

Red (2)	Blue (3)	Black (5)	No. of selection
2	1	0	${}^2C_2 \times {}^3C_1 \times {}^5C_0 = 1 \times 3 \times 1 = 3$
1	2	0	${}^2C_1 \times {}^3C_2 \times {}^5C_0 = 2 \times 3 \times 1 = 6$
0	1	2	${}^2C_0 \times {}^3C_1 \times {}^5C_2 = 1 \times 3 \times 10 = 30$
0	2	1	${}^2C_0 \times {}^3C_2 \times {}^5C_1 = 1 \times 3 \times 5 = 15$
1	0	2	${}^2C_1 \times {}^3C_0 \times {}^5C_2 = 2 \times 1 \times 10 = 20$
2	0	1	${}^2C_2 \times {}^3C_0 \times {}^5C_1 = 1 \times 1 \times 5 = 5$
Total number of selection = 79			

$\therefore$ $P(\text{only two balls are same colour}) = \dfrac{n(\text{only two balls are same colour})}{n(S)}$

$79/120 = 0.658$

iii. $P(\text{All the balls are same colour}) = \dfrac{n(\text{All the balls are same colour})}{n(S)}$

$= \dfrac{n(\text{3 blue balls or 3 black balls})}{{}^{10}C_3}$

$= \dfrac{{}^3C_3 + {}^5C_3}{120} = \dfrac{1 + 10}{120} = 0.092$

12. **Two-third of the students in a class is boys and rests are girls. It is known that the probability of a girl getting a first class is 0.25 and that of boy getting a first class is 0.28. Find the probability that a student chosen at random will get first class marks in the subject.**

Solution 12

E_1 : The students is a boy. (Cause 1)

E_2 : The student is a girl. (Cause 2)

A : The student will get first class (Effect)

From the data given in the above problem,

$P(E_1) = \dfrac{2}{3}, \quad P(E_2) = 1 - \dfrac{2}{3} = \dfrac{1}{3}, \quad p(A/E_1) = 0.28, \quad P(A/E_2) = 0.25$

$P(A) = ?$

Then, by the total probability theorem, we have,

$P(A) = P(E_1) \times P(A/E_1) + P(E_2) \cdot P(A/E_2)$

Scanned by CamScanner

$$= \frac{2}{3} \times 0.28 + \frac{1}{3} \times 0.25 = 0.187 + 0.083 = 0.27$$

13. A bag contains 3 red and 5 black balls and a second bag contains 6 red and 4 black balls. A ball is drawn from each bag. Find the probability that

i. both the balls are red and ii. one is red and other is black.

Solution 13

We have given B_1 (bag one) contains 3 red and 5 black balls. B_2 (second bag) contains 6 red and 4 black balls:

(i) P(Both red) $= P(B_1 \text{ Red and } B_2 \text{ Red})$

$$= P(B_1 \text{ Red}) \times P(B_2 \text{ Red}) = \frac{3C_1}{8C_1} \times \frac{6C_1}{10C_1} = \frac{3}{8} \times \frac{6}{10} = 0.225$$

(ii) P(One Red and other Black)

$= P[(B_1 \text{ Red and } B_2 \text{ Black}) \text{ or } (B_2 \text{ Red and } B_1 \text{ Black})]$

$$= P(B_1 \text{ Red}) \times P(B_2 \text{ Black}) + P(B_2 \text{ Red}) \times P(B_1 \text{ Black}) = \frac{3}{8} \times \frac{4}{10} + \frac{6}{10} \times \frac{5}{8} = \frac{42}{80} = 0.525$$

14. If six bullets, of which three are blanks, are randomly inserted into a gun, what is the probability that the first three bullets fired will be blanks?

Solution 14

There are total six bullets $n(S) = 6$

Among them 3 are blanks $n(B) = 3$

For the probability that the first three bullets fired will be bank

i.e. $P(B_1 \cap B_2 \cap B_3)$

Since the firing bullets are dependent events.

$$\therefore \quad P(B_1 \cap B_2 \cap B_3) = P(B_1)\, P(B_2/B_1)\, P(B_3)/B_1 \cap B_2) = \frac{3}{6} \times \frac{2}{5} \times \frac{1}{4} = \frac{1}{20} = 0.05$$

15. The probability that a man will be alive 25 years hence is 0.3 and the probability that his wife will be alive 25 year hence is 0.4. Find the probability that 25 years hence

i. both will be alive ii. only the man will be alive

iii. only the woman will be alive iv. at least one of them will be alive?

Solution 15

Let A and B denote the events of a man will be alive 25 years hence and his wife will be alive 25 years hence respectively. Then; $P(A) = 0.3$ and $P(B) = 0.4$

(i) P(Both will be alive) $= P(A \text{ and } B) = P(A) \times P(B) = 0.3 \times 0.4 = 0.12$

(ii) P(Only the man will be alive) $P(A \cap \bar{B} = P(A) - P(A \cap B) = 0.3 - 0.12 = 0.18$

(iii) P(Only the women will be alive) $P(B \cap \bar{A}) = P(B) - P(A \cap B) = 0.4 - 0.12 = 0.28$

(iv) P(at least one of them will be alive)

$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.4 - 0.12 = 0.58$

Scanned by CamScanner

16. **An urn contains 8 balls of which 3 are white, three balls are drawn at random and laid aside their color unnoted.**

Solution 16

Let us define the events as follows:

E_0 - selecting no white balls $\quad$ E_1 - selecting one white balls

E_2 - selecting two white balls $\quad$ E_3 - selecting three white balls

Then, $P(E_0) = \frac{5C_3}{8C_3} = \frac{10}{56}$, $P(E_1) = \frac{3C_1 \times 5C_2}{8C_3} = \frac{30}{56}$

$P(E_2) = \frac{3C_2 \times 5C_1}{8C_3} = \frac{15}{56}$, $P(E_3) = \frac{3C_3}{8C_3} = \frac{1}{56}$

Let A denote the event of drawing a white ball after 3 balls with out replacement.

$\therefore$ $P(A/E_0) = \frac{3}{5}$, $P(A/E_1) = \frac{2}{5}$, $P(A/E_2) = \frac{1}{5}$, $P(A/E_3) = \frac{0}{5}$

Now by the total probability the probability of drawing a white ball is given by

$P(A) = P(E_0)\,P(A/E_0) + P(E_1)\,.\,P(A/E_1) + P(E_2)\,P(A/E_2) + P(E_3)\,P(A/E_3)$

$= \frac{10}{56} \times \frac{3}{5} + \frac{30}{56} \times \frac{2}{5} + \frac{15}{56} \times \frac{1}{5} + \frac{1}{56} \times \frac{0}{5} = \frac{6+12+3}{56} = \frac{21}{56} = 0.375$

17. **Suppose a jar contains two white balls and three black balls. Five balls are drawn from the jar placed on the table in the order drawn**

a. What is the probability that they are drawn in order: white, black, white black?

b. What is the probability that the third ball drawn is black one?

Solution 17

Do yourself

18. **A bag contains 4 white, 8 black, 6 red and 2 green balls. Find the probability of getting either a white or a black or a green ball in a single draw.**

Solution 18

Let W, B, R, G denote the event of selecting White, Black, Red and Green balls respectively.

Then, $n(W) = 4$, $n(B) = 8$, $n(R) = 6$, $n(G) = 2$.

Now the probability of getting either a white or a black of a green ball in a single draw,

i.e. $P(W \text{ or } B \text{ or } G) = P(W) + P(B) + P(G)$ [being disjoint events]

$= \frac{n(W)}{n(S)} + \frac{n(B)}{n(S)} + \frac{n(G)}{n(S)} = \frac{4}{20} + \frac{8}{20} + \frac{2}{20} = \frac{14}{20} = 0.7$

19. **From a group of 8 children, 5 boys and 3 girls; 3 children are selected at random. Calculate the probabilities that the selected group contains:**

i. no girl $\quad$ ii. any one girl

iii. only one particular girl $\quad$ iv. at least one girl

v. more girls than boys

Scanned by CamScanner

Solution 19

Let B and G denote the events of selecting boy and girl respectively. Then $n(B) = 5$, $n(G) = 3$.

(i) The probability of no girls is given by

$$P(B_1 \text{ and } B_2 \text{ and } B_3) = P(B_1) \times (B_2/B_1) \times P(B_3/B_1 \cap B_2) = \frac{5}{8} \times \frac{4}{7} \times \frac{3}{6} = \frac{5}{28}$$

(ii) The probability of selecting only one girl:

$$P(\text{Only one girl}) = P(\text{2 boys and a girl}) = \frac{5C_2 \times 3C_1}{8C_3} = \frac{10 \times 3}{56} = \frac{15}{28}$$

(iii) The probability of only one particular girls

$$P(\text{2 boys and one particular girl}) = \frac{5C_2 \times 1}{8C_3} = \frac{10}{56} = \frac{5}{28}$$

(iv) For the probability of at least one girl

$$P(\text{at least one girl}) = 1 - P(\text{all three boys}) = 1 - \frac{5}{28} = \frac{23}{28}$$

(v) P(more girl and boys)

$$= P(\text{2 girl and boy}) + P(\text{All three girls}) = \frac{3C_2 \times 5C_1}{8C_3} + \frac{3C_3}{8C_3} = \frac{3 \times 5}{56} + \frac{1}{56} = \frac{16}{56} = \frac{2}{7}$$

20. A university has to select an examiner from a list of 50 persons, 20 of them women and 30 men, 10 of them knowing English and 40 not, 15 of them being teachers and remaining 35 not. What is the probability of the university selecting English knowing woman teacher?

Solution 20

Here we are going to find the probability of selecting the English knowing women teacher is

Probability that the selected candidate is English knowing $= \frac{10}{50}$

Probability that the selected candidate is women $= \frac{20}{50}$

Probability that the selected candidate is teacher $= \frac{15}{50}$

Hence probability of selected candidate will be

English knowing women teacher $= \frac{10}{50} \times \frac{20}{50} \times \frac{15}{50}$ [Being the event independent]

$$= \frac{1}{5} \times \frac{2}{5} \times \frac{15}{50} = \frac{3}{125}$$

21. Ramesh and Suresh are asked to solve a problem. The probability of Ramesh solving it is $\frac{2}{3}$ and that of Suresh solving it is $\frac{3}{4}$. Find the probability that

i. both can solve the problem

ii. Ramesh solves it but Suresh cannot.

iii. Suresh solves it but Ramesh cannot

Scanned by CamScanner

iv. None of them can solve

v. at least one of them will solve it

vi. only one of them can solve

Solution 21

Let R and S denote the event that Ramesh and Suresh will solve the problem. Then by the given problem we have:

$P(R) = \frac{2}{3}, P(S) = \frac{3}{4}$

(i) Probability that both can solve the problem

i.e. $P(R \text{ and } S) = P(R) \times P(S) = \frac{2}{3} \times \frac{3}{4} = 0.5$

(ii) Probability that Ramesh solve it but Suresh cannot

i.e. $P(R \cap \bar{S}) = P(R) - P(R \cap S) = \frac{2}{3} - \frac{1}{2} = \frac{2 \times 2 - 1 \times 3}{2 \times 3} = \frac{1}{6}$

(iii) Probability that Suresh solve it but Ramesh cannot

i.e. $P(S \cap \bar{R}) = P(S) - P(S \cap R) = \frac{3}{4} - \frac{1}{2} = \frac{3-2}{4} = \frac{1}{4}$

(iv) Probability that none of them can solve the problem

i.e. $P(\overline{S \cup R}) = 1 - (S \cup R) = 1 - \{P(S) + P(R) - P(S \cap R)\} = 1 - \left(\frac{3}{4} + \frac{2}{3} - \frac{1}{2}\right) = 1 - \frac{11}{12} = \frac{1}{12}$.

(v) Probability that least one of them solve it, i.e. $P(S \cup R) = \frac{11}{12}$

(vi) Probability that only one of them solve it.,

i.e. $P(R \cap \bar{S})$ or $P(S \cap \bar{R}) = P(R \cap \bar{S}) + P(S \cap \bar{R}) = \frac{1}{6} + \frac{1}{4} = \frac{2+3}{12} = \frac{5}{12}$.

22. A problem of statistics is given to three students A, B and C whose chance of solving it are $\frac{1}{2}, \frac{1}{3}$ and $\frac{1}{4}$ respectively. Find the probability that

i. all of them can solve the problem ii. none of them can solve

iii. problem will be solved

Solution 22

Here we have given,

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

(i) Probability that all of them solve the problem

i.e. $P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{4} = \frac{1}{24}$

(ii) Probability that none of them can solve the problem,

i.e. $P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C) = 1 - \frac{3}{4} = \frac{1}{4}$

Scanned by CamScanner

(iii) Probability that the problem will be solved

(i.e. The problem will be solved if one of them solve the problem)

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

$$= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \frac{1}{2} \times \frac{1}{3} - \frac{1}{3} \times \frac{1}{4} - \frac{1}{2} + \frac{1}{4} + \frac{1}{24} = \frac{3}{4}$$

23. In a certain college, 25% of the students failed in Mathematics, 15% of the student failed in Chemistry and 10% failed in both Mathematics and Chemistry. A student is selected at random.

i. What is the probability that he has failed in Mathematics given that he has failed in Chemistry?

ii. What is the probability of his failure in Chemistry given that he has failed in Mathematics?

Solution 23

Let M and C denote the events failed in Mathematics and Chemistry respectively.

Then we have $P(M) = \frac{25}{100} = 0.25$, $P(C) = \frac{15}{100} = 0.15$, $P(M \cap C) = 0.1$

(i) The probability of a student failed in Mathematics given that he fail in Chemistry is given by $P(M/C) = \frac{P(M \cap C)}{P(C)} = \frac{0.1}{0.15} = \frac{10}{15} = \frac{2}{3}$

(ii) The probability of a student failed in Chemistry given that he has failed in Mathematics is given by $P(C/M) = \frac{P(C \cap M)}{P(M)} = \frac{0.1}{0.25} = \frac{10}{25} = \frac{2}{5}$.

24. A box contains 3 red and 5 blue balls. Two balls are drawn one at a time at random without replacement. Find the probability of given 1 red and 1 blue ball.

Solution 24

Here we have given $n(R) = 3$ and $n(B) = 5$

Then probability of getting 1 red and 1 blue balls is given by, P(R and B)

$$= P(R) \,.\, P(B/R) + P(B) \,.\, P(R/B) = \frac{3}{8} \times \frac{5}{7} + \frac{5}{8} \times \frac{3}{7} = \frac{30}{56} = \frac{15}{28}.$$

25. From a vessel containing 3 white and 5 black ball, 4 balls are transferred into an empty vessel. From this vessel a ball is drawn and is found to be white. What is the probability that out of 4 balls transferred 3 are white and 1 black?

Solution 25

Let us define the events as follows

E_0 - transfer of 0 white and 4 black balls.

E_1 - transfer of 1 white and 3 black balls.

E_2 - transfer of 2 white and 2 black balls.

E_3 - transfer of 4 white and 0 black balls.

A - event of drawing a white ball from the empty vessel.

Then $P(E_0) = \frac{5C_4}{8C_4} = \frac{1}{14}$, $\quad P(E_1) = \frac{3C_1 \times 5C_3}{8C_4} = \frac{3}{7}$,

Scanned by CamScanner

$P(E_2) = \frac{3C_2 \times 5C_2}{8C_4} = \frac{3}{7}$, $P(E_3) = \frac{3C_3 \times 5C_1}{8C_4} = \frac{1}{14}$ and $P(E_4) = 0$

Also, $P(A/E_0) = \frac{0}{4} = 0$, $P(A/E_1) = \frac{1}{4}$, $P(A/E_2) = \frac{2}{4}$, $P(A/E_3) = \frac{3}{4}$

Hence the probability that out of 4 balls transformed, 3 are white and 1 is black is, $P(E_3/A) = ?$

Then by Baye's theorem we have:

$$P(E_3/A) = \frac{P(E_3) \cdot P(A/E_3)}{P(E_0) \times P(A/E_0) + P(E_1)\, P(A/E_1) + P(E_2)\, P(A/E_2) + P(E_3) \times P(A/E_3)}$$

$$= \frac{\frac{1}{14} \times \frac{3}{4}}{\frac{1}{14} \times 0 + \frac{3}{7} \times \frac{1}{4} + \frac{3}{7} \times \frac{2}{4} + \frac{1}{14} \times \frac{3}{4}} = \frac{3}{0 + 16 + 12 + 3} = \frac{3}{21} = \frac{1}{7} = 0.1428$$

26. Urn A contains 2 white, 1 black and 3 red balls, urn B contains 3 white, 2 black and 4 red balls and urn C contains 4 white, 3 black and 2 red balls. One urn is chosen at random and 2 balls are drawn. They happen to be red and black. What is the probability that both balls come from urn 'B'?

Solution 26

Let E_1, E_2, E_3, denote the events that the urn A, B and C is chosen resp.

Then $P(E_1) = \frac{1}{3}$, $P(E_2) = \frac{1}{3}$, $P(E_3) = \frac{1}{3}$

Let A denote the events that the two balls drawn from the selected urn are white and red.

Then $P(A/E_1) = \frac{2C_1 \times 3C_1}{6C_2} = \frac{2 \times 3}{15} = \frac{2}{5}$

$$P(A/E_2) = \frac{3C_1 \times 4C_1}{9C_2} = \frac{3 \times 4}{36} = \frac{1}{3}$$

$$P(A/E_3) = \frac{4C_1 \times 2C_1}{9C_2} = \frac{4 \times 2}{36} = \frac{2}{9}.$$

Now the probability that the two balls red and white will come from urn B is, $P(E_2/A) = ?$

Then by the Baye's law we have,

$$P(E_2/A) = \frac{P(E_2) \cdot P(A/E_2)}{P(E_1) \cdot P(A/E_1) + P(E_2) \cdot P(A/E_2) + P(E_3) \cdot P(A/E_3)}$$

$$= \frac{\frac{1}{3} \times \frac{1}{3}}{\frac{1}{3} \times \frac{2}{5} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{9}} = \frac{1}{9} \times \frac{135}{43} = \frac{15}{43}.$$

27. Urns X_1, X_2, X_3, each contain 5 red and 3 white balls. Urns Y_1, Y_2, each contain 2 red and 4 white balls. An urn is selected at random and a ball is drawn. It is found to be red. Find the probability that the ball comes out of the urns of the first type.

Solution 27

Do yourself

Scanned by CamScanner

28. There are three candidates for the position of principal Mr. A, Mr. B and Mr. C, whose chances of getting the appointment are in the proportion 1 : 2 : 3 respectively? The probability that Mr. A if selected will introduce co-education in the college is 0.3. The probability of Mr. B and Mr. C doing the same are respectively 0.5 and 0.8. What is the probability that there was co-education in the college?

Solution 28

Let E_1, E_2, E_3 denote the events that the three candidates Mr. A, Mr. B and Mr. C will get the appointment of principal. Then we have:

$$P(E_1) = \frac{1}{1+2+3} = \frac{1}{6} = 0.167,\ P(E_2) = \frac{2}{6} = 0.333 \text{ and } P(E_3) = \frac{3}{6} = 0.5$$

Let A denote the event that three candidate introduce co-education in the college.

Then the probability that Mr. A will introduce co-education $P(A/E_1) = 0.3$

Similarly $P(A/E_2) = 0.5$, $P(A/E_3) = 0.8$

Then the probability that there was co-education in the college, i.e. P(A) is given by the total probability

$$P(A) = P(E_1) \,.\, P(A/E_1) + P(E_2) \,.\, P(A/E_2) + P(E_3)\, P(A/E_3)$$
$$= 0.167 \times 0.3 + 0.333 \times 0.5 + 0.5 \times 0.8 = 0.615$$

29. Two urns contain 1 white, 6 red and 4 white, 3 red balls. One of the urns is selected at random and a ball is drawn from it. Find

i. the probability of drawing a white ball.

ii. the probability of drawing the ball from the first urn, if the ball drawn is white.

Solution 29

Let E_1 and E_2 denote the events selecting the two given urns I and II respectively.

Then $P(E_1) = \frac{1}{2}$ and $P(E_2) = \frac{1}{2}$.

Let W denote the event of selecting the white ball.

The probability of selecting white ball from the urn I is, $P(W/E_1) = \frac{1}{6+1} = \frac{1}{7}$

Also, $P(W/E_2) = \frac{4}{4+3} = \frac{4}{7}$.

(i) Then the probability of draw a white ball, P(W) = probability of selecting from I or from II $= P(E_1) \times P(W/E_1) + P(E_2) \times P(W/E_2 = \frac{1}{2} \times \frac{1}{7} + \frac{1}{2} \times \frac{4}{7} = \frac{5}{14}$

(ii) Also the probability of drawing the ball from the first urn, if the drawn ball is white is,

$$P(E_1/W) = \frac{P(E_1) \,.\, P(W/E_1)}{P(E_1) \,.\, P(W/E_1) + P(E_2) \,.\, P(W/E_2)} = \frac{\frac{1}{2} \times \frac{1}{7}}{\frac{5}{14}} = \frac{1}{14} \times \frac{14}{5} = \frac{1}{5}$$

Scanned by CamScanner

30. **There are 3 bags, each containing 5 white balls and 3 black balls. Also, there are bags, each containing 2 white balls and 4 black balls. A white ball is drawn at random. Find the probability that this white ball is from a bag of the first group.**

Solution 30

Do yourself

31. **A manufacturing firm produces still pipes in three plants with daily production volume of 500, 1000 and 2000 units respectively. According to past experience, it is known that the fractions of defective outputs produced by the three plants are respectively 0.005, 0.008 and 0.010. If a pipe is selected from a day's total production are found to be defective. From which plant the defective pipe is expected to have been produced?** [P.U. 2016/ T.U 2068/2072]

Solution 31

Let E_1, E_2 and E_3 denote the events manufacturing firm produces steel pipes in three plants with daily production volume resp.

Then $P(E_1) = \frac{500}{500 + 1000 + 2000} = \frac{500}{3500} = \frac{1}{7} = 0.143$

Similarly $P(E_2) = \frac{1000}{3500} = \frac{2}{7} = 0.2861$ and $P(E_3) = \frac{4}{7} = 0.571$

Let D denote the events producing the defective outputs by the three plants then,

$P(D/E_1) = 0.005$, $P(D/E_2) = 0.008$ and $P(D/E_3) = 0.010$

The probability of selecting first plant is $P(E_1/D) = \frac{P(E_1) \cdot P(D/E_1)}{\sum_{i=1}^{3} P(E_i) \cdot P(D/E_i)}$

$$= \frac{0.143 \times 0.005}{0.143 \times 0.005 + 0.286 \times 0.008 + 0.571 \times 0.0010} = \frac{0.0007}{0.00087} = 0.08.$$

Similarly the probability of selecting second plant is given by

$$P(E_1/D) = \frac{P(E_2) \cdot P(D/E_2)}{\sum_{i=1}^{3} P(E_i) \cdot P(D/E_i)} = \frac{0.0023}{0.0087} = 0.26.$$

Also the probability of selecting third plant is given by,

$$P(E_3/D) = \frac{P(E_3) \cdot P(D/E_3)}{\sum_{i=1}^{3} P(E_i) \cdot P(D/E_i)} = \frac{0.0057}{0.0087} = 0.66.$$

Since the probability of selecting is higher than the other two. So it could be selected.

32. **One bag contains 5 white and 4 black balls. Another bag contains 7 white and 9 black balls. A ball is transferred from the first bag to the second bag and then a ball is drawn from the second. Find the probability that the ball is white.**

Solution 32

Do yourself

Scanned by CamScanner

33. There are three machines A, B and C producing 1000, 2000 and 3000 articles per hour respectively. Those machines are known to be producing 1%, 2% and 3% defectives respectively. One article is selected at random from an hour production of the three machines and found to be defective. What is the probability that the article is produced from

i. Machine A, ii. Machine B, iii. Machine C?

[P.U. 2010 Spring/ 2008 Fall/ T.U 2061/ 2058/ 2068]

Solution 33

Let E_1, E_2, E_3 denote the events of producing articles per hour from three machines A, B and C respectively. Then,

$$P(E_1) = \frac{n(E_1)}{n(S)} = \frac{1000}{1000 + 2000 + 3000} = \frac{1000}{6000} = 0.167$$

Similarly $P(E_2) = \frac{2}{3} = 0.33$ and $P(E_3) = \frac{3}{6} = 0.5$

Also D denote the events of producing defective articles by the three machines A, B, C. Then the probability of machine A producing defective article $P(D/E_1) = 1\% = 0.01$.

Similarly, $P(D/E_2) = 0.02$ and $P(D/E_3) = 0.03$

(i) The probability of getting defective article from machine A is given by

$$P(E_1/D) = \frac{P(E_1) \cdot P(D/E_1)}{P(E_1) \cdot P(D/E_1) + P(E_2) \cdot P(D/E_2) + P(E_3) \cdot P(D/E_3)}$$

$$= \frac{0.167 \times 0.01}{0.167 \times 0.01 + 0.33 \times 0.02 + 0.5 \times 0.03} = \frac{0.00167}{0.02327} = 0.072$$

(ii) Similarly $P(E_2/D) = \dfrac{P(E_2) \cdot P(D/E_2)}{\sum_{i=1}^{3} P(E_i) \cdot P(D/E_i)} = \dfrac{0.33 \times 0.02}{0.02327} = 0.284$

(iii) Also $P(E_3/D) = \dfrac{P(E_3) \cdot P(D/E_3)}{\sum_{i=1}^{3} P(E_i) \cdot P(D/E_i)} = \dfrac{0.5 \times 0.03}{0.0237} = 0.645$

Exam Asked Questions

Long Questions

1. A can hit a target 4 times in 5 shots. B can hit the target 3 times in 4 shots and can hit the target 2 times in 3 shots. They fire a volleyball, what is the chan that the target is damaged if two hits are needed for it? [PU 2008 Fa

Solution 1

Let E_1, E_2, E_3 be the events of hitting target by A, B and C respectively. Then P($= \frac{4}{5}$, $P(E_2) = \frac{3}{4}$ and $P(E_3) = \frac{2}{3}$. Since, 2 hits are needed to damage the target so target will be damaged in the following ways.

Either, $(E_1 \cap E_2 \cap \bar{E}_3)$ or $(E_1 \cap E_3 \cap \bar{E}_2)$ or $(E_2 \cap E_3 \cap \bar{E}_1)$ or $(E_1 \cap E_2 \cap E_3)$

Scanned by CamScanner

i.e. $P[(E_1 \cap E_2 \cap \bar{E}_3) \cup (E_1 \cap E_3 \cap \bar{E}_2) \cup (E_2 \cap E_3 \cap \bar{E}_1) \cup (E_1 \cap E_2 \cap E_3)]$

Since, the given events are mutually exclusive

$= P(E_1) \times P(E_2) \times P(\bar{E}_3) + P(E_1) \times P(E_3) \times P(\bar{E}_2) + P(E_2) \times P(E_3) \times P(\bar{E}_1) + P(E_1) \times P(E_2) \times P(E_3)$

$= \frac{4}{5} \times \frac{3}{4} \times \left(1 - \frac{2}{3}\right) + \frac{4}{5} \times \frac{3}{4} \times \left(1 - \frac{3}{4}\right) \times \frac{3}{4} \times \frac{2}{3} + \left(1 - \frac{4}{5}\right) + \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3}$

$= \frac{1}{5} + \frac{2}{15} + \frac{1}{10} + \frac{2}{5} = \frac{25}{30} = \frac{5}{6}$

2. **The probability that a new airport will get an award for its design is 0.16, the probability that it will get an award for the efficient use of materials is 0.24 and the probability that it will take both awards is 0.11.**

   i. What is the probability that, it will get at least one of the award?

   ii. What is the probability that it will get only one of the two awards?

[PU 2008 Spring/TU 2068]

Solution 2

Let E_1 and E_2 denote the events that a new airport will get an award for its design and efficient use of materials respectively. Then,

$P(E_1) = 0.16$, $P(E_2) = 0.24$ and $P(E_1 \cap E_2) = 0.11$

i. The probability of getting at least one of the awards

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 0.16 + 0.24 - 0.11 = 0.29$

ii. The probability of getting only one of the award is

$P[(E_1 \cap \bar{E}_2) \text{ or } (E_2 \cap \bar{E}_1)] = P(E_1 \cap \bar{E}_2) + P(E_2 \cap \bar{E}_1)$

$= P(E_1) - P(E_1 \cap E_2) + P(E_2) - P(E \cap E_2)$

$= 0.16 - 0.11 + 0.24 - 0.11 = 0.18$

3. **If the probability that a married man will vote in a given election is 0.5 the probability that a woman will vote in the election given that her husband will vote is 0.90. Find:**

   i. The probability that a husband and wife will both vote in the election.

   ii. The probability that at least one of a married couple will vote. [PU 2002]

Solution 3

Let E_1 and E_2 denote the married man and woman will vote in the election. Then we have,

$P(E_1) = 0.5$, $P(E_2) = 0.6$ and $P(E_2/E_1] = 0.90$

i. The probability that both husband and wife will vote for election is

$P(E_1 \cap E_2) = P(E_1) \times P(E_2/E_1) = 0.5 \times 0.9 = 0.45$

ii. The probability that a man will vote inn the election given that his wife will vote is

$P(E/E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{0.45}{0.6} = 0.75$

iii. The probability that at least one of the married couple will vote is given by

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 0.5 + 0.6 - 0.45 = 0.65$

Scanned by CamScanner

4. **There is a cricket match between two teams A and B. The probability that team A will win the match is 1/4. If A and B play four matches, what is the probability that:**

   i. Team A will lose all four matches?

   ii. Team A will win at least one match? [PU 2007 Fall]

Solution 4

Let us suppose team 'A' will loose the match is taken as success then we have,

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

i. Probability that team A will lose all 4 matches is

$$P(4) = {}^4C_4 \left(\frac{3}{44}\right)^4 \left(\frac{1}{4}\right)^4 (4-4) = \frac{3^4}{4^4} = 0.32$$

ii. Probability that team A will win at least one match

$P(\bar{4}) = 1 - P(4) = 1 - 0.32 = 0.68$

5. **There are 4 possible answers to each question and of which one is correct. (i) an intelligent student knows 80% of the answers. If intelligent student gets the correct answer, what is the probability that he is guessing? (ii) A weak student knows 10% of the answer. If a weak student gets the correct answer, what is the probability that he is guessing?** [PU 2015 Fall]

Solution 5

1st Part

If $E_1, E_2, E_3, \ldots, E_n$ be mutually disjoint events of sample space S with $P(E_i) \neq 0$, i = 1, 2, 3, ..., n, then for any arbitary A, which is a subset of S such that $P(A) > 0$, we have,

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

2nd Part

Let E_1 denote the event of know the correct answer and E_2 denote the event of guessing the correct answer and A denote the event of getting the correct answer. Then, we have,

$P(A/E_1) = 1 \qquad P(A/E_2) = 1/4 = 0.25, \quad P(E_2/A) = ?$

Since, an intelligent student knows 80% of the answer then

$P(E_1) = 80\% = 0.8 \;\; P(E_2) = 1 - 0.8 = 0.2$

Now, by Bayes' theorem we have,

$$P(E_2/A) = \frac{P(E_2)\, P(A/E_2)}{P(E_1) \times P(A/E_1) + P(E_2) \times P(A/E_2)} = \frac{0.2 \times 0.25}{0.8 \times 1 + 0.2 \times 0.25} = \frac{0.05}{0.85} = 0.06$$

6. **Two production Plants, A and B, make wire Cables that are sent to a common distributors. 40% of the cables sent to the distributors come from plant A, and the remaining 60% from Plant B. among the cables product at plant A, 99% meet the strength specifications: among the cables produced at plant B 98% meet the strength specifications. (i) The distributor selects one cable at random**

Scanned by CamScanner

from among all cables in stock. If the cable selected is found to meet the strength specifications, what is the probability that the cable was product at plant A? (ii)Let A be the event that a cable is manufactured at plant A; similarly, B be the event that a cable is manufactured at plant B; let Y be the event that a cable meets the strength specifications; and let N be the event that a cable does not meet the strength specifications, are events A and Y independent? Give reason? [PU 2015 Spring]

Solution 6

Let E_1 and E_2 denote the events that a randomly selected wire cable are produced from plant A and plant B respectively. Also A denote the event that the product produced from plant A and B meet the strength specification then we have,

$P(E_1)) = 40\% = 0.4$, $P(E_2) = 60\% = 0.6$

$P(A/E_1) = 99\% = 0.99$ $P(A/E_2) = 98\% = 0.98$

i. The probability that thee randomly selected cable found to meet the strength specification, which was produced from plant A is denoted by $P(E_1/A)$. Then by Baye's theorem we have,

$$P(E_1/A) = \frac{P(E_1)\,P(A/E_1)}{P(E_1)\,P(A/E_1) + P(E_2)\,P(A/E_2)} = \frac{0.4 \times 0.99}{0.4 \times 0.99 + 0.6 \times 0.98} = \frac{0.396}{0.984} = 0.40$$

ii. Since A and B denote the events that a cable manufactured at plant A and B respectively. Then,

$P(A) = 40\% = 0.4$, $P(B) = 60\% = 0.6$

Also, Y denote the event that they meet the strength specification and N denote the event that they do not meet the strength specification. Then,

$P(Y/A) = 99\% = 0.99$ $P(N/A) = 0.01$

$P(Y/B) = 98\% = 0.98$ $P(N/B) = 0.02$

Also, we know that,

$$P(A/Y) = \frac{P(A)\,P(Y/A)}{P(A)\,P(Y/A) + P(B)\,P(Y/B)} = 0.4\% = P(A)$$

Since, $P(A/Y) = P(A)$ so the events A and Y are independent.

7. **The HAL Corporation wishes to improve the resistance of its personal computer to disk driver and key board failures. At present, the design of computer is such that disk -drive failures occur only on third as often as keyboard failures. The probability of simultaneous disk-drive and key board failures is 0.05.**

If the computer is 80% resistant to disk drive and /or keyboard failure, how low must the disk-drive failure.

If the key board is improved so that it fails only twice as often as the disk-drive (and the simultaneous failure probability is still 0.05), will the disk-drive failure probability from part (I) yield a resistance to disk-drive and /or keyboard failure higher or lower the 90%. [PU 2015 Spring]

Solution 7

Do yourself

Scanned by CamScanner

8. In an organization, there are 2 civil engineers, 4 IT officers and 2 Accountants. A committee of 4 members has to be formed. What is the probability that a committee contains i) 1 civil Engineer and 3 It officers? ii) No civil engineer? iii) at least one IT Officer? [PU 2013 Fall]

Solution 8

Let A, B, C denote the events that the selection of civil engineers, IT officers and Accountants respectively. Then,

n(A) = 2, n(B) = 4, N(C) =2, N(S) = 8

Since, we have to make the committee of 4 members,

i. The probability that a committee contains 1 civil engineer and 3 IT officers

$$P(\text{one A and three B}) = P(\text{one A}) \times P(\text{three B}) = \frac{{}^2C_1 \times {}^4C_3}{{}^8C_4} = \frac{2 \times 4}{70} = \frac{4}{35}$$

ii. The probability that a committee contains no engineers

P(no civil engineers) = n(4 IT and 0 Accountant) or (3 IT and 1 Accountant) or (2 IT and 24 Accountants)

$$= \frac{{}^4C_4 \times {}^2C_0}{{}^8C_4} + \frac{{}^4C_3 \times {}^2C_1}{{}^8C_4} + \frac{{}^4C_2 \times {}^2C_2}{{}^8C_4}$$

$$= \frac{1 + 4 \times 2 + 6 \times 1}{70} = \frac{15}{70} = \frac{3}{14}$$

iii. P(at least one IT officer) = 1 – P(no IT officer)

$$= 1 - p(\text{2 engineers and 2 Accountants}) = 1 - \frac{{}^2C_2 \times {}^2C_2}{{}^8C_4} = 1 - \frac{1 \times 1}{70} = \frac{69}{70}$$

9. In a Company there are 4 civil engineers,5 IT officers and 6 Finance officers. A committee of 3 members has to be formed from random selection. What is the probability that committee consists of

i) All civil engineers. ii) one civil engineer iii)at least one civil engineer [PU 2015 Fall]

Solution 9

Let A, B and C denotes the event of selecting civil engineers, IT officers and Finance officers. Then,

n(A) = 4, n(B) = 5, n(C) = 6 and n(S) = 15

i. Probability that the committee consists of all civil engineers is

$$P(\text{all civil engineers}) = \frac{{}^4C_3}{{}^{15}C_3} = \frac{4}{455} = 0.09$$

ii. $$P(\text{one civil engineer}) = \frac{\text{No. of ways off one civil engineer}}{\text{Total no. of ways}}$$

We can select one civil engineer in the following ways:

A(4)	B(5)	C(6)	No. of ways
1	2	0	${}^4C_1 \times {}^5C_2 \times {}^6C_0 = 4 \times 10 \times 1 = 40$
1	0	2	${}^4C_1 \times {}^5C_0 \times {}^6C_2 = 4 \times 1 \times 15 = 60$
1	1	1	${}^4C_1 \times {}^5C_1 \times {}^6C_1 = 4 \times 5 \times 6 = 120$
			220

Scanned by CamScanner

$\therefore$ P(one civil engineer) $= \frac{220}{{}^{15}C_3} = \frac{220}{455} = 0.48$

iii. P (at least one civil engineer) = 1 – P(no. civil engineer)

= 1 – P(2 IT and 1 Finance) or (1 IT and 2 Finance)

$$= 1 - \frac{{}^5C_2 \times {}^6C_1 + {}^5C_1 \times {}^6C_2}{{}^{15}C_3} = 1 - \frac{10 \times 6 + 5 \times 15}{455} = 0.7$$

10. A company has to plants to manufactures the screws. Out of 10000 screws, plant-I manufactures 80% of the screws and Plant-II manufactures 20%and Plant-I, 85% screws are rated standard quality. At Plant-II only 65% screws are rated standard quality. i) One screw is selected at random from the whole consignment and was found standard quality, what is the probability that the selected screw was manufactures by Plant-I? ii) What is the probability that the selected screws being defective? [PU 2013 Fall]

Solution 10

Let E_1 and E_2 denote the events that the screws is manufactured by plant-I and Plant II respectively. Then,

$P(E_1) = 80\% = 0.8$ and $P(E_2) = 20\% = 0.2$

Let A denote the event that screws are rated standard quality. Then,

$P(A/E_1) = 85\% = 0.85$ and $P(A/E_2) = 65\% = 0.65$

i. Probability that the randomly selected screw found to be standard quality was manufactured by Plant-I is $P(E_1/A)$. Then, by Bayes' law

$$P(E_1/A) = \frac{P(E_1)\,P(A/E_1)}{P(E_1)\,P(A/E_1) + P(E_2)\cdot P(A/E_2)} = \frac{0.8 \times 0.85}{0.8 \times 0.85 + 0.2 \times 0.65} = \frac{0.68}{0.81} = 0.84$$

ii. Probability that the selected screws being defective

= 1 – Probability that the selected screws being standard quality

$= 1 - P(A) = 1 - [P(E_1)\,P(A/E_1) + P(E_2)\,P(A/E_2)] = 1 - 0.81 = 0.19$

11. Three groups of children contain respectively 3 girls and 1 boy, 2 girls and 2 boys, 1 girl and 3 boy. One child is selected at random from each group. Show that the chance that the three selected consists of 1 girl and 2 boys is 13/22. [PU 2011 Fall]

Solution 11

We have given,

Group A = 1 boy 3 girls

Group B = 2 boys 2 girls

Group C = 3 boys 1 girl

If one child is drawn at random from each group. The possible outcomes of 2 boys and 1 girls can be drawn and shown by the following table:

Group A	Group B	Group C	Probabilities
1 Girl	1 Boy	1 Boy	3/4 × 2/4 × 3/4 = 9/32
1 Boy	1 Girl	1 Boy	1/4 × 2/4 × 3/4 = 3/32
1 Boy	1 Boy	1 Girl	1/4 × 2/4 × 1/4 = 1/32
		Total no. of probabilities	9/32 + 3/32 + 1/32 = 13/32

Scanned by CamScanner

12. From a group of 3 Nepalese, 4 Indians and 5 Americans, a subcommittee of 4 persons are selected at randomly. Find the probability that the sub-committee will consist of i) 3 Nepalese and 1 Indian ii) 1 Nepalese, 1 Indian and 2 American iii) 4 American **[PU 2016 Spring]**

Solution 12

Let N, I, A denotes the events of selecting Nepalese, Indian and American. Then,

$n(N) = 3, \quad n(I) = 4, \quad n(A) = 5, \quad n(S) = 12$

i. A sub committee consists of 3 Nepalese and 1 Indian is

$$P(3 \text{ Nepalese and } 1 \text{ Indian}) = \frac{{}^3C_1 \times {}^4C_1}{{}^{12}C_4} = \frac{1 \times 4}{495} = 0.008$$

ii. $$P(1 \text{ Nepalese, } 1 \text{ Indian and } 2 \text{ American}) = \frac{{}^3C_3 \times {}^4C_1 \times {}^5C_2}{{}^{12}C_4} = \frac{3 \times 4 \times 10}{495} = 0.24$$

iii. $$P(4 \text{ Americans}) = \frac{{}^5C_4}{{}^{12}C_4} = \frac{5}{995} = \frac{1}{99} = 0.01$$

13. A student knew only 60% of the questions in a test each with 5 answers. He simply guessed while answering the test. What is probability that he knew the answers to a question given that he answered it correctly? **[PU 2014 Spring]**

Solution 13

Let E_1 and E_2 denote the events that of knowing the correct answer and guessing the correct answer respectively and A denote the event of getting the correct answer. Then we have,

$P(E_1) = 60\% = 0.6$ and $P(E_2) = 40\% = 0.4$

Also, $P(A/E_1) = 1$ and $P(A/E_2) = 1/5 = 0.2$

The probability that he knew the answers to a question given that he answer correctly is given by

$P(E_1/A) = ?$

Now, by the Baye's law we have,

$$P(E_1/A) = \frac{P(E_1)\,P(A/E_1)}{P(E_1)\,P(A/E_1) + P(E_2)\,P(A/E_2)} = \frac{0.6 \times 1}{0.6 \times 1 + 0.4 \times 0.2} = 0.88$$

14. A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is 1/7 and that of wife's selection is 1/5. What is the probability that i) both of them will be selected ii) only one of them will be selected iii) none of them will be selected iv) at least one of them will be selected. **[PU 2011 Spring/2010 Spring/Purbanchal 2005/2010 Fall/TU 2065/2071]**

Solution 14

Let H and W denote the events that the selection off husband and wife in a interview. Then,

$P(H) = 1/7, P(W) = 1/5$

Since, they appear in the interview of same post. So, the event is mutually exclusive.

i. Probability of selecting both of them

$P(H \cap W) = P(H) \times P(W) = 1/7 \times 1/5 = 1/35$

Scanned by CamScanner

ii. Probability of selecting only one of them is,

$P[(H \cap \bar{W} \text{ or } (\bar{H} \cap W)] = P(H) \times P(\bar{W}) + P(\bar{H}) \times P(\bar{W})$

$$= \frac{1}{7} \times \left(1 - \frac{1}{5}\right) + \left(1 - \frac{1}{7}\right) \times \frac{1}{5} = \frac{1}{7} \times \frac{4}{5} + \frac{6}{7} \times \frac{1}{5} = \frac{2}{7}$$

iii. Probability off selecting at least one of them is

$P[H \cup W) = P(H) + P(W) - P(H \cap W) = \frac{1}{7} + \frac{1}{5} - \frac{1}{35} = \frac{11}{35}$

iv. Probability of selecting none of them is

$P(\overline{H \cup W}) = 1 - P(H \cup W) = 1 - \frac{11}{35} = \frac{24}{35}$

15. The probability of X, Y and Z becoming managers are 4/9, 2/9 and 1/3 respectively. The probability that the bonus scheme will be introduced if X, Y, and Z becoming managers are 3/10, 1/2 ,and 4/5 respectively.

i. What is the probability that the bonus scheme will be introduced?

ii. If the bonus scheme has been introduced what is the probability, the manager appointed was Y? [PU 2011 Spring]

Solution 15

Let E_1, E_2, E_3 denote the events that X, Y and Z being the managers, then

$P(E_1) = 4/9$ $P(E_2) = 2/9$ and $P(E_3) = 1/3$

Also, A denote the event that the bonus scheme will be introduced then,

$P(A/E_1) = 3/10$ $P(A/E_2) = 1/2$ and $P(A/E_3) = 4/5$

i. Probability that he bonus scheme will bee introduced is

$P(A) = P(E_1) \times P(A/E_1) + P(E_2)\, P(A/E_2) + P(E_3)\, P(A/E_3)$

$$= \frac{4}{9} \times \frac{3}{10} + \frac{2}{9} \times \frac{1}{2} + \frac{1}{3} \times \frac{4}{5} = \frac{2}{15} + \frac{1}{9} + \frac{4}{15} = \frac{23}{25}$$

ii. Probability of appointing Y when bonus scheme was introduced is

$P(E_2/A) = \frac{P(E_2) \times P(A/E_2)}{P(A)} = \frac{2/9 \times 1/2}{23/45} = 1/9 \times 45/23 = 5/23$

16. The chances of X, Y and Z becoming managers of a certain company are 4:2:3. The probability that bonus scheme will be introduced if X,Y and Z becomes managers are 0.3, 0.5 and 0.8 respectively. If the bonus scheme has been introduced what is the probability that i) X is appointed as the manager? ii) Y is appointed as the manager? [PU 2012 Fall]

Solution 16

Process is similar to Q. 18

17. Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4%, and 5% respectively. If an item is selected at random, find: i) the probability of defective item ii) the probability that the defective item is produced by machine A, produced by machine B and produced by machine C. [PU 2010 Spring]

Solution 17

Process is similar to Q. 18

Scanned by CamScanner

18. **A consignment of 15 pens contains 2 defective pens. Two pens are taken at random from the consignment without replacement. What is the probability that i) both are defective ii) both are non-defective** **[PU 2005 Fall]**

Solution 18

Here, we have given total no. of pens n(S) = 15

Out of 15 pens 2 are defective. Since 2 pens are taken at random without replacement then,

i. The probability that both pens will be defective is

$$P(\text{both defective}) = \frac{2}{15} \times \frac{1}{14} = \frac{1}{105} = 0.0095$$

ii. The probability that both are non-defective is

$$1 - p(\text{both defective}) = 1 - \frac{1}{105} = 0.99$$

19. **There are three machines A, B and C producing 1000, 2000 and 3000 articles per hour respectively. These machines are known to be producing 1%, 2% and 3% defective articles respectively. One article is selected at random from an hour production of the three machines and found to be defective. What is the probability that the article is produced from i) Machine A ii) Machine B** **[PU 2010 Spring]**

Solution 19

Same question of Ex. 7.2 Q. 33

20. **A manufacturing firm produces steel pipes in three plants with daily production of volume 500,1000, and 2000 units respectively. According to the past experience it is known that the fraction of defective outputs produced by three plants are respectively 0.005, 0.008 and 0.010. if pipe is selected from a day's total production and found to be defective. From which plant the defective pipe is expected to have been produced?** **[PU 2016 Spring]**

Solution 20

Same question of Ex. 7.2 Q.No. 31

21. **Problem of statistics is given to three students A, B, and C whose chances of solving it are 1/3, 1/4 and 1/5 respectively. Then find the probability that i) All of them can solve ii) none of them can solve iii) Exactly on can solve iv) Problem will be solved.** **[PU2010 Spring]**

Solution 21

Let A, B and C denote the events that the students A, B and C solve the given statistics problem. Then,

$$P(A) = 1/3, \qquad P(B) = 1/4 \qquad \text{and} \qquad P(C) = 1/5$$

i. Probability that all of them solve the problem is

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C) = \frac{1}{3} \times \frac{1}{4} \times \frac{1}{5} = \frac{1}{60}$$

ii. probability that none of them solve the problem is

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

Scanned by CamScanner

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

iii. Probability that exactly one can solve the problem is

$P[(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)]$

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

$$= \frac{1}{3} \times \left(1-\frac{1}{4}\right)\left(1-\frac{1}{5}\right) + \left(1-\frac{1}{3}\right) \times \frac{1}{4}\left(1-\frac{1}{5}\right) + \left(1-\frac{1}{3}\right)\left(1-\frac{1}{4}\right) \times \frac{1}{5}$$

$$= \frac{1}{3} \times \frac{3}{4} \times \frac{4}{5} + \frac{2}{3} \times \frac{1}{4} \times \frac{4}{5} + \frac{2}{3} \times \frac{3}{4} \times \frac{1}{5} = \frac{1}{5} + \frac{2}{15} + \frac{1}{10} = \frac{13}{30}$$

iv. Probability that the problem will be solved is given by

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

$$= \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{3} \times \frac{1}{4} - \frac{1}{4} \times \frac{1}{5} - \frac{1}{5} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{4} \times \frac{1}{5} = \frac{3}{5}$$

22. A Problem of statistics is given to three students A, B, and C whose chances of solving it are 1/2, 3/4 and 2/3 respectively. if they independently solve it, what is the probability that i) All of them can solve ii) None of them can solve iii) Exactly on can solve)Problem will be solved [PU 2017 Spring]

Solution 22

Process is similar to Q. 24

23. The contents of urns I, II and III are as follows: 1 white, 2 black and 3red balls 2 white balls, 1 black and 1red balls, 4 White, 5 black and 3 red balls. I) One urn is chosen at random and two balls are drawn. They happen to be white and red. What is the probability that they come from urns I, II or III? [PU 2009 Spring]

Solution 23

Let E_1, E_2 and E_3 denote the events of selecting the first, second and third urn respectively. Also let A denote the event that two balls taken from selected urn happen to be white and red. Then, we must have

$P(E_1) = \frac{1}{3}$, $P(E_2) = \frac{1}{3}$ and $P(E_3) = \frac{1}{3}$

And $P(A/E_1) = \frac{C(1,1) \times C(3,1)}{C(6,2)} = \frac{1 \times 3}{15} = 0.2$

$P(A/E_2) = \frac{C(2,1) \times C(5,1)}{C(4,2)} = \frac{2 \times 1}{16} = 0.33$

$P(A/E_3) = \frac{C(4,1) \times C(3,1)}{C(12,2)} = \frac{4 \times 3}{66} = 0.18$

Now, $P(A) = P(A/E_1) + P(A/E_2) + P(A/E_3) = 0.2 + 0.33 + 0.18 = 0.71$

The probability that getting red and white balls that was from urn I is

$$P(E_1/A) = \frac{P(E_1)\,P(A/E_1)}{P(A)} = \frac{0.33 \times 0.1}{0.71} = \frac{0.066}{0.71} = 0.093$$

Similarly, $P(E_2/A) = \frac{P(E_2)\,P(A/E_2)}{P(A)} = \frac{0.33 \times 0.33}{0.71} = 0.153$

Also, $P(E_3/A) = \frac{P(E_3)\,P(A/E_3)}{P(A)} = \frac{0.33 \times 0.18}{0.71} = 0.084$

Scanned by CamScanner

24. An oil exploration company currently has two active projects, one in Asian and the other in Europe. Let A be the event that the Asian project is successful and B be the event that the European projects is successful. Suppose that A and B are independent event with P(A) = 0.4 and P(B) = 0.7

i. If the Asian project is not successful, what is the probability that the European project is also not successful? Explain your reasoning.

ii. What is the probability that at least one of the two projects will be successful?

iii. Given that at least one of the two projects is successful, what is the probability that only the Asian project is successful? [PU 2016 Fall]

Solution 24

Let A and B denote the event that the Asian Project and the European Projects are successful respectively. Then,

$P(A) = 0.4$ and $P(B) = 0.7$

Since, the events A and B are independent then,

i. Probability that the European project is not successful given that the Asian project is not successful.

$$P(\bar{B}/\bar{A}) = P\frac{P(\bar{A} \cap \bar{B})}{P(\bar{A})} = \frac{P(\bar{A}) \times P(\bar{B})}{P(\bar{A})} = 1 - 0.7 = 0.3$$

ii. The probability that at least one of two projects be successful is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7 + 0.4 - 0.7 \times 0.4 = 1.1 - 0.28 = 0.82$$

iii. $$P(A \cap \bar{B}/ A \cup B) = \frac{P(A \cap \bar{B}) \cap n(A \cup B)}{P(A \cup B)} = P(A) \times P(\bar{B}) = 0.4 \times (1 - 0.7) = 0.12$$

25. Three roads construction firms X,Y and Z, bid for a certain contract. From past experience, it is estimated that the probability that X will be awarded the contract is 0.40, while for Y and Z the probabilities are 0.35 and 0.25. If X does receive the contract, the probability that the work will be satisfactorily completed on time is 0.75. for Y and Z these probabilities are 0.80 and 0.70

i. What is the probability that the work will be completed satisfactorily?

ii. It turns out that the work was done satisfactorily. What is the probability that Y was awarded the contract? [PU 2016 Fall]

Solution 25

Do yourself

26. Two brothers: Mr. X and Mr. Y appear in an interview for getting scholarship The scholarship can be provided for two persons. The probability of getting scholarship by Mr. X and is 1/7 and getting by Mr. Y is 1/5. What is the probability that i) Both of them will get scholarship ii) Only one of them will get scholarship iii) None of them will get scholarship. [PU 2010 Fall]

Solution 26

Process is similar to Q. 24

Scanned by CamScanner

27. From a lot containing 50 items, 10 of which are defective, 4 are chosen at random. Find the probability that (i) at least one will be defective (ii) one or two will be defective. **[PU 2006 Spring]**

Solution 27

Here, we have given a lot contains 50 items out of them 10 are defective. Let getting and defective item them

$$P = \frac{10}{50} = 0.2, \qquad q = 1 - p = 1 - 0.2 = 0.8, n = 4$$

i. Probability that at least one will defective is

$P(\geq 1) = 1 - P(0) = 1 - {}^4C_0(0.2) \times (0.8)^{4-0} = 1 - 1 \times 0.2 \times 0.4096 = 1 - 0.08192 = 0.92$

ii. Probability that one or two will be defective is

$$\begin{aligned} P(1 \text{ or } 2) &= P(1) + P(2) \\ &= {}^4C_1 (0.2)^1 (0.8)^{4-1} + {}^4C_2 (0.2)^2 \times (0.8)^{4-2} \\ &= 4 \times 0.2 \times 0.512 + 6 \times 0.04 \times 0.64 = 0.4096 + 0.1536 = 0.5632 \end{aligned}$$

28. Define sample space. A card is drawn from a bag containing 10 cards numbered 1 to 10. Let A denotes the event that the number drawn in the card is less than 7. B is the event that the number on the card is multiple of 2. Find the probabilities of the following events. I) AC (ii) B (iii) (A, ∩BC) (iv) (A ∪ B) (v) (AC∩ B) (vi (B/A) . **[PU 2003 Fall]**

Solution 28

First Part

The set of all possible outcomes of a statistical experiment is called the sample space.

Second part

$P(A) = 6/10 = 0.6$ and $P(B) = 5/10 = 0.5$

i. $P(A^C) = 1 - P(A) = 0.4$

ii. $P(B) = 0.5$

iii. $P(A \cap B^C) = P(A) - P(A \cap B)$
$= 0.6 - 0.3 = 0.3$

iv. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.6 + 0.5 - 0.3 = 0.8$

v. $P(A^C \cap B) = P(B) - P(A \cap B)$
$= 0.5 - 0.3 = 0.2$

vi. $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.6} = 0.5$

29. Prove that: $P(B/A) = \frac{P(A \cap B)}{P(A)}$ Where P (B/A) is conditional probability of B when A has already happened. **[PU 2010 Spring]**

Solution 29

See in theoretical part of this chapter

Scanned by CamScanner

30. Define complementary event. For two events, A and B, Prove that: $P(\bar{A} \cap B) = P(B) - P(A \cap B)$. **[PU 2005 Fall/2011 Fall]**

Solution 30

First part

Any two events are said to be complementary events if (i) they are mutually exclusive and (ii) they are collectively exhaustive.

The complement of an event E, means the non-occurrence of event E and is denoted by $\bar{E}$.

Second part

For any two sets A and B we have,

$B = (A \cap B) \cup (\bar{A} \cap B)$ [[From Venn-diagram]

$P(B) = P[(A \cup B) \cup (\bar{A} \cap B)] - P[(A \cap B) \cup (\bar{A} \cap B)]$

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

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

31. State and prove addition law of probability. **[PU 2007 Fall]**

Solution 31

See in thee theoretical part of this chapter.

32. What do you mean by dependent variable? If A and B are two events defined in sample space and are not disjoint. Prove $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. **[PU 2009 Spring]**

Solution 32

See in the theoretical part of this chapter.

33. Write a short note about conditional probability. State and prove Bayes' theorem. **[PU 2010 Fall/PU 2012 Fall/2006 Spring/2003 Fall/2016 Spring/2017 Spring/2009Fall]**

Solution 33

First part: See in Q. 38

Second part: See in theoretical part of this chapter

34. State Bayes' theorem. In a class of 75 students, 15 students were considered to be very intelligent, 45 as medium and rest below the average. The probability that a very intelligent student fail in examination is 0.005, the medium student failing has probability 0.05 and corresponding probability for a below average is 0.15. If a student is known to have passed the examination, what is the probability that he is below the average? **[PU 2013 Spring/2017 Fall]**

Solution 34

First part: If $E_1, E_2, E_3, \ldots, E_n$ be mutually disjoint events of sample space S with $P(E_i) \neq 0$, i = 1m 2m 3, ..., n then for any arbitrary event A which is subset of S such that $P(A) > 0$ we have

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

Scanned by CamScanner

Second part

let E_1, E_2 and E_3 denote the events that the students of a class being very intelligent, medium and below the average respectively then we have,

$$P(E_1) = \frac{15}{75} = 0.2, \quad P(E_2) = \frac{45}{75} = 0.6 \text{ and } P(E_3) = \frac{15}{75} = 0.2$$

Let A denote the event that the students passed the exam then we have,

$P(A/\bar{E}_1) = 0.05 \Rightarrow 1 - P(A/E_1) = 0.05 \Rightarrow P(A/E_1) = 0.995$

$P(A/\bar{E}_2) = 0.05 \Rightarrow 1 - P(A/E_2) = 0.05 \Rightarrow P(A/E_2) = 0.95$

$P(A/E_3) = 0.15 \Rightarrow 1 - P(A/E_3) = 0.15 \Rightarrow P(A/E_3) = 0.85$

Then, the probability that the below average will pass the examination is

$$P(E_3/A) = \frac{P(E_3)\,P(A/E_3)}{P(E_1) \times P(A/E_1) + P(E_2) \times P(A/E_2) + P(E_3) \times P(A/E_3)}$$

$$= \frac{0.2 \times 0.85}{0.2 \times 0.995 + 0.6 \times 0.95 + 0.2 \times 0.85} = \frac{0.17}{0.939} = 0.18$$

35. State Bayes' theorem. In a certain recruitment test, there are multiple choice question. There are 4 possible answers to each question and of which one correct. An intelligent student knows 80% of the answers. If intelligent student gets the correct answers, what is the probability that he is guessing? [PU 2015 Fall]

Solution 35

Do yourself

Short Questions

36. Write short notes on conditional probability

[PU 2006 Spring/2007 Spring/2009 Fall/2009 Spring]

Let A and B are any two events of sample space S. If the probability of occurrence of event A is influenced by the occurrence or non-occurrence of the other events is called the dependent events. The probabilities associated with such events are called conditional probabilities. Also we can define the conditional probability as follows:

If A and B are any events in sample space S and $P(B) \neq 0$ then the conditional probability off A given B is already occurred is denoted by $P(A/B)$ and defined as

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \; P(B) > 0$$

Similarly, the conditional probability of B given A is already occurred is denoted and defined by $P(B/A) = \frac{P(A \cap B)}{P(A)}$, $P(A) > 0$.

37. Write short notes on depend and independent events. [PU 2007 Fall]

Events are said to be independent of each other if the occurrence or non-occurrence off any one of them does not effect the occurrence or non-occurrence of the other. For example, In a tossing a two or more coins the occurrence of head or tail off the first coin do not effect the occurrence of head or tail of the other coins. So, the events are independent of each side.

Scanned by CamScanner

Events are said to be dependent of each other, if the occurrence or non-occurrence off any one of them effects the occurrence or non-occurrence of the other.

For example: If two cards are drawn from the deck of 52 cards without replacement, then the occurrence of first card effects the occurrence of second card also we can say the result of second draw is depends upon the results of first card. So, they are called dependent events.

38. Write short notes on approaches to probability. **[TU 2010 Spring]**

See in the theoretical part of this chapter

39. Write short notes on sample space. **[TU 2005 Spring/2011 Spring/2017 Spring]**

The set of all possible outcomes of a statistical experiment is called the sample space. It is denoted by S. Each outcomes in a sample space is called an element or a member of the sample space or sample point. Sample space is also called the universal set or probability set. Each element of the sample space S denotes the possible outcomes of the experiment.

There are different types of sample space.

i. A sample space S is said to be finite if it contains a finite number of outcomes.

ii. A sample space S is said to be discrete if it contains finite or countable infinite set of outcomes.

iii. A sample space S is said to be continuous if it contains an interval of outcomes.

Examples:

i. The sample space of tossing two coins is

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

ii. The sample space of rolling a die is

$S = \{1, 2, 3, 4, 5, 6\}$, etc.

Scanned by CamScanner

8
Chapter

Random Variable and Probability Distribution

Random Variable

A random variable is a real valued function f(x), of the elements of the sample space S, where x is an element of S. The domain of the random variable is sample space and range of the random variable is set of real number R. A random variable is usually denoted by capital letters and their possible values by lower case letters.

For example;

In a toss of two coins, if we define the random variable X as the number of heads then the random variable is given as,

$$X(x) = \begin{cases} 0 & \text{if } x = TT \\ 1 & \text{if } x = TH, HT \\ 2 & \text{if } x = HH \end{cases}$$

Properties of Random Variable

Some important and most useful properties of random variables are as follow:

i. If X is a random variable and a, b are any two constants,

Then aX, $aX \pm b$, $\frac{1}{X}$, X^2 are also random variables.

ii. If X and Y are two random variables then

$X + Y$, $X - Y$, aX, bY, XY, $\frac{X}{Y}$ are also random variables.

iii. If $X_1, X_2, \ldots, X_n$ are n random variables, then

$U_n = \max\{X_1, X_2, \ldots, X_n\}$ and $V_n = \min\{X_1, X_2, \ldots, X_n\}$ are also random variable.

Scanned by CamScanner

Types of Random Variables

There are two types of random variables

i. Discrete random variables

ii. Continuous random variables

Discrete Random Variables

A random variable which can take a finite number of values in the interval of domain is called the discrete random variables.

For example;

If we toss a coin then the variable can take only two values defined by

$$X(x) = \begin{cases} 0 & \text{if } x = T \\ 1 & \text{if } x = H \end{cases}$$

In a rolling of a dice, only six values of the variable is possible defined as

$$X(x) = \{x: x = 1, 2, 3, 4, 5, 6\}$$

Continuous Random Variables

A random variable X, which can take any values in its domain or in an interval or the Union of intervals on the real line, is called continuous random variables.

For example,

The weight of middle-aged people in Nepal lying between 40 kg and 150 kg is continuous variable i.e. $X(x) = \{x: 40 \le x1 \le 50\}$.

Probability Distribution

A 'probability distribution' of a random variable is a table or an equation or a graph or a formula that links each possible value that a random variable can assume with its probability of occurrence.

Here we study two types of probability distribution.

i. Discrete probability distribution.

ii. Continuous probability distribution.

Discrete Probability Distribution

Let X be a one-dimensional discrete random variable with possible values $x_1, x_2, x_3, \ldots$ in the sample spaces S band the probabilities of each x_i be denoted and defined by

$P(X = x_i) = P(x_i) = p_i$

The function $P(X = x_i) = P(x_i) = p_i$ is said to be the probability mass function (pmf) of discrete random variable X if it satisfy the condition

i. $P(x_i) \ge 0$ for all i, i = 1, 2, 3, ... ii. $\sum_{i=0}^{\infty} P(x_i) = 1$

The probability distribution of a discrete random variable X is a list of possible values of X together with there probabilities $P(X = x_i)$

Scanned by CamScanner

Probability Distribution Function

Let X be one dimensional discrete random variable with probability mass function (pmf) P (x). Then the cumulative discrete function of discrete random X denoted by F_x (x) or F (x) and defined by

$$F(x_i) = P(X \le x_i) = P(x_0) + P(x_1) + P(x_2) + \ldots + P(x_i) = \sum_{j=0}^{i} P(x_i)$$

is called the discrete probability distribution function.

Properties of probability distribution function (cumulative distribution function)

If a and b are two constants such that a < b and F is distribution function then,

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

ii. $0 \le F(x_i) \le 1$

iii. $x_i \le x_j \Rightarrow F(x_i) < F(x_j)$

iv. $F(-\infty) = \lim_{x_i \to -\infty} F(x_i) = 0$

v. $F(\infty) = \lim_{x_i \to \infty} F(x_i) = 1$

Continuous Probability Distribution

Let X be a continuous random variable with possible values $x \in A$ and the probabilities of each $x \in A$ be denoted and defined by $P(x \in A) = \int_A f(x)\,dx$

Then the function f (x) is said to be probability density function (pdf) of X if it satisfy the condition

i. $f(x) \ge 0$ for all x and

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

The probability density function of a random variable X usually denoted by f_x (x) or f (x).

Properties of continuous probability density function

i. $f(x) \ge 0$ for all $x \in R$

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

iii. $P(E) = \int_E f(x)\,dx$, for any well defined event E.

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

Continuous Probability Distribution Function

If X is a continuous random variable with p.d.f f (x) then the function denoted and defined by $F_x = P(X \le x)$

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

Scanned by CamScanner

is called probability distribution function or simply distribution function or cumulative distribution function (c.d.f) of the continuous random variable X.

Properties of continuous distribution function

i. $0 \le F(x) \le 1, \qquad -\infty < x < \infty$

ii. $F'(x) = f(x) \ge 0$

iii. $F(-\infty) \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 0$

iv. $F(+\infty) \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \int_{\infty}^{x} f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 1$

v. F (x) is continuous and non decreasing function.

Exercise - 8

1. What do you mean by random variable? Write down the types of random variables.

Solution 1

A rule that assigns a real number to each outcome (sample point) of a random experiment is called a random variable. Hence the random variable is a real valued function X(x) of the elements of the sample space S where x is the elements of S.

For example, consider a random experiment as tossing a coin. Let X be a real valued function of getting a head and x be the out comes of random experiment then,

$$X = X(x) = \begin{cases} 1 & \text{if} \quad x = H \\ 0 & \text{if} \quad x = T \end{cases}$$ is called the random variable.

There are two types of random variables namely (i) Discrete random variables (ii) Continuous random variable

2. What are the differences between continuous and discrete random variables?

Solution 2

There are two types of random variables namely (i) Discrete random variable (ii) Continuous random variables

Discrete Random Variable	Continuous Random Variable
(i) A real valued function defined on a discrete sample space is called a discrete random variable.	(i) A real valued function defined on a continuous sample space is called a continuous Random variable.
(ii) It can take a finite number of values in an interval of the domain.	(ii) It can take infinite number of values in an interval of the domain.
(iii) It represents the counted data.	(iii) It represents the measured data.
(iv) It takes only integral value.	(iv) It takes all the real value.

Scanned by CamScanner

3. **What are differences between p.m.f and cumulative probability distribution (discrete)?**

Solution 3

Probability Mass Function	Cumulative Probability distribution
(i) Let X be a one dimensional discrete random variable which associates with a finite number of values satisfying the conditions.	(i) A function of one dimensional discrete random variable X denoted and defined by $F(x_i) = P(X \le x_i$
(a) $P(x_i) \ge 0$ for all i, (b) $\sum_{i=0}^{n} P(x_i) = 1$ is called the probability mass function	$= P(x_0) + P(x_1) + ... + P(x_i)$ is called the cumulative probability distribution.
(ii) Denotes the single probability.	(ii) It denoted the sum of probabilities.
(iii) It is frequency distribution function.	(iii) It is probability distribution function.
(iv) It is empirical and more experimental.	(iv) It is theoretical and depends on the assumptions.

4. **What are the difference between p.d.f and cumulative probability distribution (continuous)?**

Solution 4

Probability density function	Cumulative Probability distribution
(i) If X is continuous variable and f(x) is continuous function of x then $\int f(x)\,dx$ gives the probability of the event that x lies in the interval $\left(x-\frac{dx}{2}, x+\frac{dx}{2}\right)$ then f(x) is said to be probability density function.	(i) If X is continuous random variable with probability density function f(x) then the function defined by $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$ is called the cumulative frequency distribution.
(ii) It is also known as the frequency function.	(ii) It is also known as the probability distribution function.
(iii) It denotes the single probability.	(iii) It denotes the cumulative probability.
(iv) It is empirical and the result of experiment.	(iv) It is theoretical and depends up the assumptions.
(v) It is shows the additional random variations.	(v) It is shows predictable variation with the values of variable.

5. **A random variable X has following probability function.**

Value of X, x:	−2	−1	0	1	2	3
P (x):	0.1	K	0.2	2K	0.3	K

i. Find the value of K.

ii. Find $P(X < 2)$, $P(-2 < X < 2)$, $P(0 \le X < 2)$, $P(X \ge -2)$

iii. Find F (− 2), F (0) and F (3)

iv. Construct the c.p.f F(X)

Solution 5

(i) We know that the sum of all Probabilities is equal to 1. Therefore,

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

or, $4k = 1 - 0.6$

or, $k = \frac{0.4}{4} = 0.1$

(ii) (a) For $P(X < 2) = P(-2) + P(-1) + P(0) + P(1)$

Scanned by CamScanner

$= 0.1 + 0.1 + 0.1 + 0.2 + 0.2 = 0.6$

(b) $P(-2 < x < 2) = P(-1) + P(0) + P(1)$

$= k + 0.2 + 2k = 0.1 + 0.2 + 0.2 = 0.5$

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

$= 0.2 + 0.2 = 0.4$

(d) $P(X \ge -2) = P(-2) + P(-1) + P(0) + P(1) + P(2) + P(5) = 1$

(iii) (a) $F(-2) = P(X \le -2) = P(X = -2) = 0.1$

(b) $F(0) = P(X \le 0) = P(X = -2) + P(X = -1) + P(X = 0)$

$= 0.1 + 0.1 + 0.2 = 0.4$

(c) $F(3) = P(X \le 3) = P(X = -2) + P(X = -1) + P(X=0) + P(X=1) + P(X=2) + P(X = 3) = 1$

(iv) The cumulative probability function.

X	−2	−1	0	1	2	3
P(X)	0.1	0.1	0.2	0.2	0.3	0.1
P(X)	0.1	0.2	0.4	0.6	0.9	1.0

6. **From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Answer the following when the sample is drawn without replacement.**

i. Find the probability distribution of X

ii. Find $P(X \le 1)$, $P(X < 1)$ and $P(0 < X < 2)$

Solution 6

Let X denote the random variable of number of defective item in the sample. Then,

the set of values of random variable of getting defective lot, $X = \{0, 1, 2, 3\}$.

$$P(X = 0) = \frac{{}^3C_0 \times {}^7C_4}{{}^{10}C_4} = \frac{35}{210} = \frac{1}{6},\ P(X = 1) = \frac{{}^3C_1 \times {}^7C_3}{{}^{10}C_4} = \frac{3 \times 35}{210} = \frac{1}{2}.$$

$$P(X = 2) = \frac{{}^3C_2 \times {}^7C_2}{{}^{10}C_4} = \frac{3 \times 21}{210} = \frac{3}{10},\ P(X = 3) = \frac{{}^3C_3 \times {}^7C_1}{{}^{10}C_4} = \frac{7}{210} = \frac{1}{30}.$$

The probability distribution function of X is given below:

X	0	1	2	3
P(X = x)	1/6	1/2	3/10	1/30

(ii) (a) $P(X \le 1) = P(0) + P(1) = \frac{1}{6} + \frac{1}{2} = \frac{1+3}{6} = \frac{2}{3}$

(b) $P(X < 1) = P(0) = \frac{1}{6}$

(c) $P(0 < X < 2) = P(X = 1) = \frac{1}{2}$

7. **A random variable X has the following probability functions.**

Value of X, x:	0	1	2	3	4	5	6	7
P (x):	0	K	2K	2K	3K	K2	2K2	7K2 + K

i. Find K

ii. Evaluate $P(X < 6)$, $P(X \ge 6)$ and $P(0 < X < 5)$

Scanned by CamScanner

iii. Find the distribution function of X.

Solution 7

(i) We know that the sum of all probabilities is equal to 1. So,

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

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

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

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

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

Either $k + 1 = 0$ or, $10k - 1 = 0$

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

Since k cannot be negative. So, $k = \frac{1}{10}$.

(ii) (a) $P(X < 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

$$= 0 + k + 2k + 2k + 3k + k^2 = k^2 + 5k = \frac{1}{100} + \frac{8}{10} = \frac{18}{100} = 0.18$$

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

$$= 2k^2 + 7k^2 = 9k^2 + k = \frac{9}{100} + \frac{1}{10} = \frac{19}{100} = 0.19$$

(c) $P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

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

$$= 8k = \frac{8}{10} = 0.8$$

(iii) The distribution function of X is given as follows.

X	0	1	2	3	4	5	6	7
P(X = x)	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17
f(X)	0	0.1	0.3	0.5	0.8	0.81	0.83	1

8. Let P (x) be the probability function of a discrete random variable which assumes the value x_1, x_2, x_3, and x_4 such that

$$2P(x_1) = 3P(x_2) = P(x_3) = 5P(x_4)$$

Find the probability distribution and probability distribution function.

Solution 8

Here we have given the values of the random variable are $X = \{x_1, x_2, x_3, x_4\}$

Also, $2P(x_1) = 3P(x_2) = P(x_3) = 2P(x_4)$

Suppose $P(x_3) = k$ then $2P(x_1) = k \Rightarrow P(x_1) = \frac{k}{2}$

$3P(x_2) = k$ then $P(x_2) = \frac{k}{3}$, $5P(x_4) = k \Rightarrow P(x_4) = \frac{k}{5}$

Since the sum of all probabilities is 1.

So, $P(x_1) + P(x_2) + P(x_3) + P(x_4) = 1$,

Scanned by CamScanner

$$\frac{k}{2}+\frac{k}{3}+\frac{k}{5}=1 \qquad \Rightarrow \frac{61k}{30}=1 \Rightarrow k=0.49$$

The probability distribution and probability distribution function is given as follows:

X	x_1	x_2	x_3	x_4
P(X)	0.245	0.163	0.49	0.098
f(X)	0.245	0.408	0.898	1

9. If $P(x)=\frac{x}{15}$, $\quad x=1,2,3,4,5$

$\quad =0 \quad$ else where

Find i. $P\{x=1 \text{ or } 2\}$ and ii. $P\left\{\frac{1}{2}<X<\frac{5}{2}/X>1\right\}$

Solution 9

We have given $P(X)=\begin{cases} =\frac{x}{15}, & x=1,2,3,4,5 \\ =0 & \text{else where} \end{cases}$

(i) $P([X=1 \text{ or } 2]=P(X=1)+P(X=2) \quad =\frac{1}{15}+\frac{2}{15}=\frac{3}{15}=\frac{1}{5}$

(ii) $$P\left\{\frac{1}{2}<X<\frac{5}{2}/X>1\right\}=\frac{P\left[\left(\frac{1}{2}<X<\frac{5}{2}\right)\cap(X>1)\right]}{P(X>1)}$$

$$=\frac{P\left(1<X<\frac{5}{2}\right)}{P(X>1)}=\frac{P(X=2)}{1-[(X=0)+P(X=1)]}=\frac{215}{1-\left[0+\frac{1}{15}\right]}=\frac{2}{15}\times\frac{15}{14}=\frac{1}{7}$$

10. A random variable X has the density

$$f(x)=a+bx^2, \qquad 0<x<1$$

Determine 'a' and 'b' so that $a+b=2$

Solution 10

We have given probability density function:

$f(x)=a+bx^2, 0<x<1 \qquad$ so that $a+b=2$.

Then, we know that,

(i) $0 \le f(x) \le 1$

(ii) $\int_R f(x)\,dx=1 \qquad \Rightarrow \int_0^1 (a+bx^2)=1$

$$\Rightarrow \left[ax+\frac{bx^3}{3}\right]_0^1=1$$

Scanned by CamScanner

$$\Rightarrow a + \frac{b}{3} = 1 \Rightarrow 3a + b = 3 \Rightarrow (a + b) = 2a = 3 \Rightarrow 3 + 2a = 3$$

$$\Rightarrow a = 0.5 \qquad \text{then } b = 1.5$$

11. Let X be a random variable with p.d.f.

$$\begin{aligned} f(x) &= ax && 0 \le x \le 1 \\ &= a && 1 \le x \le 2 \\ &= -ax + 3a && 2 \le x \le 3 \\ &= 0 && \text{otherwise} \end{aligned}$$

Compute

i. the constant 'a' and ii. $P(X \le 1.5)$ [P.U. 2010 Spring]

Solution 11

By the given the probability density function (p.d.f) we have,

i. (a) $0 \le f(x) \le 1$ for all values of x.

(b) $\int_R f(x)\,dx = 1$

or, $\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$

or, $0 + \int_{0}^{1} ax\,dx + \int_{1}^{2} a\,dx + + \int_{2}^{3} (-ax + 3a)\,dx + 0 = 1$

or, $\left[\frac{ax^2}{2}\right]_0^1 + [ax]_1^2 + \left[\frac{-ax^2}{2} + 3ax\right]_2^3 = 1$

or, $\frac{a}{2} + a[2-1] - \frac{a}{2}[3^2 - 2^2] + 3a(3-2) = 1$

or, $\frac{a}{2} + a - \frac{5a}{2} + 3a = 1$

or, $2a = 1 \quad \Rightarrow \quad a = ½$

(ii) $P(X \le 1.5) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{1.5} f(x)\,dx$

$$= 0 + \int_{0}^{1} \frac{1}{2} x\,dx + \int_{1}^{1.5} \frac{1}{2}\,dx$$

$$= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 + \left[\frac{x}{2}\right]_1^{1.5} = \frac{1}{4} + \frac{1}{2}[1.5 - 1] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

12. Given the p.d.f,

Scanned by CamScanner

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

Show that

i. $k = 6$

ii. $P(0.3 < X < 0.5) = 0.104$

iii. $P(X < 0.4) = 0.352$

Solution 12

By the given probability density function, we have

i. $\int_R f(x)\,dx = 1$

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

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

or, $k\left[\frac{x^2}{2}\right]_0^1 - k\left[\frac{x^3}{3}\right]_0^1 = 1$

or, $k\left[\frac{1}{2} - \frac{1}{3}\right] = 1$

or, $k = 6.$

(ii) $P(0.3 < X < 0.5) = \int_{0.3}^{0.5} f(x)\,dx = \int_{0.3}^{0.5} (6x - 6x^2)\,dx$

$= 6\left[\frac{x^2}{2}\right]_{0.3}^{0.5} - 6\left[\frac{x^3}{3}\right]_{0.3}^{0.5}$

$= 3 \times)0.2)^2 - 2(0.2)^3$

$= 0.12 - 0.016 = 0.104$

(iii) $P(X < 0.4) = \int_{0}^{0.4} f(x)\,dx = \int_{0}^{0.4} (6x - 6x^2)\,dx$

$= 6\left[\frac{x^2}{2}\right]_{0.4}^{0} - 6\left[\frac{x^3}{3}\right]_{0.4}^{0}$

$= 3 \times (0.4)^2 - 2 \,.\, (0.4)^3 = 0.352$

13. **Suppose X is a r.v having p.m.f p (x) given as**

x	−3	−1	0	1	2	3	5	8
P (x)	0.1	0.2	0.15	0.2	0.1	0.15	0.05	0.05

Compute the probability

i. X is negative ii. X is even

iii. X take a value between 1 and 8 inclusive

iv. $P(X = 3/X \le 0)$ v. $P(X \ge 3/X > 0)$

Scanned by CamScanner

Solution 13

(i) P(X is negative) = P(X = – 3) + P(P(X = – 1) = 0.1 + 0.2 = 0.3

(ii) P(X is even) = P(X = 0) + P(P(X = 2) + P(X = 8) = 0.15 + 0.1 + 0.05 = 0.3

(iii) P(X is take a value between 1 and 8 inclusive)

$= P(1 \le X \le 8) = P(X = 1) + P(X = 2) = P(X = 3)\ P(X = 5) + P(X = 8) = 0.55$

iv. $P(X = 3/X \le 0) = \dfrac{P[(X = 3) \cap X \le 0]}{P(X \le 0)} = 0$

v. $P(X = 3)/X \le 0) = \dfrac{P[(X = 3) \cap P(X > 0)]}{P(X > 0)}$

$= \dfrac{P(3 \le X \le 8)}{1 - [P(X = -3) + P(X = -1)} = \dfrac{P(3) + P(5) + P(8)}{1 - (0.1 + 0.2)} = \dfrac{0.15 + 0.05 + 0.05}{1 - 0.3} = \dfrac{0.25}{0.7} = 0.357.$

14. A probability density function is given, **[PU 2010 Spring]**

$$f(x) = \begin{cases} \dfrac{x+2}{6} & 0 \le x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find i. $P(X \not< 0)$ ii. $F(15)$ iii. $P(0 < X \le 1)$

Solution 14

From the given condition, we have

(i) $P(X > 0) = \int_0^2 f(x)\,dx + \int_2^\infty f(x)\,dx = \int_0^2 \frac{x+2}{6}\,dx + 0 = \frac{1}{6}\left[\frac{x^2}{2} + 2x\right]_0^2 = 1$

(ii) $F(1.5) = P(X \le 1.5) = \int_0^{1.5} \frac{X+2}{6}\,dx = \frac{1}{6}\left[\frac{x^2}{2} + 2x\right]_0^{1.5} = \frac{1}{6}[1.125 + 3] = 0.6875$

(iii) $P(0 < X \le 1) = \int_0^1 \frac{x+2}{6}\,dx = \frac{1}{6}\left[\frac{x^2}{2} + 2x\right]_0^1 = \frac{1}{6} \times \frac{5}{2} = 0.417$

15. A probability curve $f(x) = e^{-x}$, $0 < x < \infty$.

Find i. $P(X < 0)$ ii. $P(X > 0)$ iii. $P(20 < X < \infty)$

iv. $P(-\infty < X < 0)$ v. $P(0 < X < \infty)$.

Solution 15

We have given Probability curve $f(X) = e^{-X}$, $0 < x < \infty$

(a) $P(X < 0) = 0$

(b) $P(X > 0) = \int_0^\infty e^{-}dx = \lim_{a \to \infty}\left[\frac{e^{-x}}{-1} + 2x\right]_0^a = \lim_{a \to \infty}[1 - e^{-a}] = 1$

(c) $P(20 > X < \infty) = \int_{20}^\infty e^{-X}\,dx = \lim_{a \to \infty}\left[\frac{e^{-x}}{-1}\right]_{20}^a = [e^{-20} - e^{-\infty}] = e^{-20} = \frac{1}{e^{20}}$

Scanned by CamScanner

(d) $P(-\infty < X < 0) = 0$

(e) $P(0 < X < \infty) = \int_0^{\infty} e^{-X}\,dx = 1.$

16. The diameter of an electric cable, say X, is assumed to be a continuous random variable with $f(x) = 6x(1-x), \quad 0 \le x < 1$

i. Check the above is p.d.f.

ii. Find $P(-3 < X < 0.2)$

iii. Compute $P(X < 0.5)$

Solution 16

Let X denote the diameter of an electric cable defined by, $f(X) = 6x(1-X), 0 \le X \le 1.$

(i) For it to be a P.d.f. we must have,

$$\int_0^1 f(X)\,dx = 1 \quad \Rightarrow \quad \int_0^1 6x(1-X)\,dx = 1 \Rightarrow 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right] = 1 \Rightarrow 1 = 1$$

Hence it is a P.d.f.

(ii) $$P(-3 < X < 0.2) = \int_{-3}^{0} f(X)\,dx + \int_0^{0.2} f(X)\,dx$$

$$= 0 + \int_0^{0.2} 6x(1-X)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{0.2} = 0.104$$

(iii) $$P(X < 0.5) = 0 + \int_0^{0.5} 6x(1-X)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{0.5} = 0.5$$

17. A continuous s.v. X follows the probability law **[P.U. 2006 Spring]**

$f(x) = Ax^2, \qquad 0 \le x < 1$

i. Determine A ii. Find the probability that X lie between 0.2 and 0.5

Solution 17

We have given P.d.f. $f(x) = Ax^2 \qquad 0 \le x < 1.$

(i) If to be the P.d.f. we must have, $\int_0^1 Ax^2 dx = 1 \quad \Rightarrow A\left[\frac{x^2}{3}\right]_0^1 = \quad \Rightarrow A = 3.$

(ii) $$P(0.2 < X < 0.5) = \int_{0.2}^{0.5} 3x^2 dx = [x^3]_{0.2}^{0.5} = 0.027.$$

Scanned by CamScanner

18. **If a random variable X has the density function.**

$$f(x) = \begin{cases} \frac{1}{4} & -2 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

Obtain i. $P(X < 1)$ ii. $P(|X| > 1)$ iii. $P(2X + 3 > 5)$

Solution 18

We have given P.d.f. $f(X) = \begin{cases} 1/4 & \text{for} \quad -2 < X < 2 \\ 0 & \text{otherwise} \end{cases}$

Then, (i) $P(X < 1) = \int_{-\infty}^{1} f(X)\,dx = \int_{-\infty}^{-2} f(X)\,dx + \int_{-2}^{1} f(X)\,dx$

$$= 0 + \int_{-2}^{1} \frac{1}{4}\,dx = \frac{1}{4}[X]_{-2}^{1} = \frac{3}{4}$$

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

$$= 1 - \int_{-1}^{1} \frac{1}{4}\,dx = 1 - \frac{1}{4}dx\,[X]_{-1}^{1} = 1 - \frac{1}{2} = \frac{1}{2}$$

(iii) $P(2X + 3 > 5) = P((X > 1) = \int_{1}^{\infty} f(x)\,dx$

$$= \int_{1}^{2} \frac{1}{4}\,dx + \int_{2}^{\infty} f(x)\,dx = \frac{1}{4}[x]_{1}^{1} = \frac{1}{4}.$$

19. **A continues distribution of a variable X in the range (– 3, 3) is defined by**

$$f(x) = \begin{cases} \frac{1}{16} & (x+3)^2 & \text{for} & -3 \le x < -1 \\ \frac{1}{16} & (6 - 2x^2) & \text{for} & -1 \le x \le 1 \\ \frac{1}{16} & (3 - x)^2 & \text{for} & 1 \le x \le 3 \end{cases}$$

i. Verify that the area under the curve is unity

ii. Determine $P(-0.5 < X < 1)$

Solution 19

According to the given questions, we have

Then, (i) $\int_{-3}^{3} f(x)\,dx = \int_{-3}^{-1} \frac{(x+3)^2}{16}\,dx + \int_{1}^{1} \frac{6 - 2x^2}{16}\,dx + \int_{1}^{3} \frac{(3-x)^2}{16}\,dx$

Scanned by CamScanner

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

$$= \frac{1}{16}\left[\left(\frac{8}{3} - 0\right) + 6 \times 2 - \frac{2}{3}(1+1) - \left(0 - \frac{8}{3}\right)\right]$$

$$= \frac{1}{16}\left[\frac{8}{3} + 12 - \frac{4}{3} + \frac{8}{3}\right] = \frac{1}{16}\left[\frac{12}{3} + 12\right] = \frac{1}{16} \times \frac{48}{3} = 1$$

(ii) $P(-0.5 < X < 1) = \int_{0.5}^{1} f(x)\,dx$

$$= \int_{0.5}^{1}\left(\frac{6 - 2x^2}{16}\right)dx = \frac{1}{16}\left[6x - \frac{2x^3}{3}\right]_{-0.5}^{1} = \frac{1}{16} \times \left[6 \times 1.5 - \frac{2}{3}(1.125)\right] = 0.5156.$$

20. Find K so that the following can serve as the probability density of a random variables.

$$f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ Kxe^{-4x^2} & \text{for } x > 0 \end{cases}$$

i. Determine k.,

ii. Determine F (120)

Solution 20

Here given p.d.f. $f(x) = \begin{cases} 0 & \text{for} \quad x \le 0 \\ kx \cdot 4x^2 & \text{for} \quad x < 0 \end{cases}$

(i) Since the given function is p.d.f. then we must have,

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

or, $0 + \int_{0}^{\infty} kxe^{-4x^2}\,dx = 1$

or, $k\left[\frac{1}{8}\right] = 1 \Rightarrow \quad k = 8.$

(ii) $F(120) = P(X \le 120) = \int_{0}^{120} 8xe^{4x^2}\,dx = [1 - e^{-480}]$

21. For the continues distribution $dF = y_0 (x - x^2)\,dx; \quad 0 \le x \le 1$

Prove that $f(x) = 6(x - x^2), \quad 0 \le x \le 1$

Scanned by CamScanner

Solution 21

We have given function $dF = y_0(x - x^2)\,dx: 0 \le x \le 1$

In order that F(x) is a probability distribution function, f(x) must be a p.d.f. Then we have $f(x) = \frac{dF(x)}{dx} = y_0(x - x^2), 0 \le x \le 1$

$$\therefore \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 f(x)\,dx = \int_0^1 y_0(x - x^2)\,dx = 1$$

$$y_0\left(\frac{x^2}{2} - \frac{x^3}{3}\right)_0^1 = 1 \Rightarrow y_0 = 6$$

$\therefore$ $f(x) = 6(x - x^2)$

22. A batch of small caliber ammunition is accepted as satisfactory if none of a sample of 5 shot falls more then 2 feet from the center of the target at a given range. If X, the distance from the center of the target to a given impact point, actually has the density

$$f(x) = k.2xe^{-x^2}, \qquad 0 < x < 3$$

Solution 22

If X denote the random variable with the distance from the centre of the target. Also the batch will be accepted if it falls with in 2 feet from the centre of the target.

Also, $f(x) = k\,2xe^{-x^2} \quad 0 < x < 3.$

the function f(x) will be p.d.f. if.

$$\int_0^3 f(x)\,dx = 1 \quad \Rightarrow \int_0^3 k2xe^{-x^2} = 1$$

or, $k\left[\frac{e^{-y}}{-1}\right]_0^9 = 1$

or, $k[1 - e^{-9}] = 1$

$\therefore$ $k\frac{e^9 - 1}{e^9} = \frac{8102.1}{8103.1} = 0.9998 \approx 1$

The probability that the batch will be accepted in a 5 shots is given by.

$$P\{(X < 2)^5\} = P\{(-2 < X < 2)^5\} = P\{(0 < X < 2)^5\}$$

$$= \left(\int_0^2 2xe^{-x^2}dx\right)^5 = \left\{\left(\frac{e^{-y}}{-1}\right)_0^4\right\}^5 = \{1 - e^{-4}\}^5 = \left(\frac{e^4 - 1}{e^4}\right)^5 = 0.91227.$$

23. Find the P $(0 < X < \pi/2)$, P $(0 < X < \pi/4)$ for the distribution.

$dF(x) = \text{Sin}\,x\,dx, \qquad 0 < x \le \pi/2$

Scanned by CamScanner

Solution 23

Here given distribution is $d\,F(x) = \sin x,\ dx,\ 0 < x \le \frac{\pi}{2}$

Then we know that $f(x) = F'(x) = \sin x,\ 0 < x \le \pi/2$

Now (a) $P(0 < X < \frac{\pi}{2} = \int_0^{\pi/2} \sin x dx = [-\cos x]_0^{\pi/2} = 1$

(b) $P(0 < X < \frac{\pi}{4} = \int_0^{\pi/4} \sin x dx = [-\cos x]_0^{\pi/4} = 1 - \frac{1}{\sqrt{2}} = \frac{\sqrt{2}-1}{\sqrt{2}}$

24. $F(x) = \begin{cases} 0 & \text{if} \quad x \le 1 \\ k(x-1)^4 & \text{if} \quad 1 < x \le 3 \\ 1 & \text{if} \quad x > 3 \end{cases}$

Continuous random variable X has the distribution function. Find p.d.f. and k.

Solution 24

Here given $f(x) = \begin{cases} 0 & \text{if} \quad x \le 1 \\ k(x-1)^4 & \text{if} \quad 1 < x \le 3 \\ 1 & \text{if} \quad x > 3 \end{cases}$

Then we know the probability density function is given by

$f(x) = F'(x) \begin{cases} 0 & \text{if} \quad x \le 1 \\ 4k(x-1)^3 & \text{if} \quad 1 < x \le 3 \\ 0 & \text{if} \quad x > 3 \end{cases}$

The given function is will be probability density function

If, $\int_1^3 f(x)\,dx = 1. \Rightarrow 4k\int_1^3 (x-1)^3 dx = \left[4k\frac{(x-1)^4}{4}\right]_1^3 = 8k = 1 \Rightarrow k = 1/8.$

25. The time one has to wait for a bus at a downtown bus stop is observed to b random phenomenon X with the following probability density function.

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{9}(x+1) & \text{for } 0 \le x < 1 \\ \frac{4}{9}\left(x - \frac{1}{2}\right) & \text{for } 1 \le x \le 3/2 \\ \frac{4}{9}\left(\frac{5}{2} - x\right) & \text{for } 3/2 \le x < 2 \\ \frac{1}{9}(4-x) & \text{for } 2 \le x < 3 \\ \frac{1}{9} & \text{for } 3 \le x < 6 \\ 0 & \text{for } 6 \le x \end{cases}$$

Scanned by CamScanner

Let A and B be defined as follows

A: One waits between 0 to 2 minutes inclusive.

B: One waits between 0 to 3 minutes inclusive.

Find P (B/A) and $P\left(\bar{A}\cap\bar{B}\right)$

Solution 25

Let A denote the event one has waits for a down town between 0 to 2 minutes inclusive. Then,

$P(A) = P(0 \le X \le 2)$ similarly

$P(B) = P(0 \le x \le 3)$.

(a) $P(B/A) = \dfrac{P(A\cap B)}{P(A)} = \dfrac{P[(0 \le X \le 2) \cap (0 \le X \le 3)]}{P(0 \le X \le 2)} = \dfrac{P(0 \le X \le 2)}{P(0 \le X \le 2)} = 1$

(b) $P(\bar{A} \cap \bar{B}) = P[(\overline{A\cap A})]$

$= 1 - P(A \cup B) = 1 - P[(0 \le X \le 2) \cup (0 \le X \le 3)] = 1 - P[0 \le X \le 3)$

$= 1 - \int_0^3 f(x)\,dx = \int_3^6 f(x)\,dx = \int_3^6 \frac{1}{9}\,dx = \frac{1}{9}\,[x]_3^6 = \frac{1}{9} \times 3 = \frac{1}{3}.$

Scanned by CamScanner

9

Chapter

Mathematical Expectation

Expected Value

The expected value of a random variable is simply the long run average of the given variable over an indefinite number of samples.

Expectation of Discrete Random Variable

Let X is a discrete random variable that takes on the values $x_1, x_2, \ldots, x_n$ with probabilities $p_i = P(X = x_i)$, $i = 1, 2, 3, \ldots, n$ then its mathematical expectation is denoted and defined by

$$E(X) = \sum_{i=1}^{n} x_i p_i \quad \text{where} \quad \sum_{i=1}^{n} p_i = 1$$

Properties of expected values:

1. i. $E(X) = \Sigma x p_i = \mu$

   ii. If a is constant then $E(a) = a$

2. If a and b are any two constants and X, Y are random variables then

   i. $E(aX) = a\,E(X)$

   ii. $E(aX + b) = a\,E(X) + b$

   iii. $E(aX + bY) = a\,E(X) + b\,E(Y)$

   iv. $E(X^2) = \Sigma x^2 p(x) = \sum_{i=1}^{n} x_i^2 \cdot p(x_i)$

3. The mathematical expectation of the sum of number of r.v. is equal to the sum of their expectation. That is

Scanned by CamScanner

i. Let X and Y are two discrete random variables then

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i + y_j)\, p_{ij} = E(X) + E(Y)$$

ii. Let X, Y and Z are three discrete random variables then

$E(X+Y+Z) = E(X) + E(Y) + E(Z)$

iii. Similarly $E(X_1 + X_2 + X_3 + \ldots + X_n) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)$

4. The mathematical expectation of the product of numbers of independent random variables is equal to the product of their expectations. That is

i. If X and Y are two discrete independent random variables then

$$E(X.Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(x_i, y_j) = E(X).E(Y)$$

ii. If X, Y and Z are three discrete independent random variables then

$E(XYZ) = E(X) . E(Y) . E(Z)$

iii. If $X_1, X_2, \ldots, X_n$, $E(X_1) . E(X_2) \ldots E(X_n)$

Variance of Random Variable

Let X be a random variable with expectation E (X). Then the variance of random variable X is denoted and defined by $V(X) = E[X - E(X)^2] = E(X)^2 - [E(X)]^2$.

Also the standard deviation (S.D) of x is denoted and defined by $\sigma_x = \sqrt{var(X)}$

Properties of variance:

1. i. $V(X \pm C) = V(X)$, where C is any constant.

ii. $V(aX) = a^2 V(X)$

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

iv. $V(a) = 0$ where a is any constant.

2. i. If X and Y are any two random variables then the measure of simultaneous variation between X and Y is called the covariance between X and Y denoted by

cov (X, Y) and given by

$cov(X, Y) = E(XY) - E(X) . E(Y)$

ii. If X and Y are independent then

$cov(X, Y) = E(X) . E(Y) - E(X) . E(Y) = 0$

$[\because E(XY) = E(X) . E(Y)]$

3. If X and Y are two random variables then

i. $V(X \pm Y) = V(X) + V(Y) \pm 2\,Cov(X, Y)$

ii. $V(aX \pm bY) = a^2 V(X) + b^2 V(Y) \pm 2ab\,Cov(X, Y)$

4. If X and Y are independent random variables then

i. $V(X \pm Y) = V(X) \pm V(Y)$

ii. $V(aX \pm bY) = a^2 V(X) \pm b^2 V(Y)$

Scanned by CamScanner

Expected value of Continuous Random Variables

Let X is a continuous random variable with probability density function (p.d.f) f(x) then the expected value of x is denoted by E (X) and defined as,

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

Remarks:

i. The expected value of r.v. X gives the mean or average of the given random variable.

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

iii. The variance of the continuous random variable X is denoted and defined by:

$$V(X) = \sigma^2$$
$$= E(X^2) - [E(X)]^2$$

iv. The standard deviation is $S.d = \sigma = \sqrt{var(X)}$

Exercise 9

1. Prove E (X + Y) = E (X) + E (Y), if X and Y are r. vs.

Solution 1

Let X be a random variable with the values $x_1, x_2, x_3, \ldots, x_n$ with probabilities $P(x_1), P(x_2), P(x_3), \ldots, P(x_n)$ and let Y be a random variable which assumes the values $y_1, y_2, y_3, \ldots, y_m$ along with corresponding probabilities $P(y_1), P(y_2), P(y_3), \ldots, P(y_m)$.

Then, $\Sigma(x) = \sum_{i=1}^{n} x_i p_i$ and $E(Y) = \sum_{i=1}^{n} y_i P_i$

Let the joint p.m.f. be $P(x_i, y_j) = P(X = x_i, Y = y_j)$. The sum X + Y is also a random variable and this can take m × n values $(x_i + y_j)$, i = 1, 2, 3, ..., n, j = 1, 2, 3, ..., m

Thus

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i + y_j)\, P(X = x_i, Y = y_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i + y_j)\, P(x_i, y_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i P(x_i, y_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} y_j P(x_i, y_j)$$

Scanned by CamScanner

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} P(x_i, y_j) + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} P(x_i, y_j)$$

$$= \sum_{i=1}^{n} x_i P(x_i) + \sum_{j=1}^{m} y_j P(y_j) = E(X + E(Y).$$

2. **Prove, E (XY) = E (X) . E (Y), if X and Y are independent random variables.**

Solution 2

Let X be the random variable which assumes the values $x_1, x_2, x_3, \ldots, x_n$ along with corresponding probabilities $P(x_1), P(x_2), P(x_3), \ldots, P(x_n)$ and let Y be a random variable which assumes the values $y_1, y_2, y_3, \ldots y_m$ along with corresponding probabilities $P(y_1), P(y_2), P(y_3), \ldots, P(y_m)$. Then

$$E(X) = \sum_{i=1}^{n} x_i p_i \text{ and } E(Y) = \sum_{j=1}^{m} y_j pj$$

Then the joint p.m.f. $P(X = x_i, Y = y_j) = P(x_i) \times P(y_j)$

Then also the product XY is also a random variable and this can take mxn values of $x_i y_j$, $i = 1, 2, 3, \ldots, n$, $j = 1, 2, 3, \ldots, m$. Thus,

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j P(x_i) P(y_j) = \sum_{i=1}^{n} x_i P(x_i) . \sum_{j=1}^{m} y_j . P(y_j) = E(X) \Sigma (Y)$$

3. **If X is a r.v. and 'a' and 'b' are constant show that**

a. $E(aX + b) = aE(X) + b$ b. $Var(aX + b) = a^2 Var(X)$

c. $Cov(X, Y) = 0$, if X and Y are independent

Solution 3

(a) Let X be the random variable which assumes the values $x_1, x_2, x_3, \ldots, x_n$ along with corresponding probabilities $P_1, P_2, P_3, \ldots, P_n$ respectively.

Then $E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots + x_n p_n$

Also for any two constants a and b,

$E(ax + b) = (ax_1 + b) p_1 + (ax_2 + b) p_2 + (ax_3 + b) p_3 + \ldots, (ax_n + b) p_n$

$= ax_1p_1 + bp_1 + ax_2p_2 + bp_2 + ax_3p_3 + bp_3 + \ldots, + ax_np_n + bp_n$

$= a(x_1p_1 + x_2p_2 + \ldots, + x_np_n) + b(p_1 + p_2 + p_3 + \ldots, + p_n) = aE(X) + b.1$

(b) Let X be any random variable and a and b any two constants.

Let us suppose $Y = aX + b$

Then, $E(Y) = E(aX + b) = XE(X) + b$

Then, $Y - R(Y) = (aX + b) - [aE(X) + b)] = a[X - E(X)]$

Now, squaring and taking expectation of both sides we get,

$E[Y - E(Y)]^2 = E[[a[X - E(X)]]^2] \Rightarrow V(Y) = a^2 E(X - E(X)]^2 = a^2 V(X)$

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

(c) Let X and Y are any two random variables.

Then, covariance between X and Y is denoted and defined by

Scanned by CamScanner

$$COV(X, Y) = E(XY) - E(X) . E(Y) \quad ...(I)$$

If S and Y are two independent random variables then,

$$V(X, Y) = E(XY) - E(X) E(Y) = E(X) . E(Y) - E(X) E(Y) = 0.$$

4. **What do you mean by mathematical expectation in continuous and discrete random variable?**

Solution 4

If X is a discrete random variable that takes on the values $x_1, x_2, x_3, ..., x_n$ with probabilities $p_1, p_2, p_3, ..., p_n$ respectively, then expected value of X is denoted and defined by $E(X) = \sum_{i=1}^{n} x_i p_i$ Where, $\sum_{i=1}^{n} p_i = 1$.

Again, if X is a continuous random variable with p.d.f. f(x) then the expected value of X is denoted and defined by $E(X) = \int_a^b xf(x)\,dx, \quad a \le x \le b.$

5. **State and prove additional theorem of expectation.**

Solution 5

Addition Theorem of Expectation:-

Statement: If X and Y are random variables, then E(X + Y) = E(X) + E(Y) provided that all the expectations exists.

Proof: Let X and Y be the continuous random variables with joint p.d.f. $f_{XY}(x, y)$ and marginal pdf $f_X(x)$ and $f_Y(y)$ respectively.

Then, $E(X) = \int_{-\infty}^{\infty} xf_X(x)\,dx$ and $E(Y) = E(X) \int_{-\infty}^{\infty} yf_Y(x)\,dy$

$$\text{Now, } E(X + Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x + y) f_{XY}(x, y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{XY}(x, y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf_{XY}(x, y)\,dxdy$$

$$= \int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f_{XY}(x, y)\,dy\right]dx + \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f_{XY}(x, y)\,dx\right]dy$$

$$= \int_{-\infty}^{\infty} xf_X(x)\,dx + \int_{-\infty}^{\infty} y . f_Y(y)\,dy = E(X) + E(Y)$$

6. **Prove that expectation of the product of number of independent random variables equal to the product of their expectation.**

Scanned by CamScanner

Solution 6

Let the random variable X associates with values $x_1, x_2, x_3, \ldots, x_m$ with probabilities $p(x_1), p(x_2), \ldots, p(x_n)$ respectively. Let the random variable Y associates with the values $y_1, y_2, y_3, \ldots, y_n$ with probabilities $p(y_1), p(y_2), p(y_3), \ldots, p(y_n)$ respectively. Then

$$E(X) = \sum_{i=1}^{m} x_i P(x_i), \qquad E(Y) = \sum_{i=1}^{n} y_j P(y_j)$$

Also the joint probability mass function

$P(X = x_i, Y = j) = P(X = i), Y = j) = p(x_i) P(y_j)$ being the variable X and Y are independent.

Also the product XY is a random variable which can mxn values $x_i y_j$, $i = 1, 2, 3 \ldots, n$, $j = 1, 2, 3, \ldots, m$.

Then by definition, $E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j P(x_i) P(y_j)$

$$= \sum_{i=1}^{n} x_i P(x_i) \sum_{j=1}^{m} y_j P(y_j) = E(X) E(Y).$$

7. **What is our mathematical expectation if we stand to win Rs. 400 if a balanced coin comes up head?**

Solution 7

Here given X = 400

Since balance coin is tossed and probability of coming head $p = \frac{1}{2} = 0.5$

$E(X) = xp = 400 \times 0.5 = 200.$

8. **If a service cube sells 4000 raffle tickets for a cash price of Rs. 800, what is the mathematical expectation of a person who buys one of the tickets?**

Solution 8

Here given variable value is X = Rs 800

The probability of a person buy one ticket $p = \frac{1}{4000}$

The expectation value is given by $= x \times p = 800 \times \frac{1}{4000} = 0.2$

9. **What is our mathematical expectation if we win Rs. 50 if a balanced coin comes up heads and loss Rs. 50 if it comes up tail?**

Solution 9

Let $x_1 = 50$, $x_2 = -50$ then the probability of occurring head $p_1 = 0.5$ and $p_2 = 0.5$

$\therefore$ $E(X) = x_1p_1 + x_2p_2 = 50 \times 0.5 + (-50) \times 0.5 = 0$

10. **The probability that a man fishing at a particular place will catch 1, 2, 3, 4 fish are 0.4, 0.3, 0.2, 0.1 respectively. What is the expected number of fish caught?**

Scanned by CamScanner

Solution 10

Let x_1, x_2, x_3, x_4 denote the no of fish catched and p_1, p_2, p_3, p_4 denote the probability of catching them.

Then, $E(X) = x_1p_1 + x_2p_2 + x_3p_3 + x_4p_4 = 1 \times 0.4 + 2 \times 0.3 + 3 \times 0.2 + 4 \times 0.1 = 2$

11. Find the expectation of the sum of the number on a dice when thrown two times.

Solution 11

Let X denote the sum of the numbers of points on the two die. Then the values of X may lies between 2 and 12 inclusive with probability distribution given as below.

Sample Points	Sum of numbers (X)	Probability (p)
{1, 1}	2	1/36
{1, 2}, {2, 1}	3	2/36
{1, 3}, {3, 2}, {3,1}	4	3/36
{1, 4}, {2, 3}, {3, 2}, {4, 1}	5	4/36
{1, 5}, {2, 4}, {3, 3}, {4, 2}	6	5/36
{1, 6}, {2, 5}, {3, 4}, {4, 3}, {5, 2}, {6, 1}	7	6/36
{2, 6}, {3, 5}, {4, 4}, {5, 3}, {6, 2}	8	5/36
{3, 6}, {4, 5}, {5, 4}, {6, 3}	9	4/36
{4, 6}, {5, 5}, {6, 4}	10	3/36
{5, 6}, {6, 5}	11	2/36
{6, 6}	12	1/36

Then the expectation value is given by

$E(X) = x_1p_1 + x_2p_2 + \ldots, + x_np_n$

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

$= 252/36 = 7.$

12. In a given business venture a man can make a profit of Rs. 1000 with probability 0.8 or suffer a loss of Rs. 400 with probability 0.2. Determine his expectation.

Solution 12

Let x_1 denote profit of Rs. 1000 and x_2 denote the loss of 400 i.e. (Rs. – 400) then

$p_1 = 0.8 \qquad p_2 = 0.2$

Expected value $= x_1p_1 + x_2p_2 = 1000 \times 0.8 + (-400) \times 0.2 = 800 - 80 =$ Rs. 720.

13. A random variable X has the following probability function.

value of X, x:	–2	–1	0	1	2	3
P (x)	0.1	k^2	0.2	$2k^2$	0.3	k^2

Solution 13

We know that sum of all probabilities equal to unity i.e. $\Sigma P(x_i) = 1$

i.e. $0.1 + k^2 + 0.2 + 2k^2 + 0.3 + k^2 = 1$

$4k^2 + 0.6 = 1$

$4k^2 = 0.4 \quad \Rightarrow \quad k^2\ \frac{0.4}{4} = 0.1$

Scanned by CamScanner

Also $E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \ldots + x_np_n$

$= (-2) \times 0.1 + (-1) \times 0.1 + 0 \times 0.2 + 1 \times 0.2 + 2 \times 0.3 + 3 \times 0.1$

$= -0.2 - 0.1 + 0 + 0.2 + 0.6 + 0.3 = 0.8$

14. A random variable X has the probability distribution

X:	−1	0	1	2
P (x):	1/3	1/6	1/6	1/3

Compute i. Expectation of random variable X

ii. Variance of random variable X.

Solution 14

(i) We know that $E(X) = \Sigma x_i P(x_i) = -1 \times \frac{1}{3} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 2 \times \frac{1}{3} = \frac{1}{2}$

(ii) We know that variance of random variable X is given by $Var(X) = E(X^2) - (EX)^2$

$$= \left[(-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{6} + 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{3}\right] - \left(\frac{1}{2}\right)^2 = \frac{1}{3} + \frac{1}{6} + \frac{4}{3} - \frac{1}{4} = \frac{19}{12} = 1.58$$

Solution 15

Let X denote the no. of boys in 3 children. Which is given as follows.

Sample space	Number of boys (X)	Probability p (x)
(G, G, G)	0	1/8
(G, G, B), (G, B, G), (B, G, G)	1	3/8
(G, B, B), (B, G, B), (B, B, G)	2	3/8
(B, B, B)	3	1/8

16. Referring to question No.15, find the mathematical expectation of these children being boys.

Solution 16

We know that the expectation value is

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

17. A and B enter into a bet according to which A will get Rs. 200 if it rains on that day and will lose Rs. 100 if it does not rain. The probability of raining on that day is 0.7. What is the mathematical expectation of A?

Solution 17

Let $x_1 = 200$ and $x_2 = -100$

$p_1 = 0.7$ and $p_2 = 1 - 0.7 = 0.3$

The expectation of A $= x_1p_1 + x_1x_2 = 200 \times 0.7 + (-100) \times 0.3 = 140 - 30 =$ Rs. 110

18. A union wage negotiator feels the probabilities are 0.4, 0.3, 0.2 and 0.1 that the union members will get a Rs. 150 an hour raise, a Rs. 100 an hour raise, a Rs. 50 an hour raise or no raise at all. What is their expected raise?

Solution 18

Let X denote the random variable of raise in hours then we have given

Scanned by CamScanner

$x_1 = 150, x_2 = 100, x_3 = 50$ and $x_4 = 0$

Then their probabilities are given as,

$p_1 = 0.4, p_2 = 0.3, p_3 = 0.2, p_4 = 0.1$

Then the expected value od X is given by

$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + x_4p_4 = 150 \times 0.4 + 100 \times 0.3 + 50 \times 0.1 + 0 \times 0.1 =$ Rs. 100.

19. In four tosses of a coin, let X be the number of heads. Tabulate the 16 possible outcomes with the corresponding value of X. By simple counting, derive the distribution of X and hence calculate the expected value of X.

Solution 19

Let H represents a head, T a tail and X denotes the random variable denoting the number of heads. Then we have,

S.N.	Possible outcomes	No. of Heads (X = x)	Probabilities P(X = x)
1.	TTTT	0	1/16
2.	HTTT, THTT, TTHT, TTTH	1	4/16
3.	HHTT, THHT, TTHH, HTHT, THTHHTTH	2	6/16
4.	HHHT, THHH, HHTH, HTHH	3	4/16
5.	HHHH	4	1/16

The above table shows the all possible outcomes and also gives the distribution of X.

Now, expected value of X is given by $E(X) = \Sigma x_i P(x_i)$

$$= 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{4 + 12 + 12 + 4}{16} = \frac{32}{16} = 2.$$

20. A coin is tossed until a head appears. What is the expectation of number of tosses required?

Solution 20

On tossing a coin, let H denote the event occurring head and T denote the event occurring tail. Suppose X be the number of tosses required until head occurs. Then head can occur in the following way.

Outcomes	No. of tosses (X = x)	P(x)	xp
H	1	1/2	1/2
TH	2	1/4	2/4
TTH	3	1/8	3/8
TTTH	4	1/16	4/16

Now the expected value of X is given by

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \ldots, \quad \ldots(i)$$

$$\text{Let } S = 1/2 + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{5}{32} + \ldots, \quad \ldots(ii)$$

$$\text{Then } \frac{S}{2} = \frac{1}{4} + \frac{2}{8} + \frac{4}{38} + \frac{5}{64} + \ldots \quad \ldots(iii)$$

Now (ii) – (iii) we get,

Scanned by CamScanner

$$S-\frac{S}{2}=\frac{1}{2}+\left(\frac{2}{4}-\frac{1}{4}\right)+\left(\frac{3}{8}-\frac{2}{8}\right)+\left(\frac{4}{16}-\frac{3}{16}\right)+\ldots$$

or, $\frac{S}{2}=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots$

or, $\frac{S}{2}=\frac{\frac{1}{2}}{1-\frac{1}{2}} \quad \Rightarrow \frac{S}{2}=1 \Rightarrow S=2$

21. Five urns contains respectively 5 green and 2 white balls, 7 green and 6 white balls, 2 green and 1 white ball, 5 green and 2 white balls and 2 green and 4 white balls. One ball is drawn from each urn. Find the expected number of white balls drawn out.

Solution 21

Do yourself

22. Urn A contains 5 cards numbered from 1 to 5 and urn B contains 4 cards numbered from 1 to 4. One card is drawn from each of two urns. Find the probability function of sum of the number which appears on the cards drawn and its mathematical expectation.

Solution 22

Let X denote the random variable of the sum of the number on the two urns. If one card is drawn from each of two urns. The probability function is given as below.

S.N.	Sample space	Sum of numbers (X)	Probabilities (P)
1.	(1, 1)	2	1/20
2.	(1, 2), (2, 1)	3	2/20
3.	(1, 3), (2, 2), (3, 1)	4	3/20
4.	(1, 4), (2, 3), (3, 2), (4, 1)	5	4/20
5.	(2, 4), (3, 3), (4, 2), (5, 1)	6	4/20
6.	(3, 4), (4, 3), (5, 2)	7	3/20
7.	(4, 4), (5, 3)	8	2/20
8.	(5, 4)	9	1/20

Also the mathematical expectation is given as:

$$E(X)=2\times\frac{1}{20}+3\times\frac{2}{20}+4\times\frac{3}{20}+5\times\frac{4}{20}+6\times\frac{4}{20}+7\times\frac{3}{20}+8\times\frac{2}{20}+9\times\frac{1}{20}=\frac{110}{20}=\frac{11}{2}.$$

23. Find the expectation of the number on a dice when thrown.

Solution 23

Let X denote the random variable representing the numbers on the die when thrown. Then X can take any one of the numbers (values) 1, 2, 3, 4, 5 and 6. Also each with the equal probability $\frac{1}{6}$. Hence the expectation is

$$\Sigma(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 + x_6 p_6$$

$$=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}\times 6\times\frac{1}{6}=\frac{21}{6}=7/2.$$

Scanned by CamScanner

24. Suppose X is a r.v. having p.m.f. p(x) given

x	−3	−1	0	1	2	3	5	8
p (x)	0.1	0.2	0.15	0.2	0.1	0.15	0.05	0.05

Find mean and variance.

Solution 24

Here given X is discrete random variable. Then the calculation of mean and variance is given as follows:

$X = x_i$	P(x)	$x_i P(x_i)$	x^2	$x_i^2 P(x_i)$
−3	0.1	−0.3	9	0.9
−1	0.2	−0.2	1	0.2
0	0.15	0	0	0
1	0.2	0.2	1	0.2
2	0.1	0.2	4	0.4
3	0.15	0.45	9	1.35
5	0.05	0.25	25	1.25
8	0.05	0.4	64	3.2
		$\Sigma x_i P(x_i) = 1$		$\Sigma[x_i^2 . P(x_i)] = 7.5$

We know that Mean = $E(X) = \Sigma x_i P(x_i) = 1$

Also variance = $E(X^2) - [E(X)]^2 = \Sigma x_i^2 . P(x_i) - 1^2 = 7.5 - 1 = 6.5$

25. An appliance dealer sells different models of freezers having 13.5, 15.9 and 19.1 cu. ft. of storage space respectively. Let X be the amount of storage space purchased by the next customer to buy a freezer and p (x) be the p.m.f. given as

x	−3	−1	0
p (x)	0.1	0.2	0.15

a. Compute E (X) and V (X)

b. If the price of the freezer having capacity x cubic feet is 25x − 8.5, what is the expected price paid by the next customer to buy a freezer?

Solution 25

Let X denote the discrete random variable

X; x	P(x)	x P(x)	x^2	$x_i^2 P(x)$
13.5	0.2	2.7	182.25	36.45
15.9	0.5	7.95	252.81	126.405
19.1	0.3	5.73	364.81	109.443
		$\Sigma x P(x) = 16.38$		$\Sigma[x^2 P(x) = 272.298$

(a) Now $E(X) = \Sigma x P(x) = 16.38$

$$V(X) = E(X^2) - [E(X)]^2 = \Sigma x^2 P(x) - (16.38)^2 = 272.298 - 268.3044 = 4$$

(b) If the price of freezer having capacity x is Rs 25x − 8.5. Then the expected price is given by E(price) = 25 × 16.38 = Rs 401

26. Find the mean and variance of the random variable with probability density function $f(x) = xe^{-x}, x \geq 0$

Scanned by CamScanner

Solution 26

Here given probability density function $f(x) = xe^{-x}$, $x \geq 0$

for a continuous random variable x.

Then mean $= E(X) = \int_a^b xf(x)\,dx$

$$= \int_0^\infty x.\,xe^{-x}\,dx \quad \left[\because \int_0^\infty x^{n-1}\,dx = \Gamma(x)\right] = \int_0^\infty e^{-x}.\,x^{3-1}\,dx = \Gamma(3) = 3! = 2$$

Also for the variance,

$$E(X^2) = \int_a^b x^2.xf(x)\,dx = \int_\infty^0 e^{-x}.\,x^{4-1}\,dx = \Gamma(4) = 3! = 6$$

Now variance $V(X) = E(X^2) - [E(X)]^2 = 6 - 2^2 = 6 - 4 = 2$

27. **The p.d.f. of a continuous r.v. is given as**

$f(x) = A\,x(1-x)^2 \qquad 0 \leq x \leq 1 \quad$ elsewhere.

Find E (X) and V(X)

Solution 27

Let X be the continuous random variable with p.d.f. given by

$$f(x)\begin{cases} A\,x\,(1-x)^2 & 0 \leq x \leq 1 \\ 0 & \text{else where} \end{cases}$$

Since the $f(x)$ is a p.d.f of continuous random variable X then we have.

$$\int_{-\infty}^{\infty} f(x) = 1 \Rightarrow \int_{-\infty}^{0} f(x) + \int_0^1 f(x) + \int_1^\infty f(x)\,dx = 1$$

or, $0 + \int_0^1 Ax\,(1-x)^2\,dx + 0 = 1.$

or, $A\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1 \Rightarrow A\left[\frac{1}{2} - \frac{1}{3}\right] = 1 \Rightarrow A.\frac{1}{6} = 1 \Rightarrow A = 6.$

Now, we know that

$$E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^1 x.\,6x\,(1-x)^2\,dx = 6\int_0^1 x^2\,(1-x)^2\,dx$$

$$\left[\because \int_0^1 x^{m-1}(1-x)^{n-1}\,dx = b\,/\,(m,n)\right] = 6.\beta\,(3,3) = 6.\frac{\Gamma(3)\,\Gamma\,(3)}{\Gamma(3+3)} = \frac{6\times 2!2!}{5!} = \frac{24}{120} = \frac{1}{5}$$

Scanned by CamScanner

Also, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 6x^3 (1-x)^2\,dx = 6.\beta(4,3)$

$$= \frac{6.\Gamma(4)\Gamma(3)}{\Gamma(4+3)} = \frac{6 \times 3! \times 2!}{720} = \frac{1}{10}$$

Now $V(X) = E(X^2) - (\Sigma X)^2 = \frac{1}{10} - \left(\frac{1}{5}\right)^2 = \frac{3}{50}$ #.

28. **In four tosses of a coin, let X be the number of heads. Tabulate the 16 possible outcomes with the corresponding value of X. By simple counting, derive the distribution of X and hence calculate the expected value of X.**

Solution 28

Please refer solution 19

29. **A random variable X has the density $f(x) = a + bx^2$, $0 < x < 1$, determine 'a' and 'b' so that mean is 2/3. Also find the variance of the distribution.**

Solution 29

Let X denote the continuous random variable with p.d.f $f(x) = a + bx^2$, $0 < x < 1$.

Since $f(x)$ is p.d.f then we most have;

$$\int_a^b f(x)\,dx = 1 \quad \Rightarrow \int_0^1 (a + bx^2)\,dx = 1 \Rightarrow \left[ax + \frac{bx^3}{3}\right]_0^1 = 1$$

$$\Rightarrow 3a + b = 3 \qquad \ldots(A)$$

Also the mean $E(X) = \int_a^b x\,f(x)\,dx$

$$\text{or,} \quad \frac{2}{3} = \int_0^1 x\,(a + bx^2)\,dx$$

$$\text{or,} \quad \frac{2}{3} = \left[\frac{ax^2}{2} + \frac{bx^4}{4}\right]_0^1$$

$$\text{or,} \quad \frac{2}{3} = \frac{2a + b}{4}$$

$$\text{or,} \quad 6a + 3b = 8 \qquad \ldots(B)$$

Solving (A) and (B) we get, $a = \frac{1}{3}$ $\quad b = 2$

Then $E(X^2) = \int_a^b x^2 f(x)\,dx = \int_0^1 x^2 (a + bx^2)\,dx = \left[\frac{ax^3}{2} + \frac{bx^5}{4}\right]_0^1 = \left(\frac{1}{3}\cdot\frac{1}{3} + 2.\frac{1}{5}\right) = \frac{23}{45}$

Now $V(X) = E(X^2) - [E(X)]^2 = \frac{23}{45} - \left(\frac{2}{3}\right)^2 = \frac{1}{15}$

Scanned by CamScanner

30. **Suppose that the number of plants of a particular type which are found in a rectangular region in certain geographical area is a r.v. X with p.m.f.**

$$p(x) = \begin{cases} \frac{a}{x^3} & x = 1 \\ 2 & 3 \\ 0 & \text{otherwise} \end{cases}$$

Obtain mean, variance and standard deviation.

Solution 30

Since the given function is p.m.f.,

Then $\Sigma P(x) = 1 = \frac{a}{1^3} + \frac{a}{2^3} + \frac{a}{3^3} = 1 \Rightarrow \frac{251a}{216} = 1 \Rightarrow a = \frac{216}{251}$

Now mean $= E(X) = \Sigma x P(x) = 1 . \frac{4}{1^3} + 2 . \frac{a}{2^3} + 3 . \frac{a}{3^3}$

$$a\left(1 + \frac{1}{4} + \frac{1}{9}\right) = \frac{49a}{36} = \frac{49}{36} \times \frac{216}{251} = 1.1713$$

Also for variance:

$$E(X)^2 = \Sigma x^2 P(x) = 1^2 \cdot a + 2^2 \cdot \frac{a}{8} + 3^2 + \frac{a}{27} = \frac{11a}{6} = 1.58$$

$$E(X^2) - [E(X)]^2 = 1.58 - (1.1713)^2 = 0.2080$$

Also standard deviation $= \sqrt{V(X)} = \sqrt{0.2080} \quad = 0.456.$

31. **On the average how many times a dice throw to get a 6?**

Solution 31

Do Yourself

32. **A bag contains 5 white and 3 red balls. Three balls are taken at random. Find the expected number of white balls drown in the three.**

Solution 32

Let X denote the discrete random variable that the number of white balls which are selected. Then X is r.v with values. 1, 2, 3, respectively. Then the probabilities of drawing white balls in the three draws is given by.

$$P(X = 0) = P[\text{no white ball}) = \frac{3C_3}{8C_3} = \frac{1}{56}$$

$$P(X = 1) = P[\text{one white ball}) = \frac{5C_1 \times 3C_2}{8C_3} = \frac{5 \times 3}{56} = \frac{15}{56}$$

$$P(X = 2) = P[\text{two white ball}) = \frac{5C_2 \times 3C_1}{8C_3} = \frac{3 \times 10}{56} = 30/56$$

$$P(X = 3) = P[\text{three white ball}) = \frac{5C_3}{8C_3} = \frac{10}{56}.$$

Now, $E(X) = \Sigma x P(x) = x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4$

$$= \frac{0.1}{56} + 1 \times \frac{15}{56} + 2 \times \frac{30}{56} + 3 \times \frac{10}{56} = 105/56 = 1.875$$

Scanned by CamScanner

33. Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} ax & 0 \le x \le 2 \\ a & 2 \le x \le 4 \\ ax + 3a & 4 \le x \le 6 \\ 0 & \text{elsewhere} \end{cases}$$

i Determine the constants 'a' and find.

ii. $p(X \le 1.5)$ and $p(1.5 \le X \le 2.5)$ iii. obtain $E(X)$

Solution 33

(a) Since f(x) is p.d.f. then we must have $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

or, $$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{4} f(x)\,dx + \int_{4}^{6} f(x)\,d(x) + \int_{6}^{\infty} f(x)\,dx = 1$$

or, $$0 + \int_{0}^{2} ax\,dx + \int_{2}^{4} a\,dx + \int_{4}^{6} (ax + 3a)\,dx = 1$$

or, $$\left(\frac{ax^2}{2}\right)_0^2 + [ax]_2^4 + \left[\frac{ax^2}{2} + 3ax\right]_4^6 = 1$$

$$2a + 2a + 6a + 6a = 1$$

$$20\,a = 1$$

$$\Rightarrow \quad a = \frac{1}{20}$$

(ii) $$P(X \le 1.5) = \int_{0}^{1.5} ax\,dx = \left[a\frac{x^2}{2}\right]_0^{1.5} = \frac{a}{160}$$

$$P(1.5 \le X \le 2.5) = \int_{1.5}^{2.5} f(x)\,dx$$

$$= \int_{2}^{1.5} f(x)\,dx + \int_{2}^{2.5} f(x)\,dx$$

$$= \int_{2}^{1.5} ax\,dx + \int_{2}^{2.5} a\,dx$$

$$= \left[\frac{ax^2}{2}\right]_{1.5}^{2} + [ax]_2^{2.5}$$

Scanned by CamScanner

$$= a\left[\frac{2^2}{2} - \left(\frac{3}{2}\right)^2 \cdot \frac{1}{2} + (2.5 - 2)\right]$$

$$= a\left[2 - \frac{9}{8} + \frac{1}{2}\right] = \frac{1}{20} \times \frac{11}{8} = \frac{11}{160}$$

(iii) $$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = 0 + \int_0^2 x\,ax\,dx + \int_2^4 x.adx + \int_4^6 x\,(ax + 3a)\,dx + 0$$

$$= \left[\frac{ax^3}{3}\right]_0^2 + \left[\frac{ax^2}{2}\right]_2^4 + \left[\frac{ax^3}{2} + \frac{3a\,x^2}{2}\right]_4^6$$

$$= a\left[\frac{2^3}{3} + \frac{(4^2 - 2^2)}{2} + \frac{6^3 - 4^3}{3} + \frac{3}{2}(6^2 - 4^2)\right]$$

$$= \frac{1}{20}\left[\frac{8}{3} + 6\frac{152}{3} + \frac{3}{2} \times 20\right] = 4.466$$

34. The distribution function of a r.v. x in given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \frac{1}{4}e^{-x} & \text{for } x \geq 0 \end{cases}$$

Determine f (x) and P (X > 0)

Solution 34

Here given $F(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - \frac{1}{4} \times e^{-x} & \text{for} \quad x \geq 0 \end{cases}$

We know that $f(x) = F(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{1}{4} \times e^{-x} & \text{for} \quad x \geq 0 \end{cases}$

Also, $P|X > 0| = \int_0^{\infty} \frac{e^{-x}}{4}\,dx = \frac{1}{4}\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = \frac{-1}{4}[e^{-\infty} - e^{-0}] = \frac{1}{4}$

35. The length (in hours) X of a certain type of light bulbs may be supposed to be a continuous r.v. with probability density function.

$$f(x) = \frac{a}{x^3} \qquad 1500 < x < 2500$$

i. Determine a

ii. Find the cumulative distribution of X

iii. Compute the probability of the event $1700 \leq x \leq 1900$

Solution 35

Let X denote continuous random variable that is length (in hrs) of a certain type of light bulbs may be supposed.

Scanned by CamScanner

$$f(x) = \frac{a}{x^3}, \qquad 1500 < X < 2500$$

(i) Since f (x) is pdf then we must have $\int_a^b f(x)\,dx = 1$

$$\int_{1500}^{2500} \frac{a}{x^3}\,dx = 1 \Rightarrow a\left[\frac{x^{-3+1}}{-3+1}\right]_{1500}^{2500} = \Rightarrow \left[\left(\frac{1}{2500}\right)^2\left(\frac{1}{1500}\right)^2\right] = 1$$

$$\Rightarrow \frac{-a}{2}\left[\frac{156660}{15625}\right] = 1 \Rightarrow a = \frac{31250}{15660} = 1.99$$

Let F denote the distribution function of X. By the definition of f, F(x) is zero for, < 1500 and X > 2500.

Next if $1500 \le X \le 2500$ them.

$$F(x) = \int_{1500}^{x} f(x)\,dx = \int_{2500}^{2500} \frac{a}{x^3}\,dx = \left[\frac{-a}{2x^2}\right]_{1500}^{x} = \frac{a}{2}\left[\frac{a}{2250000} - \frac{1}{x^2}\right]$$

(iii) $P(1700 < X < 1900) = \int_{1700}^{1900} \frac{a}{x^3}\,dx = -\frac{a}{2}\left[\frac{1}{x^2}\right]_{1700}^{1900} = \frac{a}{2}\left[\frac{1}{2890000} - \frac{1}{3610000}\right]$

36. A random variable x has the p.d.f.

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

Find

i. $P\left(X < \frac{1}{2}\right)$ ii. $P\left(\frac{1}{4} < X < \frac{1}{2}\right)$

iii. $P\left(X > \frac{3}{4} / X > \frac{1}{2}\right)$ iv. $P\left(X < \frac{3}{4} / X > \frac{1}{2}\right)$

Solution 36

Here X is continuous random variable with p.d.f $f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

(i) $P\left(X < \frac{1}{2}\right) = \int_0^{1/2} f(x)\,dx = \int_0^{1/2} 2x\,dx = [x^2]_{1/2}^{0} = \frac{1}{4}$

(ii) $P\left(\frac{1}{4} < X < \frac{1}{2}\right) = \int_{1/4}^{1/2} 2x\,dx = [x^2]_{1/4}^{1/2} = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$

(iii) $$P\left(X > \frac{3}{4} / X > \frac{1}{2}\right) = \frac{P\left[\left(X > \frac{3}{4}\right) \cap \left(X > \frac{1}{2}\right)\right]}{P\left(X > \frac{1}{2}\right)}$$

Scanned by CamScanner

$$= \frac{P(X > \frac{3}{4})}{1 - P\left(X \le \frac{1}{2}\right)} = \frac{\int_{3/4}^{1} 2x\,dx}{\frac{3}{4}} = \frac{4}{3}\left[1^2 - \left(\frac{3}{4}\right)^2\right] = \frac{4}{3} \times \frac{7}{16} = \frac{7}{12}$$

(iv) $$P\left(X < \frac{3}{4} \mid X > \frac{1}{2}\right) = \frac{P\left[\left(X < \frac{3}{4}\right) \cap \left(X > \frac{1}{2}\right)\right]}{P\left(X > \frac{1}{2}\right)} = \frac{P\left(\frac{1}{2} < x \frac{3}{4}\right)}{\frac{3}{4}} = \frac{4}{3}\int_{1/2}^{3/4} 2x\,dx$$

$$= \frac{4}{3}\left[\left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2\right] = \frac{4}{3} \times \frac{5}{16} = \frac{5}{12}$$

37. A continuous distribution of a r.v. X in range (– 3, 3) is defined by

$$f(x) = \frac{1}{16}(3+x)^2, \qquad -3 \le x \le -1$$

$$= \frac{1}{16}(6 - 2x^2), \qquad -1 \le x \le 1$$

$$= \frac{1}{16}(3-x)^2, \qquad 1 \le x \le 3$$

i. Verify that the area under the curve is unity.

ii. Find mean and variance.

Solution 37

(i) The area under the given curve is given

$$A = \int_a^b f(x)\,dx = \int_{-3}^{3} f(x)\,dx$$

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

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

$$= \frac{1}{16} \times \frac{1}{3}[8 + 18(1+1) - 2(1+1) - (0-8)] = \frac{48}{48} = 1$$

(ii) $$\text{Mean } E(X) = \int_a^b x f(x)\,dx = \int_{-3}^{3} x f(x)\,dx$$

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

Scanned by CamScanner

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

$$= \frac{1}{16}\left[\left(\frac{9x^2}{x} + 6 \times \frac{x^3}{3} + \frac{x^4}{4}\right)_{-3}^{-1} + 0 + \left(\frac{9x^2}{2} - 6 \times \frac{x^3}{3} + \frac{x^4}{4}\right)_{1}^{3}\right]$$

$$= \frac{1}{16}\left[\frac{9}{2}(1+9) + 2(-1+2) + \frac{1}{4}(1-81) + \frac{9}{2}(9-1) - 2(27-1) + \frac{81-1}{4}\right]$$

$$= \frac{1}{16}\left[\frac{9\times 8}{2} + 2 \times 26 - \frac{80}{4} + \frac{8\times 9}{2} - 2\times 26 + \frac{80}{4}\right] = 0$$

Also, $E(X^2) = \int_{a}^{b} x^2 f(x)\,dx = \int_{-3}^{3} x^2 f(x)\,dx$

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

$$= \frac{1}{16}\left[\int_{-3}^{-1} (9x^2 + 6x^3 + x^5)\,dx + \int_{-1}^{1} (6x^2 - 2x^4)\,dx + \int_{1}^{3} (9x^2 - 6x^3 + x^4)\,dx\right]$$

$$= \frac{1}{16}\left[\left(\frac{9.x^3}{3} + \frac{6.x^4}{4} + \frac{x^5}{5}\right)_{-3}^{-1} + \left\{4\left(3.\frac{x^3}{3} - \frac{x^3}{3}\right)\right\}_{0}^{1} + \left(\frac{9x^3}{3} - \frac{6x^4}{4} + \frac{x^5}{5}\right)_{1}^{3}\right]$$

$$= \frac{1}{16}\left[3(-1+27) + \frac{3}{2}(1-81) + \frac{1}{5}(-1+243) + \left(1 - \frac{1}{5}\right) + 3(27-1) - \frac{3}{2}(81.1) + \frac{1}{5}(243-1)\right]$$

$$= \frac{1}{16}\left[78 - 120 + 48 + \frac{16}{5} + 78 - 120 + 48\right] = \frac{1}{16} \times \left(12 + \frac{16}{5}\right) = 0.95$$

$V(X) = E(X^2) - [E(x)]^2 = 0.95 - 0 = 0.95 = 1.$

38. **Is there any relationship between expectation of random variable and mean of the distribution? If so, what is that? The probability mass function of random variable X is as given in the table.**

X = x	−6	−3	0	3	6	9
P (X = x)	0.1	0.1	0.2	0.2	0.3	0.1

Using law of expectation,

Determine, i. $E(8X - 10)$ ii. $E(6X + 3)^2$ iii. $V(4X + 7)$

Solution 38

X be the discrete random variable then

$E(X) = -6 \times 0.1 - 3 \times 0.1 + 0 \times 0.2 + 3 \times 0.2 + 6 \times 0.3 + 9 \times 0.1$

$= -0.6 - 0.3 + 0.6 + 1.8 + 0.9 = 2.4$

Since the given distribution is discrete so mean = E(X) = 2.4

Scanned by CamScanner

(i) $E(8x + 10) = 8E(X) - 10 = 8 \times 2.4 - 10 = 9.2$

(ii) $E(6x + 3)^2 = E[36x^2 + 36x + 9] = 36\,E(X^2) + 36E(X) + 9$

$E(X^2) = (-6)^2 \times 0.1 + (-3)^2 \times 0.1 + 0^2 \times 0.2 + 3^3 \times 0.2 + 6^2 \times 0.3 + 9^2 \times 0.1$

$= 3.6 + 0.9 + 0 + 1.8 + 10.8 + 8.1 = 25.2 = 36 \times 25.2 + 36 + 2.4 + 9 = 1002.6$

(iii) $V(4x + 7) = 4^2\,V(X) = 16 \times 19.44 = 311.04$

Where $V(X) = E(X^2) = [E(X)]^2 = 25.2 - (2.4)^2 = 19.44$

39. Find $E(X^2)$, where X is the number obtained in a throw of a dice.

Solution 39

Since x is discrete random which represent the number in a dice. Then we have:

No of a dice (x)	P(X = x)	X^2	X^2 P(x)
1	1/6	1	1/6
2	1/6	4	4/6
3	1/6	9	9/6
4	1/6	16	16/6
5	1/6	25	25/6
6	1/6	36	36/6

We know that $E(X^2) = \Sigma x^2 P(x) = \frac{1}{6} + \frac{4}{9} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} = 15.17$

40. If X is the number of points in a balance die, find the expected value of random variable $f(x) = 2x^2 + 1$

Solution 40

Also $E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6}$

Also, $E(X^2) = 15.17$.

$\therefore$ $E(2x^2 + 1) = 2E(x^2) + 1 = 2 \times 15.17 + 1 = 31.34$

Exam asked Questions

1. Two balanced dice are thrown, find the probability distribution and the expected value of the sum of numbers of point on them. Also find the variance.

[P.U 2009 spring/ POU 2016]

Solution 1

Let X denote the random variable of sum of numbers on the two dice. Then the value of X (i.e. discrete random variable X) has the values 2, 3, 4, ..., 12 with probability distribution.

Sample Points	Sum of numbers (x_i)	Probability p (x_i)
(1,1)	2	1/36
(1,2),(2,1)	3	2/36
(1,3), (2,2), (3,1)	4	3/36
(1,4), (2,3), (3,2), (4,1)	5	4/36
(1,5), (2,4), (3,3), (4,2), (5,1)	6	5/36
(1,6), (2,5),(3,4), (4,3), (5,2), 6,1)	7	6/36
(2,6), (3,5), (4,4), (5,3), (6,2)	8	5/36
(3,6), (4,5), (5,4), (6,3)	9	4/36
(4,6), (5,5), (6,4)	10	3/36
(5,6), (6,5)	11	2/36
(6,6)	12	1/36
Total probability $\Sigma P(x)$		1

Scanned by CamScanner

2nd part:

The expected value of the above discrete random variable X is E (X) and given by $E(X) = \Sigma x_i P(x_i)$

$$\therefore \quad E(X) = 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36} + 6\times\frac{5}{36} + 7\times\frac{6}{36} + 8\times\frac{5}{36} + 9\times\frac{4}{36} + 10\times\frac{3}{36} + 11\times\frac{2}{36} + 12\times\frac{1}{36} = \frac{252}{36} = 7$$

3rd part:

Also the variance X is given by

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

where,

$$E(X^2) = \Sigma x_i^2 P(x_i)$$

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

$$= \frac{1974}{36} = 54.83$$

$$\therefore \quad V(X) = E(X^2) - [E(X)]^2 = 54.83 - 7^2 = 5.83$$

2. **From a lot of 10 items containing 3 defectives a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Answer the following:**

   i. E (X) ii. V (X)

   iii. The probability distribution of X [P.U 2016 spring/ T.U 2065 Kartik]

Solution 2

Let X denote the number of defective items in the sample. Then the possible value of X can be 0, 1, 2, 3. Then

Probability of getting 0 defective

$$P(X=0) = \frac{C(3,0)\times C(7,4)}{C(10,4)} = \frac{35}{210} = \frac{1}{6}$$

Probability of getting 1 defective

$$P(X=1) = \frac{C(3,1)\times C(7,3)}{C(10,4)} = \frac{105}{210} = \frac{1}{2}$$

Probability of getting 2 defective

$$P(X=2) = \frac{C(3,2)\times C(7,2)}{C(10,4)} = \frac{63}{210} = \frac{3}{10}$$

Probability of getting 3 defective

$$P(X=3) = \frac{C(3,3)\times C(7,1)}{C(10,4)} = \frac{7}{210} = \frac{1}{30}$$

iii. The probability distribution of X is given by

Scanned by CamScanner

$X = (x_i)$	0	1	2	3	Total
$P(x)$	1/6	1/2	3/10	1/30	1

i. $E(X) = \Sigma x_i P(x_i) = 0 \times \frac{1}{6} + 1 \times \frac{1}{2} + 2 \times \frac{3}{10} + 3 \times \frac{1}{30} = \frac{6}{5} = 1.2$

ii. $V(X) = E(X^2) - [E(X)]^2 = 2 - (1.2)^2 = 2 - 1.44 = 0.56$

where $E(X^2) = \Sigma x_i^2 P(x_i) = 0^2 \times \frac{1}{6} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{3}{10} + 3^2 \times \frac{1}{30} = 2$

3. A coin is tossed until head appears. What is the expectation number of tosses required? [P.U. 2017 Fall/ 2013 Spring]

Solution 3

See Ex.9 Q.No. 20

4. A random variable X has the following probability function

X	2	3	4	5	6	7
F (x)	0.1	k	0.2	2k	0.3	k

Find the value of k and calculate mean and variance. [P.U. 2017 Spring]

Solution 4

Since the given probability distribution function is a discrete probability distribution function then we must have,

$\Sigma P(x) = 1 \Rightarrow 0.1 + k + 0.2 + 2k + 0.3 + k = 1$

$\Rightarrow 4k + 0.6 = 1 \quad \Rightarrow k = 0.1$

The mean of the given distribution function,

Mean $(\mu) = E(X) = \Sigma x_i P(x_i)$

$= 2 \times 0.1 + 3 \times 0.1 + 4 \times 0.2 + 5 \times 0.2 + 6 \times 0.3 + 7 \times 0.1 = 4.8$

Also variance of the given distribution is given by

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

where, $E(X^2) = \Sigma x_i^2 P(x_i)$

$= 2^2 \times 0.1 + 3^2 \times 0.1 + 4^2 \times 0.2 + 5^2 \times 0.2 + 6^2 \times 0.3 + 7^2 \times 0.1 = 25.2$

$\therefore \sigma^2 = 25.2 - (4.8)^2 = 2.16$

Hence the mean and variance of the given distribution is 4.8 and 2.16 respectively.

5. An appliance dealer sells different models of freezers having 13.5, 15.9 and 19.1 cu. ft. of storage space purchased by the next customer to buy a freezer and P (x) be the p.m.f. given below as

X : x	13.5	15.9	19.1
P (x)	0.2	0.5	0.3

i. Compute E (X) and V (X)

ii. If the price if the freezer having capacity x cubic feet is 25 x – 8.5, what is the expected price paid by the next customer to buy a freezer?

iii. What is the variance of the price 25 x – 8.5 paid by the next customer.

[P.U. 2016 Spring]

Scanned by CamScanner

Solution 5

See EX 9 Q.No. 25

6. **On the average how many times must a dice 4 thrown until one gets a 6?** [P.U. 2015 Spring]

Solution 6

Do yourself

7. **A coin is tossed until a head appears. What is the expectation of number of tosses required?** [P.U. 2013 Spring]

Solution 7

See Ex. 9. Q.No.20

8. **A random variable X is defined to be the difference between the higher value and lower value when two dice are thrown, if they have the same value X is zero. Find E (2X + 10)** [P.U. 2012 Fall]

Solution 8

Since X is the random variable, which is defined to be the different between the higher value and lower value when two dice are thrown. Then

Difference (X)	Out-comes	P (x)
0	{1,1}, {2,2}, {3,3}, {4,4}, {5,5}, {6,6}	6/21
1	{2,1}, {3,2}, {4,3}, {5,4}, {6,5}	5/21
2	{3,1}, {4,2}, {5,3}, {6,4}	4/21
3	{4,1}, {5,2}, {6,3}	3/21
4	{5,1}, {6,2}	2/21
5	{6,1}	1/21

Now, $E(2X + 10) = E(2X) + E(10) = 2E(X) + 10 = 2 \times 1.67 + 10 = 13.34$

where $E(X) = \Sigma x P(x) = 0 \times \frac{6}{21} + 1 \times \frac{5}{21} + 2 \times \frac{4}{21} + 3 \times \frac{3}{21} + 4 \times \frac{2}{21} + 5 \times \frac{1}{21} = 1.67$

9. **Define mathematical expectation of a discrete random variable. The probability distribution for the number of defective items (X) in a sample is as follows:**

X : x :	0	1	2	3	4
P (x) :	0.35	0.39	a	0.06	0.01

Find: i. value of a ii. $P(X \geq 2)$ and F (3)

iii. E (X) and V (X) [P.U. 2011 Spring]

Solution 9

Do yourself

10. **A probability density function is given as**

$$F(x) = \begin{cases} \frac{x+2}{6} & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find

i. $P(X > 0)$ ii. F (1.5)

iii. $P(0 < X \leq 1)$ iv. E (X) [P.U. 2010 Spring]

Scanned by CamScanner

Solution 10

See Ex.8 Q.14

Remaining part:

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

$$= \int_0^2 x.\left(\frac{x+2}{6}\right) dx = \frac{1}{6}\int_0^2 (x^2+2x)\, dx = \frac{1}{6}\left[\frac{x^3}{3}+x^2\right]_0^2 = \frac{1}{6}\left[\frac{8}{3}+4\right] = \frac{10}{9}$$

11. Let X be continuous random variable with probability density function

$$f(x) = \begin{cases} ax & 0 \le x \le 2 \\ a & 2 \le x \le 4 \\ ax+3a & 4 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

i. Find the value of a

ii. Find $P(X \le 1.5)$ and $P(1.5 \le X \le 2.5)$

iii. Find $E(X)$ [P.U. 2010 Spring]

Solution 11:

Process is similar to Ex.8 Q.No. 11

Remaining part

ii. $P(1.5 \le X \le 2.5) = \int_{1.5}^{2.5} f(x)\, dx$

$$= \int_{1.5}^{2} \frac{1}{20}x\, dx + \int_{2}^{2.5} \frac{1}{20}\, dx = \frac{1}{20}\left[\frac{x^2}{2}\right]_2^{1.5} + \frac{1}{20}[x^2]_2^{2.5}$$

$$= \frac{1}{40}[2^2-(1.5)^2] + \frac{1}{20}[2.5-2] = \frac{1}{40}\times 1.75 + \frac{1}{20}\times 0.5 = 0.06875$$

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

$$= \int_0^2 x.\frac{1}{20}x\, dx + \int_2^4 x.\frac{1}{20}x\, dx + \int_4^6 \frac{x}{20}(x+3)\, dx$$

$$= \frac{1}{20}\times\left[\frac{x^3}{3}\right]_0^2 + \frac{1}{20}\left[\frac{x^2}{2}\right]_2^4 + \frac{1}{20}\left[\frac{x^3}{3}+\frac{3x^2}{2}\right]_4^6$$

$$= \frac{1}{20}\left[\frac{2^3-0^3}{3}+\frac{4^2-2^2}{2}+\frac{6^3-4^3}{3}+\frac{3(6^2-4^2)}{2}\right] = \frac{1}{20}\left[\frac{8}{3}+6+\frac{152}{3}+60\right] = 5.97$$

12. A random variable X has the following probability mass function;

Scanned by CamScanner

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

Find:

i. The value of k ii. Expected value of X

iii. Standard deviation of X [P.U. 2010 Fall]

Solution 12

i. Since the given function is discrete probability mass function so we must have $\Sigma P(x_i) = 1$

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

$\Rightarrow$ $4k + 0.6 = 1$

$\Rightarrow$ $k = \frac{0.4}{4} = 0.1$

ii. Expected value of X is given by

$E(X) = \Sigma x_i P(x_i) = -2 \times 0.1 - 1 \times k + 0 \times 0.2 + 1 \times 2k + 2 \times 0.3 + 3 \times k$

$= -0.2 - 0.1 + 0 + 0.2 + 0.6 + 0.3 = 0.8$

iii. Standard deviation $(\sigma) = \sqrt{va(X)} = \sqrt{E(X^2) - [E(X)]^2}$

Also, $E(X^2) = \Sigma x_i^2 P(x_i) = (-2)^2 \times 0.1 + (-1)^2 \times 0.1 + 0^2 \times 0.2 + 1^2 \times 0.2 + 2^2 \times 0.3 + 3^2 \times 0.1$

$= 0.4 + 0.1 + 0 + 0.2 + 1.2 + 0.9 = 2.8$

$\therefore$ S.D $(\sigma) = \sqrt{2.85 - (0.8)^2} = \sqrt{2.16} = 1.48$

13. Find the expectation of number on two dices when they are thrown. [P.U. 2007 Fall]

Solution 13

Do yourself.

14. A continuous random variable x follows the probability law $f(x) = Ax^2$, $0 \le x 1$

i. Determine A.

ii. Find the probability that X lies between 0.2 and 0.5.

iii. Mean and variance of the distribution. [P.U 2006 Spring]

Solution 14

See Ex. 8 Q.No. 17

Remaining part

iii. Mean of the distribution is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{1} x.\, Ax^2 dx = 0.3 \left[\frac{x^4}{4}\right]_0^1 = 0.075$$

Also the variance of the distribution is given by

$V(X) = E(X)^2 - [E(X)]^2 = 0.06 - (0.075)^2 = 0.054$

Scanned by CamScanner

where $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2 . Ax^2 dx = 0.3\left[\frac{x^5}{5}\right]_0^1 = 0.06$

Hence the mean and variance of the distribution is 0.075 and 0.054 respectively.

15. **Define the mathematical expectation of a discrete random variable. The probability distribution for the number of defective items (X) is a sample of 4 is as follow**

X:	0	1	2	3	4
P (X):	0.35	0.39	0.19	0.06	0.01

Calculate the expected value, variance and standard deviation for X.

[P.U. 2005 Fall, P.U. 2011 Spring]

Solution 15

Process is similar to Q.12.

16. **The length of time (in minute) that is a certain lady speaks on the telephone is found to be a random phenomenon with the probability function specified by the probability density function** $f(x) = \begin{cases} Ae^{-x/5} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

i. Find the value of A.

ii. What is the probability that the number of minute that she will talk over the phone is

a. more than 10 minute b. less than 5 minute

c. between 5 and 10 minute

iii. Find mean and variance of talk time on the telephone

[P.U. 2011 Fall/ 2015 Fall/ 2017 Fall]

Solution 16

i. Since the given function f (x) is probability density function then we must have

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_0^{\infty} Ae^{-x/5}\,dx = 1$$

$$\Rightarrow A\left[\frac{e^{-x/5}}{-1/5}\right]_0^{\infty} = 1 \Rightarrow -5A\left[e^{-\infty} - e^{-0}\right] = 1 \Rightarrow A = \frac{1}{5}$$

ii. a. Probability that the number of minutes that she will talk over the phone more than 10 minute is,

$$P(X > 10) = \int_{10}^{\infty} f(x)\,dx = \int_{10}^{\infty} \frac{1}{5} e^{-x/5}\,dx = -\left[e^{-\infty} - e^{-10/5}\right] = e^{-2}$$

b. Probability that the number of minutes that she will talk between 5 and 10 minute is given by

$$P(X < 5) = \int_0^5 f(x)\,dx = \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_0^5 = -\left[e^{-5/5} - e^{-5/5}\right] = e^{-1}.e^{-2}$$

Scanned by CamScanner

iii. The mean talk time is given by

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

$$= \int_{0}^{\infty} x\left(\frac{1}{5}e^{-x/5}\right) dx = \frac{1}{5}\left[x.\frac{e^{-x/5}}{(-1/5)} - 1.\left(\frac{-5}{1}\right).\frac{e^{-x/5}}{-1/5}\right]_0^{\infty}$$

$$= \frac{1}{5}[5x\, e^{-x/5} + 25\, e^{-x/5}] = -\left[x\, e^{-x/5} + 5\, e^{-x/5}\right]_0^{\infty} = -\left[x\, e^{-x/5}\right]_0^{\infty} + 5\left[e^{-x/5}\right]_0^{\infty}$$

$$= 0 - 5\left[e^{-\infty} - e^{-0}\right] = 5$$

Also the variance of x is given by:

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

where, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_{0}^{\infty} x^2 \frac{1}{5} e^{-x/5}\, dx$

$$= \frac{1}{5}\left[x^2.\frac{e^{-x/5}}{(-1/5)} - 2x\,(-5)\frac{e^{-x/5}}{(-1/5)} + 2.(-5).(-5)\frac{e^{-x/5}}{(-1/5)}\right]_0^{\infty}$$

$$= \frac{1}{5}[0 + 0 - 250\,(e^{-\infty} - e^{-0})] = \frac{250}{5} = 50$$

$\therefore$ $V(X) = 50 - (5)^2 = 50 - 25 = 25$

17. Suppose that X is a random variable whose probability density function is given by $f(x) = \begin{cases} c(4x - 2x2) & 0<x<2 \\ 0 & \text{otherwise} \end{cases}$

i. Determine the value of C

ii. Determine the cumulative distribution function of X

iii. Find $P(X > 1)$

iv. Find the $E(X)$ and $V(X)$ [P.U. 2016 Fall]

Solution 17

i. Since the given function f (x) is probability density function so we must have,

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

$$\Rightarrow \int_{0}^{2} c\,(4x - 2x^2)\, dx = 1$$

$$\Rightarrow c\left[\frac{4 \times x^2}{2} - \frac{2 \times x^3}{3}\right]_0^2 = 1 \Rightarrow c\left[2 \times 2^2 - \frac{2}{3} \times 2^3\right] = 1 \Rightarrow c \times \frac{8}{3} = 1 \Rightarrow c = \frac{3}{8}$$

Scanned by CamScanner

ii. Cumulative distribution function of X is given by

$F(X) = P(X \le x)$

$$= \int_0^x f(x)\,dx = \int_0^x \frac{3}{8}(4x - 2x^2)\,dx = \frac{3}{8}\left[\frac{4 \times x^2}{2} - \frac{2 \times x^3}{3}\right]_0^x = \frac{3}{8}\left[2x^2 - \frac{2x^3}{3}\right]$$

$$\therefore \quad F(X) = \begin{cases} \frac{3}{4}x^2 - \frac{x^3}{4} & \text{when } 0<x<2 \\ 0 & \text{otherwise} \end{cases}$$

iii. $P(X > 1) = \int_1^\infty f(x)\,dx = \int_1^2 \frac{3}{8}[4x - 2x^2]\,dx = \left[\frac{3x^2}{4} - \frac{x^3}{4}\right]_1^2 = \frac{1}{2}$

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

$$= \int_0^2 x \cdot \frac{3}{8}[4x - 2x^2]\,dx = \frac{3}{8}\left[\frac{4x^3}{3} - \frac{2x^4}{4}\right]_0^2 = \frac{3}{8}\left[\frac{4 \times 8}{3} - \frac{2 \times 16}{4}\right] = 1$$

Also, $V(X) = E(X^2) - (E(X))^2 = 1.2 - (1)^2 = 0.2$

where, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^2 x^2 \frac{3}{8}(4x - 2x^2)\,dx = \left[\frac{3}{8}\left(4 \times \frac{4x^4}{4} - 2 \times \frac{x^5}{5}\right)\right]_0^2 = 1.2$

18. The diameter of an electric cable say X is assumed to be continuous random variable with $f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

i. Check whether the function is p.d.f or not

ii. Find E (X) and V (X). [P.U. 2013 Fall]

Solution 18

The process is same as Q. 17

19. Define the mathematical expectation of a discrete random variable. A probability distribution is given as

X	0	1	2	3	4	5
P (X)	0.26	0.25	0.11	0.02	0.25	0.11

Find

i. $P(X = 4)$ ii. $P(0 < X < 4)$ iii. $F(3)$

iv. $E(X)$ v. $V(X)$ [P.U. 2099 Fall]

Solution 19

Process is similar to Q. 12

Scanned by CamScanner

20. A random variable X has the density function $f(x) = a + bx^2$, $0 < x < 1$. Determine a and b show that, mean is $\frac{2}{3}$. Also find the variance of the distribution. [P.U. 2009 Fall]

Solution 20

1st part

Since the given function is a probability density function then we must have

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

$$\Rightarrow \int_0^1 (a + bx^2)\,dx = 1 \Rightarrow \left[ax + \frac{bx^3}{3}\right]_0^1 = 1 \Rightarrow 3a + b = 3 \quad \text{...(i)}$$

Also given that mean is $\frac{2}{3}$

$$\text{Then, } E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^1 x\,(a + bx^2)\,dx$$

$$= \left[\frac{ax^2}{2} + \frac{bx^3}{3}\right]_0^1 = \frac{2}{3} = \frac{2a+b}{4} = 6a + 3b = 8 \quad \text{...(ii)}$$

Solving (i) and (ii) we get,

$a = \frac{1}{2}$ $\qquad$ $b = 2$

2nd part

Also we know that the variance of the given distribution is,

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

where,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2\left(\frac{1}{3} + 2x^2\right)dx = \left[\frac{1}{3}\cdot\frac{x^3}{3} + 2\frac{x^5}{5}\right]_0^1 = \frac{1}{9} + \frac{2}{5} = \frac{23}{45}$$

$$\therefore \quad V(X) = \frac{23}{45} - \left(\frac{2}{3}\right)^2 = \frac{23}{45} - \frac{4}{9} = \frac{1}{15}$$

21. A random variable X has the following probability distribution

X : x	0	1	2	3	4	5	6	7	8
P (X)	a	3a	5a	7a	9a	11a	13a	15a	16a

i. Determine the value of a

ii. Find $P(X < 3)$ and $P(0 < X < 5)$

iii. Mean and variance of X. [P.U. 2003 Fall]

Solution 21

The process is similar Q.12

Scanned by CamScanner

Exam asked short question

22. **Write short notes on probability distribution function.** [P.U. 2005 Fall]

Solution 22

See in theoretical part of this part

23. **Write short notes on probability mass function.** [P.U. 2005 Fall]

Solution 23

See in theoretical part of this part

24. **Write short notes on Random variable.**
[P.U. 2003 Fall/ 2009 Fall/ 2010 Fall/ 2010 Spring]

Solution 24

See in theoretical part of this part

25. **Write short notes on distribution function.** [P.U. 2003 Fall]

Solution 25

See in theoretical part of this part

Scanned by CamScanner

10
Chapter

Theoretical Discrete Distribution

There are many types of theoretical discrete probability or discrete frequency distributions. Here we discuss about the more commonly used discrete probability distribution which are given as follows,

i. Bernoulli Probability Distribution
ii. Binomial Probability Distribution
iii. Hyper geometric Probability Distribution
iv. Poisson Probability Distribution
v. Negative Binomial Probability Distribution
vi. Geometric Distribution

"Bernoulli Probability Distribution"

Bernoulli trial:

An experiment consisting only two possible outcomes either success or failure is called a Bernoulli trial.

Bernoulli Probability Distribution:

A random variable 'X' is said to have Bernoulli distribution with parameter p, if its probability mass function (p.d.f) is given by

$$P(X = x) = \begin{cases} p & \text{if} \quad x = 1 \\ q & \text{if} \quad x = 0 \end{cases}$$

where $p \rightarrow$ probability of success in a single trial

$q \rightarrow$ probability of failure in a single trial

Remarks:

1. We can use Bernoulli distribution if
   i. The experiment consists of independent repeated trials.
   ii. In the trial there are only two possible outcomes successes or failure.
   iii. Probability of success (p) and probability of failure (q) must remain constant for each trial.
   iv. The random variable X is occurrence of success in the trial.

Scanned by CamScanner

2. The mean of the Bernoulli distribution, $\mu = E(X) = p$
3. The variance of Bernoulli distribution, $V(X) = \sigma^2 = pq$
4. $q = 1 - p$ and $p + q = 1$
5. The moment generating function of Bernoulli distribution is
$M_x(t) = E[e^{tx}] = q + e^t.p$

Binomial Distribution

A discrete random variable X is said to have Binomial probability distribution with parameter n and p if its probability mass function (p.d.f) is given by,

$$B(x; n, p) = P(X = x) = \begin{cases} c(n, x)\, p^x.q^{n-x} & x = 0,1,2,3,\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

where, n = the total number of trials.
p = probability of success in a single trial.
q = probability of failure in a single trial.
P (X = x) = probability of getting exactly x success in n trials.

Remarks:

1. i. Binomial distribution follows all the condition of Bernoulli distribution.
   ii. It has two variables (parameters) n and p.
   iii. n, the number of trial must be finite and fixed.
   iv. P (X = x) gives the probability of exactly x success in n independent trials.
2. The mean of the Binomial distribution, $\mu = E(X) = np$
3. The variance of Binomial distribution, $V(X) = \sigma^2 = npq$
4. The moment generating function of Binomial distribution
$M_x(t) = E(e^{tx}) = (q + pe^t)^n$
5. The cumulative binomial distribution function (c.d.f) of the binomial distribution is given by, $B(x; n, p) = P(X \le x) = \sum_{k=0}^{x} b(k; n, p)$

Hyper geometric Distribution

A discrete random variable X is said to have hyper geometric distribution with parameters n, N_1, N_2 if its probability mass function (p.m.f) is denoted and defined by

$$P(X = x) = h(x; n, N_1, N_2) = \frac{C(N_1, x)\, C(N_2, n - x)}{C(N, n)}, \; x = 0, 1, 2, \ldots, N_1 \le n$$

where, n = sample size of the distribution.
N_1 = no. of type one items in the distribution.
N_2 = no. of type two items in the distribution.
$N = N_1 + N_2$
P (X = x) is the probability of exactly x – success and (n – x) failures.

Remarks:

1. There are three independent constants (parameters) n, N_1, N_2
2. The set of N objects are divided into two different parts N_1 objects classified as success and N_2 objects classified as failure.
3. The mean of the hyper geometric distribution $= \mu = E(X) = \frac{nN_1}{N}$
4. The variance of the hyper geometric distribution
$$V(X) = \sigma^2 = \frac{nN_1(N - N_1)(N - n)}{N^2(N - 1)}$$

Scanned by CamScanner

Poisson Distributions

A discrete random variable X is said to have Poisson distribution with parameter $m > 0$, if its probability mass function (p.m.f) is given by

$$P(X = x) = \frac{e^{-m}m^x}{x!}, x = 0, 1, 2, 3, \ldots$$

where, m = mean of the distribution.
x = number of events.
P (X = x) = the probability of x success.

Remarks:

1. Poisson distribution is used in the following conditions.
   i. When the number of trials is infinitely large (i.e $n \to \infty$)
   ii. When the probability of success in each trial is indefinitely small (i.e $p \to 0$ or $q \to 1$)
   iii. The mean of the distribution is constant and finite. [i.e $np = m$]
2. Poisson distribution is limiting case of binomial distribution.
3. We use the notation X – P (m) to denote that X is a Poisson distribution.
4. In this distribution m denotes the rate per unit time or per unit area.
5. The mean of the Poisson distribution = m.
6. The variance of the Poisson distribution = m.
7. The moment generating function of Poisson distribution is
   $M_x(t) = E[e^{tx}] = e^{m(e^t-1)}$
8. If X is a Poisson random variable with p.m.f
   $$P(X = x) = \frac{e^{-m}.m^x}{x!}, x = 0, 1, 2, 3, \ldots$$
   then Poisson distribution function (c.d.f) is defined as
   $$P(X \le x) = F(x, m) = \sum_{k=0}^{m} \frac{e^{-m}.k^x}{k!}$$

Negative Binomial Distribution : [Pascal distribution]

A discrete random variable X is said to follow a negative binomial distribution with parameter r and p if its probability mass function (pm.f) is given by
$P(X = x) = nb(x : r, P) = C(x + r - 1, r - 1)p^r q^x = C(x + r - 1, x)\, p^r q^x$
P(X = x) – probability of x failures and and (r + 1) success in (x + r – 3) trial
x + r – the total no. of trial
r – the total no. of success.

Remarks:

1. We use the negative binomial distribution
   i. If the experiment consists of x repeated and sequence of independent trials.
   ii. Each trial has only two possible outcomes either success or failure and their probabilities for each trial is fixed and same.
2. The mean of the negative binomial distribution = $E(X) = \frac{rq}{p}$
3. The variance of negative binomial distribution = $V(X) = \frac{rq}{p^2}$
4. If we put r = 1 in the negative binomial distribution it is called the geometric distribution. That is,

Scanned by CamScanner

Geometric Distribution :

A discrete random variable X is said to follow an geometric distribution with parameter p if its p.m.f is given by

$p(X = x) = p \times q^x, x = 0,1,2,3,\ldots$

Remarks:

1. The mean of geometric distribution $= \frac{q}{p}$
2. The variance of geometric distribution $= \frac{q}{p^2}$
3. The moment generating function of geometric distribution $= p(1 - qe^t)^{-1}$.

Exercise 10

1. **What is Bernoulli distribution? Find moment generating function of Bernoulli distribution.**

Solution 1

See in theoretical part of this chapter

Moment generating function of Bernoulli Distribution:

The moment generating function of random variable x is defined as.

$$M_x(t) = E(e^{tx}) = \sum_{X=1}^{0} e^{tx} P(x) = e^{t \times 0} p(X = 0) _ e^{t \times 1} P(X = 1) = q + e^t P$$

2. **If p is the probability of success at a single trial. Obtain the probability of x successes out of n independent trial. Prove that $^nC_x p^x q^{n-x}$ is probability mass function.**

Solution 2

It p is the probability of success at a single trial as well as probability of failure is $q = 1 - p$. Now for the probability of x success out of n independent trial is given as follows. If we get x success out of n independent trial there will be $(n - x)$ failure. Suppose S and F denote the success and failure in a single trial, then p and q be the P(S) and P(F). Suppose x Successes occurs in the specific order SFSSFFSSS ..... FF SSS (say).

Then the probability of x success and consequently $(n - x)$ failure in n independent trial is given by

P(SFSSFF...SSFF) = P(S) P(F) F(S) F(S) P(F) P(F) ... P(S) P(S) . P(P) P(F) [being each trial are independent]

$= pqppqq \ldots ppqq$

$= \underbrace{ppp \ldots p}_{x \text{ times}} \times \underbrace{qqq \ldots q}_{(n-x) \text{ times}}$

$= p^x \cdot q^{n-x}$

Also the probability for each such case is $p^x q^{n-x}$.

But out of n trials, x success can occurs in n_{c_x} Hence the probability of x success in n trials in any order what ever is given by the additional theorem of probability. The required probability is given by

Scanned by CamScanner

$P(X = x) = p^x . q^{n-x} + p^x q^{n-x} + p^x . q^{n-x} + \ldots + \ldots$ to ${}^nC_x = {}^nC_x p^x q^{n-x}$

Which is the probability mass function

3. **Define the binomial distribution with parameter p and n and also find mean, standard deviation and moment generating function.**

Solution 3

A discrete random variable x is said to have binomial distribution with parameters n and p if its p.m.f. is given by

$$b(x; n, p) = P(X = x) \begin{cases} {}^nC_x\, p^x q^{x-x}, & n = 0,1,2,3,\ldots\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

Mean, s.d and moment generating function of Binomial Distribution:-

We have know that:

$$\text{Mean} = E(X) = \sum_{X=0}^{n} xP(x)$$

$$= \sum_{X=0}^{n} x\, {}^nC_x\, p^x q^{n-x} = \sum_{X=0}^{n} x . \frac{n!}{(n-x)!\, x!} \times p^x q^{n-x}$$

$$= \sum_{X=0}^{n} \frac{x\, n(n-1)!}{x(x-1)!\,(n-x)!} \times p . p^{x-1} . q^{n-x}$$

$$= np \sum_{X=0}^{n} {}^{n-1}C_{x-1} p^{x-1} q^{(n-1-x+1)} = np\,(q+p)^{n-1}$$

$$= np.1\ [\because p + q = 1]$$

$$\text{Also, } \Sigma(X^2) = \sum_{X=0}^{n} x^2 p(x)$$

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

$$= \sum_{X=0}^{n} x(x-1) \frac{n(n-1)(n-2)!}{x(x-1)(x-2)!\,(n-x)!}\, p^2 . p^{x-2} . q^{n-x} + \sum_{X=0}^{n} x \frac{n!}{(n-x)!\, x!}\, p^x q^{n-x}$$

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

$$= n(n-1)\left[{}^{n-2}C_0\, q^{x-2} + {}^{n-2}C_1\, pq^{n-3} + \ldots + {}^{n-2}C_{x-2}\, p^{n-2} + np \right]$$

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

$$\text{s.d} = \sqrt{Var(x)} = \sqrt{E(X^2) - (E(X))^2}$$

$$= \sqrt{n(n-1)\, p^2 + np - (np)^2} = \sqrt{n^2p^2 - np^2 + np - n^2p^2} = \sqrt{np(1-p)} = \sqrt{npq}$$

Also the moment generating function is

Scanned by CamScanner

$$M_X(t) = E(e^{tx}) = \sum_{X=0}^{n} e^{tx} P(x) = \sum_{X=0}^{n} e^{tx}, {}^nC_x (e^t p)^x q^{n-x}$$

$$= {}^nC_0 q^n + {}^nC_1 C^t p\ q^{n-1} + {}^nC_2 (e^t p)^2. q^{n-2} + \ldots + {}^nC_n (e^t p)^n$$

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

4. **Prove that Poisson distribution is limiting case of binomial distribution under certain conditions. Write down the conditions.**

Solution 4

The probability of x success in a series of n independent trials is given by a Binomial distribution as:

$$b(x : n, p) = P(X = x) = \begin{cases} {}^nC_x\, p^x q^{n-x} & X = 0,1,2,\ldots\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

Now we obtain the probability mass function as a limiting case as $x \to \infty$, $p \to 0$ and $np = \lambda$ Hence:

$$b(x; n, p) = \lim_{n \to \infty} \frac{n!}{(n-x)!\,x!} \left(\frac{1}{n}\right)^x . \left(1 - \frac{1}{n}\right)^{n-x}$$

$$= \lim_{n \to \infty} \frac{n(n-1)(n-2)\ldots\{n-(x-1)\}(n-x)}{(x-x)!\,x!} \left(\frac{1}{n}\right)^x . \left(1 - \frac{1}{n}\right)^{n-x}$$

$$= \lim_{n \to \infty} \frac{n^x \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{x-1}{n}\right)}{n^x . x!} . \lambda^x . \left(1 - \frac{1}{n}\right)^{n-x}$$

$$= \frac{x}{x!} \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{x-1}{n}\right) \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^{n-x} \times \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^x$$

$$= \frac{e^{-\lambda} \lambda^x}{x!} \left[\because \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^x = e^{-\lambda}, \lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^x = 1\right]$$

Which is the Possion distribution.

Hence the Possion distribution is the limiting case of Binomial distribution.

The conditions are:

(i) n, the number of trials is infinitely large i.e. $n \to \infty$.

(ii) p, the constant probability of success in each trial is indefinitely small i.e. $p \to 0$.

(iii) $np = m$ (say) is finite.

5. **Define the Poisson distribution. Prove that mean and variance of Poisson distributions are equals, also find m.g.f.**

Solution 5

A discrete random variable X is said to have Poisson distribution with parameter $m>0$ is its p.m.f. is given by $P(X = x) = \begin{cases} \dfrac{e^{-m} m^x}{x!} & X = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$

Scanned by CamScanner

We use the notation $X \sim P(m)$ to denote that X is a Poisson variable with parameter m.

"Mean and variance of Possion Distribution"

$$\text{Mean} = E(X) = \sum_{X=0}^{\infty} xP(x) = \sum_{X=0}^{\infty} \frac{x\, e^{-m} m^x}{x!}$$

$$= e^{-m} \sum_{X=0}^{\infty} \frac{x\, m.\, m^{x-1}}{x(x-1)!} \quad = me^{-m}\left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots\right] = m.e^{-m}.e^{m} = m$$

$$\text{Also, } E(X^2) = \sum_{X=0}^{\infty} x^2 p(x) = \sum_{X=0}^{\infty} [x(x-1) + x] \frac{e^{-m} m^x}{x!}$$

$$= \sum_{X=0}^{\infty} \frac{x(x-1)\, e^{-m} m^2 m^{x-2}}{x(x-1)(x-2)!} + \sum_{X=0}^{\infty} \frac{x\, e^{-m}.\, m.\, m^{x-1}}{x(x-1)!}$$

$$= e^{-m}.\, m^2 \sum_{X=2}^{\infty} \frac{m^{x-2}}{(x-2)!} + e^{-m} \times m \sum_{X=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

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

$$= m^2.e^{-m} e^{m} + e^{-m}.m = m^2 + m$$

$$\because \text{Variance} = V(X) = E(X^2) - [E(X)]^2 = m^2 + m - m^2 = m$$

Hence mean = variance = m.

Moment generating function:

We know that the moment generating function is given by:

$$M_X(t) = E(e^{tx}) = \sum_{X=0}^{\infty} e^{tx}.\, P(x)$$

$$= \sum_{X=0}^{\infty} e^{tx}.\frac{e^{-m} m^x}{x!} = e^{-m} \sum_{X=0}^{\infty} \left(\frac{e^t.m}{x!}\right)^x$$

$$= e^{-m}\left[1 + \frac{me^t}{1!} + \frac{(me^t)^2}{2!} + \frac{(me^t)^3}{3!} + \ldots\right] = e^{-m}.\, e^{me^t} = e^{m(e^t - 1)}$$

6. **An urn contains, 'a + b' items of which 'a' are number of first type and 'b' are of second type. Find the probability of drawing x of first type out of n drawn randomly.**

Solution6

Let p and q denote the probability of success and failure of a single trial.

Then, $p = \frac{a}{a+b}$, $q = \frac{b}{a+b}$.

Probability of x success in n trials is given by:

$$P(X = x) \quad = {}^nC_x \left(\frac{a}{a+b}\right)^x \left(\frac{b}{a+b}\right)^{n-x} = {}^nC_x \frac{a^x (b)^{n-x}}{(a+b)^n}$$

Scanned by CamScanner

7. **Define negative binomial distribution. Find mean and variance of negative binomial distribution.**

Solution 7

A discrete random variable X is said to follow a negative binomial distribution with parameter r and p if its p.m.f. is given

$$\text{by } P(X = x) = \begin{cases} \binom{x+r-1}{r-1} p^r q^x & X = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Mean and Variance

$$\text{Mean} = E(X) = \left\{\frac{d}{dt}[M_x(t)]\right\}_{t=0}$$

$$= \left[\frac{d}{dt}\{p^r (1 - qe^t)^{-r}\}\right]_{t=0}$$

$$= [p^r.(-r)(1 - qe^t)^{-r-1} \times (-qe^t)]_{t=0} = rq\, p^r (1 - q)^{-r-1} = rq\,.\,p^r.\,p^{-r}.\,p^{-1} = \frac{rq}{p}$$

$$\text{Also, } E(X^2) = \left[\frac{d^2}{dt^2}(M_x(t))\right]_{t=0}$$

$$= \left[\frac{d^2}{dt^2}\{p^r (1 - qe^t)^{-r}\}\right]_{t=0}$$

$$= p^r.\,rq \times \frac{d}{dt}\left[e^t (1 - qe^t)^{-r-1}\right]_{t=0}$$

$$= p^r.rq\left[e^t\{-(r+1)(1 - eq^t)^{-r-2} \times (-te^t) + e^t (1 - qe^t)^{-r-1}\right]_{t=0}$$

$$= p^r r^q\left[q(r+1)(1 - q)^{-r-2} + (1 - q)^{-r-1}\right]$$

$$= p^r r^q.\,q(r+1).p^{-r} p^{-2} + p^r.\,rq\,,\,p^{-r}.p^{-1}$$

$$= \frac{r(r+1)q^2}{p^2} + \frac{rq}{p}$$

$$V(X) = E(x^2) - (E(x)^2$$

$$= \frac{r(r+1)q^2}{p^2} + \frac{rq}{p} - \left(\frac{rq}{p}\right)^2$$

$$= \frac{r^2 q^2}{p^2} + \frac{rq^2}{p^2} + \frac{rq}{p} - \frac{r^2 q^2}{p^2} = \frac{rq^2 + rpq}{p^2} = \frac{rq(q+p)}{q^2} = \frac{rq}{p^2}$$

8. **Write short notes on.**

i. difference between binomial and hyper geometric distribution.

ii. difference between binomial and negative binomial distribution.

iii. difference between Bernoulli and binomial distribution.

iv. What is the relation between binomial and distribution and Poisson distribution.

v. In which fields, we may use Poisson distribution.

Scanned by CamScanner

Solution 8

(a) The difference between binomial and hyper-geometric distribution is given as follows.

Binomial distribution	Hyper-geometric distribution
(i) In this distribution items are replaced after each trial.	(ii) In this distribution items are replaced after each trial.
(ii) Entire system is single.	(ii) Entire system is divided into two group.
(iii) The probability of success/failure is constant throughout the experiment.	(iii) The probability of success/failure does not remain constant throughout the experiment.
(iv) The size of population remain constant after each trial.	(iv) The size of population decrease after the each trial.
(v) Each trial are independent.	(v) Each trials are not independents.

(b) **The difference between binomial and negative binomial distribution is given as follows.**

Binomial Distribution	Negative Binomial Distribution
(i) The last trial may be either success/failure in binomial experiment.	(i) The last trial must be success in negative binomial experiment.
(ii) Binomial distribution is converse of negative binomial distribution.	(ii) Negative binomial distribution is the converse of binomial distribution.
(iii) The number of trial is fixed.	(iii) The number of trial is not fix.
(iv) The number of success or failure to occur is a random variable.	(iv) The number of trial is a random variable

(c) **The difference between Bernoulli and Binomial distribution is given as follows.**

Bernoulli Distribution	Binomial Distribution
(i) Bernoulli distribution has single parameter p.	(i) Binomial distribution has two parameters n and p.
(ii) It is used for only two values of x.	(ii) It can be used for finite values of x.
(iii) It is the particular case of Binomial distribution.	(iii) It is general form of Bernoulli distribution.

(d) The relation between Binomial and Poisson distribution is given as follows.

We know that poison distribution is the limiting case of the Binomial distribution. Both are same but we get the poison distribution if we use the following conditions in the Binomial distribution.

(i) If we make number of trials infinite in the binomial distribution it becomes the Poisson distribution.

(ii) If the probability of success p is very small in the binomial we use Poisson distribution are constant and finite.

(iii) The mean = np is both distribution are constant and finite.

Scanned by CamScanner

(e) **The use of Poisson distribution is given as follows:**

(i) It is used to model the number of times that a certain event occurs per unit of times distance or volume.

(ii) It is used as a mathematical model describing behaviour such as a radioactive, decay, arrivals of peoples in a line, planes arriving at an airport cars pulling into a gas station. etc.

9. **Four coins are tossed simultaneously what is the probability of getting**

i. 2 head and 2 tails ii. at least 2 heads

iii. at least one head.

Solution 9

The total number of a coin tossed (n) = 4

The probability of getting head in each toss (p) = ½

The probability of getting tail in each toss (q) = ½

(i) Probability of getting 2 heads and 2 tails P(2H and 2T)

$P(X = 2) = {}^4C_2p^2 . q^{4-2} = 6 \times (0.5)^2 \times (0.5)^2 = 0.375$

(ii) Probability of getting at least 2 heads $P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$

$= {}^4C_2p^2 . q^{4-2} + {}^4C_3p^3q^{4-3} + {}^4C_4p^4q^{4-4}$

$= 6 \times (0.5)^2 \times (0.5)^2 + 4 \times (0.5)^3 \times 0.5 + 1 \times (0.5)^4 \times 1$

$= 0.375 + 0.25 + 0.0625 = 0.6875$

(iii) Probability of getting at least one head $P(X \geq 1) = 1 - P(X < 1)$

$= 1 - P(X = 0) = 1 - {}^4C_0(0.5)^0 (0.5)^{4-0} = 1 - 0.0625 = 0.9375$

10. **Six dice are thrown 729 times. How many times do you expected at least three dice to shown a five or six?**

Solution 10

Let X denote the random variable falling 5 or 6 in a dice.

Then probability of occurring 5 or 6 $(P) = \frac{2}{6} = \frac{1}{3}$

Probability of not occurring 5 or 6 $(P) = 1 - \frac{1}{3} = \frac{2}{3}$.

The number of dice thrown (n) = 6

The number of times dice thrown (N = 729).

Now the probability of getting at least three dice to show a 5 or 6 is given by

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

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

$= 1 - {}^6C_0\left(\frac{1}{3}\right)^0 \times \left(\frac{2}{3}\right)^6 + 6C_1\left(\frac{1}{3}\right)^1 . \left(\frac{2}{3}\right)^{6-1} + 6C_2 . \left(\frac{1}{3}\right)^2 . \left(\frac{2}{3}\right)^{6-2}$

$= 1 - \left[\frac{1 \times 64 + 6 \times 1 \times 32 + 15 \times 1 \times 16}{729}\right] = 1 - 0.680 = 0.32$

Now the expected number of times at least three dice to shown 5 or 6

$= NP(X \geq 3) = 729 \times 0.32 = 233.$

Scanned by CamScanner

11. **12% of the items produced by a machine are defective. What is the probability that out of a random sample of 20 items produced by the machine, 5 are defective?**

Solution 11

Let X denote the random variable of a machine producing item.

The number of sample items (n) = 20

Probability of getting defective (success) in a single process p = 12% = 0.12

Then q = 1 – p = 1 – 012 = 0.88

Now probability of getting 5 defective items

$$P(X = 5) = {}^{20}C_5 (0.12)^5 \times (0.88)^{20-5} = 0.0567$$

12. **The phone lines to an airline reservation system are occupied 40% of the time. Assume that the events that the lines are occupied on successive calls are independent. Assume that 10 calls are placed to the airline,**

a. What is the probability that for exactly three calls the lines are occupied?

b. What is the probability that for at least one call the lines are not occupied?

c. What is the expected number of calls in which the lines are occupied?

Solution 12

Let X denote the random variable the phone lines occupied.

The total sample calls (n) = 10

The probability of success in a single trial (p) = 40% = 0.4

Then q = 1 – p = 1 – 0.4 = 0.6

(a) The probability that for exactly three calls the lines are occupied

$$P(X = 3) = {}^{10}C_3(0.4)^3 . (0.6)^{10-3} = 0.215$$

(b) For the occupied case,

p = 0.6 and q = 0.4, n = 10

Now, $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - {}^{10}C_3(0.6)^0 . (0.4)^{10} = 0.998.$

(c) For the expected number of calls E(X) = np = 10 × 0.4 = 4

13. An insurance sales man sells polices to five men, all of identical age and good health. According to the actuarial table, the probability that a man of the particular age will be alive 30 year hence is $\frac{2}{3}$. Find the probability that in 30 years

i. all five men ii. at least one man

iii. at most three men, will be lived.

Solution 13

Let X denote the random variable that a man of particular age will be alive in 30 year hence.

The probability of being alive in 30 yrs $P = \frac{2}{3}$

Then, $q = 1 - p - \frac{2}{3} = \frac{1}{3}$, n = 5

Scanned by CamScanner

(i) The probability of all five men will be alive i.e.

$$P(X = 5) = {}^5C_5 \left(\frac{2}{3}\right)^5 . \left(\frac{1}{5}\right)^{5-5} = 0.1316$$

(ii) The probability of at least one man alive

$$P(Xx \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - {}^5C_0 \left(\frac{2}{3}\right)^0 . \left(\frac{1}{3}\right)^{5-0} = 0.9959$$

(iii) The probability of at most three man will be alive

i.e. $P(X \geq 3) = P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)1$

$$= 1 - [P(X = 4) + P(X = 5)]$$

$$= 1 - \left[{}^5C_4 \left(\frac{2}{3}\right)^4 . \left(\frac{1}{3}\right)^{5-4} + {}^5C_5 \left(\frac{2}{3}\right)^5 . \left(\frac{1}{3}\right)\right]$$

$$= 1 - [0.3292 + 0.1316 = 0.5392$$

14. Out of 9000 families 4 children each, how many families would you expect to have,

a. 2 boys and 2 girls b. at least one boy c. no girl

d. 3 boys and 1 girl e. at most 2 girls

f. at least one boy and one girl. If boys and girls are equally probable.

Solution 14

Let X denote the random variable denoting the number of boys in the family with 4 children each:

The probability of getting a child boy $p = 0.5$

Also $q = 1 - p = 1 - 0.5 = 0.5$, $N = 9000$, $n = 4$.

Let r be the number of success (boy) in the experiment then by binomial distribution.

$P(X = r) = {}^nC_r\, p^r q^{n-r}$, $r = 0, 1, 2, 3, \ldots, n$

i. The number of family expected to have 2 boys and 2 girls is given by = E(X = 2)

$N . P(X = 2) = 9000 \times C(4, 2)\, (0.5)^2 . (0.5)^{4-2} = 3375.$

ii. The number of families to have at least one boys = E (X ≥ 1)

$= N . P(X \geq 1) = N[1 - P(X = 0)]$

$= 9000\, [1 - C(4, 0)\, (0.5)^2 . (0.5)^{4-0}] = 9000\, [1 - 0.0625] = 8437.5 \approx 8438.$

iii. The number of families to have no girls (means all 4 boys) = E (X = 4)

$= N . P(X = 4\ 1) = 9000 \times C(4,0)\, (0.5)^4 . (0.5)^0 = 9000 \times 0.0625 = 562.5 \approx 563$

iv. The number of families expected to have 3 boys and 1 girl = E (X =3)

$= N . P(X = 3) = 9000 \times c\, (4, 3)\, (0.5)^3 . (0.5)^{4-3} = 2250.$

v. The number of families expected to have at most 3 girls (at least 2 boys)

$= E\, (X \geq 2) = N . P(X \geq 2) = 9000 \times [P(X = 2) + P(X = 3) + P(X = 4)]$

$= 9000\, [C(4, 2)\, (0.5)^2 . (0.5)^{4-2} + C\, (4, 3)\, (0.5)^3 (0.5)^{4-3} + C(4, 4)^0 (0.5)^4]$

$= 9000\, [0.375 + 0.25 + 0.0625] = 6187.5 = 6188$

vi. The number of families expected to have at least one boy and one girl

$= E(X \geq 1) = NP(X \geq 1) = 900\, [1 - P(X = 0)] = 8438$

Scanned by CamScanner

15. **Between the hours of 2 and 4 P.M., the average number of phone calls per minute coming into the switch-board of a company is 2.5. Find the probability that during one particular minute there will be**

i. no call at all. ii. exactly 3 calls. iii. at least 5 calls.

Solution 15

Let X be a random variable which denotes the number of telephone calls per minute. Also x follows Poisson distribution with parameter m = 2.5.

(i) Now the probability of coming no calls at all is given by

$$P(X=3)=\frac{e^{-m}.m^{o}}{0!}=e^{-2.5}=0.082$$

(ii) Probability of coming exactly 3 calls is given by

$$P(X=3)=\frac{e^{-m}.m^{3}}{3!}\frac{e^{-2.5}.(2.5)^{3}}{6}=\frac{15.625}{73.095}=0.214$$

(iii) The probability of getting at least 5 calls is given by

$$P(X\geq 5)=1-P(X<5)$$

$$=1-[P(X=0)+p\,[X=1)+P(X=2)+P(X=3)+P(X=4)]$$

$$=1-\left[\frac{e^{-m}.m^{0}}{0!}+\frac{e^{-m}.m^{1}}{1!}+\frac{e^{-m}\,m^{2}}{2!}+\frac{e^{-m}.m^{3}}{3!}+\frac{e^{-3}.m^{4}}{4!}\right]$$

$$=1\,e^{-2.5}\left[\frac{(2.5)^{0}}{0!}+\frac{(2.5)^{1}}{1!}+\frac{(2.5)^{2}}{2!}+\frac{(2.5)^{3}}{3!}+\frac{(2.5)^{1}}{4!}\right]$$

$$=1-0.0821\,[1+2.5+3.125+2.604+1.628]=0.109$$

16. **It is known that average number of that suicide per week in Kathmandu is 1.5. Express the number of suicide in a month. Find the probability that there is 5 or more suicide in a month. (4 week = months)**

Solution 16

Let X denote the number of suicides per month Also λ denote the average number of suicides per month.

We have given the average number of suicides per week = 1.5. Therefore, $\lambda = 1.5 \times 4 = 6$. Then, using Poisson distribution, the probability of exactly x suicides in a given month is $P(X=x)\frac{e^{-\lambda}\lambda^{x}}{x!}$, $X = 0, 1, 2, \ldots$

Then the probability of there is 5 or more suicides in a month is given by $P(X\geq 5)$

$$=\sum_{k=5}^{\infty}\frac{e^{-6}.6^{k}}{k!}=1-\sum_{k=0}^{4}\frac{e^{-6}.6^{k}}{k!}=1-\left[\frac{e^{-6}.6^{0}}{0!}+\frac{e^{-6}.6^{2}}{1!}+\frac{e^{-6}6^{1}}{2!}+\frac{e^{-6}.6^{3}}{3!}+\frac{e^{-6}.6^{4}}{4!}\right]$$

$$=1-e^{-6}\,[1+6+18+36+54]=1-0.285=0.715$$

17. **A typist made 2.8 mistakes on an average per page. Find the probability that in the page typed by him,** [P.U. 2010 Fall]

i. there is no mistake ii. there are two or less mistakes

iii. more than 10 mistakes

Scanned by CamScanner

Solution 17

Let X denote the number of mistakes per page and λ denote the average mistakes per page. then $\lambda = 2.8$

Then using Poisson distribution the probability of exactly x mistakes in the given page is $P(X = x) = \frac{e+ \lambda x}{x!}$, X = 0, 1, 2, 3, ...

(i) The probability of getting no mistakes is given by $P(X = 0) \frac{e^{-2.8}(2.8)^0}{0!} = 0.0608$

(ii) The probability of getting two or less mistakes is

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

$$= \frac{e^{-2.8}(2.8)^0}{0!} + \frac{e^{-2.8}.(2.8)^1}{1!} + \frac{e^{-2.8}(2.8)^2}{2!} = e^{-2.8}[1 + 2.8 + 3.92] = 0.469$$

(ii) The probability of getting more then 10 mistakes is given by

$$P(X \ge 10) = 1 - P(X < 10) = 1 - \sum_{k=0}^{9} \frac{e^{-2.8}.(2.8)^k}{k!}$$

$$= 1 - e^{-2.8}\left[\frac{(2.8)^0}{0!} + \frac{(2.8)^1}{1!} + \frac{(2.8)^2}{2!} + \frac{(2.8)^3}{3!} + \frac{(2.8)^4}{4!} + \frac{(2.8)^5}{5!} + \frac{(2.8)^6}{6!} + \frac{(2.8)^7}{7!} + \frac{(2.8)^8}{8!} + \frac{(2.8)^9}{9!}\right]$$

$$= 1 - e^{-2.8}[1 + 2.8 + 3.92 + 3.66 + 2.56 + 1.434 + 0.669 + 0.268 + 0.0937 + 0.02915]$$

$$= 1 - 0.999 = 0.0001$$

18. If 2 percent of electric bulbs manufactured by a certain company are defective, find the probability that in a sample of 200 bulbs.

i. less then 2 bulbs

ii. more then 3 bulbs, are defective.

Solution 18

Let X denote the number of defective bulbs manufactured by a company and λ denote average number of defective bulbs manufactured by a company Then.

$P = 2\% = 0.02, n = 200$

$\because \lambda = np = 0.02 \times 200 = 4$

(i) The probability that less than 2 bulbs to be defective is given by

$$P(X < 2) = P(X = 0) + P(X = 1) = \frac{e^{-4}.4^0}{0!} + \frac{e^{-4}.4^1}{1!} = 5 \times e^{-4} = 0.0915$$

(ii) The probability that more than 3 bulbs are defective is $P(X \ge 3)$

$$= 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$= 1 - \left[\frac{e^{-4}.4^0}{0!} + \frac{e^{-4}.4^1}{1!} + \frac{e^{-4}.4^2}{2!}\right] = 1 - e^{-4} \times 13 = 1 - 0.238 = 0.762$$

19. It is known that 5% of books bound at a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective binding using.

i. the formula for the binomial distribution

ii. the Poisson approximation to the binomial distribution

iii. find the error in percentage. We would make by using Poisson approximation.

Scanned by CamScanner

Solution 19

Let X denote the random variable which represents the number of defective binding.

Then, total number of sample books (n) = 100

Probability of success in a single trial (p) = 5% = 0.05

Then $q = 1 - p = 1 - 0.05 = 0.95$

i. The probability of 2 books have defective binding is given by P(X = 2)

$$= {}^{100}C_2 (0.05)^2 . (0.95)^{100-2} = 4950 \times 0.0025 \times 0.0066 = 0.0812$$

ii. Also, $\lambda = np = 100 \times \frac{5}{100} = 5$

$$P(X = 2) = \frac{e^{-5} 5^2}{2!} = 0.0842$$

iii. $\text{Error} = \frac{0.0812 - 0.0842}{0.0812} \times 100\% = 3.7\%$

20. An aircraft has 116 seats the airline has found. From long experience, that, on average, 2.5% of the people who have bought the tickets for the flight do not arrive for the flight. The airliner sells 120 tickets for a particular flight.

[PU 2011 Spring]

a. Calculate the probability that more than 118 people arrive for the flight?

b. Calculate also the probability that there re empty seats on the flight?

Solution 20

See in the exam asked question

21. A well-known baseball player has a lifetime batting average of 0.7. What is the probability that 40 or fewer times at bat are required for him to achieve 36 hits?

Solution 21

Do yourself

22. Assume that each of your calls to a popular radio station has a probability of 0.02 of connecting that is of not obtaining a busy signal. Assume that your calls are independent.

i. What is the probability that your first call that connects is your tenth call?

ii. What is the probability that it requires more than five calls for you to connect?

iii. What is the mean no. of calls needed to connect?

Solution 22

Let X denote the number of calls that connects to a popular radio station.

Then $p = 0.02$, $q = 1 - 0.02 = 0.98$

(i) The probability that the first call that connects is 10th cell is given by

P(X = 1) where, $x + r = 10$, $r = 1$, $\therefore x = 9$.

Now by negative binomial distribution

Scanned by CamScanner

$$P(X=1) = \binom{x+r-1}{r-1} p^r q^{x-r} \binom{10-1}{1-1} p^1 q^{x-r} \binom{9}{0} (0.02).(0.98)^9 = 0.0167$$

(ii) The probability that it requires more than 5 calls for to connects is given by:

$$P(X>5) = 1 - P(\leq 5) = 1 - [P(X=5) + P(X=4) + P(X=3) + P(X=2) + P(X=1)]$$

$$= 1 - \left[\binom{9}{4}(0.02)^5(0.980)^5 + \binom{9}{3}(0.02)^4(0.98)^6 + \ldots + \binom{9}{0}(0.02)^1.(0.98)^9\right]$$

$$= 1 - 0.9 = 0.981$$

(iii) The mean of number of calls is given by $E(X) = \frac{r(1-p)}{p} = \frac{1 \times 0.98}{0.02} = 49$

23. From a lot containing 17 items, 7 of which are defective, 6 are chosen at random. Find the probability of

i. 4 or 5 defective
ii. at least 4 defective
iii. at most 2 defective
iv. 4 defectives and 2 non-defective items.

Solution 23

Let X denote the number defective items.

Total number of items N = 17

The number of defective M = 7

The number of sample n = 6

(i) The probability of getting 4 or 5 defective items is given by

$$P(X = 4 \text{ or } 5) = P(X=4) + P(x=5)$$

$$= \frac{C(7,4) \times C(17-7, 6-4)}{C(17,6)} + \frac{C(7,5) \times C(17-7, 6-5)}{C(17,6)}$$

$$= \frac{35 \times 45}{12376} = \frac{21 \times 10}{12376} = \frac{1785}{12376} = 0.1442.$$

(ii) The probability of getting at least 4 defective items is given by:

$$P(X \leq 4) = P(X=4) + P(X=5) + P(X=6)$$

$$= \frac{35 \times 45}{12376} + \frac{21 \times 10}{12376} = \frac{{}^7C_6 \times {}^{10}C_0}{{}^{17}C_6} = \frac{1792}{12376} = 0.1448.$$

(iii) The probability of getting at most 2 defective items is given by:

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

$$= \frac{{}^7C_4\,{}^{10}C_4}{{}^{17}C_4} + \frac{{}^7C_1\,{}^{10}C_5}{{}^{17}C_4} + \frac{{}^7C_0\,{}^{10}C_6}{{}^{17}C_4} = \frac{21 \times 210}{12376} + \frac{7 \times 252}{12376} + \frac{210}{12376} = 0.516.$$

(iv) The probability of getting 4 defective items is given by:

$$P(X=4)\ \frac{{}^7C_4\,{}^{10}C_2}{{}^{17}C_6} = \frac{35 \times 45}{12376} = 0.127.$$

24. In a ward, there are 121 houses in which 65 of them are new. Find the probability that 7 houses are old out of 25 houses selected randomly.

Solution 24

Let X denote the number of old house in a ward of 121 houses. Then

The total no. of houses N = 121

Scanned by CamScanner

The total no. of old houses $M = 121 - 65 = 56$

The number of sample houses $n = 25$

Then the probability that 7 houses are old is given by

$$P(X = 7) = \frac{C(56, 7) \times C(121 - 56, 25 - 7)}{C(121, 25)} = 0.022$$

25. An agricultural cooperative claims that 90% of the watermelons shipped out are ripe and ready to eat. Find the probabilities that among 18 watermelons shipped out

a. all 18 are ripe and ready to eat

b. at least 16 are ripe and ready to eat

c. at most 14 are ripe and ready to eat

Solution 25

Let X denote the number of water melons shipped out that are ripe and ready to eat.

Then, $p = 90\% = 0.9$

$q = 1 - p = 1 - 0.9 = 0.1$

$n = 18$

(i) The probability of all 18 are ripe and ready to eat is given by

$P = (X = 18) = {}^{18}C_{18} (0.9)^{18} . (0.1)^0 = 0.1501$

(ii) The probability of at least 16 are ripe and ready to eat is given by

$P = (X \geq 16) = P(X = 16) + P(X = 17) + P(X = 18)$

$= {}^{18}C_{16} (0.9)^{16} . (0.1)^2 + {}^{18}C_{17}(0.9)^{17} \times (0.1)^1 + 0.1501$

$= 0.284 + 0.3002 + 0.1501 = 0.7343$

(iii) The probability of at most 14 are ripe and ready to eat is given by

$P(X \leq 14) = 1 - P(X > 14) = 0.977$

26. Among the 300 employees of a company, 240 are union members, while the other are not. If 8 of the employees are chosen by lot to serve on the committee, which administers the pension fund, find the probability that five of them will be union member while the other are not, by using

a. the formula for the hypergeometric distribution.

b. the formula for the binomial distribution as an approximation.

Solution 26

(i) Let X denote the random variable with the number of union members among the employees of a company. Then,

Total no, of employees (N) = 300

Total no. of union members (M) = 240

Sample size (n) = 8

Then by the hyper-geometric distribution the probability x union members out of n is given as.

$$P(X = x) = h(x; n, M, N) = \frac{{}^{M}C_x \, {}^{N-M}C_{n-x}}{{}^{N}C_n}, X = 0, 1, 2, 3, \ldots, 8$$

Scanned by CamScanner

Now the probability of selecting 5 union members out of 8 is given by;

$$P(X = 5) = h(5' 8, 240, 300) = \frac{{}^{240}C_5\ {}^{60}C_3}{{}^{300}C_8} = 0.1470$$

(ii) Let X denote the random variable with the number of among the employees of a company. Then suppose selecting a union member is success then,

$$p = \frac{240}{300} = 0.8.$$

$q = 1 - p = 1 - 0.8 = 0.2$

$n = 8, r = 5$

Then probability of r success is given by

$P(X = r) = b(x; n, p) = {}^{n}C_r p^r q^{n-r}, r = 0, 1, 2, 3, \ldots$

Then the probability that 5 of 8 members will be union members is given by

$P(X = 5) = {}^{8}C_5 . (0.8)^5 . (0.2)^{8-5} = 0.1468$

27. If 6 and 18 new buildings in a city violate the building code, what is the probability that a building inspector, who randomly selects 4 of the new buildings, for inspection, will catch

a. none of the new buildings that violate the building code.

b. one of the new building that violet the building code.

c. two of the new buildings that violate the building code?

d. at least 3 of the new buildings that violate the building code?

Solution 27

Let x denote the number of building that violate the building code. Then N = 18, M = 6, n = 4

(a) The probability that none of the new building that violate the building code is given by $P(X = 0) = h(0;4, 6, 18) = \frac{{}^{6}C_0\ {}^{18-6}C_{4-0}}{{}^{18}C_4} = \frac{1 \times 495}{3060} = 0.162$

(b) The probability that one of the new building that violate the building code is given by $P(X = 1) = h(1;4, 6, 18) = \frac{{}^{6}C_1\ {}^{18-6}C_{4-1}}{{}^{18}C_4} = \frac{6 \times 220}{3060} = 0.4314$

(c) The probability that two of the new building that violate the building code is given by $P(X = 2) = h(2; 4, 6, 18) = \frac{{}^{6}C_2\ {}^{18-6}C_{4-2}}{{}^{18}C_4} = \frac{15 \times 66}{3060} = 0.323$

(d) The probability that at least 3 of the building that violate the building code is given by: $P(X \geq 1) = P(X = 3) + P(X = 4) = \frac{20 \times 12}{3060} + \frac{15 \times 1}{3060} = \frac{255}{3060} = 0.0833.$

28. If on an average 8 ships out of 10 arrive safely at a port, find the mean and variance of the number of ships arriving safely out of a total of 1600 ships.

Solution 28

Let X denote the random variable that the number of ships arriving safely at a port.

Since on an average 8 ships out of 10 arrive safely at a port then $p = \frac{8}{10} = 0.8$

Scanned by CamScanner

Also the total number of ships n = 1600

Then by binomial theorem we have:

Mean = np = 1600 × 0.8 = 1280

Variance = npq = 1600 × 0.8 × 0.2 = 256

29. For a binomial distribution, mean = 7 and variance = 11. Comment on the rest.

Solution 29

We have given mean = 7 and variance = 11.

Since the given distribution is a binomial distribution having parameter n and p then,

mean = np = 7 (i) Also variance = npq = 11 (ii)

Now dividing (ii) by (i) we get,

$$\frac{npq}{np} = \frac{11}{7} \rightarrow q = \frac{11}{7}$$

Also, $p = 1 - q = 1 - \frac{11}{7} = \frac{-11}{7} = \frac{-4}{7}$

Also, np = 7

$$n\left(\frac{-4}{7}\right) = 7 \Rightarrow n = \frac{-49}{4} \Rightarrow n = -12.25$$

Here we see that in the given distribution $\left(p = \frac{-4}{7} < 0\right)$ p is negative which is impossible. $q = \frac{11}{7} > 0$, q is greater than 1. Also n is negative. Which all are impossible. So the given statement is wrong.

30. If the probability of a defective bolt is 1/10 find mean, variance and P (X > 0) if number of items = 400

Solution 30

Here given p = 1/10, n = 400

Then mean (np) = $400 \times \frac{1}{10} = 40$

Variance (npq) = $400 \times \left(\frac{1}{10}\right)\left(1 - \frac{1}{10}\right) = 400 \times \frac{1}{10} \times \frac{9}{10} = 36$

Also, P(X>0) = 1 – P(X = ≤ 0)

$$= 1 - P(X = 0) = 1 - {}^{400}C_0 \left(\frac{1}{10}\right)^0 \cdot \left(\frac{9}{10}\right)^{400} = 1 - (\approx 0) \approx 1.$$

31. Fit a Binomial distribution on the following data

x	:	0	1	2	3	4
f	:	28	62	46	10	4

Solution 31

In the binomial distribution we know that,

Scanned by CamScanner

$n = 5, N = \Sigma f = 150.$

Now for the value of p, let us find the mean,

$$\text{mean} = \frac{\Sigma fx}{N} = \frac{0 \times 28 + 1 \times 62 + 2 \times 46 + 3 \times 10 + 4 \times 4}{150} = \frac{200}{150} = \frac{4}{3}$$

$$\Rightarrow np = \frac{4}{3} \Rightarrow p = \frac{4}{3 \times 5} = \frac{4}{15}$$

Also, $q = 1 - \frac{11}{15}$.

Now by the Binomial probability law the probability of x success is given by:

$$P(X = x) = P(n) = {}^nC_x p^x q^{n-x} = {}^5C_x \left(\frac{4}{15}\right)^x \cdot \left(\frac{11}{15}\right)^{5-x}$$

Thus the calculation of expected frequencies is given below.

Success (x)	P(x)	Expected frequency f(n) = N. P(x)
0	${}^5C_0\left(\frac{4}{15}\right)^0\left(\frac{11}{15}\right)^{5-0} = 0.2$	$0.2 \times 150 = 32$
1	${}^5C_1\left(\frac{4}{15}\right)^1\left(\frac{11}{15}\right)^{5-1} = 0.386$	$0.386 \times 150 = 58$
2	${}^5C_2\left(\frac{4}{15}\right)^2\left(\frac{11}{15}\right)^{5-2} = 0.280$	$0.280 \times 150 = 42$
3	${}^5C_3\left(\frac{4}{15}\right)^3\left(\frac{11}{15}\right)^{5-3} = 0.102$	$0.102 \times 150 = 15$
4	${}^5C_4\left(\frac{4}{15}\right)^4\left(\frac{11}{15}\right)^{5-4} = 0.0185$	$0.0185 \times 150 = 3$
	Total	150

32. Over a 10 minute period, counter records an average number of 1.3 Poisson particles per millisecond coming from a radioactive substance. Determine

a. m; b. the probability of one or more Poisson particles coming in per millisecond and; c. variance

Solution 32

Let X denote the random variable that the number of Poission particles per millisecond coming from a radioactive substance. m denote the average number of that particle per millisecond coming from a radioactive substance. Then,

(a) $m = 1.3$

(b) The probability of one or more Poission particle coming from a radioactive substance per millsecond is given by

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{e^{-1.3} \cdot (1.3)^0}{0!} = 1 - 0.273 = 0.727.$$

(c) The variance of the Poission distribution is given by $Var(X) = m = 1.3$

33. At a checkout counter customers arrive at a average number of 2.5 per minute. Find the probability that at most four will arrive in any given minute.

Scanned by CamScanner

Solution 33

Let X denote the random variable that the number of customers checkout from counter per minute also m denote the average number of customers checkout from a counter per minute. Then,

$m = 2.5$

Now the probability that at most four will arrive in any given minute is given by,

$$P(X \le 4) = \frac{e^{-2.5}.(2.5)^0}{0!} + \frac{e^{-2.5}.(2.5)^1}{1!} + \frac{e^{-2.5}.(2.5)^2}{2!} + \frac{e^{-2.5}.(2.5)^3}{3!} + \frac{e^{-2.5}.(2.5)^4}{4!}$$

$$= e^{-2.5}[1 + 2.5 + 3.125 + 2.604 + 1.628] = 0.891$$

34. Six coin are tossed 64 times. Using the Poisson distribution, find the approximate probability of getting 6 heads 1 time.

Solution 34

Let X denote the number of getting 6 heads tossing 6 coins.

Then probability of getting 6 heads (p) = $\left(\frac{1}{2}\right)^6 = \frac{1}{64}$

Now the number of tossing coins (n) = 64

Then the average number of getting 6 heads

$$m = np = 64 \times \left(\frac{1}{2}\right)^6 = 64 \times \frac{1}{4} = 1$$

Now the probability of getting 6 heads 1 time by Poission distribution is given by,

$$P(X = 1) = \frac{e^{-1}.1^1}{1!} = \frac{1}{e} = 0.3678.$$

35. Fit a Poisson distribution of the following data and calculate the expected frequencies:

x:	0	1	2	3	4	5	6	7	8
f:	56	156	132	92	37	22	4	0	1

Solution 35

Let X- denote the Poission distribution such that,

$n = 9$, and $N = \Sigma f = 56 + 156 + 132 + 92 + 37 + 22 + 4 + 0 + 1 = 500$

Also mean $m = \frac{\Sigma fx}{N}$

$$= \frac{0 \times 56 + 1 \times 56 + 2 \times 132 + 3 \times 92 + 4 \times 37 + 5 \times 22 + 6 \times 4 + 7 \times 0 + 8 \times 1}{500}$$

$$= \frac{986}{500} = 1.972.$$

Also, the probability of x success by Poission distribution is given by:

$$P(X) = \frac{e^{-m}.(m)^x}{x!} = \frac{e^{-1.972}.(1.972)^x}{x!}$$

The expected frequencies are given as follows:-

Scanned by CamScanner

Success (x)	P(x)	Expected frequency f(n) = N. P(x)
0	$\frac{e^{-1.972}(1.972)^0}{1!} = 0.139$	500 × 0.139 = 70
1	$\frac{e^{-1.972}(1.972)^1}{1!} = 0.274$	137
2	0.2705	135
3	0.17793	89
4	0.10964	44
5	0.3459	17
6	0.01137	6
7	0.00320	2
8	0.00078	0
		500

36. Among the 12 solar collectors on display at a trade show, 9 are flat plate collectors and other is concerned collectors. If a person visiting the show randomly selects four of the solar collectors to check out, what is the probability that 3 of them will be flat-plate collectors? [P.U. 2007 Fall]

Solution 36

Let X- denote the numbers of flat plate collectors.

The total number of solar collectors on display N = 12

The number of flat-plate collectors m = 9

The number of sample size n = 4

Then by the hyper geometric x success out n trials is given by

$$P(X = x) = \frac{{}^{M}C_x \, {}^{N-M}C_{n-x}}{{}^{N}C_n}$$

Then the probability that 3 of them will be flat-plate collectors is given by,

$$P(X = 3) = \frac{{}^{9}C_3 \, {}^{3}C_1}{{}^{12}C_4} = \frac{84 \times 3}{495} = 0.5091$$

37. At a checkout counter customers arrive at an average number of 1.5 per minute. Find the probability that: [P.U. 2007 Fall/ 2009 Fall]

a. At most four will arrive in any given minute,

b. At least 3 will arrive during an interval 2 minutes,

c. At most 5 will arrive during an interval of 6 minutes.

Solution 37

Let X- denote the numbers of customers checkout from the counter per minute and m denote the average number of customers checkout from the counter per minute.

Then m = 1.5

(a) Then the probability that at most four will arrive in any given minute is given by the Poisson distribution:

i.e. $P(X \le 4) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)$

Scanned by CamScanner

$$= \frac{e^{-1.5}.(1.5)^4}{4!} + \frac{e^{-1.5}.(1.5)^3}{3!} + \frac{e^{-1.5}.(1.5)^2}{2!} + \frac{e^{-1.5}.(1.5)^1}{1!} + \frac{e^{-1.5}.(1.5)^0}{0!}$$

$$= e^{-1.5}[0.211 + 0.5625 + 1.125 + 1.5 + 1] = 0.9814$$

(b) Let X denote the number of customers who will arrive in 2 minutes. Now the average number of customers arrive in 2 minute $m = 1.5 \times 2$

Then the probability that at least 3 will arrive during the interval of 2 minutes by Poisson distribution is given by

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

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

$$= 1 - \left[\frac{e^{-3}.3^2}{2!} + \frac{e^{-3}.3^1}{1!} + \frac{e^{-3}.3^0}{0!}\right]$$

$$= 1 - e^{-2}[4.5 + 3 + 1] = 1 - 0.4232 = 0.5768.$$

(c) Let X denote the number of customers that arrive in 6 minutes. Then the average number of customers that will arrive in the interval of 6 minutes, $m = 1.5 \times 6 = 9$

Then the probability that at most 5 will arrive in 6 minutes by Poisson distribution is given by,

$$P(x \leq 5) = P(X = 5) + P(X = 4) + P(x = 3) + P(X = 2) + P(X = 1) + P(X = 0)$$

$$= \frac{e^{-9}5^5}{5!} + \frac{e^{-9}.9^4}{4!} + \frac{e^{-9}.9^3}{3!} + \frac{e^{-9}.9^2}{2!} + \frac{e^{-9}.9^0}{1!}$$

$$= 1 - e^{-9}[492.075 + 273.375 + 121.5 + 40.5 + 9 + 1] = 0.1157$$

38. Suppose vehicles arrive at a signalized road intersection at an average of 360 per hour and the cycle of the traffic lights is set at 40 seconds. In what percentage of cycles will the number of vehicles arriving be?

a. exactly '5' b. less than '5'?

c. if after the lights change to green, there is time to clear only 5 vehicles before the signal changes to red again, what is the probability that waiting vehicles are not cleared in one cycle?

Solution 38

Let X denote the number of vehicles arriving in traffic signal. m denote the average number of vehicles in the interval of 40 seconds. Then $m = 40 \times \frac{360}{60 \times 60} = 4$

(a) The probability that exactly 5 vehicles will be arriving in 40 seconds is given by

$$P(X = 5) = \frac{e^{-4}.4^5}{5!} = 0.1563$$

In percentage $0.1563 \times 100\% = 15\% = 15.6\%$

(b) Similarly $P(X<5) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)$

$$= \frac{e^{-4}.5^4}{4!} + \frac{e^{-4}.4^3}{3!} + \frac{e^{-4}.4^2}{2!} + \frac{e^{-4}.4^1}{1!} + \frac{e^{-4}.4^0}{0!}$$

$$= 1 - e^{-4}[10.67 + 10.67 + 8 + 4 + 1] = 0.629 = 62.3\%$$

Scanned by CamScanner

(c) The probability that the waiting vehicles are not cleared in one cycle is given by,

$P(X > 5) \quad = 1 - P(X = \leq 5)$

$= 1 - [P(X < 5) + P(X = 5)] = 1 - [0.629 + 0.1563] = 0.2147.$

39. **The number of visitors to a web server per minute follows a Poisson distribution. If the average number of visitors per minute is four. What is the probability that,**

a. There are two or fewer visitors in one minute?

b. There are exactly two visitors in 30 seconds?

Solution 39

(a) Let X denote the number of visitors to a web server per minute and m denote the average number of visitors per minute. Then m = 4.

Then the probability that there are 2 or fewer visitors in one minute is given by.

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

$$= \frac{e^{-4} . 4^2}{2!} + \frac{e^{-4} . 4^1}{1!} + \frac{e^{-4} . 4^0}{0!} = e^{-4}(8 + 4 + 1] = 0.238$$

(b) Let X denote the number of visitors to a web server per 30 seconds and m denote the average number of visitors per 30 minutes. Then $m = \frac{4}{2} = 2$.

Now the probability that there are exactly two visitors in 30 seconds is given by.

$$P(X = 2) = \frac{e^{-2} . 2^2}{2!} = 0.2707.$$

40. Assume that the probability of an individual coal miner being killed in a mine accident during a year in 1/2400. Use appropriate statistical distribution to calculate the probability that in a mine employing 200 miners there will be,

a. At least one fatal accident. b. Exactly two fatal accidents.

c. At least two fatal accidents in a year.

Solution 40

Let X denote the number of fatal accidents in mine accident in a year.

Then we have, $n = 200$, $p = \frac{1}{2400}$ So, $m = np = 200 \times \frac{1}{2400} = 0.083$.

(a) The probability that at least one fatal accidents is given by,

$$P(X \geq 1) \quad = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{e^{-0.083} \times (0.083)^0}{0!} = 1 - 0.9204 = 0.0769$$

(b) The probability that exactly two fatal accidents is given by,

$$P(X = 2) \quad = \frac{e^{-0.083} \times (0.083)^2}{2!} = 0.00317$$

(c) The probability of at least two fatal accidents is given by,

$P(X \geq 2) \quad = 1 - P(X < 2) = 1 - [P(X = 1) + P(X = 0)]$

$$= 1 - \left(\frac{e^{-0.083} \times (0.083)^1}{1!} + \frac{e^{-0.083}(0.083)^0}{0!}\right) = 1 - [0.0764 + 0.9204] = 0.0032$$

Scanned by CamScanner

Exam asked questions

"Long questions"

1. **A bag contains 5 white and 3 red balls. Three balls are drawn at random**

(i) Find the expected numbers of white ball drawn is three.

(ii) If each white balls drawn carries a reward of Rs. 4 and each red balls Rs. 6 find the expected reward of the draw. **[PU 2002/ TU 2065]**

Solution 1

Here given n(W) = 5, n(R) = 3 total balls = 8. We have given three balls are drawn at random, So three balls can be drawn out of 8 in C (8, 3) = 56 ways. If X denotes the random variable that the number of white balls drawn then X can be 0,1,2,3.

i. Now, the probability of no. white balls is given by

$$P(X=0) = \frac{C(5,0)\times C(3,3)}{(18,3)} = \frac{1}{56}$$

$$\text{Similarly, } P(X=1)) = \frac{C(15,2)\times(3,2)}{(18,3)} = \frac{5\times 3}{56} = \frac{15}{56}$$

$$\text{Similarly, } P(X=2) = \frac{C(15,2)\times C(3,1)}{(18,3)} = \frac{10\times 3}{56} = \frac{15}{28}$$

$$P(x=3) = \frac{C(5,3)\times C(3,0)}{(18,3)} = \frac{10}{56} = \frac{5}{28}$$

Now the probability distribution function of X is.

X : x	0	1	2	3
P(x)	1/56	15/56	15/28	5/28

Now, the expected number of white balls drawn

$$E(X) = \Sigma x_i\, p(x_i) = 0\times\frac{1}{56} + 1\times\frac{15}{56} + 2\times\frac{15}{28} + 3\times\frac{5}{28} = 1.875$$

ii. The reward function R(x) for drawing x white balls and (3 – x) red balls is given by R(x) = Rs.[4x + 6(3 – x)] = Rs. [18 – 2x]

Now the expected value of reward is

$$E(R(x)) = E[10 - 2X] = 18 - 2E(X) = 18 - 2\times 1.875 = 14.25$$

2. **Three defective items are mixed with seven good ones. If three items are drawn at random. Find the probability distribution of number of defective items. And also find the mean and variance of the distribution. [PU. 2016 Spring / TU 2065]**

Solution 2

Let X denote the number of defective items. The X can be assumed 0,1,2,3. Then the probability distribution function can be obtained as, probability of getting 0 defective item is given by

$$P(X=0) = \frac{C(3,0)\times C(7,3)}{C(10,3)} = \frac{1\times 35}{120} = \frac{7}{24}$$

Probability of getting 1 defective item is given by

$$P(X=1) = \frac{C(3,1)\times C(7,2)}{C(10,3)} = \frac{3\times 21}{120} = \frac{21}{40}$$

Scanned by CamScanner

Similarly, $P(X=2) = \frac{C(3,2) \times C(7,3)}{C(10,3)} = \frac{3 \times 7}{120} = \frac{1}{120}$

and $P(X=3) = \frac{C(3,3) \times C(7,0)}{C(10,3)} = \frac{1}{120}$

Hence the probability distribution function of X is

X : x	0	1	2	3
P(x)	7/24	21/40	7/40	1/120

2nd part

The mean of the distribution is given by.

$E(X) = \Sigma x_i\, p(x_i) = 0 \times \frac{7}{24} + 1 \times \frac{21}{40} + 2 \times \frac{7}{40} + 3 \times \frac{1}{120} = 0.9$

Also the variance of the distribution is given by.

$V(X) = E(X^2) - [E(X)]^2\ 1.3 - (0.9)^2 = 0.49.$

Where, $E(X^2) = \Sigma x_i^2\, p(x_i) = 0^2 \times \frac{7}{24} + 1^2 \times \frac{21}{40} + 2^2 \times \frac{7}{40} + 3^2 \times \frac{1}{120} = 1.3$

3. **Six dice are thrown 729 times. How many times do you expect at least three dice to shown a five or six?** **[PU 2016 Spring]**

Solution 3

See in Exercise 10 Q.no.10

4. **A food packing apparatus under fills 10% of the containers. Find the probability that for any five containers the number of under filled will be (i) exactly 3 (ii) exactly 2 (iii) Zero (iv) at least one.** **[PU 2001]**

Solution 4

We know that the sampling process represents a binomial experiment with $n = 5$, $p = 10\% = 0.1$, $q = 1 - p = 0.9$.

Let X denote the number of containers under filled.

i. The probability that exactly 3 containers under filled is

$P(X=3) = C(5,3)(0.1)^3 \times (0.9)^{5-3} = 10 \times (0.1)^3 \times (0.9)^2 = 0.0081.$

ii. The probability that exactly 2 containers under filled is

$P(X=3) = C(5,2)(0.1)^2 \times (0.9)^{5-2} = 10 \times (0.1)^2 \times (0.9)^3 = 0.0729.$

iii Similarly $P(X=0) = C(5,0)(0.1)^0 \times (0.9)^5 = 0.5905$

iv $P(X \le 1) = 1 - P(X=0) = 1 - 0.5905 = 0.4095$

5. **A taxi cab company had 12 Ambassador and 8 fiats. If 5 of those taxi- cabs are in shops for repair and Ambassador is as likely to be in for repairs as a fiat, what is the probability that;**

(i) 3 of them are Ambassador and 2 are fiats.

(ii) 3 of them are fiats and 3 are ambassadors

(iii) at least 3 of them are Ambassadors

(iv) all of them are of the same make. **[PU 2002]**

Scanned by CamScanner

Solution 5

Let X denote the random variable that the number of Ambassadors cars which are in the shop for repairs.

i. Probability that 3 of them are Ambassadors and 2 of them are fiats is given by the hypergeometric distribution.

$$P(X=3) = \frac{C(12,3)\times C(8,2)}{C(20,5)} = \frac{220\times 28}{15504} = 0.397$$

ii Probability that 3 of them are fiats and 2 of them are Ambassadors is

$$P(X=2) = \frac{C(12,2)\times C(8,3)}{C(20,5)} = 0.238$$

iii Probability that at least 3 of them are ambassadors is

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

$$= \frac{C(12,3)\times C(8,2)}{C(20,5)} + \frac{C(12,4)\times C(8,1)}{C(20,5)} + \frac{C(12,5)\times C(8,0)}{C(20,5)}$$

$$= \frac{220\times 28}{15504} + \frac{495\times 8}{15504} + \frac{729\times 1}{15504} = 0.704$$

iv Let Y be the random variable that the umber of fiat cars which are in the shop for repairs. Since X and Y are mutually exclusive.

$$P(X=5 \text{ or } Y=5) = P(X=5)+P(Y=5)$$

$$= \frac{C(12,5)}{C(20,5)} + \frac{C(8,5)}{C(20,5)} = 0.051 + 0.036 = 0.0055$$

6. It is known that 5% of the screws manufactured by an automatic machine of defective. If a sample of 20 screws is selected at random. Find the probability that the sample contains

(i) Exactly two defective screws

(ii) At least two defectives screws. [PU 2016 Fall]

Solution 6

Let X denote the number of defective screws manufactured by an automatic machine. Also getting an defective screws is taken as success then,

$P = 5\% = 0.05,\ q = 1 - 0.05 = 0.95,\ n = 20$

i Probability that the sample contains exactly two defective screws;

$$P(X=2) = C(20,2)\,(0.05)^2 \times (0.9)^{20-2} = 0.315$$

ii Probability that the same contains at least two defective screws is given by

$$P(X\geq 2) = 1 - P(X<2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - C(20,0)(0.05)^0 \times (0.95)^{10} - C(20,1)(0.05)^1 \times (0.95)^{10-1} = 0.97$$

7. A box contains 10 red, 10 black and 3 green balls. Three balls are drawn at random for the box. using hyper geometric distribution. Find the probability that

(i) There are 2 red and one black balls.

(ii) There are all green balls.

(iii) There are one red and two black balls. [PU 2013]

Solution 7

Scanned by CamScanner

Let X denote the number of green balls are drawn. Then by Hyper-Geometric distribution we have,

$$P(X = x) = \frac{C(3, x) \times C(20, 3 - x)}{C(23, 3)} \Rightarrow x = 0,1,2,3$$

i Probability of getting all three green balls is given by

$$P(X = 3) = \frac{C(3, 3) \times C(20, 0)}{C(123, 3)} = \frac{1}{1771} = 0.00057$$

ii Probability of getting two red and one black ball is

$$P(2 \text{ red ball and } 1 \text{ black ball}) = \frac{C(10, 2) \times C(10, 1) \times C(3,0)}{C(23, 3)} = 0.254$$

iii Probability of getting one black and two red ball is

$$P(1 \text{ black and } 2 \text{ red}) = \frac{C(10, 1) \times C(10, 2) \times C(3, 0)}{C(23,3)} = 0.254$$

8. **The number of road accidents in a year attributed to taxi drivers follows Poisson distribution with mean 3 out of 1000 taxi drivers, find approximately the number of taxi drivers with**

(i) No accidents in a year

(ii) More than three accidents in a year [PU 2009 Spring]

Solution 8

Let X denote the random variable that the number of road accidents in a year, X follow the Poisson distribution with mean m = 3. Also, N = 1000

Now the probability of X road accidents in a year is given by

$$P(X = x) = \frac{e^{-3} x^x}{x!}, x = 0, 1, 2, \ldots$$

i Probability that there is no accident in a year is

$$P(X = 0) = \frac{e^{-3} \times 30}{0!} = e^{3} = \frac{1}{e^{3}}$$

The approximate number of taxi drivers with no accidents in a year is given by

$$N \times P(x = 0) = 1000 \times \frac{1}{e^{3}} = 49.79 = 50$$

ii Similarly the approximate number of taxi drivers with more than 3 accidents in a year is given by

$$\begin{aligned} N \times P(X \geq 3) &= 1000[1 - P(X < 3)] \\ &= 1000[1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1000\left[1 - \frac{e^{-3}.3^0}{0!} - \frac{e^{-3}.3^1}{1!} - \frac{e^{-3}.3^2}{!}\right] \\ &= 1000[1 - 0.05 - 0.15 - 0.224] = 575 \end{aligned}$$

9. **At a checkout, counter customers arrive at an average of 1.5 per minute. Find the probability that**

(i) At most 4 will arrive in any given minute.

(ii) At most 3 will arrive during an interval of 2 minutes.

(iii) At most 15 will arrive in an interval of 6 minutes. [PU. 2007 Fall/ 2009 Fall]

Solution 9

See in Exercise 10 Q.No.37

Scanned by CamScanner

10 **In a certain factory turning out optical lenses, there is a small chance 1/500 for any lens to be defective. The lenses are supplied in packets of 10 each. What is the probability that the packet will contain** **[PU 2005 Spring]**

(i) No defective lens (ii) Exactly two defective lenses

(iii) At least one defective lenses (iv) At most two defective lenses.

Solution 10

Let X denotes the number of defective lens in a packet of 10. then we have,

$n = 10, p = 1/500, m = np = 10 \times 1/500 = 0.02$

Then by Poisson distribution, the probability of x defective lens in a packet of 10 is given by $P(X = x) = \frac{e^{-m} \times m^x}{x!}, x = 0,1,2...$

i The probability that a pocket will certain no defective lens is

$P(X = 0) = \frac{e^{-0.02} \times (0.02)^0}{0!} = 1/e^{0.02} = 0.98$

ii $P(X = 2) = \frac{e^{-0.02} \times (0.02)^2}{2!} = 0.0002$

iii $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.98 = 0.2$

iv $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.98 + \frac{e^{-0.02} \times (0.02)^1}{1!} + 0.0002 = 0.99$

11. **A marks man firing bullets at a target and probability of hitting the target at any trial is 0.7. Find the probability that his seventh shot is his forth hit.** **[PU. 2015 Fall]**

Solution 11

We have given the probability of hitting the target in each trial

$(p) = 0.7 \quad (q) = 1 - p = 0.3$

Total no of firing $x + r = 7$

no. of hitting target $r = 4$

Let X denote the random variable that the number of not hitting target. Then the probability that his 7th shot is his 4th hit is given by

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

$P(X = 4) = C(6,3)(0.7)^4 \times (0.3)^3 = 20 \times (0.7)^4 \times (0.3)^3 = 0.13$

12 **Customers arrive at complaint department of a store at the rate of 5 per hour. If arrivals follow Poisson distribution calculate the probability that**

(i) No customers will arrive in an hour.

(ii) At least two customers will arrive an hour.

(iii) At most two customers will arrive in 30 minutes.

[PU 2007 Fall/2017 Spring/2013 Spring]

Solution 12

Let X denote the random variable which associates the number of customers arriving at complaint department.

Also average rate per hour (m) = 5

Then the probability that customer will arrive in a hour is given by

Scanned by CamScanner

$P(X = x) = e^{-5} \cdot 5^x/x!,\ x = 0,1,2,3...$

i The probability that no customers will arrive in a hour is

$P(X = 0) = e^{-5} \times 5^0/0! = 1/e^5 = 0.007$

ii. $P(X \geq 2) = 1 - P(X < 2) \quad = 1 - P(X = 0) - P(X = 1) = 1 - \frac{e^{-5} \times 5^0}{0!} - \frac{e^{-5} \times 5^1}{1!} = 0.96$

iii Since the given rate is for an hour. So the average rate of customer for 30 minute $= 5/2 = 2.5$.

The probability that at most two customer will arrive in 30 minute is

$$P(X \leq 2) \quad = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \frac{e^{-2.5} \times (2.5)^0}{0!} + \frac{e^{-2.5} \times (2.5)^1}{1!} + \frac{e^{-2.5} \times (2.5)^2}{2!} = 0.54$$

13. **Warranty records show that the probability that a new car needs a warranty repair in the first 90 days is 0.20. If a sample of new six car is selected, What is the probability of that in the first year:** **[PU 2012 Fall]**

(i) None needs a warrant repair.

(ii) At least one needs a warrant repairs.

Solution 13

Solve as above problems.

14. **Customers arrive at complaint department of a store at the rate of 5 per hour. If arrivals follow Poisson distribution calculate the probability that**

(i) at most 4 customers will arrive in an hour.

(ii) No customers will arrived in 300 minute period. [PU 2012 Fall]

Solution 14

The process is similar to Q.12

15. **If a mean number if serious accidents per month in a large factory is one. Find the probability that in the current year there will be at least accidents. The annual serious accidents in the factory follow a Poisson distribution.** **[PU 2011 Fall]**

Solution 15

Let X denote the random variable, that represents the number of accidents in a month.

Also given m = 1.Then by the Poission distribution,

For the probability that there will be at least 10 accidents in the current year is given by

$m = 4 \times 1 = m$

$$P(X \geq 10) \quad = 1 - P(X < 10) = 1 - \sum_{k=0}^{9} \frac{e^{-4} \times 4^k}{k!} = 1 - \frac{1}{e^4}\left[\frac{4^0}{0!} + \frac{4^1}{1!} + \ldots + \frac{4^9}{9!}\right]$$

$$= 1 - \frac{1}{e^4} \times 54.15 = 0.0081$$

16. **A certain screw machine produces on the average 4% of the defected screws. The screws are packed in boxes of 500. Find the probability that** **[PU 2013]**

Scanned by CamScanner

(i) A box contains 15 defective screws.

(ii) A box contains no defective screws.

(iii) A box contains at least 5 defective screws.

Solution 16

Let X denote the random variable that represents the number of defective screws.

Probability of producing a defective screws is

$P = 4\% = 0.04, n = 500$

$m = np = 500 \times 0.04 = 20$

i. The probability that a box containing 500 screws 15 are defective

$$P(X = 15) = \frac{e^{-20} \times (20)^{15}}{15!} = 0.052$$

ii The probability of getting no defective screws is

$$P(X = 0) = \frac{e^{-20} \times (20)^{0}}{0!} = 2 \times 10^{-9}$$

iii The probability of getting at least 5 defective screws is

$$P(X \geq 5) = 1 - P(X < 5)$$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \frac{1}{e^{20}}\left[\frac{(20)^0}{0!} + \frac{(20)^1}{1!} + \frac{(20)^2}{2!} + \frac{(20)^3}{3!} + \frac{(20)^4}{4!}\right] = 1 - \frac{1}{e^{20}} \times 8221 = 0.99$$

17. It is given that 2% of the electric bulbs manufactured by a company are defective. Find the probability that a sampling of 200 bulbs will contain [PU 2010 Spring]

(i) 2 defective bulbs (ii) At most 3 defective bulbs

(iii) At least 3 defective bulbs.

Solution 17

Process is similar to Q.No. 16

18. If 5% of the electric bolts manufactured by a company are defective. Find the probability that in a sample of 500 bolts [PU 2011]

(i) Less than 5 bolts are defective. (ii) At least two bolts are defective.

Solution 18

Process is similar to Q.No. 16

19. The probabilities of a Poisson variate taking the values 3 and 4 are equal. Calculate the probabilities of the variate taking the value of 0 and 1. [P. U 2001]

Solution 19

Since X is a random variable having the Poisson distribution then

$$P(X = x) = \frac{e^{-m} x^{m}}{m!}$$

Also the variate taking values 3 and 4 are equal.

Then, $P(X=3) = P(X=4)$

or, $$\frac{e^{-m} \times m^3}{3!} = \frac{e^{-m} \times m^4}{4!}$$

Scanned by CamScanner

$\Rightarrow \quad 24\,m^3 - 6\,m^4 = 0 \Rightarrow 6\,m^3\,(4 - m) = 0 \Rightarrow m = 4 \quad [\because m \neq 0]$

Now the variate value taking 0 and 1 is given by:

$$P(X=0) = \frac{e^{-4} \times 4^0}{0!} = \frac{1}{e^4} = 0.018$$

$$P(X=1) = \frac{e^{-4} \times 4^1}{1!} = \frac{4}{e^4} = 0.073$$

20. **A boy is throwing stone at a target. What is the probability that his 10th throw is his 5th hit. If the probability of hitting the target at any trial is 0.5. [PU 2009]**

Solution 20

Process is similar to Q.11

21. **An item is produced in large number. The machine is known to produce 5% defective. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defective? Also find the mean and variance. [PU. 2001]**

Solution 21

Let X denote the number produced item by a machine. The probability of producing defective item = 5% = 0.05.

Here we have to get 2 defective item in at least 4 items. So it follows a negative binomial distribution.

Here, $x + r = 4$ and $r = 2$

Now the probability that at least 4 items are to be examined in order to get 2 defective items is

$$P(X + r \geq 4) = P(X + 2 \geq 4)$$
$$= P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - [{}^2C_0\,(0.05)^2 + {}^2C_0\,(0.05)^2(-0.95)] = 0.993$$

Also mean of the distribution

$$E(X) = \frac{rq}{p} = \frac{2 \times 0.95}{0.05} = 38 \qquad V(X) = \frac{rq}{p^2} = \frac{2 \times 0.95}{(0.05)^2} = 760$$

22. **Among 12 solar collectors on display at a trade show, 9 are flat - plate collectors and the others are concentrating collectors. If a person visiting the show randomly selects 4 of the solar collectors to check out, what is the probability that 3 of them will flat plate collectors? [PU 2007 Fall]**

Solution 22

See in Ex 10 Q.17

23. **A typist made 2.8 mistakes on an average per page. Find the probability that in the page typed by him**

(a) there is no mistakes (b) two or less mistakes

(c) more than 10 mistakes [PU 2010 Fall]

Solution 23

See in Ex 10 Q.36

24. **In a particular college in Nepal it ahs been found that 20% of the students withdraw without completing the Engineering course, assume that 15 students have register in a new semester. What is the probability that, [PU 2010 Spring]**

(i) none will withdraw the course

Scanned by CamScanner

(ii) at most two will withdraw the course

(iii) Find the mean of this distribution.

Solution 24

Let X denote the random variable that the number of students without completing the Engineering course. Also given that

Probability of withdraw without computing (p) = 20% = 0.2

Also no. of students (n) = 15 $q = 1 - p = 0.8$

i. Probability that none will withdraw the course is

$P(X = 0) = C(15,0)(0.2)^0(0.8)^{15} = 0.035$

ii. Probability that at least 2 will withdraw the course is

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - [P(X = 0) + P(X = 1)] \\ &= 1 - [C(15, 0)(0.2)^0(0.8)^{15} + C(15,1)(0.2)^1(0.8)^{14}] \\ &= 1 - [0.035 + 0.084] = 0.881 \end{aligned}$$

iii. The mean of the binomial distribution

$E(X) = np = 15 \times 0.2 = 3$

25. From a lot containing 50 items, 10 of which are defective 4 are chosen at random. Find the probability that

(i) at least one will be defective

(ii) one or two will be defective. [PU 2006 Spring]

Solution 25

Let X denote the random variable that the number of defective items.

The probability of selecting defective item $p = \frac{10}{50} = 0.2$

no. of items (n) = 4

i. Probability that at least one will be defective

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X = 0) = 1 - C(4, 0)(0.2)^0 \times (0.8)^4 = 0.5904 \end{aligned}$$

ii Probability of one or two will be defective is

$$\begin{aligned} P(X = 1 \text{ or } X = 2) &= P(X = 1) + P(X = 2) \\ &= C(4, 1) \times (0.2)^1 \times (0.8)^3 + C(4, 2) \times (0.2)^2 \times (0.8)^2 \\ &= 0.4096 + 0.1536 = 0.5632 \end{aligned}$$

26 An office switch board receives an average of 4 emergency calls in a 10 minutes interval. What is the probability that

(i) there is no call in a 10 minute interval.

(ii) There is at most 2 calls in 5 minute interval. [PU 2006 Spring]

Solution 26

Let X denote the random variable with the number of telephone calls received in a 10 min interval, then X follow the Poisson distribution with m =4

Scanned by CamScanner

i. Probability that there are no calls in 10 minutes is

$$P(X=0) = \frac{e^{-4} \times 4^0}{0!} = 0.018$$

ii. For five minute the average of calls $m_1 = 4/2 = 2$

Then the probability that there is at most 2 calls in 5 minutes is

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

$$= \frac{e^{-2} \times 2^0}{0!} + \frac{e^{-2} \times 2^1}{1!} + \frac{e^{-2} \times 2^2}{2!} = 0.68$$

27 What do you mean by binomial distribution? An aircraft has seats that airlines has found. From long experience that on average, 2.5% of the people who have bought the tickets for the flight do not arrive for the flight. The airlines sells tickets for a particular flight.

(i) Calculate the probability that more than 118 people arrive for the flight.

(ii) Calculate also the probability that there are empty seats on the flight.

[PU 2011 Spring]

Solution 27

1st Part

A discrete random variable X is said to have binomial probability distribution with parameter n and p if its probability mass function (p.m.f) is given by

$$b(x; n, p) = P(X=x) = \begin{cases} C(n, x)\, p^x q^{n-x} & n = 0,1,2,3,\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

where

p is probability of success in a single trial.

q is probability of failure in a single trial.

n is probability of independent trial.

2nd Part

Let X denote the number of people do not arrive for the flight.

The probability that the people do not arriving for the flight

$p = 2.5\% = 0.025$

The number of ticket sold (n) = 120

$\therefore$ λ = Average number of people do not arrive for the flight = $np = 120 \times 0.025 = 3$

Let X denote the number of people do not arrive for the flight. Then using Poisson distribution, the probability of exactly 'r' people do not arrive for the flight is given by

$$P(X=r) = \frac{e^{-m} \times m^r}{r!} = \frac{e^{-3} \times (3)^r}{r!}\ ;\ r = 0, 1, 2, 3, \ldots$$

i. The probability that more than 116 people arrive for the flight is given by

P (more than 116 people arrive for the flight)

= P (less than 4 people do not arrive for a flight)

$= P(X < 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

Scanned by CamScanner

$$= \frac{e^{-3} \times 3^0}{0!} + \frac{e^{-3} \times 3^1}{1!} + \frac{e^{-3} \times 3^2}{2!} + \frac{e^{-3} \times 3^3}{3!} = e^{-3}\left[1 + 3 + \frac{9}{2} + \frac{27}{6}\right] = e^{-3} \times 13 = 0.6472$$

ii. The probability that there are empty seats on the flight is given by,

P (empty seats)

= P (less than 116 people arrive for a flight)

= P (5 or more people do not arrive for a flight)

= P ($X \geq 5$)

$$= 1 - P(X < 5) = 1 - [P(X < 4) + P(X = 4)] = 1 - \left[0.6472 + \frac{e^{-3} \times 3^4}{4!}\right] = 0.1847$$

Exam asked short question:

28. **Write short notes on Negative Exponential Distribution.** [P.U. 2010 Spring/ 2011 Fall]

Solution 28

See theoretical part of this chapter

29. **Write short notes on Assumptions/ conditions of Binomial Distribution** [P.U. 2010 Spring/ 2011 Spring]

Solution 29

See theoretical part of this chapter

30. **Write short notes on Hyper Geometric distribution.** [P.U. 2012 Fall]

Solution 30

See in theoretical part of this chapter

31. **Write short notes on negative Binomial distribution.** [P.U. 2012 Fall]

Solution 31

See in theoretical part of this chapter

Scanned by CamScanner

11 Chapter

Theoretical Continuous Distribution

Some of the important and most useful theoretical continuous probability distribution is given below:

1. Uniform (Rectangular) Distribution.
2. Gama Distribution.
3. Beta Distribution.
4. Normal Distribution.
5. Log-Normal Distribution.
6. Exponential Distribution.

Uniform (or Rectangular) Distribution:-

A continuous random variable X is said to have a Uniform (or Rectangular) distribution over an interval [a, b] if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a<x<b \\ 0 & \text{otherwise} \end{cases}$$

Remarks: -

i Uniform distribution has parameter a and b.

ii If F(x) is probability distribution function of continuous random variable X. Then $F'(x) = f(x)$ is the pd.f of x.

iii The cumulative distribution function (cdf) is given by : -

$$F(x) = \begin{cases} 0 & \text{for } x \le a \\ \frac{x-a}{b-a} & \text{for } a<x<b \\ 1 & \text{for } b \le x \end{cases}$$

iv The mean of the distribution $\mu = E(X) = \frac{b+a}{2}$

v The variance of the distribution $V(X) = \sigma^2 = \frac{(b-a)^2}{12}$

Gamma Distribution: -

Scanned by CamScanner

A continuous random variable X is said to have gamma distribution with parameter α and β if its probability density function (pd.f) is given by

$$f(x) = \begin{cases} \frac{1}{\beta^{\alpha}.\Gamma(\alpha)} .x^{\alpha-1}e^{(-x/\beta)} & \text{for } x>0, \ \alpha>0, \ \beta>0 \\ 0 & \text{otherwise} \end{cases}$$

Remarks: -

1. The gamma function

   i $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}\,dx \ \alpha>0$ ii $\frac{\Gamma(\alpha)}{a^{\alpha}} = \int_0^{\infty} e^{-ax}x^{\alpha-1}\,dx$

   iii $\Gamma(\alpha+1) = \alpha\Gamma(x) = \alpha!$ iv $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

2. Gamma distribution useful in the study of the length of life industrial of industrial equipments, electrical supply in certain area, distribution of petrol and other fields.
3. This distribution is also applied to waiting time models in life testing, waiting time etc.
4. Gamma distribution tends to normal distribution as $\lambda = \frac{1}{\beta} \to \infty$
5. If $\alpha = 1$ and $\beta = \frac{1}{\lambda}$ gamma distribution reduces to exponential distribution.
6. The mean of the gamma distribution $\mu = E(X) = \alpha-\beta$
7. The variance of gamma distribution $\sigma^2 = V(X) = \alpha\beta^2$

"Beta Distribution"

A continuous random variable X is said to have a Beta distribution with parameter α and β if its probability density function (pdf) is given by

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \alpha>0, \beta>0 \\ 0 & \text{otherwise} \end{cases}$$

Remarks: -

1. In Beta function the r.v. X can take values between 0 and 1 only.
2. **The Beta function**

   i $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx, \ \alpha>0, \beta>0$

   ii $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

3. The mean of the distribution is $\mu = E(X) = \frac{\alpha}{\alpha+\beta}$
4. The variance of the distribution is $V(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Normal Distribution: -(Bell distribution)

Scanned by CamScanner

A continuous random variable X is said to have a normal distribution with parameters μ and 6^2 if its probability density function (pdf) is given by

$$f(x) = f(x; \mu, \sigma^2) = P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \begin{matrix} -\infty<x<\infty \\ -\infty<\mu<\infty \\ \sigma>0 \end{matrix}$$

Where, $\mu = E(X)$ = mean, $\sigma^2 = V(X)$ = variance, e=exp=exponential function.

Remarks: -

1. A random variable X with mean μ and variance σ^2 and following the normal distribution is expressed by $X \sim N(\mu, \sigma^2)$
2. The mean of the normal distribution $E(X) = \mu$
3. The mean of the normal distribution $V(X) = \sigma^2$
4. The curve is bell shaped and symmetrical about the line $X = \mu$.
5. If the given distribution is normal then mean = median= mode.
6. If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{x-\mu}{\sigma}$ the standard normal variable follows standard normal distribution with mean zero and variance unit [i.e. $E(Z) = 0$, $V(Z) = 1$] and its probability density function is given by $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \quad -\infty < Z < \infty$.
7. The graph of $f(z) = f(z; 0, 1)$ is called the standard normal curve (or z curve). The corresponding distribution function of z is

$$F(z) = P(Z \le z) = \int_{-\infty}^{z} f(z)\, dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2}\, dz$$

Properties of Normal distribution: -

Let $X \sim N(\mu, \sigma^2)$ and it has the following properties:

1. The total area under a normal distribution is 1.0 or 10%.
2. The normal curve is bell-shaped and symmetrical about the ordinate at $x = \mu$
3. The ordinate at $x = \mu$ divides the normal curve into two equal parts. Further more the 50% of the total area under a normal curve lies on the left side of the mean and 50% lies on the right side of the mean.
4. The normal curve has a single peak; thus it is unimodel.
5. Mean, median and mode of the distribution
   Coincide i.e. Mean = median = mode.
6. The quartiles are equidistance from the median.
7. X-axis is an asymptote to the curve.
8. No portion of the curve lies below the x-axis.
9. Since the distribution is symmetrical, the moment coefficient of skewness is zero.
10. Since the distribution is mesokurtic, the coefficient of kurtosis is given by
    $\beta_2 = 3 \Rightarrow r_2 = 0$
11. The maximum probability of the normal distribution occurs at $x = \mu$ and the maximum value is $\frac{1}{\sigma\sqrt{2\pi}}$.

Standard Normal Distribution

Standard normal distribution is a special case of the normal distribution. If a random variable X follows normal distribution with mean μ and standard deviation σ, define a random variable Z as $Z = \frac{x-\mu}{\sigma}$ is called the Standard Normal

Scanned by CamScanner

Variate. The expected value and variance of the standard normal variate Z is given by, $E(Z) = 0$, $V(Z) = 1$

The probability density function of standard normal variate Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} ; -\infty < Z < \infty.$$

Symbolically, we write $Z \sim N(0,1)$.

"Log-Normal Distribution"

A continuous random variable 'X' has a log-normal distribution if the random variable $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ. The resulting density function of X is

$$f(x, \mu, \sigma) = \begin{cases} \dfrac{1}{x\,\sigma\sqrt{2\pi}} \operatorname{Exp}\left[-\dfrac{1}{2}\left(\dfrac{\ln x - \mu}{\sigma}\right)^2\right] ; & 0<x<\infty \\ 0 & \text{otherwise} \end{cases}$$

[Where, $\ln x = \log_e x$ (natural logarithm of x)]

The mean of this distribution with $E(X) = e^{\mu+\sigma^2/2}$

The variance this distribution $V(X) = e^{(2\mu+\sigma^2)}(e^{\sigma^2}-1)$

Exponential Distribution

A continuous random variable X is said to have exponential distribution with parameter λ if its probability density function (pdf) is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Remarks: -

1. Exponential distribution is the particular case of gamma distribution

   When $\alpha = 1$ and $\frac{1}{\beta} = \lambda$.

2. The mean of this distribution is, $E(X) = \frac{1}{\lambda}$

3. The variance of this distribution is $V(X) = \frac{1}{\lambda^2}$

4. The cumulative distribution function is given by

   $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx \,.\, x \geq 0.$$

5. This distribution is used as a model for the distribution of times between the occurrences of successive events such as the waiting time between successive arrivals, the amount of time. etc.

Scanned by CamScanner

Exercise 11

1. **Show that following distribution is rectangular distribution**

$$f(x) = \frac{1}{2a}, -a < x < a$$

Solution 1

Here given probability distribution function is $f(x) = \frac{1}{2a} - a < x < a$.

Let us find the total probability of given function

$$\int_a^b f(x)\frac{1}{2a}dx = \int_{-a}^{a}\frac{1}{2a}[x]_{-a}^{a} = \frac{1}{2a}[a-(-a)] = \frac{2a}{2a} = 1.$$

Also it is define as. $f(x)\begin{cases} \frac{1}{a-(-a)} & -a < x < a \\ 0 & \text{otherwise} \end{cases}$

So it is a Rectangular [or Uniform] distribution.

2. **A random variable X has a uniform distribution over (-3, 3) computer**

i. $P(X = 2)$ ii. $P(|X| \geq 2)$ iii. $P(|X-2| < 2|$

Solution 2

Since X has a Uniform distribution over (- 3, 3)

Then , we have, $P(X = x) = f(x) = \begin{cases} \frac{1}{b-a} & -a < x < b \\ 0 & \text{otherwise} \end{cases}$

$$= \begin{cases} \frac{1}{6} & -3 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

(i) $P(X = 2) = \int_2^3 f(x)\,dx = \int_2^3 \frac{1}{6}dx = \frac{1}{6}[x]_2^3 = \frac{1}{6}\times 1 = \frac{1}{6}$

(ii) $P(|X| \geq 2) = P(-2 \leq X \leq 2) = \int_{-2}^{2} f(x)\,dx = \int_{-2}^{2}\frac{1}{6}.dx$

$$= \frac{1}{6}[x]_{-2}^{2} = \frac{1}{6}[2-(-2)] = \frac{4}{6} = \frac{2}{3}$$

(iii) $P(X-2| < 2) = P(-2 < X-2 < 2) = P(0 < \times X < 4)$

$$= \int_0^4 f(x)\,dx = \int_0^3 f(x)\,dx + \int_3^4 f(x)\,dx$$

$$= \int_0^3 \frac{1}{6}\,dx + 0 = \frac{1}{6}[x]_0^3 = \frac{3}{6} = \frac{1}{2}$$

Scanned by CamScanner

3. **If a random variable has the standard normal distribution, find the probability that it will be take on a value.**

   a. less than 1.5 b. greater than –1.75

   c. between 0.5 to 1.5

Solution 3

Since X is the standard normal distribution. Then the probability density function is given by $P(X = Z) = f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}, -\infty < Z < \infty.$

(a) Then the probability that it will take on a value less than 1.5 is given by:

$$P(X < 1.5) = \int_{-\infty}^{1.5} f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{1.5} e^{-z^2/2}\,dx = 0.9322$$

(b) $P(X > -1.75) = 1 - P(X \le 1.75) = 1 - \int_{-\infty}^{1.75} \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}\,dx = 1 - 0.9599 = 0.0401.$

(c) $P(0.5 < X < 1.5) = P(X < 1.5) - P(X < 0.5)$

$= 0.9332 - 0.6915$ [form Z-table] $= 0.2417$

4. **Find the value of Z if the probability that a random variable having the standard normal distribution will take on the value**

   a. less than Z is 0.9911 b. greater than Z is 0.1093

   c. greater than Z is 0.6443 d. less than Z is 0.0217

   e. between –Z and Z is 0.9298

Solution 4

Let Z is standard normal distribution.

(a) From the given table of normal distribution.

$F(2.37) = P(X \le Z) = P(X \le 2.37) = 0.9911$

$\therefore$ $z = 2.37$

(b) From the given table of normal distribution.

$1.33) \cdot F(= P(X \le Z) = P(X \le 1.23) = 0.1093 \Rightarrow Z = 0.37$

$\therefore$ $P(X \le 1.23) = P(X \le 1.23)$

$\therefore$ $Z = 1.23$

(c) We have,

$F(0.37) = P(X \le Z) = P(X \le 0.37) = 0.6443 \Rightarrow Z = 0.37$

(d) $\cdot F(2.02) = P(P \le Z) = P(X \le 2.02) = 0.0217$

$\therefore$ $z = 2.02$

(e) $P(|X| = 0.9298$

$P(Z \le X \le Z) = 0.9298$

or, $P(X \le Z)\ P(X \le Z) = 0.9298$

or, $P(X \le Z)\ [1\ P(X \le Z)] = 0.9298$

or, $2P(X \le Z) = 1.9298$

$P(Z \le Z) = \frac{1.9298}{2} = 0.9649$

$\therefore$ $Z = 1.81$

Scanned by CamScanner

5. **Given a random variable having the normal distribution with $\mu = 16.2$ and $\sigma = 1.5625$. Find the probability that it will take on a value,**

a. greater than 16.8 b. less than 14.9

c. between 13.6 and 18.8 d. between 16.5 and 16.7

Solution 5

Let X denote the normal distribution with $\mu = 16.2$, $\sigma = 1.5625$, and $X \sim N(16.2, 1.5625^2)$. Then,

(a) $P(X > 16.8)$

$$= P\frac{X-m}{s} > \frac{16.8-m}{s} = P(Z > 0.384) = 1P(Z \le 0.384) = 1 - 0.6480 = 0.352$$

(b) $1(X < 14.9) = P\left(\frac{X-m}{s} < \frac{14.9-m}{s}\right) = P(Z < -0.83) = 0.2033$ [form table]

(c) $P(13.6 < x < 18.8) = P\left(\frac{13.6-m}{s} < \frac{X-m}{s} < \frac{18.8-m}{s}\right)$

$$= P(-1.66 < Z < 1.66) = P(Z < 1.66) - P(Z < -1.66) = 0.9515 - 0.0485 = 0.903$$

(d) $P(16.5 < x < 16.7) = P\left(\frac{16.5-m}{s} < \frac{X-m}{s} < \frac{16.7-m}{s}\right)$

$P(0.19 < Z < 0.32) = P(Z < 0.32) - P(Z < 0.19) = 0.6255 - 0.5753 = 0.0502$

6. **A random variable has a normal distribution with mean $\mu = 62.4$. Find its s.d. if the probability is 0.2 that it will take on a value greater than 79.2.**

Solution 6

Let X denote the random variable has a normal distribution with mean.

$\mu = 62.4$

Also, $P(X > 79.2) = 0.2$

$$P\left(\frac{X-m}{s} > \frac{79.2-62.4}{s}\right) = 0.2$$

or, $P\left(Z > \frac{16.8}{s}\right) = 0.2$

or, $1 - P\left(Z < \frac{16.8}{s}\right) = 0.2$

or, $P\left(Z < \frac{16.8}{s}\right) = 0.8$

or, $P\left(Z < \frac{16.8}{s}\right) = P(Z < 0.84)$

Equating we get,

$$\frac{16.8}{s} = 0.84$$

$$\Rightarrow \sigma = \frac{16.8}{0.84} = 20$$

Scanned by CamScanner

7. **Suppose the time it takes a data collection operator to fill out an electronic form for a database is uniformly between 1.5 and 2.2 minutes.**

   a. What is the mean and variances of the time it takes an operator to fill out the form?

   b. What is the probability that it will take less than two minutes to fill out the form?

   c. Determine the cumulative distribution function of the time it takes to fill out the form?

Solution 7

Let X denote the time taken to fill out an electronic form is a uniform distribution defined in the interval [1.5, 2.2]. Then, a = 1.5, b = 2.2

(a) The mean of uniform distribution is given by

$$\text{Mean} = E(X) = \frac{a+b}{2} = \frac{2.2+1.5}{2} = 1.85 \text{ min}$$

$$\text{Also variance } V(X) = \frac{(b-a)^2}{12} = \frac{(2.2-1.5)^2}{12} = 0.041 \text{ min}^2$$

(b) The probability that it will take less than two minute to fill out the form is given by $P(X < 2) = \int_{105}^{2} f(x)\,dx = \frac{1}{0.7}[x]_{1.5}^{2} = \frac{0.5}{0.7} = 0.714$

(c) The density function of uniform distribution is

$$f(x) = \frac{1}{b-a} = \frac{1}{0.7} = 1.43 \quad \text{for, } 1.5 < X < 2.2.$$

Then the cumulative distribution of the time it takes to fill out the form is given by $F(x) = \left\{ \frac{x-a}{b-a} \quad a < x < b \right. = \left\{ \frac{x-1.5}{0.7} \quad 1.5 < x < 2.2 \right.$

8. **The bus company A schedules a north bound bus every 30 minutes at a certain bus stop. A man comes to the stop at random time. Let the random variable X count the number of minutes he has to wait for the next bus. Assume X has a uniform distribution over the interval (0, 30). This is how we interpret the statement that he enters the station at the random time.**

   For each K = 5, 10, 15, 20, 30, compute the probability that he has to wait at least k minutes for the next bus.

Solution 8

Let X denote the no. of time a man has to wait a next bus in a bus station which is uniformly distributed in the interval 0 to 30 min.

Then the probability density function is given by $f(x) = \frac{1}{30}$ $0 < x < 30$

$$\text{Then, } P(X \le 5) = \int_{0}^{5} f(x)\,dx = \frac{5}{30} = \frac{1}{6} = 0.167$$

Scanned by CamScanner

$$P(X \le 10) = \int_0^{10} f(x)\,dx = \frac{10}{30} = \frac{1}{3} = 0.33$$

$$P(X \le 15) = \int_0^{15} f(x)\,dx = \frac{15}{30} = \frac{1}{2} = 0.5$$

$$P(X \le 20) = \int_0^{10} f(x) = \frac{20}{30} = \frac{2}{3} = 0.67$$

$$P(X \le 30) \int_0^{30} f(x)\,dx = \frac{30}{30} = 1$$

9. **If a random variable has the gamma distribution with $\alpha = 2$ and $\beta = 3$. Find the mean and standard deviation of this distribution.**

Solution 9

Let X denote the random variable has the gamma distribution with $\alpha = 2$ and $\beta = 3$.

We know that the mean and standard deviation of the random variable is given by

mean $E(X) = \alpha.\beta = 2 \times 3 = 6$

s.d $= \sigma = \alpha\beta^2 = 2 \times 3^2 = 18$

10. **In certain experiments, the error made in determining the solubility of a substance is a random variable having the uniform density with a = –0.025 and b = 0.025. What are the probabilities that such an error will be:**

a. between 0.010 and 0.015? b. between –0.012 and 0.012?

Solution 10

Let X denote the error made in determining the solubility of a substance uniformly distributed between a = – 0.025 to b = 0.025. Then the probability a = – 0.025 to b = 0.025. Then the probability distribution function is given by

$$f(x) = \frac{1}{b-a} = \frac{1}{0.025 - (-0.025)} = \frac{1}{0.05} = 20$$

(a) Then the probability of error having between 0.010 and 0.015 is given by

$$P(0.010 < X < 0.015) = \int_{0.010}^{0.015} f(x)\,dx = 20\,[x]_{0.010}^{0.015} = 20 \times 0.005 = 0.1$$

(b) The probability of error having between – 0.012 and 0.012 is given by

$$P(-0.012 < \times X < 0.012) = \int_{0.012}^{0.012} f(x)\,dx = [x]_{0.012}^{0.012} = 20 \times 0.024 = 0.48$$

11. **Suppose that the life time of a certain kind of computer is random variable X having the Gamma distribution with $\alpha = 2$ and $\beta = 6$. Find:**

a. mean lifetime of computers

b. The probability that such a computer will last more than 10 years.

Scanned by CamScanner

Solution 11

Let X is the life time of a certain kind of computer having Gamma distribution with $\alpha = 2$ and $\beta = 6$.

(a) Then mean of the gamma function is given by $= E(X) = \alpha\beta = 2 \times 6 = 12$

(b) The probability that such a computer will last more than 10 years is given by $P(X) > 10 \quad = 1 - P(X \le 10)$

$$= 1 \int_0^9 f(x)\,dx = 1 - \int_0^9 \frac{1}{\sigma\Gamma(2)} (10)^{2-1} \cdot e^{\frac{-10}{6}}\,dx = 1 - \frac{10 \times e^{-1.67}}{36 \times 1} \cdot [x]_0^9$$

$$= 1 - 0.471 = 0.529$$

12. **If the annual proportion of erroneous income tax refunds filed with the I.R.S. can be looked upon as a random variable having a beta distribution with $\alpha = 2$ and $\beta = 3$, what is the probability that in any given years there will be fewer than 10% erroneous returns?**

Solution 12

Let X denote the annual proportion of erroneous income tax return filled with the IRS. has the Beta distribution with $\alpha = 2, \beta = 3$. Then the p.d.f of Beta function is given by:

$$f(x) = \frac{1}{\beta(a,b)} x^{\alpha-1} (1-x)^{\beta-1} \quad \alpha > 0, \beta > 0, 0 < x < 1$$

$$= \frac{\Gamma(a,b)}{\Gamma(a)\,\Gamma(b)} . x^{\alpha-1} (1-x)^{\beta-1} = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} . x^1 . (1-x)^2 = 12x\,(1-x)^2$$

Here $X = 10\% = 0.1$

The probability that in any given year there will be fewer than 10% erroneous returns is given by $P(X < 0.1) = \int_0^{01} f(x)\,dx$

$$= \int_0^{0.1} 12x\,(1 - 2x + x^2)\,dx = \int_0^{0.1} (12x - 24^2 + 12x^3)\,dx$$

$$= 12\left[\frac{x^2}{2}\right]_0^{0.1} - 24\left[\frac{x^3}{3}\right]_0^{0.1} + 12\left[\frac{x^4}{4}\right]_0^{0.1}$$

$$= 6 \times 0.01 - 8 \times 0.001 + 3 \times 0.0001 = 0.0523$$

13. **Suppose that the proportion of defectives ship by a vendor, which varies somewhat from shipment to shipment, may be looked upon as a random variable having the beta distribution with $\alpha = 1$ and $\beta = 4$.**

a. Find the mean of this beta distribution namely, the average proportion of defectives in a shipment from this vendor.

b. Find the probability that a shipment from this vendor will contain 25% or more defectives.

Scanned by CamScanner

Solution 13

Let X denote the proportion of defective in a shipment from this vender has Beta distribution with $\alpha = 1$ and $\beta = 4$. Then

(a) The mean of Beta distribution (the average proportion of defectives in a shipment from this vendor) $= \frac{a}{a+b} = \frac{1}{1+4} = \frac{1}{5} = 0.2$

(b) Then the p.d.f of Beta function is given by:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(5)}{\Gamma(1)\,\Gamma(4)} \cdot x^0 (1-x)^3 = 4(1-x)^3$$

Then the probability that a shipment from his vendor will contain 25% or more defective is given by $P(X \geq 0.25)$

$$= \int_{0.25}^{1} 4(1-x)^3\,dx = \left[\frac{(1-x)^4}{-1}\right]_{0.25}^{1}$$

$$= -[0-(1-0.24)^4] = 0.316$$

14. In a certain country, the proportion of highway sections requiring repairs in any given year is a random variable having the beta distribution with $\alpha = 5$ and $\beta = 2$. Find,

a. On the average what percentage of the highway sections required repairs in any given years.

b. The probability that at most half of the highway section will required repairs in any given year.

Solution 14

Let X denote the proportion of highways sections requiring repairs in a given year having Beta distribution with $\alpha = 5$ and $\beta = 2$.

(a) Now the mean of the Beta distribution is given by $E(X) = \frac{a}{a+b} = \frac{5}{5+2} = \frac{5}{7} = 0.715$

Hence the average percentage of the highway section requiring repairs in a given. Year is 71.5%

(b) The p.d.f of Beta function is given by

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(5+2)}{\Gamma(5)\,\Gamma(2)} \cdot x^4(1-x) = \frac{6!}{4!\,2!}(x^4 - x^5) = 30(x^4 - x^5)$$

The probability that at most half of the high way section will required repairs in any given year is

$$P\left(X \leq \frac{1}{2}\right) = \int_0^{1/2} f(x)\,dx = \int_0^{1/2} 30(x^4 - x^5)\,dx = 30\left[\frac{x^5}{5} - \frac{x^6}{6}\right]_0^{1/2} = 6 \times \frac{1}{32} - 5 \times \frac{1}{64} = 0.11$$

15. The probability density function of exponential distribution f(x)

$$f(x) = \begin{cases} 0.25e^{-0.25x} & 0 < x \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability that:

a. the time to observe a particle is more than 200 microseconds.

b. the time to observe a particle is less than 10 microsecond.

Scanned by CamScanner

Solution 15

Let X denote the time to observe α particle has exponential distribution with probability density function $f(x) = \begin{cases} 0.25\, e^{-0.25x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$

(a) The probability that the time to observe a particle is more than 200 microseconds $P(X \geq 200)$

$$= 1 - P(X \leq 200) \quad = 1 - \int_0^{200} f(x)\, dx = 1 - \int_0^{200} 0.25\, e^{-0.25x}\, dx$$

$$= 1 - 0.25 \cdot \left[\frac{e^{-0.25x}}{-0.25}\right]_0^{200} = 1 - \{-[e^{-50} - e^0] = 1 - [1 - e^{-50}] = e^{-50}$$

(b) The probability that the time to observe a particle is less that 10 microsecond is

$$P(X < 10) \quad = \int_0^{10} f(x)\, dx = \int_0^{10} 0.25\, e^{-0.25x}\, dx = -(e^{-0.25x})_0^{10} = -(e^{-2.5} - 1) = 0.92$$

16. The burning time of an experimental rocket is a random variable having the normal distribution with $\mu = 4.76$ seconds and $\alpha = 0.04$ second. What is the probability that this kind of rocket will burn.

a. less than 4.66 seconds b. more than 4.80 seconds

c. anywhere from 4.70 to 4.82 seconds?

Solution 16

Let X denote the time that burning an experimental rocket has the normal distribution with $\mu = 4.76$ seconds. and $\sigma = 0.04$ seconds.

(a) The probability that the rocket will burn less than 4.66 second is given by

$$P(X < 4.66) = P\left(\frac{X - m}{\sigma} < \frac{4.66 - 4.76}{0.04}\right) = P(Z < -2.5) = 0.0062 \text{ (by Z-table).}$$

(b) The probability that the rocket will burn more than 4.80 seconds is given by.

$$P(X > 4.80) = P\left(\frac{X - m}{\sigma} > \frac{4.80 - 4.76}{0.04}\right) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

(c) $$P(4.70 < X < 4.82) \quad = P\left(\frac{4.70 - m}{\sigma} < \frac{X - m}{\sigma} < \frac{4.82 - m}{\sigma}\right)$$

$$= P(-1.5 < Z < 1.5) = P(Z < 1.5) - P(Z < -1.5)$$

$$= 0.9332 - 0.0668 \quad = 0.8664.$$

17. A normal pollution of 700 wages earners has a mean income of Rs. 220 per month and variance is 332. Find the number of persons who earn (i) between Rs. 150 and Rs. 200 per month, (ii) more than Rs.300 per month and (ii) less than Rs.100 per month.

Solution 17

Let X denote the monthly wages in Rs. has the normal distribution with mean μ = Rs. 220 and variance = $\sigma^2 = 332$. Then the standard deviation is given by s.d $(\sigma) = \sqrt{V(X)} = \sqrt{332} = 18.22$

Scanned by CamScanner

(i) The number of person who earn between Rs. 150 and Rs. 200 per month is given by

$$N.P\ (150 \times X < 200) = 700 \times P\left(\frac{150-m}{\sigma} < \frac{X-m}{\sigma} < \frac{200-m}{\sigma}\right)$$

$$= 700 \times P(-3.84 < Z < -1.11)$$

$$= 700\ [\{p\ (Z < -0.11] - P\ (Z < -3.84)]$$

$$= 700\ [F\ (-0.11) - F\ (-3.84)]$$

$$= 700\ [0.4562 - 0.0002] = 319.2 \approx 320$$

(ii) The number of persons who earn more than Rs. 300 per month is given by

$$N.\ P(X > 300) = 700\ P\left(\frac{X-m}{\sigma} > \frac{300-m}{\sigma}\right)$$

$$= 700\ P\ (Z > 4.39) = 700\ [1 - P\ (Z < 4.39)]$$

$$= 700 \times 0.00003 = 0.021 \approx 0$$

(iii) The number of persons who earn less than Rs 100 per month is given by

$$N.P(X < 100) = N.P\left(\frac{X-m}{\sigma} < \frac{100\ K\ m}{\sigma}\right) = 700\ P(Z < 6.59) = \text{(all most none)}$$

18. In a normal population with mean 15 and standard deviation 3.5, it is known that 647 observation exceed 16.25. What is the total number of observations in the population.

Solution 18

Let X denote the number of observation in the given population has the normal distribution with mean (μ) = 15 and standard deviation $\sigma = 3.5$.

Also the number of observations in the population exceed 16.25 is 647 i.e.

$N \times P\ (X > 16.25) = 647$

or, $N.P\left(\frac{X-\mu}{\sigma} > \frac{16.25-15}{3.5}\right) = 647$

or, $N.P\ (Z > 0.360) = 647$

or, $N\ (1 - 0.6406) = 647$

or, $N = \frac{647}{0.3594} = 1800.22 \approx 1800$

19. Assume the mean heights of soldiers to be 68.22 inches with variance of 10.8 inches². How many soldiers in a regiment of 1,000 would you expect to be over 6 feet tall?

Solution 19

Let X denote the random variable that the heights of soldiers has normal distribution with mean $\mu = 68.22$ inches and s.d (σ) = $\sqrt{10.8} = 3.29$ inches. Also we have given N = 1,000.

Then expected number of soldiers to be over 6 feet (6×12 = 72 inches)

Is given by E (X > 72)

Scanned by CamScanner

$= N \times P(X > 72)$

$= 1000 \times P\left(\frac{X - m}{\sigma} > \frac{72 - 68.22}{3.29}\right)$

$= 1000 \times P(Z > 1.15) = 1000 \times [1 - P(Z \le 1.15)] = 1000 \times (1 - 0.87 + 49) = 125.1 \approx 125$

20. If X is normally distributed with mean 2 and variance 1. Find $P(|X - 2| < 1)$.

Solution 20

Let X is the random variable, which has normal distribution with mean $\mu = 2$ and variance $(\sigma^2) = 1$. So, s.d $= \sqrt{s^2} = 1$

Then for the value of

$P(X - 2| < 1) = P(-1 < X - 2 < 1)$

$= P(1 < X < 3) = P\left(\frac{1-m}{6} < \frac{X - m}{6} < \frac{3 - m}{6}\right)$

$= P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826$

21. If X is normal variable with mean 50 and s.d. 10, find $P(Y \le 3137)$ where $Y = X^2 + 1$.

Solution 21

Let X denote the random variable having normal distribution with mean $(\mu) = 50$ and s.d $(\sigma) = 10$.

Also we have given $Y = X^2 + 1$.

Now, $P(Y \le 3137) = P(X^2 + 1 \le 3137) = P(X^2 \le 3136) = P(|X| \le 56)$

$= P(X \le 56)$ [Since $Y = X^2 + 1$ is positive function.]

$= P\left(\frac{X - m}{s} \le \frac{56 - 50}{10}\right) = P(Z \le 0.6) = 0.7257$ [By Z - table]

22. X_1, X_2, X_3, X_4 is a random sample from normal distribution with mean 100 and variance 25 and $\bar{X} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$. Find expected value and variance of each of the following:

a. $4\bar{X}$ b. $X_1 - 2X_2 + X_3 - 3X_4$

Solution 22

Do yourself

23. X is a normal variate with mean 1 and variance 4, Y is another normal variate independent of X with mean 2 and variance 3. Find the mean and variance of $7X + 12Y + 15$.

Solution 23

Do yourself

24. In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution.

Solution 24

Let X denote the random variable having normal distribution. If $X \sim N(\mu, \sigma^2)$ then, 7% of the items are under 35. That is $P(X < 35) = 7\% = 0.07$ Also, 89% of the items are under 63.

Scanned by CamScanner

i.e. $P(X < 63) = 80\% = 0.89$

Suppose the points $X = 35$ located on the left of the ordinate at $X = \mu$ also the point $X = 63$ lies on the right of $X = \mu$. Then when $X = 35$, $Z = \frac{35 - m}{s} = -z_1$ ...(1)

Also when $X = 63$, $Z = \frac{63 - m}{s} = z_2$ ...(2)

Then from above, we have,

$P(Z < -z_1) = 0.07$ and $P(Z < z_2) = 0.89$

Then by the Z – table we get,

$P(Z < -1.48) \approx 0.07$ and $P(Z < 1.23) = 0.89$

$\Rightarrow z_1 = 1.48$ and $z_2 = 1.23$.

Now form (1) and (2) we get,

$$\frac{35 - m}{s} = -1.48 \Rightarrow \mu - 1.48\sigma = 35 \quad ...(3)$$

and $$\frac{63 - m}{s} = 1.23 \Rightarrow \mu + 1.23\sigma = 63 \quad ...(4)$$

Solving (3) and (4) we get

$\mu = 50.3$ and $\sigma = 10.33$

25. **In an aptitude test administered to 900 college student the mean was found to be 50 and S.D. 20.**

a. Find the number of students securing between 30 and 70.

b. Find the value of the score exceeded by the top 90 student.

Solution 25

Let X denote the score securing by the college students in the aptitude test, has normal distribution with mean $(\mu) = 50$, s.d. $(\sigma) = 20$ and $N = 900$

(a) The number of students securing between 30 and 70 is given by

$$NP(30 < X < 70) = NP\left(\frac{30 - 50}{20} < \frac{X - m}{s} < \frac{70 - 50}{20}\right)$$

$$= NP(-1 < Z < 1)$$

$$= NP(Z < 1) - P(Z < -1) = 900\,(0.8413 - 0.1587) = 614.34 \approx 614$$

(b) For the value of the score exceeded by top 90 students

The proportion of 90 students $= \frac{90}{900} \times 100\% = 10\%$

Let x_1 be the minimum marks of 90 students.

When $X = x_1$, $Z = \frac{X - m}{s} \geq \frac{x_1 - 50}{20} = z_1$ ..... (1)

Now, $P(X \geq x_1) = P\left(\frac{X - m}{s} \geq \frac{x_1 - 50}{20}\right)$

$\Rightarrow P(Z \geq z_1) = 0.1$

$\Rightarrow 1 - P(Z < z_1) = 0.1$

or, $P(Z < z_1) = 0.9$

Scanned by CamScanner

Also form the Z – table we know that $P(Z \le 1.29) = 0.9015 \approx 0.9$

$\because$ $z_1 = 1.29$.

Now form (1) we have, $\frac{x_1 - 50}{20} = 1.29$

$\because$ $x_1 = 1.29 \times 20 + 50 \approx 76$ #.

26. The mean weight of products is 68.22 grams with a variance of 10.8 grams². How many products in a batch of 1000 would you expect?

a. to be over 72 grams b. between 70 and 72 grams

Solution 26

Let X denote the weight of products in grams and Z be its standard normal variable. With mean (μ) = 68.22 and standard deviation (σ) = $\sqrt{10.8} = 3.29$ Also, N = 1000.

(a) The expected number of weights over 72 grams is given by N.P (X > 72)

$$= 1000\,[1 - P(X \le 72)] = 1000\left[1 - P\left(\frac{X-m}{s} \le \frac{72-68.22}{3.29}\right)\right]$$

$$= 1000\,[1 - P(Z \le 1.15)] = 1000 \times 0.1251 = 125.1 \approx 125$$

(b) The expected number of weights between 70 and 72 grams is given by N.P (70 < X < 72)

$$= 1000\,P\left(\frac{70-68.22}{3.29} < \frac{X-m}{s} < \frac{72.68.22}{3.29}\right) = 1000\,P(0.54 < Z < 1.15)$$

$$= 1000\,[P(Z < 1.15) - P(Z < 0.54)] = 1000\,[0.8749 - 0.7054] = 169.5 \approx 170$$

27. Given a normal distribution $\mu = 50$ and $\sigma = 10$, find the value of X that has (a) 13% of the area to its left and (b) 14% of the area to its right.

Solution 27

Let X be the random variable which contains the area of the normal distribution with mean (μ) = 50 and standard deviation (σ) = 10.

(a) The value of X that has 13% of the area to its left. Let x_1 is the minimum value of 13% of the area to its left.

When $X = x_1$, $\frac{X-m}{s} = \frac{x_1 - m}{s} = -z_1$ ...(i)

Now, $P(X \ge x_1)$ $\Rightarrow P(Z \ge -z_1) = 0.13$

$\Rightarrow 1 - P(Z < z_1) = 0.87$

Now from the Z – table we have,

$P(Z < 1.13) = 0.8708 \approx 0.87$

Then we have, $Z_1 = 1.13$

None form (1), $\frac{x_1 - 50}{10} = -1.13 \Rightarrow x_1 = -11.3 + 50 = 38.7$

(b) For the value of X that has 14% of the area to its right. Let x_2 is the minimum value to its right 14% of the area to its right.

Scanned by CamScanner

When $X = x_2$, $\frac{X-m}{s} = \frac{x_2 - 50}{10} = z_2$ ...(ii)

None, $P(X \geq z_2) = 0.14$

$\Rightarrow 1 - P(Z < z_2) = 0.14$

$\Rightarrow P(Z < z_2) = 0.86$

Now form the Z - Table we have,

$P(Z \leq 1.07) = 0.8599 \approx 0.86$

$\therefore$ $z_2 = 1.07$.

Now form (II) we have, $\frac{x_2 - 50}{10} = 1.07$

$\Rightarrow x_2 = 10.7 + 50 = 60.7$.

28. **In an intelligence test administered to 1000 students the average score was 42 and standard deviation 24. Find (a) the number of student exceeding a score 50, (b) the number of student lying between 30 and 54, (c) the value of score exceeded by the top 100 student.**

Solution 28

Let X denote the score of an intelligence test which has normal distribution with mean (μ) = 42, s.d (σ) = 24 and N = 1000.

(a) The number of students exceeding a score 50 is given by

$$N.P(X \geq 50) = 1000\, P\left(\frac{X-m}{s} \geq \frac{50-42}{42}\right)$$

$$= 1000 \times P(Z \geq 0.33)$$

$$= 1000\,[1 - P(Z < 0.33)] = 1000\,[1 - 0.6293] = 370.7 = 371.$$

(b) The number of students scoring between 30 and 54 is given by

$$N.P(30 < X < 54) = 1000 \times P\left(\frac{30-42}{24} < \frac{X-m}{s} < \frac{54-42}{24}\right)$$

$$= 1000 \times P(-0.5 < Z < 0.5)$$

$$= 1000\,[P(Z < 0.5) - P(Z < -0.5)]$$

$$= 1000\,(0.6915 - 0.3085) = 383$$

(c) For the score exceed by top 100 students. The proportion of 100 students

$$= \frac{100}{1000} \times 100\% = 10\%$$

Let x_1 be the minimum marks obtained by 100 students. Then

When $X = x_1$, $\frac{X-m}{s} = \frac{x_1 - m}{s}$

$\Rightarrow Z = \frac{x_1 - 42}{24} = z_1$ (say) ...(i)

Now $P(X \geq x_1) = 0.1$

or, $P(Z \geq z_1) = 0.1$

or, $1 - P(Z < Z_1) = 0.1$

Scanned by CamScanner

or, $P(Z < Z_1) = 0.9$

Now form the Z - table we get,

$P(Z < 1.29) = 0.9015 \approx 0.9$

Thus $z = 1.29$

Now form equation (I) we get,

$$\frac{x_1 - 42}{24} = 1.29 \Rightarrow x_1 = 24 \times 1.29 + 42 = 72.96 \approx 73$$

29. **At a certain examination, 10% of the student who appeared for the paper in statistics got less than 30 marks and 97% of students got less than 62 marks. Assuming the distribution to be normal, find the mean and the standard deviation of the distribution.**

Solution 29

Let X denote the marks obtained by the students in the paper of statistics. Let Z denote the standard normal variable having normally distributed.

Since 10% of the students get less than 30

marks i.e. $P(X \le 30) = 0.1$

Also, 97% of the students go less than 62 marks,.

i.e, $P(X \le 62) = 0.97$.

Let μ be the mean and σ be the standard deviation of the normal distribution.

$X = 30$ lies on the left and $X = 62$ lies on the right of the ordinate $X = \mu$.

Then,

When $X = 30$, $\frac{X - m}{s} = \frac{30 - m}{s} = -zz$ ...(i)

When $X = 62$, $\frac{X - m}{s} = \frac{62 - m}{s} = z$ ...(ii)

Since $P(X \le 30) = 0.1$

$\Rightarrow P(Z \le -z_1) = 0.1$

Then by Z table we have,

$P(Z \le -1.28) = 0.1$

$\because z_1 = 1.28$

$P(x \le 62) = 0.97$

$\Rightarrow P(Z \le z_2) = 0.97$

Then by Z table we have,

$P(Z \le 1.89) \approx 0.97$

$z_2 = 1.89$

Now from (i) we get,

$$\frac{30 - m}{s} = 1 - 1.28 \Rightarrow \mu + 1.28\sigma = 30 \quad ...(iii)$$

Also form (ii) we have,

$$\frac{60 - m}{s} = 1.89 \Rightarrow \mu + 1.89\sigma = 60 \quad ...(iv)$$

Solving (iii) and (iv) we get,

Mean $(\mu) = 40.78$ and $\sigma = 9.5$

30. **A normal distribution has 77 as mean. Find its standard deviation if 20% of the area under the curve lies to the right of 90.**

Scanned by CamScanner

Solution 30

Let X the normal distribution with mean (μ) = 77 and standard deviation σ.

Also the 20% of the area under the curve lies to the right of 90.

i.e. $P(X > 90) = 0.2$

or, $P\left(\frac{X-m}{s} > \frac{90-77}{s}\right) = 0.2$

or $P\left(Z > \frac{13}{s}\right) = 0.2$

or, $1 - P\left(Z \le \frac{13}{s}\right) = 0.2$

or, $P\left(Z \le \frac{13}{s}\right) = 0.8$.

Form the Z – table we get,

$P(Z \le 0.84) = 0.7995 \approx 0.8$

$\because \frac{13}{s} = 0.84$

$\sigma = \frac{13}{0.84} = 15.47$

31. **A set of examination marks is approximately normally distributed with mean of 75 and standard deviation of 5. If the top 5% of student get grade A and bottom 25% get grade F, what marks is the lowest A and what mark is the highest F?**

Solution 31

Let X denote the marks obtained by the students in an examinations is approximately normally distributed with mean (μ) = 75 and standard deviation (σ) = 5.

Since top 5% of the students get grade A. Let x_1 denote the minimum marks obtained by top 5% of the students. Then, $P(X \ge x_1) = 0.05$

For $X = x_1$, $\frac{X-\mu}{s} = \frac{x_1 - 75}{5} = z_1$ ...(i)

Also let x_2 denote the maximum marks obtained by bottom 25% students them.

$P(X \le x_2) = 0.25$

For $X = x_2$, $\frac{X-m}{s} = \frac{x_2 - 75}{5} = -z_2$ ...(ii)

Now , $P(X \ge x_1) = 0.05$

$\Rightarrow P(Z \ge z_1) = 0.05$

$\Rightarrow P(Z < z_1) = 0.95$

None from the z-table.

$P(Z < 1.65) = 0.9505 \approx 0.95$

$\because z_1 = 1.65$

Also, $P(X \le x_2) = 0.25$

$\Rightarrow P(Z \le z_2) = 0.25$

None form the Z-table we get,

$P(Z \le -0.67) \approx 0.25$

$\because z_2 = 0.67$

Now form equation (i) we have.

Scanned by CamScanner

$$\frac{x_1 - 75}{5} = 1.65 \qquad \Rightarrow x_1 = 5 \times 1.65 + 75 = 83.25 \approx 83$$

Also form equation (II) we have,

$$\frac{x_2 - 75}{5} = -0.67 \Rightarrow x_2 = 5 \times 0.67 + 75 = 71.65 \approx 72$$

$\because$ The lowest mark for grade A is 83 and the highest mark for grade F is 72.

32. **A random variable has a normal distribution with 10 as standard deviation. Find its mean if the probability that the random variable takes on a value less than 80.5 is 0.3264.**

Solution 32

Let X denote the random variable having normal distribution and Z denote the standard normal variable with mean μ and standard deviation $\sigma = 10$.

Also the probability that the random variable takes on a value less than 80.5 is 0.3264.

i.e $P(X < 8.05) = 0.3264$

$$\Rightarrow P\left(\frac{X - m}{s} < \frac{8.05 - m}{10}\right) = 0.3264 \Rightarrow P\left(Z < \frac{80.5 - m}{10}\right) = 0.3264$$

We know that form the Z - table

$$P(Z \le -0.45) = 0.3264$$

Therefore, $\frac{80.5 - m}{10} = -0.45$

$$\Rightarrow 80.5 - \mu = -4.5 \qquad \Rightarrow \mu = 85$$

Hence the mean of the distribution is 85.

33. **The mean weight of 500 male students in a certain college is 151 lbs and the standard deviation is 15 lbs. Assume the weight are the normally distributed. Find how much student weight more than 185 lbs.**

Solution 33

Let X denote the weight of the students in a certain college, Z denote the standard normal variable with mean $\mu = 151$ lbs and s.d $(\sigma) = 15$ lbs.

Also total number of students = N = 500.

The number of students having weight more than 185 lbs is given by:

$$N.P(X > 185) \quad = P\left(\frac{X - m}{s} > \frac{185 - 151}{15}\right) = 500.\ P(Z > 2.27)$$

$$= 500.\ [1 - P(Z \le 2.27)] = 500\ [1 - 0.9884] = 5.8 \approx 6.$$

34. A stamping machine produces can tops whose diameter are normally distributed with a S.D. of 0.01 inch. At what "normal" (mean) diameter should the machine to be set so that no more than 5% of the can tops produced have diameters exceeding 3 inches?

Solution 34

Let X denote the length of diameter of can top produces by a machine in inches has normally distributed with mean μ and s.d $(\sigma) = 0.01$ inches.

Scanned by CamScanner

For the at most 5% of the can tops produced having diameter at most 3 inches is gives by:

$P(Z \geq 3) \leq 0.05$

$$\Rightarrow \quad P\left(\frac{X-m}{s} \geq \frac{3-m}{0.01}\right) \leq 0.05$$

$$\Rightarrow \quad P\left(z \geq \frac{3-m}{0.01}\right) \leq 0.05$$

$$\Rightarrow \quad 1 - P\left(Z < \frac{3-m}{0.01}\right) \leq 0.05$$

$$P\left(Z < \frac{3-m}{0.01}\right) \geq 0.95$$

Now form the Z - table: $P(Z \leq 1.65) \geq 0.95$

Therefore, $\frac{3-m}{0.01} = 1.65$

$\Rightarrow \quad 3 - \mu = 0.0165$

$\mu = 3 - 0.0165 = 2.9835$

35. **Specifications for a contain job all for washers with an inside diameter of 0.300 ± 0.005 inch. If the inside diameters of the washers supplied by a given manufacture may be looked upon a random variable having the normal distribution $\mu = 0.302$ inch and $\sigma = 0.003$ inch, what percentage of these washers will meet specification?**

Solution 35

Let X denote the length of diameter for washers, which is normally distributed with the mean $(\mu) = 0.302$ and standard deviation $(\sigma) = 0.003$ inch. Also given specification is 0.300 ± 0.005

For the probability that these washers will meet the specification is given by

$P(0.3 - 0.005 < X\ 0.3 + 0.005)$

$$= \left(\frac{0.295 - 0.302}{0.003} < \frac{X-m}{s} < \frac{0.305 - 0.302}{0.003}\right)$$

$= P(-2.33 < Z < 1) = P(Z < 1) - P(Z < -2.33) = 0.8413 - 0.0099 = 0.8314$

Hence the percentage of washers will meet the specification is 83.14%.

36. **Extrude plastic rods are automatically cut into normally length of 6 inches. Actual lengths are normally distribution about a mean of 6 and their S.D. is 0.06 inch.**

a. What proportion of the rod exceeds tolerance limits of 5.9 inches to 6.1 inches?

b. To what values does the S.D. need to be reduced if 99% of the rods must be within tolerance?

Solution 36

Let X denote the actual length of extruded plastic rods. Which are normally distributed with mean $(\mu) = 6$ and s.d $(\sigma) = 0.06$ inches.

Scanned by CamScanner

The proportion of the rod exceeds tolerance limit of 5.9 inches to 6.1 inches is given by P(5.9 < X < 6.1)

$$= P\left(\frac{5.9-6}{0.06} < \frac{X-m}{s} < \frac{6.1-6}{0.06}\right)$$

$$= P(-1.67 < Z < 1.67)$$

$$= P(Z < 1.67) - P(Z < -1.67)$$

$$= F(1.67) - F(-1.67) = 0.9525 - 0.0475 = 0.905 = 90.5\%$$

37. The length of time (in seconds) that a user views a page on website before moving to another page is a log-normal random variable with parameters $\mu = 0.5$ and $\sigma^2 = 1$.

a. What is the probability that a page is viewed for more than 10 seconds?

b. What is the length of time that 50% of users view the page?

c. What is the mean and standard deviation of the time until a user moves from the page?

Solution 37

Let X denote the length of time (in second) that a user views a page on a website before moving to another page, has the log-normal distribution with parameters mean

$\mu = 0.5$ and s.d $(\sigma) = \sqrt{s^2} = \sqrt{1} = 1$.

Then $Y = \ln x \sim N(0.5, 1^2) \Rightarrow X = e^Y \sim N(0.5, 1^2)$

(a) The probability that a page is viewed for more than 10 seconds is given by

$$P(X > 10) = P(e^Y > 10) = P(Y > \ln(10)) = P\left(\frac{Y-m}{s} > \frac{\ln(10)-m}{s}\right)$$

$$= P\left(Z > \frac{2.30-0.5}{1}\right) = P(Z > 1.80) = 1 - P(Z \le 1.80) = 1 - F(1.80)$$

$$= 1 - 0.9641 = 0.0359$$

(b) Let X denote the minimum length of time that 50% of the users view the page, then we have,

$P(X > x) = 0.5$

$$P(e^Y > x) = P(Y > \ln(x) = P\left(\frac{Y-m}{s} > \frac{\ln(x)-0.5}{1}\right) = 0.5$$

$$\Rightarrow \quad P\left(Z > \frac{\ln(x)-0.5}{1}\right) = 0.5$$

$$\Rightarrow \quad P(Z < \ln(x) - 0.5) = 0.5$$

Now form the Z - table we know that

$$\Rightarrow \quad P(Z < 0) = 0.5$$

Therefore, $\ln x - 0.5 = 0$

$\ln x - 0.5 = 0$

$\Rightarrow \quad X = e^{0.5} = 1.65$ seconds.

(c) We know that mean of the log-normal distribution is given by

Scanned by CamScanner

$E(X) = e^{\mu + \sigma^2/2} = e^{0.5 + ½} = e^{0.5 + 0.5} = e^1 = 2.72$

Variance $V(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = e^{2 \times 0.5 + 1}(e^1 - 1) = 7.4 \times 1.72 = 12.72.$

$\because$ s.d $(\sigma) = \sqrt{Var(X)} = \sqrt{12.72} = 3.567.$

38. **Suppose that 'X' has a log-normal distribution and that the mean and variance of 'X' are 50 and 4000 respectively. Determine the following:**

a. The parameters μ and σ^2 of the log-normal distribution.

b. The probability that 'X' is less than 150.

Solution 38

Let X denote the random variable having log-normal distribution. Also we have gives Mean E(X) = 50 and variable V(X) = 4000.

(a) Let μ and σ^2 denote the parameters of the given log-normal distribution.

We know that,

$E(X) = e^{\mu + \sigma^2/2}$ $\qquad$ $V(X) = e^{2\mu + \sigma^2}.(e^{\sigma^2} - 1)$

or, $50 = e^{\mu + \sigma^2/2}$ $\qquad$ $4000 = e^{7.82}(e^{\sigma^2 - 1})$

or, $\mu + \frac{s^2}{2} = ln50$ $\qquad$ or, $e^{\sigma^2} - 1 = 1.61$

or, $2\mu + \sigma^2 = 7.82$ ..... (I) $\qquad$ or, $\sigma^2 = ln(2.61) = 0.96$

Substituting in (I) we get,

$2\mu = 7.82 - 0.96 \Rightarrow \mu = 3.43.$

Hence $\mu = 3.43$ and $\sigma^2 = 0.96$

(b) The probability of X is less than 150

i.e. $P(X < 150) = P(e^Y < 150) = P(Y < ln\,150)$

$$= P\left(\frac{Y - m}{s} < \frac{ln(150) - 3.43}{0.98}\right) = P(Z < 1.61) = 0.9463$$

39. **The time between arrivals of taxi at a busy intersection is exponentially distributed with a mean of 10 minutes.**

a. What is the probability that you wait longer than one hour for a taxi?

b. Suppose you have already been waiting for one hour for a tax, what is the probability that one arrives within the next 10 minutes?

c. Determine 'x' such that the probability that you wait less than 'x' minute is 0.10.

Solution 39

Let X denote the time between arrival of taxi at a busy intersection has exponential distribution with mean $(\mu) = 10$.

We know that mean $(\mu) = \frac{1}{1} = 10 \Rightarrow \lambda = 0.1$

Also, $f(x) = \lambda e^{-\lambda x} = 0.1\, e^{-0.1x}, x \geq 0.$

(a) The probability that you will wait longer than one hour for a taxi is given by

$P(X > 60) = 1 - P(X \leq 60)$

Scanned by CamScanner

$$= 1 - \int_0^{60} f(x)\,dx = 1 - \int_0^{60} (0,1)\, e^{-0.1x}\,dx = 1 + [e^{-0.1x}]_0^{60} = 1 + [e^{-6} - e^0] = 0.0025$$

(b) The probability that the taxi will arrive with in 70 minute is given by

$$P(X \le 70) = \int_0^{70} f(x)\,dx = \int_0^{70} (0.1)\, e^{-0.1x}\,dx = 0.999$$

(c) If the probability that you will wait for taxi less than x minute is 0.10 then

$P(X < x) = 0.10$

or, $\int_0^x f(x)\,dx = 0.10$

or, $\int_0^x 0.1\, e^{0.1x}\,dx = 0.10$

or, $-[e^{-0.1x} - e^0] = 0.10$

or, $-e^{-0.1X} = 0.9$

or, $-0.1X = \ln(0.9)$

or, $-0.1X = -0.105$

$\because$ $X = \frac{0.105}{0.1} = 1.05$ min.

40. The distance between major cracks in a highway follows an exponential distribution with a mean of 5 miles.

a. What is the probability that there are no major cracks in a 10-mile stretch of the highway?

b. What is the probability that there are two major cracks in a 10 mile stretch of the highway?

c. What is the standard deviation of the distance between major crack?

Solution 40

Let X denote the distance between major cracks in a high way has exponential distribution with man (μ) = 5 miles.

We know that mean $= \frac{1}{l} = 5 \Rightarrow \lambda = \frac{1}{5} = 0.2$

Also, $f(x) = \lambda\, e^{-\lambda x} = 0.2\, e^{-0.2x}\ x \ge 0$.

(a) The probability that there is no major cracks in a 10 mile is given by $1 - P(X < 10)$

$$= 1 - \int_0^{10} 0.2\, e^{-0.2x}\,dx = 1 + (e^{-2} - e^0) = \frac{1}{e^2} = 0.1353$$

(c) We know that the standard deviation of the distance between major cracks is given by $V(X) = \frac{1}{l^2} = \frac{1}{(0.2)^2} = 25$

Standard deviation s.d $= \sqrt{s^2} = \sqrt{25} = 5$

Scanned by CamScanner

41. Suppose that the service life of a bulb is a r.v. having exponential distribution with $\lambda = 0.5$. What is the probability that such bulb will be still in operating condition after 4000 hours?

Solution 41

Let X be the random variable which denotes the service life of a bulb, which has exponential distribution with parameter $\lambda = 0.5$.

Then $f(x) = \lambda e^{-\lambda x} = 0.5\, e^{-0.5x}, x \geq 0$.

Now the probability that the bulb will be still in operating condition after 4000 hours is given by.

$$P(X > 4000) = 1 - P(X \leq 4000) = 1 - \int_0^{4000} f(x)\,dx = [e^{-2000} - 1] + 1 = \frac{1}{e^{2000}} \approx 0.$$

42. The amount of time that a surveillance camera will run without having to be reset is a r.v. having exponential distribution with $\lambda = \frac{1}{50}$ day. Find the probability that such a camera will have to rest in less than 20 days.

Solution 42

Let X denote the amount of time that a surveillance camera will run without having to be reset has the exponential distribution with parameter $\lambda = \frac{1}{50} = 0.02$ day.

Then the probability that such a camera will have to reset in less than 20 days is given by

$$P(X < 20) = \int_0^{20} f(x)\,dx = \int_0^{20} 0.02\, e^{-0.02x}\,dx = -[e^{-0.4} - e^0] = 1 - e^{-0.4} = 0.3296$$

43. The mileage in thousands of miles which car owners get with a certain kind of type is a random variable having exponential distribution with $\lambda = 0.0005$.

i. At most 10,000 miles ii. Anywhere from 16000 to 24000 miles

iii. At least 30,000 miles

Solution 43

Let X denote the mileage of car by the car owners given in thousand of miles has exponential distribution with parameter $\lambda = 0.0005$

(i) The probability that one of these types will last at most 10,000 miles is given by

$$P(X \leq 10{,}000) = \int_0^{10000} f(x)\,dx = \int_0^{10000} 0.0005\, e^{-0.0005x}\,dx = [e^{-5} - e^0] = 1 - e^{-5} = 0.9932$$

(ii) The probability that one of these types will be between 1600 to 24000 miles is given by, $P(1600 < X < 24000)$

Scanned by CamScanner

$$= \int_{16000}^{24000} f(x)\,dx = \int_{16000}^{24000} 0.0005\, e^{-0.0005x}\,dx = -[e^{-12} - e^{-8}] = e^{-8} - e^{-12}$$

(iii) The probability that at least 30,000 times is given by

$P(X \geq 30{,}000) \quad = 1 - P(X < 30{,}000)$

$$= 1 - \int_0^{30000} 0.0005\, e^{-0.0005x}\,dx = 1 + [e^{-15} - e^{0}] = e^{-15}$$

44. Suppose that the proportion of defective shipped by a vendor, which varies some what shipment to shipment, may be looked upon as a random variable having a beta distribution with $\alpha = 1$ and $\beta = 4$.

a. Find the mean of this beta distribution namely, the average proportion of defectives in a shipment from this vendor.

b. Find the probability that a shipment from this vendor will contain 25% or more defectives.

Solution 44

Repeated questions of Q.13

45. Suppose that the lifetime of a certain kind of an emergency back up battery (in hours) is a r.v. X having the gamma distribution with $\alpha = 2$, and $\beta = 3$.

a. Find the mean life of these batteries.

b. Find the probability that such a battery will last more than 9 hours.

Solution 45

Let X denote the life time of a certain kind of an emergency backup battery (in hours) is a random variable having the gamma distribution with parameters $\alpha = 2$ and $\beta = 3$

(a) The mean life of these batteries is given by mean $E(X) = \alpha = \beta = 2 \times 3 = 6$ hours.

(b) The probability that such a battery will more than 9 house is given by

$P(X > 9) = 1 - P(X \leq 9)$

$$= 1 - \int_0^9 f(x)\,dx = 1 - \int_0^9 \frac{1}{3^2.\Gamma(2)}\, x^{2-1}.\, e^{-x/3}\,dx = 1 - \frac{1}{9}\left[\frac{xe^{-x/3}}{(-1/3)} - 1\frac{e^{-x/3}}{(-1/3)^2}\right]_0^9$$

$$= 1 - \frac{1}{9}[-27e^{-3} - 9(e^{-3} - e^{0})] = 1 - \frac{1}{9}[-27e^{-3} - 9e^{-3} + 9] = \frac{e^{-3}}{9} \times 36 = 4e^{-3} = 0.199$$

46. A sales tax officer has reported that the average sales of the 500 business that he has to deal with during a year amount to Rs. 36000 with a standard deviation of Rs. 10000. Assuming that the sales in these business are normally distributed, find:

i. The number of business, the sales of which are over Rs. 40000.

ii. The percentage of business, the sales of which are likely to range between Rs.300000 and Rs.40000.

iii. The probability that the sales of a business selected at random will be over Rs. 30000.

Scanned by CamScanner

Solution 46

Let X denote the amount of sales tax in a year in thousand is normally distributed with the parameter $\mu = 36{,}000 = 36$ and standard deviation $\sigma = 10{,}000 = 10$

(i) The number of business, the sales of which are over Rs. 40,000 is given by

$$N \times P(X > 40) = 500 \times P\left(\frac{X-m}{s} > \frac{40-36}{10}\right) = 500 \times P(Z > 0.4)$$

$$= 500\,[1 - P(Z \le 0.4)] = 500\,[1 - F(0.4)] = 500 \times [1 - 0.6554] = 172$$

(ii) For the percentage of business, the sales of which are likely to range between Rs. 30,000 and Rs. 40,000 is given by;

$$P(30 < X < 40) = P\left(\frac{30-36}{10} < \frac{X-m}{s} < \frac{40-36}{10}\right) = P(-0.6 < Z < 0.4)$$

$$= P(Z < 0.4) - P(Z < -0.6)$$

$$= F(0.4) - P(-0.6) = 0.6554 - 0.2743 = 0.3811 \approx 38\%$$

(iii) The probability that the sales of a business selected at random will be over Rs. 30,000 is given by

$$P(X > 30) = P\left(\frac{X-m}{s} > \frac{30-36}{10}\right)$$

$$= P(Z > -0.6) = 1 - P(Z \le -0.6) = 1 - F(-0.6) = 1 - 0.2743 = 0.7257$$

47. The arithmetic mean of purchase per day by a customer in a large store is Rs.25 with a standard deviation of Rs. 10. If on a particular day, 100 customers purchased for Rs. 37.80 or more, estimate the total number of customers who purchased from the store that day. Given that the normal area between Z = 0 and Z = 1.28 is 0.4000, where Z is a standard normal variate.

Solution 47

Let X denote the purchases per day by a customer in a large store normally distributed with the mean $\mu = 25$ and standard deviation $\sigma = 10$.

Let N denote the total no. of customers who purchases from the store.

Also the number of customer for Rs. 37.80 or more is 100.

i.e $N.\,P(X \ge 37.80) = 100$

or, $P\left(\frac{X-m}{s} \ge \frac{37.80-25}{10}\right) = \frac{100}{N}$

or, $P(Z \ge 1.28)$

or, $1 - P(Z < 1.28) = \frac{100}{N}$

or, $1 - (0.5 + 0.4000) = \frac{100}{N}$

or, $0.1\,N = 100$

$N = \frac{100}{0.1} = 1000.$

Hence the total no. of customer = 1000.

Scanned by CamScanner

48. Given a normal distribution with $\mu = 50$ and $\sigma = 10$. Find the value of X that has (i) 5% of the area to its left (ii) 14% of the area of its right.

Solution 48

Let X denote the random variable having normal distribution with parameter mean $\mu = 50$ and standard deviation $\sigma = 10$.

(i) Let x_1 denote maximum value of X that has 5% of the area to its left then.

$$P(X \le x_1) = 0.05$$

or, $$P\left(\frac{X-m}{s} \le \frac{x_1 - 50}{10}\right) = 0.05$$

or, $$P\left(Z \le \frac{x_1 - 50}{10}\right) = 0.05$$

Now form the Z - table we get, $P(Z \le -1.645) = 0.05$

Therefore $\frac{x_1 - 50}{10} = -1.645 \Rightarrow x_1 = 50 - 16.45 = 33.55$

(ii) Let x_2 denote the minimum value of X that has 14% of the area to its right. Then we have, $P(X \ge x_2) = 0.14$

$$P\left(\frac{X-m}{s} \ge \frac{x_2 - 50}{10}\right) = 0.14$$

$$P\left(Z \ge \frac{x_2 - 50}{10}\right) = 0.14 \Rightarrow P\left(Z \le \frac{x_2 - 50}{10}\right) = 0.86$$

Then by the Z-table we get, $P(Z \le 1.08) = 0.36$

Therefore we have, $\frac{x_2 - 50}{10} = 1.08 \Rightarrow x_2 = 60.8$

49. The marks obtained by students in a degree examination are normally distributed. The mean marks and S.D. of the distribution are 5000 and 10 respectively. If 674 appeared in the examination and out of those; 550 are to be declared passed, what should be the minimum pass marks?

Solution 49

Let X denote the marks obtained by the students in an examination which are normally distributed with parameter $\mu = 500$ and $\sigma = 100$.

Let x be the minimum marks obtained by 550 students.

Then the percentage of passed students is given by $\frac{550}{674} \times 100\% = 81.6\%$

Then $P(X \ge x) = 0.816$

or, $$P\left(\frac{X-m}{s} \ge \frac{x-500}{100}\right) = 0.816$$

or, $$1 - P\left(Z < \frac{x-500}{100}\right) = 0.816$$

$$\because \quad F\left(\frac{x-500}{100}\right) = 0.1839$$

Also form the Z-table we get,

$$F(-0.9) = 0.1839$$

Therefore, $\frac{x-500}{100} = -0.9$

$$\because X = 500 - 90 = 410$$

Hence the minimum pass marks is 410

Scanned by CamScanner

50. **In a certain examination, the percentages of passes and distinctions were 45 and 9 respectively. Estimate the average marks and standard deviation obtained by the candidates, the minimum pass and distinction marks being 40 and 75. (Assume the distribution of marks to be normal).**

Solution 50

Let X denote the marks obtained by students in an examination, which is normally distributed with parameter μ and σ.

Also given, pass percentage of students is 45 obtaining minimum 40 marks.

i.e. $P(X \geq 40) = 0.45$

$$P\left(\frac{X-m}{s} \geq \frac{40-m}{s}\right) = 0.45$$

or, $$1 - P\left(Z < \frac{40-m}{s}\right) = 0.45$$

$$\because \quad F\left(\frac{40-m}{s}\right) = 0.55$$

Form the Z-table we have, F (0.125) = 0.55

Therefore, $\frac{40-m}{s} = 0.125$

or, $\mu + 1.025\sigma = 40$ ...(i)

Also, From the question,

i.e. $P(X \geq 75) = 0.09$

or, $$P\left(Z \geq \frac{75-m}{s}\right) = 0.91$$

$$\Rightarrow \quad P\left(Z < \frac{75-m}{s}\right) = 0.91$$

$$\Rightarrow \quad F\left(\frac{75-m}{s}\right) = 0.91$$

Also from the Z-table we have, F (1.34) = 0.91

Therefore $\frac{75-m}{s} = 1.45$

$\Rightarrow \quad \mu + 1.34\sigma = 75$ ...(ii)

Solving (I) and (II) we get.

$\mu = 36.4 \qquad \sigma = 28.8$

Hence the mean is 36 and s.d is 29.

51. **In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution?**

Solution 51

Let X denote the random variable having normal distribution with parameter μ and σ.

Scanned by CamScanner

We have given 7% of the items are under 35 that is given by

$P(X < 35) = 0.07$

or, $P\left(\frac{X-m}{s} < \frac{35-m}{s}\right) = 0.07$

$F\left(\frac{35-m}{s}\right) = 0.07$

But by Z-table we have, $F(-1.475) = 0.07$

Therefore,

$\frac{35-\mu}{s} = -1.475$

$\Rightarrow \quad \mu - 1.475\sigma = 35 \qquad \ldots(i)$

Also, 89% of the items are under 63.

i.e. $P(X < 63) = 0.89$

$P\left(Z < \frac{63-m}{s}\right) = 0.89$

i.e $F\left(\frac{63-m}{s}\right) = 0.89$

But by Z table we get, $F(1.225) = 0.89$

Therefore, $\frac{63-m}{s} = 1.225$

$\Rightarrow \quad \mu + 1.225\sigma = 63 \qquad \ldots(ii)$

Solving (I) and (II) we get,

$\sigma = 10.37 \qquad \mu = 50.23$

Hence mean is 50.23 and s.d is 10.37.

52. i. **A normal distribution has 77 as mean. Find its standard deviation if 20 percent of the area under the curve lies to the right of 90.**

ii. **A random variable has a normal distribution with 10 as standard deviation. Find its mean if the probability that the random variable takes in a value less than 80.5 is 0.3264.**

Solution 52

Let X denote the random variable having normal distribution with mean μ and s.d σ.

(i) Since $\mu = 77$.

If 20% of the area under the curve lies on the right of 90 Then;

$P(X > 90) = 0.2 \Rightarrow P\left(Z > \frac{90-77}{s}\right) = 0.2 \Rightarrow 1 - P\left(Z < \frac{13}{s}\right) = 0.2 \Rightarrow F\left(\frac{13}{s}\right) = 0.8$

Also form the z-table we have, $F(0.845) = 0.8$

Therefore we have, $\frac{13}{s} = 0.845$

$\therefore \quad \sigma = 15.38$

Scanned by CamScanner

(ii) Since $\sigma = 10$

Also the probability that takes on a value less than 80.5 is 0.3264

i.e. $P(Z < 80.5) = 0.3264$

$$\Rightarrow \quad P\left(Z < \frac{80.5 - m}{10}\right) = 0.3264$$

$$\Rightarrow \quad F\left(\frac{80.5 - m}{10}\right) = 0.3264$$

But from the z-table we have, $F(-0.45) = 0.3264$.

Therefore, $\frac{80.5 - m}{10} = -0.45$

$$\Rightarrow \quad \mu = 80.5 + 4.5 = 85$$

53. The mean height and variance of height of 500 students were found to be 165 cm and 25 cm² respectively. Find the range of height of middle 80% of the students.

Solution 53

Let X denote the height of the students in cm which has normal distribution with parameters $\mu = 165$ cm and $\sigma^2 = 25$ cm². Also total number of students $N = 500$

We know that the range of middle 80% students is given by

The maximum height of top 90% students – The maximum height of bottom 10% students.

Let x_1 denote the maximum height of top 90% students. Then, $P(X \le x_1) = 0.9$

$$\Rightarrow P\left(\frac{X - m}{s} \le \frac{x_1 - 165}{5}\right) = 0.9 \Rightarrow F\left(\frac{x_1 - 165}{5}\right) = 0.9$$

Now by the Z table we have, $F(1.28) = 0.9$

$$\because \quad \frac{x_1 - 165}{5} = 1.28 \qquad \Rightarrow x_1 = 165 + 5 \times 1.28 = 171.4 \text{ cm.}$$

Also suppose x_2 denote the maximum height of bottom 10% students. Then

$$P(X \le x_2) = 0.1 \Rightarrow P\left(\frac{X - m}{s} \le \frac{x_2 - 165}{5}\right) = 0.1 \Rightarrow F\left(\frac{x_2 - 165}{5}\right) = 0.1$$

Now by the Z table we have, $F(-1.28) = 0.1$

$$\because \frac{x_2 - 165}{5} = -1.28 \Rightarrow x_2 = 165 - 5 \times 1.28 = 158.6 \text{ cm}$$

Hence the rang is, $x_1 - x_2 = 171.4 - 158.6 = 12.8$ cm.

54. The final examination scores in statistics are normally distributed with an average score of 70 and a variance of 25.

a. If the lowest passing grade is 58, what percentage of the class is failing?

b. If the professor given the grades on a curve and everybody getting 82 or above gets a grade of A, then what percentage of students get A grade?

c. If the highest 80% of the class are to pass the course, what is the lowest passing score?

d. What should the score be so that only 15% of the students get a score higher than this?

Scanned by CamScanner

Solution 54

Let X denote the scores obtained by the students in statistics are normally distributed with the parameter mean (μ) = 70 and standard deviation s.d (σ) = $\sqrt{25} = 5$.

(a) Since the lowest pass mark is 58 then percentage of passed students is given by

$$P(X \geq 58) = P\left(\frac{X-m}{s} \geq \frac{58-70}{5}\right) = P(Z \geq -2.4)$$

$$= 1 - P(Z \leq -2.4) = 1 - F(-2.4) = 1 - 0.0082 = 0.9918 = 99.18\%$$

Since 99.18% students passed. Then percentage of fail students $= 100 - 99.18$
$= 0.82\%$

(b) Since for A grade everybody getting 82 or above is given by

$$P(X \geq 82) = P\left(\frac{X-m}{s} \geq \frac{82-70}{5}\right)$$

$$= P(Z \geq 2.4) = 1 - P(Z < 2.4) = 1 - F(2.4) = 1 - 0.9918 = 0.0082 = 0.82\ \%$$

(c) Let x_1 denote the lowest pass score then $P(X \geq x_1) = 0.8$

$$\Rightarrow \quad P\left(\frac{X-m}{s} \geq \frac{x_1-70}{5}\right) = 0.8$$

$$\Rightarrow \quad P\left(Z \geq \frac{x_1-70}{5}\right) = 0.8 \Rightarrow 1 - P\left(Z < \frac{x_1-70}{5}\right) = 0.8 \Rightarrow F\left(\frac{x_1-70}{5}\right) = 0.2$$

Also form the Z-table we get, $F(-0.84) = 0.2$

$$\because \quad \frac{x_1-70}{5} = -0.84 \quad \Rightarrow x_1 = 70 - 5 \times 0.84 = 65.8$$

(d) Let x_2 denote the score so that only 15% of the students get a score higher then this.

$$\because \quad P(X \geq x_2) = 0.15$$

$$\Rightarrow P\left(Z \geq \frac{x_2-70}{5}\right) = 0.15$$

$$\Rightarrow 1 - F\left(\frac{x_2-70}{5}\right) = 0.15$$

$$\Rightarrow F\left(\frac{x_2-70}{5}\right) = 0.85$$

Also from the z table we see that

$F(1.04) = 0.85$

$$\because \quad \frac{x_2-70}{5} = 1.04 \qquad \Rightarrow x_2 = 70 + 5 \times 1.04 = 75.2$$

Scanned by CamScanner

Exam Asked Long Question

1. **In a certain examination test 2000 students appeared in statistics. The average deviation was 5%. How many students do you expect to obtain more than 60% marks? What is the minimum marks of the top 100 students? Assume that the marks are normally distributed.** [PU 2016 Spring]

Solution 1

Do yourself

2. **The length of time (in seconds) that a user views a page on website before moving to another page is log normal random variate with parameters $\mu = 0.5$ and $\sigma^2 = 1$.**

(i) What is the probability that a page is viewed for more than 10 seconds?

(ii) What is the length of time 50% of user view the page?

(iii) What is the mean and S.D of the time until a user moves from the page?

[PU 2016 Spring]

Solution 2

(Same as Ex.11) Q.No. 37.

3. **The magnitude of earth quakes recorded in Sindupalchowk of Nepal during the year 2072 can be modeled as has an exponential distribution with mean 4.5, as measured on the Rector scale. Suppose that the magnitude of the earthquake striking that region has a gamma distribution with $\alpha = 1.5$ and $\beta = 3$.**

(i) What is the mean magnitude of earthquakes striking that region?

(ii) What is the probability that the magnitude an earthquake striking that region will exceed 5.0 on the Rector scale?

(iii) What is the probability that an earthquake striking the region will Fall between 5.0 and 6.0? [PU 2016 Fall]

Solution 3

Let X-denote the magnitude of the earth-quakes striking the region is a random variable having Gamma distribution with $\alpha = 1.5$, $\beta=3$.

Then,

i The mean magnitude of earthquakes striking the region $E(X) = \alpha\times\beta = 1.5\times3 = 4.5$

ii The probability that the magnitude an earthquakes striking the region will exceed 5.0 on the rector scale is

$$P(X>5.0)= 1- P(X<5.0) = 1-\int_0^5 f(x)dx$$

$$= 1-\int_0^5 \frac{1}{\lambda}e^{-\lambda x}dx \quad \text{where, } \lambda=\frac{1}{\beta}=\frac{1}{3} \quad =1-\int_0^5 \frac{1}{3}e^{-x/3}dx = 1+\frac{1}{3}\times\frac{3}{1}[e^{-5/3}-e^{-0/3}] = 0.188$$

iii The probability that the earth quake striking the region fall between 5.0 to 6.0 is given by

Scanned by CamScanner

$$P(5.0 \le X \le 6.0) = \int_{5.0}^{6.0} f(x)dx = \int_{5}^{6} \frac{1}{3} e^{-x/3} dx = -[e^{-6/3} - e^{-5/3}] = 0.188 - 0.135 = 0.053$$

4. **The lifetime of semiconductor laser has a log- normal distribution with μ = hrs and σ = 1.5 hrs.**

   (i) What is the probability that the lifetime exceeds 10,000 hrs?

   (ii) What lifetime is exceeded by 99% of the laser? [PU 2016 Fa

Solution 4

Let X-denote the life time of semi-conductor laser is a random variable havin log-normal distribution with parameter μ=10 and σ=1.5 (in hrs).

Then, $Y = \ln(X) - N(10, 1.5^2) \Rightarrow X = e^Y - N(10, 1.5^2)$

i The probability that the life time exceeds 10000 hours is given by.

$P(X>10000) = P(e^Y>10000) = P(\ln(e^Y) > \ln 10000)$

$$= P\left[\frac{P(Y-\mu)}{6} > \frac{\ln(10000)-10}{1.5}\right] = P(Z > -0.53) = 1 - P(Z \le -0.53) = 0.6955$$

ii Let X denote the life time which is exceed by 99% of laser, then,

$P(X > x) = 0.99$

or, $P(e^Y > x) = 0.99$

or, $P(Y > \ln x) = 0.99$

or, $P\left(\frac{Y-\mu}{6} > \frac{\ln x - 10}{1.5}\right) = 0.99$

or, $P\left(Z > \frac{\ln x - 10}{1.5}\right) = 0.99$

or, $1 - P\left(Z < \frac{\ln x - 10}{1.5}\right) = 0.99$

or, $P\left(Z > \frac{\ln x - 10}{1.5}\right) = 0.01$

or, $\frac{\ln x - 10}{1.5} = -2.33$ $[\because P(Z \le -2.33) = 0.11]$

$\Rightarrow$ $\ln x = -2.33 \times 1.5 + 10 = 6.505$

$\therefore$ $x = 668.48$ hours

5 **Unisys. com is one of the most frequently business to business websites; assume that the length of a visit on the unisys website is distributed as a normal random variable with a mean of 65.7 minutes and standard deviation of 15 minutes.** [PU 2015 spring]

(i) What is the probability that the randomly visited it will last more than 90 minutes?

(ii) Only 20% of the visits will last less than how many minutes?

Scanned by CamScanner

Solution 5

Let X denote the length of time (in minute) of a visit on the Unisys Website, is a random variable having normal distribution with parameters

$\mu = 65.7$ and $\sigma = 15$

i. Probability that randomly visited will last more than 90 minutes.

i.e. $P(X > 90) = P\left(\frac{X-\mu}{\sigma} > \frac{90-65.7}{15}\right)$

$= P(Z > 1.62) = 1 - P(Z \le 1.62) = 1 - 0.9474 = 0.0526$

ii. Let X denote minute which the 20% of the person visited the website then,

$P(X \le x) = 0.2$

$P\left(\frac{X-\mu}{\sigma} \le \frac{x-65.7}{15}\right) = 0.2$

$P\left(Z \le \frac{x-65.7}{15}\right) = -0.84$ $[\because P(Z \le -0.84) = 0.2]$

$\therefore \frac{x-65.7}{15} = -0.84$

$\Rightarrow x = -0.84 \times 15 + 65.7 = 53.1$ min

6. **In a certain city, the daily consumption of electric power (in millions of kilowatt - hours) can be treated as random variable having a gamma distribution with $\alpha = 3$ and $\beta = 2$. If the power plant of this city has a daily capacity of 12 millions kilowatt hours. What is the mean of this gamma distribution? What is the probability that this power supply will be inadequate on any given day?** **[PU 2013 Spring /2014 Spring/ 2013 Fall]**

Solution 6

Let X denote the daily consumption of electric power (in million of kilowatt-hours) is a random variable having gamma distribution with parameters

$\alpha = 3$ and $\beta = 2$

Then the mean of the gamma distribution is given by

$E(X) = \alpha \times \beta = 3 \times 2 = 6$

Also we know that the probability density function of gamma function is given by $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} = \frac{1}{2^3 \Gamma(3)} x^{3-1} e^{-x/2} = \frac{1}{16} x^2 e^{-x/2}$

Now, the probability that the power supply will be inadequate

$$P(X > 12) = 1 - P(X \le 12) = 1 - \int_0^{12} f(x)\,dx$$

$$= 1 - \frac{1}{16}\int_0^{12} x^2 e^{-x/2}\,dx = 1 - 0.053 = 0.995$$

Scanned by CamScanner

7 In a certain examination test 2000 students appeared in statistics. The full marks of the exam is 100. The average marks obtain was 50 and the standard deviation was 5? How many students do you expect to obtain more than 60 marks? What is the minimum marks of the top 10% students? Assume that the marks are normally distributed. [PU 2012 Fall]

Solution 7

Let X denote he marks obtained by the students, is a random variable also normally distributed with parameters $\mu = 50$ and $\sigma = 5$. Also, $N = 2000$. Now the number of students having marks more than 60

i.e $NP(X > 60) = NP\left(\frac{X-\mu}{\sigma} > \frac{60-50}{5}\right)$

$= NP(Z > 2) = N(1 - P(Z \geq 2)) = N(1 - 0.9772) = 2000 \times 0.0228 = 46$

Let x_1 denote the minimum marks of 10% students.

i.e $P(X > x_1) = \frac{10}{100}$

$$\Rightarrow P\left(\frac{X-\mu}{\sigma} > \frac{x_1 - 50}{5}\right) = 0.1$$

$$\Rightarrow P\left(Z > \frac{x_1 - 50}{5}\right) = 0.1$$

$$\Rightarrow P\left(Z \leq \frac{x_1 - 50}{5}\right) = 0.99$$

$$P(Z \leq 2.33) = 0.99$$

$$P\left(Z \leq \frac{x_1 - 50}{5}\right) = P(Z \leq 2.33)$$

$$\Rightarrow \frac{x_1 - 50}{5} = 2.33$$

$$\Rightarrow x_1 = 5 \times 2.33 + 50 = 61.65$$

9. A minimum height is to be prescribed for eligibility to government services such that 60% of the young men should have a fair chance of coming up to the standard. The height of the young man are normally distributed with mean σ.' and S.d. 2.55" Determine the minimum specification. [PU 2011 Spring]

Solution 9

Let X denote the height of the young man is a random variable also normally distributed with parameter $\mu = 6.6$ and $\sigma = 2.55$

Let x_1 denote the minimum specification. Then according to question

$P(X > x_1) = 60\%$

$$\Rightarrow P\left(\frac{X-\mu}{\sigma} > \frac{x_1 - 6.6}{2.55}\right) = 0.6$$

$$\Rightarrow P\left(Z \leq \frac{x_1 - 6.6}{2.55}\right) = 0.4$$

Scanned by CamScanner

$$\Rightarrow P\left(Z \le \frac{x_1 - 6.6}{2.55}\right) = P(Z \le -0.25)$$

$$\Rightarrow \frac{x_1 - 6.6}{2.55} = -0.25$$

$$\Rightarrow x_1 = 6.6 - 0.25 \times 2.55 = 5.9625'.$$

Hence the minimum specification is $6'$(approximately)

10. **The internal examination scores of engineering students of certain college in numerical methods is normally distributed with mean 65 and standard deviation 10. Find the lowest score of top 20% students.** **[PU 2011 Fall]**

Solution 10

Process is similar to Q.7 or Q.9

11. **The average number of network errors experienced in 2 hours period is 6 what is the probability that in given hour :** **[PU 2010 Spring]**

(i) Exactly two network errors will occurs.

(ii) At least two network errors will occur.

Solution 11

Do yourself

12. **Incomes of a group of 1000 persons were found to be normally distributed with the mean of Rs. 520 and standard deviation Rs. 60 Find the number of persons whose income is lies in between Rs. 160 to Rs. 640.** **[PU 2010 Spring]**

Solution 12

Let X denote the income of a group of people is a random variable which is normally distributed with parameters $\mu = 520$ and $\sigma = 60$. Also, $N = 1000$. Now the number of persons whose income lies between 460 to 640.

$= N \times P(460 \le X \le 640)$

$$= N \times P\left(\frac{460 - 520}{60} \le \frac{X - \mu}{\sigma} \le \frac{640 - 520}{60}\right)$$

$= N \times P(-1 \le Z \le 2)$

$= N[P(Z \le 2) - P(Z \le -1)] = N[0.9772 - 0.1587] = 1000 \times 0.8185 = 818.5 \approx 819$

13. **A variate x has the uniform distribution with pdf given by**

$$\begin{cases} 1/100 & \text{for } 0<x<200 \\ 0 & \text{otherwise} \end{cases}$$

Compute (i) $P(X>160)$ (ii) $P(20 \le X \le 40)$

(iii) $E(X)$ **[PU 2007 Fall]**

Solution 13

Here, we have given $f(x) = \begin{cases} 1/100 & 0 < x < 200 \\ 0 & \text{otherwise} \end{cases}$

Then

Scanned by CamScanner

i. $P(X \geq 160) = \int_{160}^{200} f(x)\,dx = \int_{160}^{200} \frac{1}{100}\,dx = \frac{1}{100}$

14. At a certain examination 10% of the students who appears for the paper of statistics got less than 30 marks and 97% of the students got less than 80 marks. Assuming that the distribution of marks is Normal. find

i. Mean and Variance of marks

ii. Find the probability that randomly selected students get more than 70 marks. PU 2013 Fall

Solution 14

Same as Ex-11 Q.No. 29

Remaining part

The probability that the randomly selected students get more than 70 marks is given by $P(X > 70) = \frac{P(X-\mu)}{\sigma} > 70$

15. Of a large group of men, 5% are under 60 inches in height and 10% are 72 inches in height. If height of men is normally distributed find the mean height and standard deviation. PU 2013 Spring

Solution 15

Process is similar to Q.No. 14

16. Suppose the time taken for a data collection operator to fill out an electronic form for a data base is uniformly distributed between 1.5 and 2.2 minutes. Then

i. What is the mean and variance of the time it takes to operator to fill out the form?

ii. What is the probability it takes less than two minutes to fill out the form?

iii. Determine the cumulative frequency distribution function of the time it takes to fill out thee form? PU 2014 Spring

Solution 16

See in Ex.11 Q.No. 7

17. In an examination, 30% of the candidate obtained marks below 40 and 10% of candidates got above 75 marks. Assuming the marks are normally distributed find the mean and standard deviation of the distribution? PU 2014 Spring

Solution 17

Let X denotes the marks obtained by the students is a random variable which is normally distributed. let Z denote the standard normal variable corresponding to X.

Since 10% students got less than 30 marks. Then,

$$P(X \leq 30) = 0.1 \quad \Rightarrow P\left(\frac{X-\mu}{\sigma} \leq \frac{30-\mu}{\sigma}\right) = 0.1$$

$$\Rightarrow P\left(Z \leq \frac{30-\mu}{\sigma}\right) = P(Z \leq -1.28)$$

Scanned by CamScanner

$$\Rightarrow \frac{30-\mu}{\sigma} = 1.28$$

$$\Rightarrow \mu - 1.286 = 30 \qquad \ldots(i)$$

Also, 97% students got less than 80 marks, then

$$P(X \le 80) = 0.97 \quad \Rightarrow P\left(\frac{X-\mu}{} \le \frac{80-\mu}{\sigma}\right) = 0.1$$

$$\Rightarrow P\left(Z \le \frac{80-\mu}{\sigma}\right) = P(Z \le 1.88)$$

$$\Rightarrow \frac{80-\mu}{\sigma} = 1.88$$

$$\Rightarrow \mu + 1.886 = 80 \qquad \ldots(ii)$$

Solving (i) and (ii), we get, mean (μ) = 59.63 and S.D. (σ) = 23.15

18. **In an examination 15% of the candidate got first class (60% marks or above) while 40% failed (Securing below 40%). Assuming the mark to be normally distributed, estimate the mean and standard deviation.** **PU 2016 Fall/2017 Spring**

Solution 18

Let X denote the marks obtained in an examination is a random variable which is normally distributed. Let Z be the standard normal variable. Since 15% students got 60 marks or above. Then,

$$P(X \ge 60) = 0.15 \quad \Rightarrow P\left(\frac{X-\mu}{} \le \frac{60-\mu}{\sigma}\right) = 0.15$$

$$\Rightarrow P\left(Z \ge \frac{60-\mu}{\sigma}\right) = 0.15$$

$$\Rightarrow P\left(Z < \frac{60-\mu}{\sigma}\right) = 0.85$$

$$\Rightarrow \left(Z < \frac{60-\mu}{\sigma}\right) = P(Z \le 1.05)$$

$$\therefore \quad \frac{60-\mu}{\sigma} = 1.05$$

$$\Rightarrow \quad \mu + 1.056 = 60 \qquad \ldots(i)$$

Also, 40% students got below 40 marks

i.e. $$P(X < 40) = 0.4 \quad \Rightarrow P\left(\frac{X-\mu}{} \le \frac{40-\mu}{\sigma}\right) = 0.4$$

$$\Rightarrow P\left(Z < \frac{40-\mu}{\sigma}\right) = P(Z \le -0.25)$$

$$\Rightarrow \frac{40-\mu}{\sigma} = 0.25$$

$$\Rightarrow \mu - 0.256 = 40 \qquad \ldots(ii)$$

Solving (i) and (ii) we get, $\mu = 46.845$, $\sigma = 15.38$

Scanned by CamScanner

19. The time (in hours) required to repair a machine is exponentially distributed with parameter $\lambda = 1/3$. What is the probability that the repair time exceeds 3 hours? Also find the probability that machine will required within 3 hours. PU 2015 Fall

Solution 19

Let X denote the time (in hours) to repair a machine is a random variable which is exponentially distributed with parameter $\lambda = 1.3$. Then, the probability density function is given by, $f(x) = \lambda e^{-\lambda x}$

The probability that the repair time exceeds 3 hours is

$$P(X > 3) = 1 - P(X \le 3) = 1 - \int_0^3 f(x)\,dx = 1 - \int_0^3 \frac{1}{3} e^{-1/3x}\,dx$$

$$= 1 - \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_0^3 = 1 + (e^{-1} - e^0) = 0.368$$

The probability that machine will repair with in 3 hour is

$$P(X \le 3) = \int_0^3 f(x)\,dx = \int_0^3 \frac{1}{3} e^{-x/3}\,dx$$

$$= \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_0^3 = -[e^{-1} - e^0] = 1 - 0.368 = 0.632$$

20. Batteries last an average of 60 days with standard deviation of 10 days. Sixty batteries are bought .what is the probability that i) Mean life of all 60 batteries exceeds 61 days ii) mean life differs from60 days more than 3 days. PU 2015 Fall

Solution 20

Let X denote the life of batteries (in days) is a random variable which is normally distributed with parameter $\mu = 60$ and $\sigma = 10$.

i. The probability that the batteries exceeds 61 days is,

$$P(X > 61) = P\left(\frac{X - \mu}{\sigma} > \frac{61 - 60}{10}\right)$$

$$= P(Z > 0.1) = 1 - P(Z \le 0.1) = 1 - 0.5398 = 0.4602$$

ii. The probability that the batteries life differs from more than 3 days is given by

$$P(X < 57) = P\left(\frac{X - \mu}{\sigma} > \frac{57 - 60}{10}\right) P(Z < -0.3) = 0.3821$$

21. Harly Davidson director of quality control for the Kyoto Motor company is conducting his monthly spot check of automatic transmission. In this procedure, 10, transmissions are removed from the pool of components and are checked for manufacturing defects. Historically. Only 2% of the transmissions have such flaws.(Assume that flaws occur independently in different transmission)

i. What is the probability that Harley's sample contains more than two transmissions with manufacturing flaws?

ii. What is the probability that none of the selected transmission has any manufacturing flaws? **PU 2015 Spring**

Solution 21

Do yourself

Scanned by CamScanner

22. **Busses arrive at a specified stop at 15-minute intervals starting at 7 AM. That is, they arrive at 7, 7:15, 7:30 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30 find the probability that he waits i) less than 5 minutes for a bus ii) at least 12 minutes for a bus.**

PU 2015 Spring

Solution 22

Let X denote thee time (in minutes) past 7 a.m. that the passenger arrive at the stop. Since, X is a uniform random variable over the interval [0, 30), (i) it follows that the passenger will have to wait less than 5 minutes if he arrives between 7 : 10 and 7 : 15 or between 7:25 and 7:30. Hence, the desired probability is,

$$P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

ii. Similarly, he would have to wait at least 12 minutes if he arrives between 7 and 7:03 or between 7:15 and 7:18 and so that the probability is

$$P(0 < X < 3) + P(5 < X < 18) = \int_0^3 \frac{1}{30}\,dx + \int_{15}^{18} \frac{1}{30-0}\,dx$$

$$= \frac{3}{30} + \frac{18-15}{30} = \frac{1}{5}$$

23. **A sample of 600 mobiles of a certain Brand is tested to find the length of battery. After each full charge, produced result of average 12 hours and standard deviation of 4.5 hours. Assuming the data to be normally distributed, find the number of mobiles that are expected to have life after each full charges:**

i. More than 15 hours ii) between 10 and 14 hours

iii. top 10% performer (on the basis of battery life) PU 2017 Fall

Solution 23

Let X denote the life of battery after each full charges is a random variable which is normally distributed with parameter $\mu = 12$ and $\sigma = 4.5$. Also $N = 600$

i. The expected number of mobiles to have life more than 15 hours is $N \times P(X > 15)$,

where, $P(X > 15) = \left(\frac{X-\mu}{\sigma} > \frac{15-12}{4.5}\right) = P(Z > 0.67) = 1 - P(Z \le 0.67)$

$= 1 - 0.7486 = 0.2514$

$\therefore$ $N \times P(X > 15) = 600 \times 0.2514 = 150.84 \approx 151$

ii. The expected number of mobiles having life between 10 and 14 hours is

$N\,P(10 \le X \le 14)$

$= 60 \times P\left(\frac{10-12}{45} \le \frac{X-\mu}{\sigma} \le \frac{14-12}{4.5}\right)$

$= 600 \times P(-0.44 \le Z \le 0.44)$

$= 600\,[P(Z \le 0.44) - P(Z \le -0.44)]$

$= 600[0.6700 - 0.300] = 204$

Scanned by CamScanner

iii. The expected number of batteries of top 10% performed

$P(X \geq x_1) = 0.1$

or, $P\left(\frac{X-\mu}{\sigma} > \frac{x_1 - 12}{4.5}\right) = 0.1$

or, $P\left(Z \geq \frac{x_1 - 12}{4.5}\right) = 0.1$

or, $\left(Z \leq \frac{x_1 - 12}{4.5}\right) = 0.9$

or, $\left(Z \leq \frac{x_1 - 12}{4.5}\right) = P(Z \leq 1.28)$

$\Rightarrow \frac{x_1 - 12}{4.5} = 1.28 \Rightarrow x_1 = 4.5 \times 1.28 + 12 = 17.76 = 18$

24. In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and S. D. of the distribution. PU 2010 Fall

Solution 24

Process is similar to Q.No. 18

25. A set of final examination grades in an introductory statistics course was found to be normally distributed with mean of 73 and a standard deviation of 8.

i. What is the probability of getting a grade less than 91 on this exam?

ii. Only 5% of the students taking the test scored higher than what grade?

PU 2007 Fall

Solution 25

Process is similar to Q.No. 23

26. In a particular College in Nepal it has been found that 20% of the students withdraw without completing the Engineering course, assume that 15 students have register in a new semester what is the probability that:

i) none will withdraw the course ii) at most two will withdraw the course

iii) find the mean of this distribution. PU2010 Spring

Solution 26

Do yourself

27. The average number of network errors experienced in two hours period is 6 What is the probability that in any given hour: i) exactly two network errors will occur ii) at least two network error will occur PU 2010 Spring

Solution

Process is similar to Q.No. 11

28. Incomes of a group of 1000 persons were found to be normally distributed with mean Rs. 520 and standard deviation Rs. 60 Find the number of persons whose income lies between Rs. 460 to Rs. 640. PU 2010 Spring

Solution

Process is similar to Q.NO. 11

Scanned by CamScanner

29. **The distribution of the wages of the employees in a firm was perfectly normal and the number of workers employed in the organization was 4000. The mean and standard deviation were calculated as Rs. 800 and Rs. 200 respectively. On the basis of the information estimate i) The number of workers getting salary between Rs. 700 and Rs. 900. ii)% of workers getting salary between Rs. 700 and 900. iii) % of workers getting salary above Rs. 1000.** **PU 2005 Fall**

Solution 29

Process is similar to Q.No. 11

30. **Mr. Zenith is a computer engineer in a Microsoft company. He has found that on an average 3% of the produced screens of laptop computer monitors are defective. A sample of one dozen monitors from a production lot was taken at random. What is the probability that in this sample fewer than 2 defectives screens will be found? Also calculate the expected number of defective screens in a random sample of 12 monitors.** **PU 2010 Spring**

Solution 30

This question is related to chapter 10

31. **Between 1 PM and 4 PM, the average number of phone calls per minute coming in to the switch board of a company is 2.5. Find the probability that during 2:30 and 2:32 PM, there will be**

i) No call at all. ii) At least 5 calls. PU 2010 Spring

Solution 31

Process is similar to Q.No. 22

32. **As a result of tests on 20000 electric bulbs manufactured by a company it was found that the life time of the bulb was normally distributed with an average of 2040 hours and the standard deviation of 60 hours. On the basis of the information, estimate the number of bulbs that are expected to burn for less than 1960 hours.** **PU 2010 Spring**

Solution 32

Process is similar as above

33. **Assuming the height of soldier is normally distributed with mean 68.52 inches and variance 10.8 sq. Inches. How many soldiers in a regiment of 1000 would you expect to be over six feet tall?** **PU 2009 Spring**

Solution 33

See in Ex-11 Q.No. 19

34. **The mean yield for one-acre plot is 662 kilos with standard deviation 32 kilos. Assuming normal distribution, how many one acre plots in batch of 1000 plots would you expect to have yield.**

i. over 700 kilos ii) below 650 kilos PU 2009 Fall

Solution 34

Process is similar to Q.No. 33

Scanned by CamScanner

35. In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution. PU 2006 Spring

Solution

Similar as Q.No. 24

36. Of a large group of men, 5% are under 60 inches in height and 40% are between 60 and 65 inches. Assuming a normal distribution find the mean height and standard deviation. *PU 2003 Fall*

Solution 36

Process is similar to Q.No. 33

Scanned by CamScanner

12 Chapter

Joint Probability Distribution

Joint Probability Distribution: Let X and Y be two discrete random variables occurs simultaneously and X associates with m values $x_1, x_2, x_3, \ldots, x_n$ and Y associates with n values $y_1, y_2, y_3, \ldots, y_n$. The probability of order pair $(x_i\ y_i)$, $i = 1, 2, 3, \ldots, m$; $j = 1, 2, 3, \ldots, n$ is denoted by

$$P_{ij} = P(x_i, y_j) = P(X = x_i, Y = y_j) = P(x = x_i \cap Y = y_j)$$

is known as joint probability distribution of discrete random variables X and Y.

Joint probability mass function or discrete density function

Let X and Y be two discrete random variables defined on the sample space S, the joint probability distribution (Px_i, y_j), $i = 1, 2, 3, \ldots, m$; $j = 1, 2, 3, \ldots, n$ is called the joint probability mass function if it satisfies the following conditions:

i. $P(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in S$

ii. $\sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) = \sum\sum_{(x_i, y_j) \in S} P(x_i, y_j) = 1$

Marginal Probability mass function

If $P(x_i, y_j) = P(X = x_i, Y = y_j)$ is the joint probability mass function of discrete random variables X and Y then the marginal probability mass function of X is denoted and defined as,

$$P_i = P_X(x_i) = P(x = x_i) = \sum_{j=1}^{n} P(x_i, y_j) = P(x_i, y_1) + P(x_i, y_2) + \ldots + P(x_i, y_n)$$

and $P_X(x_i)$ is called the marginal probability function of r.v. $X = x_i$

Similarly, the marginal probability function of r.v. $Y = y_j$ is denoted and defined by

$$P_j = P_Y(y_j) = P(Y = y_j) = \sum_{j=1}^{m} P(x_i, y_j) = P(x_1, y_j) + P(x_2, y_j) + \ldots + P(x_m, y_j)$$

Scanned by CamScanner

Note:

1. If $P(X = x_i)$ is the probability mass function then we must have $P(X = x_i) \geq 0$, for each i and $\sum_{i=1}^{m} P(X = x_i) = 1$
2. If $P(Y = y_j)$ is probability mass function then we must have $P(Y = y_j) \geq 0$ for each j and $\sum_{j=1}^{m} P(Y - y_j) = 1$.

Conditional probability mass function

If $P(x_i, y_j) = P(X = x_i, Y = y_j)$, $i = 1, 2, 3, \ldots, m$; $j = 1, 2, \ldots, n$ is joint probability mass function of discrete random variables X and Y, then the conditional probability mass function of X given $Y = y_j$ is defined as

$$P(X = x_i / Y = y_j) = \frac{P(x_i, y_j)}{P(y_j)}$$

Also, the conditional probability mass function Y given $X = x_i$ is defined as

$$P(Y = y_j / X = x_i) = \frac{P(x_i, y_j)}{p(x_i)}$$

Independence

Two random variable X and Y are said to be independent if

$$P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \text{ for all } i, j$$

Joint probability distribution function

Let $P(x_i, y_j)$, $i = 1, 2, 3, \ldots, m$; $j = 1, 2, \ldots, n$ be joint pmf of discrete random variables X and Y. The joint probability distribution function of two dimensional random variables X and Y is denoted and defined by

$$F_{XY}(x_s, y_t)) = P(X \leq x_s, Y \leq y_t) = \sum_{i=1}^{s} \sum_{j=1}^{t} P(x_i, y_j)$$

Mean and Variance from joint distribution

1. Mean of X $[E(X)] = \sum_{i=1}^{m} x_i P(x_i) = x_1P(x_1) + x_2P(x_2) + \ldots + x_m (Px_m)$
2. Mean of Y $[E(Y)] = \sum_{i=1}^{n} y_i P(y_i) = y_1P(y_1) + y_2P(y_2) + \ldots + y_n (Py_n)$
3. Variance of X $V(X) = E(X^2) - [E(X)]^2 = \sum_{i=1}^{m} x_i^2 P(x_i) - \left(\sum_{i=1}^{m} x_i P(x_i)\right)^2$
4. $E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i \times y_j) P(x_i, y_j)$
5. $Cov(X, Y) = E[X - E(X)] [Y - E(Y)] = E(XY) - E(X) (EY)$
6. $Var(X + Y) = var (X) + Var (Y) + 2Cov(X, Y)$
7. $Var(aX - bY) = a^2Var(X) + b^2Var(Y) - 2ab\, Cov(X, Y)$
8. $E[aX + bY] = aE(X) + bE(Y)$

Note: If X and Y are independent then $E(XY) = E(X) E(Y)$ i.e. $Cov(XY) = 0$

Scanned by CamScanner

Joint Probability Density Function (Joint p.d.f)

The joint probability function of two continuous random variables X and Y i.e. f(x,y) is said to be joint probability density function if it satisfies:

i. $f(x, y) \geq 0$ for all values of x and y

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

Note: If X and Y are two continuous random variables and the function f(x, y) is called the joint probability density function, if the probability that $a_1 \leq X \leq b_1$, $a_2 \leq Y \leq b_2$ is given by the integral $\int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x, y)\, dx\, dy$ such that:

i. $f(x, y) \geq 0$ for all values of x, y

ii. $\int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x, y)\, dx\, dy = 1$

Marginal Probability Density Function (Marginal p.d.f)

If f(x, y) is joint pdf of continuous random variables X and Y then the marginal probability density function of X is defined as

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad \text{for } -\infty < y < \infty$$

Similarly, the marginal probability density function of Y is defined as

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx, \qquad \text{for } -\infty < x < \infty$$

Conditional Probability Density Function

If f(x, y) is joint pdf of continuous random variables X and Y then the conditional probability density function of X given Y = y is defined as

$$F_{X/Y}(x/y) = \frac{f(x, y)}{f(y)}, \qquad \text{for } x, y \in R$$

Similarly, the conditional probability density function of Y given X = x is defined as

$$F_{Y/X}(y/x) = \frac{f(x, y)}{f(x)}, \qquad \text{for } x, y \in R$$

Joint Probability Distribution Function

If f(x, y) is joint pdf of continuous random variables X and Y then the joint probability distribution of X and Y is denoted and defined as

$$F_{XY}(x/y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\left\{\int_{-\infty}^{y} f(x, y)\, dy\right\} dx$$

Scanned by CamScanner

Marginal Probability Distribution Function

If f(x, y) is joint pdf of continuous random variables X and Y then the marginal probability distribution function of X is given by

$$f_X(x) = P(X \le x) = P(X \le X, Y < \infty) = \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f(x, y)\, dy \right\} dx$$

Similarly, marginal probability distribution function of Y is given by

$$f_Y(y) = P(Y \le y) = P(y < \infty, Y \le y) = \int_{-\infty}^{y} \left\{ \int_{-\infty}^{\infty} f(x, y)\, dx \right\} dy$$

Independence

Two continuous random variables X and Y are said to be independent if and only if their joint probability density equals the product of the corresponding values of the marginal densities.

i.e. $f(x, y) = f_X(x)\, f_Y(y)$

Mean and Variance

1. Mean of X $[E(X)] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$

2. Mean of Y $[E(Y)] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$

3. $E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x,y)\, dxdy$

4. $Var(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\int_{-\infty}^{\infty} x\, f_X(x)\, dx \right)^2$

5. $Var(Y) = E(Y^2) - [E(Y)]^2 = \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy - \left(\int_{-\infty}^{\infty} f_Y(y)\, dy \right)^2$

6. $Cov(X, Y) = E(XY) - E(X)\, E(Y)$

Exercise-12

1. **Explain the concepts (i) Joint probability function (ii) Conditional probability (iii) Independent random variables (iv) Marginal Conditional probability and probability distribution (v) Joint probability density function (vi) Joint probability distribution function.**

Solution 1

Please see in theoretical part of this chapter

Scanned by CamScanner

2. **What is the difference between joint probability mass function and joint probability density function?**

Solution 2

Please see in theoretical part of this chapter

3. **If F(x, y) be joint distribution function of continuous random variables X and Y, what will be the distribution function for the marginal distribution of X and Y?**

Solution 3

Please see in theoretical part of this chapter

4. **What is meant by conditional distribution of Y under the condition that X = x? Consider separately the cases of continuous and discrete random variables.**

Solution 4

Please see in theoretical part of this chapter

5. **What is meant by central limit theorem? Write down the case of C.L.T.**

Solution 5

Please see in theoretical part of this chapter

6. **A coin and a dice are thrown simultaneously. Let X denote the number (Head = 1, tail - 0) on a coin and Y denote the number on the dice. Find the following probability.**

i. $P(X + Y = 5)$ ii. $P(X + Y = 5/X = 0)$ iii. $P(X = Y)$

Solution 6

Let X denote the number in a tossing a coin (i.e Head = 1, Tail = 0) $\therefore X = \{0,1\}$.
Also Y denote the number in a dice.
i.e. $Y = \{1, 2, 3, 4, 5, 6\}$
$(X,Y) = \{(0,1), (0,2), (0,3), (0,4), (0,5), (0,6), (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$

(i) $P(X+Y = 5) = P(0,5) + P(1,4) = \frac{1}{12} + \frac{1}{12} = \frac{2}{12} = \frac{1}{6}$

(ii) $P(X+Y = 5/X = 0) = \frac{P\{(X+Y=5) \cap (X = 0)\}}{P(X = 0)} = \frac{1/12}{1/2} = \frac{2}{12} = \frac{1}{6}$

(iii) $P(X = Y) = P(1,1) = \frac{1}{12}$

7. **The joint probability distribution of a pair (X,Y) of random variables is given by following table.**

y \ x	1	2	3
1	0.1/1.6	0.2/1.6	0.2/1.6
2	0.2/1.6	0.3/1.6	0.1/1.6
3	0.3/1.6	0.1/1.6	0.1/1.6

Find: i. the marginal probability distribution of X

ii. the conditional probability distribution of X given Y = 1

iii. E(X), E(Y), V(X) and covariance between X and Y

iv. F(1, 3), F(2, 2) and F(3, 3)

Scanned by CamScanner

Solution 7

(i) The marginal probabilities of X are

$$P(X=1) = P(1,1) + P(1,2) + P(1,3) = \frac{0.1}{1.6} + \frac{0.2}{1.6} + \frac{0.3}{1.6} = \frac{0.6}{1.6} = 0.375$$

$$P(X=2) = P(2,1) + P(2,2) + P(2,3) = \frac{0.2}{1.6} + \frac{0.3}{1.6} + \frac{0.1}{1.6} = \frac{0.6}{1.6} = 0.375$$

$$P(X=3) = P(3,1) + P(3,2) + P(3,3) = \frac{0.2}{1.6} + \frac{0.1}{1.6} + \frac{0.1}{1.6} = \frac{0.4}{1.6} = 0.25$$

Therefore the marginal distribution of X is.

$X = x_i$	1	2	3	Total
$P(X = x_i)$	0.375	0.375	0.25	1

(ii) The conditional probability distribution of X given Y = 1 is denoted by

$$P(X = x_i / Y=1) = \frac{P(x_i,1)}{P(Y=1)}, i = 1, 2, 3$$

Therefore $P(X = 1 / Y = 1) = \frac{P(1,1)}{P(Y=1)} = \frac{0.1/1.6}{(0.1 + 0.2 + 0.2) / 1.6} = \frac{0.1}{0.5} = 0.2$

$$P(X = 2 / Y = 1) = \frac{P(2,1)}{P(Y=1)} = \frac{0.2/1.6}{0.5/1.6} = 0.4$$

$$P(X = 3) = \frac{P(3,1)}{P(Y=1)} = \frac{0.2/1.6}{0.5/1.6} = 0.4$$

(iii) We know that

(a) $E(X) = \sum_{i=1}^{3} (x_i)$

$= 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3)$

$= 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.25 = 1.875$

(b) For the E(Y) We need the marginal distribution of Y which are given as follows:

$$P(Y=1) = P(1,1) + P(2,1) + P(3,1) = \frac{0.1}{1.6} + \frac{0.2}{1.6} + \frac{0.2}{1.6} = \frac{0.5}{1.6} = 0.3125$$

$$P(Y = 2) = P(1, 2) + (2, 2) + P(2, 3) = \frac{0.2}{1.6} + \frac{0.3}{1.6} + \frac{0.1}{1.6} = \frac{0.6}{1.6} = 0.375$$

$$P(Y=3) = P(1,3) + P(2,3) + P(3,3) = \frac{0.3}{1.6} + \frac{0.1}{1.6} + \frac{0.1}{1.6} = \frac{0.5}{1.6} = 0.3125$$

Hence the marginal distribution of Y is

$Y = Y_j$	1	2	3	Total
$P(Y = Y_j)$	0.3125	0.375	0.3125	1

$\because$ $E(Y) = \sum_{j=1}^{n} y_j P(y_j)$

$= 1 \times P(Y -1) 2 \times (Y = 2) + 3 \times P(Y = 3)$

$= 1 \times 0.3125 + 2 \times 0.375 + 3 \times 0.3125 = 2$

(c) $V(X) = E(X^2) - [E(X)]^2 = \sum_{i=1}^{3} x_i^2 P(x_i) = \mu^2$

$= 1^2 \times P(X = 1) + 2^2 \times P(X = 2) + 3^2 . P(X = 3) - (1.875)^2$

$= 1 \times 0.375 + 4 \times 0.375 + 9 \times 0.25 - (1.875)^2 = 0.61$

(d) The covariance between X and Y is given by

$Cov(X, Y) = E(XY) - E(X). E(Y)$

Where $E(X,Y) = \Sigma x_i y_j P(x_i, y_j)$

Scanned by CamScanner

$= (1 \times 1)\, P(1,1) + (1 \times 2)\, P(1,2) + (1 \times 3)\, P(1,3) + (2 \times 1)\, P(2,1) + (2 \times 2) P(2,2) + (2 \times 3)\, P(2,3) + (3 \times 1)\, P(3,1) + (3 \times 2)\, P(3,2) + (3 \times 3)\, P(3,3)$

$= 1 \times \frac{0.1}{1.6} + 2 \times \frac{0.2}{1.6} + 3 \times \frac{0.3}{1.6} + 2 + \frac{0.2}{1.6} + 4 \times \frac{0.3}{1.6} + 6 \times \frac{0.1}{1.6} + 3 \times \frac{0.1}{1.6} + 6 \times \frac{0.1}{1.6} + 9 \times \frac{0.1}{1.6}$

$= \frac{5.7}{1.6} = 3.5625$

$\therefore$ Cov $(X,Y) = 3.5625 - 1.875 \times 2 = -0.1875$

(iv) (a) $F(1,3) = P(X \le 1, Y \le 3)$

$= P(1,1) + (P(1,2) + P(1,3) = \frac{0.1}{1.6} + \frac{0.2}{1.6} + \frac{0.3}{1.6} = \frac{0.6}{1.6} = 0.375$

(b) $F(2,2) = P(X \le 2, Y \le 2)$

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

$= \frac{0.1}{1.6} + \frac{0.2}{1.6} + \frac{0.2}{1.6} + \frac{0.3}{1.6} = \frac{0.8}{1.6} = 0.5$

(c) $F(3,3) = P(X \le 3, Y \le 3)$

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

$= \frac{0.6}{1.6} + \frac{0.6}{1.6} + \frac{0.4}{1.6} = \frac{1.6}{1.6} = 1$

8. Two discrete random variables X and Y have

$P(X = 0, Y = 0) = 0.15$, $P(X = 0, Y = 1) = 0.05$

$P(X = 1, Y = 0) = 0.1$, $P(X = 1, Y = 1) = 0.3$

$P(X = 2. Y = 0) = 0.2$. $P(X = 2, Y = 1) = 0.2$

Examine whether X and Y are independent.

Solution 8

Here we have given

Y \ X	0	1	2
0	0.15	0.1	0.2
1	0.05	0.3	0.2

Now, $P(X = 0) = P(0,0) + P(0,1) = 0.15 + 0.05 = 0.2$

$P(Y=0) = P(0,0) + P(1,0) + P(2,0) = 0.15 + 0.1 + 0.2 = 0.35$

Also, $P(X = 0, Y = 0) = P(0,0) = 0.15$

Since $P(X = 0, Y = 0) \neq P(X = 0) . P(Y = 0)$

Hence the discrete random variable are dependent.

9. **The joint probability distribution of the number X of cards and the number Y of buses per signal cycle of a proposed left turn lane is displayed in the accompanying joint probability table,**

P(x, y)		Y = y		
		0	1	2
X = X	0	0.025	0.015	0.010
	1	0.050	0.030	0.020
	2	0.125	0.075	0.050
	3	0.150	0.090	0.060
	4	0.10	0.060	0.040
	5	0.050	0.030	0.020

Scanned by CamScanner

i. What is the probability that there is exactly one car and exactly one bus during the cycle?

ii. What is the probability that there is at most one car and at most one bus during a cycle?

iii. What is the probability that there is exactly one car during a cycle? Exactly one bus?

iv. Suppose that left turn lane is to have a capacity of five cards, and one bus is equivalent to three cars. What is the probability of an overflow during a cycle? [TU 2058/2059/2062/2063/2064]

Solution 9

Let X – denote the number of cars and Y – denote the number of busses per signal cycle of a proposed left turn lane. Then

(i) The probability that there is exactly one car and exactly one bus during the cycle is given by.

$P(X = 1 \text{ and } Y = 1) = P(1, 1) = 0.030$

(ii) The probability that there is at most one car and at most one bus during the cycle is given by $P(X \le 1, Y \le 1)$

$= F(1,1) = P(0,0) + P(0,1) + P(1,0) + P(1,1)$

$= 0.025 + 0.015 + 0.050 + 0.030 = 0.120$

(iii) The probability of exactly one car is given by

$P(X = 1) = P(1,0) + P(1,1) + P(1,2) = 0.050 + 0.030 + 0.020 = 0.100$

Similarly the probability of exactly one bus is given by

$P(Y = 1) = P(0,1) + P(1,1) + P(2,1) + P(3,1) + P(4,1) + P(5,1)$

$= 0.015 + 0.030 + 0.075 + 0.090 + 0.060 + 0.030 = 0.300$

(iv) Since the left turn lane is to have a capacity of five cars and one bus is equivalent to three cars. The probability of an over flow during a cycle is given by

$P[(X \ge 3, Y = 1) \text{ or } [Y \ge 0, Y = 2)]$

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

$= P(3,1) + P(4,1) + P(5,1) + P(0,2) + P(1,2) + P(2,2) + P(3,2) + P(4,2) + P(5,2)$

$= 0.090 + 0.060 + 0.030 + 0.010 + 0.020 + 0.050 + 0.060 + 0.040 + 0.020 = 0.380$

10. Give the following bivariates probability distribution.

Y \ X	0	1
0	0.28	0.37
1	0.22	0.13

Solution 10

The marginal distribution of X is as follows:

$P(X = 0) = P(0,0) + P(0,1) = 0.28 + 0.22 = 0.5$

$P(X = 1) = P(1,0) + P(1,1) = 0.37 + 0.13 = 0.5$

Which can be shown as.

X = x	0	1	Total
P(X = x)	0.5	0.5	1

11. Given the following bivariates probability distribution.

Y \ X	0	1
0	0.35	0.3
1	0.15	0.2

Scanned by CamScanner

Solution 11

The marginal distribution of Y is as follow:

$P(Y = 0) = P(0,1) + P(0,1) = 0.35 + 0.3 = 0.65$

$P(Y = 1) = P(1,0) + P(1,1) = 0.15 + 0.2 = 0.35$

Which can be shown as,

$Y = y_i$	0	1	Total
$P(Y = y_i)$	0.65	0.35	1

12. **Joint distribution is given below.**

Y \ X	–0	0	1
0	1/15	2/15	1/15
1	3/15	A	1/15
2	2/15	1/15	2/15

i. Find the value of A.

ii. If $P(0, 0) = 0$, $P(1, 0) = \frac{1}{4}$ and $P(2, 0) = \frac{3}{4}$

iii. If $P(-1, 0) = 1/2$, $P(0, 0) = 1.4$ and $P(1, 0) = 1/4$: Find $P_{x/y}$ $(X = -1/Y = 0)$

Solution 12

(i) Since the sum of all probabilities must be unity so, we have,

$$\frac{1}{15} + \frac{3}{15} + \frac{2}{15} + \frac{2}{15} + A + \frac{1}{15} + \frac{1}{15} + \frac{1}{15} + \frac{2}{15} = 1$$

or, $\frac{13}{15} + A = 1 \Rightarrow A = 1 - \frac{13}{15} = \frac{2}{15}$

(ii) Form the given information

$$P(X = 1 / Y = 0) = \frac{P(1,0)}{P(Y=0)} = \frac{P(1,0)}{P(0,0) + P(1,0) + P(2,0)} = \frac{1/4}{0 + 1/4 + 3/4} = \frac{1}{4}$$

(iii) $$P(X = -1/Y = 0) = \frac{P(-1,0)}{P(-1,0) + P(0,0) + P(1,0)} = \frac{\frac{1}{2}}{1/2 + 1/4 + 1/4} = \frac{1}{2}$$

13. **X and Y are two random variables having the joint mass function.**

$$f(X, Y) = \frac{1}{27}(2x + y),\ x = 0, 1, 2,\ y = 0, 1, 2$$

i. Write down distribution in tabular form.

ii. Find the conditional distribution of Y for X = x.

Solution 13

Let X and Y are two random variables having the joint mass function

$$f(x,y) = \frac{1}{27}(2x + y),\ x = 0, 1, 2;\ y = 0, 1, 2.$$

(i) The joint probability distribution is given as follows:

P(x,y)		Y			ΣP(x,y)
		0	1	2	
X	0	0	1/27	2/27	3/27
	1	2/27	3/27	4/27	9/27
	2	4/27	5/27	6/27	15/27
ΣP(x,y)		6/27	9/27	12/27	1

Scanned by CamScanner

(ii) The Conditional probability distribution of Y for X = x is given by:

$$P(Y = y_i / X = x) = \frac{P(Y = y_i, X = x)}{P(X = x)} = \frac{P(X = x, Y = Yi)}{P(X = x)}$$

(a) If x = 0, then, $P(Y = y_1 / X = 0) = \frac{P(X = 0, Y = y_1)}{P(X = 0)} = \frac{P(X = 0, Y = 0)}{P(X = 0)} = 0$

$$P(Y = y_2 / X = 0) = \frac{P(X = 0, Y = y_2)}{P(X = 0)} = \frac{P(0,1)}{P(X = 0)} = \frac{1}{3}$$

$$P(Y = y_3 / Y = 0) \frac{P(X = 0, Y = y_3)}{P(X = 0)} = \frac{P(0,2)}{P(X = 0)} = \frac{2}{3}$$

(b) When X = 1,

$$P(Y = y_1/X = 1) = \frac{P(X = 1, Y = y_1)}{P(X = 1)} \frac{P(1,0)}{P(X=1)} = \frac{2}{9}$$

$$P(Y = y_2/X = 1) = \frac{P(X = 1, Y = y_2)}{P(X = 1)} = \frac{P(1,1)}{P(X=1)} = \frac{1}{3}$$

$$P(Y = y_3/X = 1) = \frac{P(X = 1, Y = y_3)}{P(X = 1)} = \frac{P(1,2)}{P(X=1)} = \frac{4}{9}$$

(c) When X = 2,

$$P(Y = y_3 / X = 2) = \frac{P(X = 2, Y = 0)}{P(X = 2)} = \frac{4}{15}$$

$$P(Y = y_2 / X = 2) = \frac{P(X = 2, Y = 1)}{P(X = 2)} = \frac{1}{3}$$

$$P(Y = y_3 / X = 2) \frac{P(X = 2, Y = 2)}{P(X = 2)} = \frac{2}{5}$$

14. The joint p.d.f. of two r. vs X and Y is given by:

$$f(x, y) = \frac{9(1 + x + y)}{2(1 + x)^4 (1 + y)^4}; \begin{matrix} 0 \le x < \infty \\ 0 < y < \infty \end{matrix}$$

Find the marginal distribution of X and Y, and the conditional distribution of Y for X = x.

Solution 14

Let X and Y are random variable with point p.d.f given by

$$f(x,y) = \frac{9(1+x+y)}{2(1+x)^4 (1+y)^4}; \begin{matrix} 0 \le x < \infty \\ 0 < y < \infty \end{matrix}$$

(a) Now the marginal distribution of X is given by $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

$0 \le x < \infty; 0 < y < \infty,$

$$= \int_0^{\infty} \frac{9(1+x+y)}{2(1+x)^4 (1+y)^4} dy = \frac{9}{2(1+x^4)} \int_0^{\infty} \frac{x+1+y}{(1+y)^4} dy$$

$$= \frac{9}{2(1+x)^4} \left[x \cdot \frac{(1+y)^{-4+1}}{-4+1} + \frac{(1+y)^{-3+1}}{-3+1} \right]_0^{\infty}$$

$$= \frac{9}{2(1+x)^4} \left[\frac{x}{-3(1+y)^3} + \frac{1}{-2(1+y)^2} \right]_0^{\infty}$$

$$= \frac{9}{2(1+x)^4} \left[0 - \left(\frac{-x}{3} - \frac{1}{2} \right) \right]$$

Scanned by CamScanner

$$= \frac{9}{2(1+a)^4} \times \left(\frac{2x+3}{6}\right) = \frac{3\,(2x+3)}{4\,(1+x)^4}$$

Also the marginal distribution of Y is given by $f_Y(y)$

$$= \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{\infty} \frac{9\,(1+x+y)}{2(1+x)^4.\,(1+y)^4}\,dx = \frac{3(2y+3)}{4(1+y)^4} \text{ [Similarly as above]}$$

(b) Now the conditional distribution of Y for $X = x$ is given by;

$$F_{Y/X}\,(y/x) = \frac{f(x,y)}{f_X(x)} = \frac{9(1+x+y)}{2(1+x)^4\,(1+y)^4} \times \frac{4(1+x)^4}{3(2x+3)} = \frac{6(1+x+y)}{(2x+3)\,(1+y)^4}$$

15. Two dice are rolled, S = {(a, b); a , b = 1, 2, ..., 6)}

Let X denote the sum of the two face and Y the absolute value of their difference, i.e. X is distributed over the integers 2, 3, ..., 12 and Y over 0, 1, 2, ..., 5.

Assuming the dice are fair, find the probability that:

i. $X = 5 \cap Y = 1$ ii. $X = 7 \cap Y \geq 3$

iii. $X = Y$ iv. $X + Y = 4 \cap X - Y = 2$

Solution 15

Since X - denote the sum of the two faces of two dice and Y – denote the absolute value of their differences. Then we have.

$X = \{2, 3, 4, \ldots\ldots, 12\}$, $Y = \{0, 1, 2, 3, 4, 5\}$

(i) Now the probability of $(X = 5 \cap Y = 1)$ is given by $P(X = 5 \cap Y = 1) = \frac{2}{36} = \frac{1}{18}$

[∵ Having sum 5 and difference 1 is the ordered pairs (2,3) and (3,2)]

(ii) The probability of $X = 7 \cap Y \geq 3$ is given by $P\,(X = 7n\,Y \geq 3) = \frac{4}{36} = \frac{1}{9}$

[∵ The pairs having sum 7 with difference 3 or more are, (1,6), (2,5), (6,1) and (5,2)]

(iii) The probability of $X = Y$ is given by $P\,(X = Y) = \frac{0}{36} = 0$

[Since the numbers on dice are all positive so sum and difference cannot be same]

(iv) The probability of $X + Y = 4 \cap X - Y = 2$ is given by $P(X+Y=4 \cap X-Y=2) = \frac{2}{36} = \frac{1}{18}$

[The pairs (1,2) and (2,1) satisfy the about condition].

16. Given $f(x, y) = e^{-(x+y)}$, $x \geq 0$, $y \geq 0$. Are X and Y independent?

Solution 16

Since the p.d.f of the joint probabilities of two random variables X and Y are given by, $f(x,y) = e^{-(x+y)}$, $x \geq 0$, $y \geq 0$.

Two continuous random variables X and Y are said to be independent if and only if $f(x, y) = f_X(x) \cdot f_X(x) \cdot f_Y(y)$

Where, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy,\ y \geq 0 = \int_{0}^{\infty} e^{-(x+y)}\,dy = e^{-x}\left[\frac{e^{-y}}{(-1)}\right]_0^{\infty} = e^{-x}$

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

Scanned by CamScanner

Now, $f_X(x).\ f_Y(y) = e^{-x}.\ e^{-y} = e^{-(x+y)} = f(x,y)$

Hence the random variables X and Y are independent.

17. **Two discrete r.v.s. X and Y have**

$P(X = 0, Y = 0) = \frac{2}{9}$; $P(X = 0, Y = 1) = \frac{1}{9}$

$P(X = 1, Y = 0) = \frac{1}{9}$; $P(X = 1, Y = 1) = \frac{5}{9}$

Examine whether X and Y are independent

Solution 17

We know that tw0 discrete random variables X and Y are said to be independent if and only if it satisfy the condition $P(X=x,\ Y=y) = P(X=x)\ P(Y=y)$ for all $x \in X,\ y \in Y$.

(i) Here we have given $P(X=0,Y=0) = \frac{2}{9}$

Now $P(X=0) = P(0,0) + P(0,1) = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$

$P(Y=0) = P(0,0) + P(1,0) = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$

Now $P(X=0) \times P(Y=0) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} \neq P(X=0,Y=0)$

Hence the random variables X and Y are dependent.

18. If the joint distribution form of X and Y given by

$$f(x, y) \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{for } x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

i. Find marginal p.d.f. of X and Y

ii. Are X and Y independent?

iii. $P(1 < x < 3, 1 < y < 2)$

Solution 18

Let X,Y be any two continuous random variables with join distribution

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

(i) The marginal p.d.f of X is given by $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

$$= \int_0^{\infty} (1 - e^{-x})(1 - e^{-y})\, dy = (1 - e^{-x}).\ [y + e^{-y}]_0^{\infty} = 1 - e^{-x}$$

19. **Joint probability density function of a two-dimensional random variable (x, y) is given by,**

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

i. Find the marginal density function of X and Y.

Scanned by CamScanner

ii. Find the conditional density function of Y given X = x and conditional density function of X given given Y = y.

Solution 19

Here the joint probability density function is given by

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

(i) Then the marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_0^x 2\, dy = 2\,[y]_0^x = 2x$$

$\because$ $f_X(x) = 2x, 0 < x < 1$

Also the marginal density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_y^1 2\, dx = 2\,[x]_1^y = 2\,[1 - y]$$

$\because$ $f_Y(y) = 2(1-y), 0 < y < 1.$

(ii) The conditional density function of Y given X = x is given by :

$$f_{Y/X}(y/x) = \frac{f(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}, 0 < x < 1.$$

Also the conditional density function of X when Y = y is given is,

$$F_{X/Y}\left(\frac{x}{y}\right) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}, 0 < y < 1$$

20. **Let joint p.d.f. of X and Y be**

$$P(x, y) = \frac{x + y}{21}; x = 1, 2, 3; y = 1, 2$$

Find the marginal p.d.f. of X and Y

Solution 20

Here given joint probability mass function of X and Y is

$$P(x,y) = \frac{x+y}{21}; x = 1, 2, 3; y = 1,2$$

Then the marginal probability mass function of X is given by

$$P_X(x) = \int_{-\infty}^{\infty} P(x,y)\, dy = \int_1^2 \frac{x+y}{21}\, dy$$

$$= \left[\frac{xy + y^2/2}{21}\right]_1^2 = \left[\frac{2xy + y^2}{42}\right]_1^2 = \frac{2x + 3}{42}, x = 1, 2, 3$$

$$P_Y(y) = \int_1^3 P(x,y)\, dx = \int_1^3 \left(\frac{x+y}{21}\right) dx$$

$$= \left[\frac{x^2/2 + xy}{21}\right]_1^3 = \left[\frac{x^2 + 2xy}{42}\right]_1^3 = \frac{2y+4}{21}, y = 1,2$$

Scanned by CamScanner

21. If the joint probability function of X and Y is

$$f(x, y) = A(x + y); 0 < x < 2, 0 < y < 1$$

i. Determine A ii. Find marginal p.d.f of X and Y

iii. Find conditional distribution of x given that Y = y.

Solution 21

Here given joint probability function of X and Y is

$f(x,y) = A(x+y)\,; 0 < x < 2, 0 < y < 1.$

(i) For the value of A, we must have the total probability must be unity

$$\text{i.e} \int_0^2 \int_0^1 \left\{ {}_0^1 A(x,y)\, dy \right\} dx = 1$$

$$\Rightarrow \int_0^2 \int_0^1 \{A(x+y)\, dy\}\, dx = 1$$

$$\Rightarrow \int_0^2 A\left(xy + \frac{y^2}{2}\right)_0^1 dx = 1$$

$$\Rightarrow A \int_0^2 \left(x + \frac{1}{2}\right) dx = 1 \Rightarrow A\left[\frac{x^2}{2} + \frac{x}{2}\right]_0^2 = 1 \Rightarrow 3A = 1 \Rightarrow A = \frac{1}{3}.$$

(iii) The marginal probability of X is given by $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

$$= \int_0^1 \frac{1}{3}(x+y)\, dy = \frac{1}{3}\left[xy + \frac{y^2}{2}\right]_0^1 = \frac{2x+1}{6}, 0 < x < 2$$

The marginal probability of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_0^2 \frac{1}{3}(x+y)\, dx = \frac{1}{3}\left(\frac{x^2}{2} + xy\right)_0^2 = \frac{2y+1}{3}, 0 < y < 1$$

(iii) The conditional probability distribution of X given than Y = y is given by

$$f_{X/Y}\left(\frac{x}{y}\right) = \frac{f(x,y)}{f_Y(y)} = \frac{1/3(x+y)}{1/3(2y+1)} = \frac{x+y}{2y+1}; \begin{matrix} 0 < x < 2 \\ 0 < y < 1 \end{matrix}$$

22. If $f(x, y) = \begin{cases} e^{-(x+y)} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Is the joint probability density function of random variables X and Y. Find:

i. $P(X < 1)$ ii. Marginal probability density of X and Y

iii. Examine whether X and Y are independent [TU 2071]

Solution 22

The joint probability density foundation of X and Y is given by

Scanned by CamScanner

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

(i) Then $P(X < 1) = \int_0^1 \int_0^\infty f(x,y)\, dy\, dx$

$$= \int_0^1 \left[\int_0^\infty e^{-(x+y)}\, dy \right] dx = \int_0^1 e^{-x} \cdot \left(\frac{e^{-y}}{-1} \right)_0^\infty dx = \int_0^1 e^{-x} \cdot 1\, dx = 1 - e^{-1}$$

(ii) The marginal probability density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_0^\infty e^{-(x+y)}\, dy = e^{-x} \cdot \left[\frac{e^{-y}}{-1} \right]_0^\infty = e^{-x}, x \geq 0$$

Also the marginal probability density function of Y is given by

$$f_Y(y) = \int_0^\infty f(x,y)\, dx = \int_0^\infty e^{-(x+y)}\, dx = e^{-y} \cdot \left[\frac{e^{-y}}{-1} \right]_0^\infty = e^{-y}, y \geq 0.$$

(iii) Two random variables X and Y are said to be independent if and only if $f(x,y) = f_X(x) \cdot f_Y(y)$ for all x,y

Now, $f_X(x) \cdot f_Y(y) = e^{-x} \cdot e^{-y} = e^{-(x+y)} = f(x,y)$

Hence X and Y are independent.

23. Joint distribution of X and Y are given by

$$f(x, y) = 4xy\, e^{-(x^2+y^2)};\ x > 0,\ y \geq 0$$

i. Test whether X and Y are independent

ii. Find the conditional density of x and given Y = y

Solution 23

(i) Two random variables X and Y are said to be independent if and only if $f(x,y) = f_X(x) \cdot f_Y(y)$ for all x,y.

Where $f_X(x) = \int_0^\infty f(x,y)\, dy$

$$= \int_0^\infty 4xy\, e^{-(x^2+y^2)}\, dy = 4x\, e^{-x^2} \int_0^\infty y e^{-y^2} dy$$

$$= 2xe^{-x^2}, x > 0 \quad \left[\because \int_0^\infty ye{-}y2\, dy = \frac{1}{2} \right]$$

Also, $f_Y(y) = \int_0^\infty f(x,y)\, dx = \int_0^\infty 4xy\, e^{-(x^2-y^2)}\, dx$

$$= 4y\, e^{-y^2} \int_0^\infty xe^{-x^2} dx = 2ye^{-y^2}, y \geq 0$$

Scanned by CamScanner

Now, $f_X(x).\ f_Y(y) = 2xe^{-x^2} \times 2y\ e^{-y^2}$

$= 4xye^{-(x^2+y^2)} = f(x,y)$

Hence the variables X and Y are independent.

(ii) The conditional density function of X given Y = y is

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_Y(y)} = \frac{4xy\ e^{-(x^2+y^2)}}{2y\ e^{-y^2}} = 2xe^{-y^2},\ x > 0$$

24. The joint p.d.f. of random variables X and Y is given by

$$f(x, y) = \frac{3}{2}y^2,\ 0 \le x \le 2,\ 0 \le y \le 1$$

Determine the marginal density function of x and Y. Are X and Y independent?

Solution 24

The joint probability density function of X and Y is given as

$$f(x,y) = \frac{3y^2}{2},\ 0 \le x \le 2,\ 0 \le y \le 1$$

Then the marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{3y^2}{2}\,dy = \left[\frac{3}{2}\cdot\frac{y^3}{3}\right]_0^1 = \frac{1}{2}$$

$$f_Y(y) \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{3y^2}{2}\,dy = \frac{3y^2}{2}[x]_0^2 = 3y^2,\ 0 \le y \le 1.$$

Also, $f_X(x).\ f_Y(y) = \frac{1}{2} \times 3y^2 = \frac{3y^2}{2} = f(x,y)$

Since $f(x,y) = f_X(x)\ f_Y(y)$ for all x,y.

So X and Y are independent.

25. Suppose that in a certain circuit 20 resistors are connected in series. The mean and variance of each resistor are 5 ohms and 0.2 ohm² respective. What is the probability that total resistance of the circuit will exceed 98 ohms, assuming independent?

Solution 25

Since in a certain circuit 20 resistors are connected in a series with then mean E(X) = 5 ohms and variance V(X) = 0.2 ohms².

26. A Ski lift is designed with a load limit of 18000 lbs. It claims a capacity of 100 persons. If the weight of all the people using the lift have a mean of 175 lbs and a s.d. of 30 lbs. What is the probability that a group of 100 persons will exceed the load limit?

Solution 26

Let $X_1, X_2, X_3, \ldots\ldots\ldots\ldots X_{100}$ denote the weight of 100 persons, These random variables are independent and approximately identical (large sample). Form the central limit theorem $X_i \sim N(175, 900)$, i = 1, 2, 3, ..., 100.

Scanned by CamScanner

Total weight of 100 persons is $T_\infty = \sum_{i=1}^{100} x_i$

and $E(T_o) = n\mu = 100 \times 175 = 17500$

$V(T_o) = n\sigma^2 = 100 \times 900 = 90,000$

$\because$ To ~ B (17, 500, 90, 000)

The probability that a random group of 100 persons will exceed the load limit is

$$P(T_o > 18,000) = P\left(\frac{T_o - E(T_o)}{\sqrt{V(T_o)}} > \frac{1800 - 17500}{300}\right)$$

$$= P(Z > 1.67) = 1 - P(Z \le 1.67) = 1 - 0.9525 = 0.0475$$

27. **A population has a mean of 0.112 and standard deviation of 2.12, find the probability that the mean of random sample 800 will be negative?**

Solution 27

Since any sample of the population is a normal distribution then suppose X is a normal distribution with parameters $\mu = 0.112$ and $\sigma = 2.12$.

The $X \sim N(0.112, 2.12^2)$

Here sample size n = 800. Let Z be the standard normal random variable. Then X

$$\sim N\left(m, \frac{s^2}{n}\right) = N(0.112, (2.12)^2/800)$$

Hence the probability that the mean of the random sample of size 800 will be negative is given by $P(\bar{X} < 0)$

$$\Rightarrow \quad P\left(\frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < \frac{0 - 0.112}{\frac{(2.12)}{\sqrt{800}}}\right) \Rightarrow P(Z < -1.49) = 0.0681$$

28. **Two dimensional random variable (X, Y) have joint density.**

$f(x, y) - 8xy, 0 < x < y < 1$

i. Find $P\left(X < \frac{1}{2} \cap Y < \frac{1}{4}\right)$

ii. Find marginal and conditional distributions.

iii. Area X and Y independent? Give reasons for your. [TU 2063]

Solution 28

The joint probability density function of the random variables X and Y is $f(x,y) = 8xy, 0 < x < y < 1$

(i) $$\left(X < \frac{1}{2} \cap Y < \frac{1}{4}\right) = \int_0^{1/2}\int_0^{1/4} f(x,y)\,dydx = \int_0^{1/2}\int_0^{1/4} 8xy\,dydx$$

$$= \int_0^{1/2} 8x\left[\frac{y^2}{2}\right]_0^{1/4} dx = \int_0^{1/2} 8x \times \frac{1}{32}\,dx$$

$$= \frac{1}{4}\left[\frac{x^2}{2}\right]_0^{1/2} = \frac{1}{4} \times \frac{1}{8} = \frac{1}{32}$$

(ii) The marginal probability density function of x is denoted by

Scanned by CamScanner

$$f_X(x) = \int_x^1 f(x,y)dy = \int_x^1 8xydy = 8x\left[\frac{y^2}{2}\right]_x^1$$

$$= 8x\left[\frac{1}{2} - \frac{x^2}{2}\right] = 4x(1-x^2) \quad 0 < x < 1.$$

$$f_Y(y) = \int_0^y f(x,y)\,dx = \int_0^y 8xy\,dx = 8y\left[\frac{y^2}{2}\right]_0^y = 4y^3 \quad 0 < y < 1$$

The conditional probability distribution of X given that Y = y is,

$$f_{X/Y}(X/Y) = \frac{f(x,y)}{f_Y(y)} = \frac{8xy}{4y^3} = \frac{2x}{y^2}, 0 < x < y < 1$$

Also the conditional probability distribution of Y given that X = x is

$$f_{Y/X}(Y/X) = \frac{f(x,y)}{f_X(x)} = \frac{8xy}{4x(1-x^2)}$$

iii. Now, $f_X(x).\ f_Y(y) = 4x(1-x^2).\ 4y^3 = 16xy^3(1-x^2) \neq 8xy = f(x,y)$
Since $f(x,y) \neq f(x).\ f(y)$
So the variables X and Y are dependent.

29. If three random variables have the joint density.

$$f(x,y,z) = \begin{cases} k(x+y)e^{-z} & \text{for } 0 < x < 1, 0 < y < 2, z > 0 \\ 0 & \text{elsewhere} \end{cases}$$

a. Find the value of k

b. Find the probability that $X < Y \cap Z > 1$

Solution 29

The three random variables X,Y,Z have the joint p.d.f is given by

$$f(x,y,z) = \begin{cases} k(x+y)e^{-z} & 0 < x < 1, 0 < y < 2, z > 0 \\ 0 & \text{elseshere} \end{cases}$$

(a) For the given function to be density function we must have.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\,dzdydx = 1$$

or $$\int_0^1\int_0^2\int_0^{\infty} k(x+y)e^{-z}\,dx\,dy\,dx = 1$$

or, $$\int_0^1\int_0^2 k(x+y).\left[\frac{e^{-z}}{-1}\right]_0^{\infty} dydx = 1$$

or, $$\int_0^1 \left[k\left(xy + \frac{y^2}{2}\right)\right]_0^2 dx = 1$$

or, $$\frac{k}{2}\int_0^1 (4x+4)\,dx = 1$$

Scanned by CamScanner

or $\frac{k}{2}\left[\frac{4x^2}{2}+4x\right]_0^1 = 1$

or, $\frac{k}{4}[12] = 1 \Rightarrow k\frac{1}{3}$

30. **If measurements of the length and the width of a rectangle have the joint density,**

$$f(x, y) = \begin{cases} \frac{1}{ab} & \text{for } L-\frac{a}{2} < x < L+\frac{a}{2}, W-\frac{b}{2} < y < W+\frac{b}{2} \\ 0 & \text{elsewhere} \end{cases}$$

Find the mean and variance of the corresponding distribution of the rectangle.

Solution 30

The joint probability density function of two variables X (length of rectangle) and Y (width of rectangle) is

$$f(x,y) = \begin{cases} \frac{1}{ab} & L\frac{-a}{2} < x < L+\frac{a}{2}, w\frac{-b}{2} < y < w\frac{+b}{2} \\ 0 & \end{cases}$$

The mean is given by

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

$$= \int_{L-a/2}^{L+a/2}\int_{W-b/2}^{W+b/2} xy.\frac{1}{ab}\, dy\, dx$$

$$= \frac{1}{ab}\int_{L-a/2}^{L+a/2} x \left[\frac{y^2}{2}\right]_{W-b/2}^{W+b/2} dx$$

$$= \frac{1}{ab}\int_{L-a/2}^{L+a/2} \frac{x}{2}\left[\left(w+\frac{b}{2}\right)^2 - \left(w-\frac{b}{2}\right)^2\right] dx$$

$$= \frac{2wb}{2ab}\left[\frac{x^2}{2}\right]_{L-a/2}^{L+a/2}$$

$$= \frac{W}{2a}\left[\left(L+\frac{a}{2}\right)^2 - \left(L-\frac{a^2}{2}\right)\right] = \frac{W}{2a} \times 2aL = LW.$$

Exam Asked Questions

1. **The joint probability distribution of a pair (X, Y) of random variable is given below:** **[PU 2010 Spring]**

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

i. Find the value of A.

ii. Find the marginal probability distribution of X and Y.

Scanned by CamScanner

iii. Are X and Y independent?

iv. Find E(X), E(Y) and Var(X)

Solution 1

i. Since the given probability distribution function is joint distribution function. So, we must have

$$\frac{1}{15}+A+\frac{1}{15}+\frac{3}{15}+A+\frac{1}{15}+A+\frac{1}{15}+A=1$$

or, $\frac{7}{15}+4A=1 \Rightarrow 4A=1-\frac{7}{15}=\frac{8}{15} \Rightarrow A=\frac{2}{15}$

ii. The marginal probability distribution function of X is given by

$$P(X=x_i)=P(x_i, y_1)+P(x_i, y_2)+\ldots+P(x_i, y_n)$$

$\therefore$ $P(X=-1)=P(-1,0)+P(-1,1)+P(-1,2)=\frac{1}{15}+\frac{3}{15}+\frac{2}{15}=\frac{6}{15}=\frac{2}{5}$

$P(X=0)=P(0,0)+P(0,1)+P(0,2)=\frac{2}{15}+\frac{2}{15}+\frac{1}{15}=\frac{5}{15}=\frac{1}{3}$

$P(X=1)=P(1,0)+P(1,1)+P(1,2)=\frac{1}{15}+\frac{1}{15}+\frac{2}{15}=\frac{4}{15}$

Which can be shown in the following table:

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

Similarly, the marginal probability distribution of Y is given by

$$P(Y=y_j)=P(x_1, y_j)+P(x_2, y_j)+P(x_3, y_j)+\ldots+P(x_m, y_j)$$

i.e. $P(Y=0)=P(-1,0)+P(0,0)+P(1,0)=\frac{1}{15}+\frac{2}{15}+\frac{1}{15}=\frac{4}{15}$

$P(Y=1)=P(-1,1)+P(0,1)+P(1,1)=\frac{3}{15}+\frac{2}{15}+\frac{1}{15}=\frac{6}{15}=\frac{2}{5}$

$P(Y=2_=P(-1,2)+P(0,2)+P(1,2)=\frac{2}{15}+\frac{1}{15}+\frac{2}{15}=\frac{5}{15}=\frac{1}{3}$

Which can be shown in the following table:

Y = y	0	1	2
P(Y = y)	4/15	2/5	1/3

iii. From the above table, we have

$P(X=-1, Y=0)=1/15$, $P(X=-1)=2/5$ and $P(Y=0)=4/15$

Now, $P(X=-1)\times P(Y=0)=\frac{2}{5}\times\frac{4}{15}=\frac{8}{75}\neq\frac{1}{15}=P(X=-1, Y=0)$

So, X and Y are independent.

iv. Now, $E(X)=\sum_{i=1}^{m} x_i P(x_i)=x_1P(x_1)+x_2P(x_2)+x_3P(x_3)$

$$=(-1)\times\frac{2}{5}+0\times\frac{1}{3}+1\times\frac{4}{15}=\frac{4}{15}-\frac{2}{5}=\frac{4-6}{15}=\frac{-2}{15}$$

Scanned by CamScanner

Also, Var (X) $= E(X^2) - (E(X))^2$

$$= \sum_{i=1}^{m} x_i P(x_i) - \left(\sum_{i=1}^{m} x_i P(x_i)\right)^2$$

$$= (x_1^2 P(x_1) + x_2^2 (P(x_2) + x_3^2 P(x_3)) - (-2/15)^2$$

$$= (-1)^2 \times 2/5 + 0^2 \times 1.3 + 1^2 \times 4/15 - (2/15)^2$$

$$= 10/75 - 4/225 = \frac{30-4}{225} = \frac{26}{225}$$

2. **The joint probability distribution of the number X of cards and the number Y of buses per signal cycle of a proposed left turn lane is displayed in the accompanying joint probability table,**

P(x, y)		Y = y		
		0	1	2
X = X	0	0.025	0.015	0.010
	1	0.050	0.030	0.020
	2	0.125	0.075	0.050
	3	0.150	0.090	0.060
	4	0.10	0.060	0.040
	5	0.050	0.030	0.020

i. What is the probability that there is exactly one car and exactly one bus during the cycle?

ii. What is the probability that there is at most one car and at most one bus during a cycle?

iii. What is the probability that there is exactly one car during a cycle? Exactly one bus?

iv. Suppose that left turn lane is to have a capacity of five cards, and one bus is equivalent to three cars. What is the probability of an overflow during a cycle? [TU 2058/2059/2062/2063/2064]

Solution 2

Please see ex. 11 Q.No. 9

3. **Let (X, Y) be the pair of discrete random variables each taking three values 1, 2, 3 with the following joint distribution.** **[Purbanchal 2010]**

P(x, y)		X		
		1	2	3
Y	1	5/27	4/27	2/27
	2	1/27	3/27	3/27
	3	3/27	4/27	2/27

Then find: (i) E(X) (ii) E(Y) (iii) E(XY) (iv) COV(X, Y)
(v) E(X + Y) (vi) $E(X^2 + Y^2)$

Solution 3

Here, we have the marginal probability distribution of X and Y are given as follows:

Scanned by CamScanner

Y = x	1	2	3	Total
P(Y = x)	$\frac{5}{27}+\frac{1}{27}+\frac{3}{27}=\frac{1}{3}$	$\frac{4}{27}+\frac{3}{27}+\frac{4}{27}=\frac{11}{27}$	$\frac{2}{27}+\frac{3}{27}+\frac{2}{27}=\frac{7}{27}$	1

Y = y	1	2	3	Total
P(Y = y)	$\frac{5}{27}+\frac{4}{27}+\frac{2}{27}=\frac{11}{27}$	$\frac{1}{27}+\frac{3}{27}+\frac{3}{27}=\frac{7}{27}$	$\frac{6}{27}+\frac{4}{27}+\frac{2}{27}=\frac{1}{3}$	1

i. $E(X) = \sum_{i=1}^{3} x_i P(x_i) = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) = 1 \times \frac{1}{3} + 2 \times \frac{11}{27} + 3 \times \frac{7}{27} = 1.93$

ii. $E(Y) = \sum_{j=1}^{3} y_j P(y_j) = y_1P(y_1) + y_2P(y_2) + y_3P(y_3) = 1 \times \frac{11}{27} + 2 \times \frac{7}{27} + 3 \times \frac{1}{3} = 1.93$

iii. $E(XY) = \sum_{i=1}^{3} \sum_{j=1}^{3} (x_i \times by_j) P(x_i, y_j)$

$= (1 \times 1) P(1, 1) + (1 \times 2) P(1, 2)) + 1 \times 3 P(1, 3) + 2 \times 1 P(2, 1) + 2 \times 2 P(2, 2) + 2 \times 3 P(2, 3) + 3 \times 1 P(3, 1) + 3 \times 2 P(3, 2) + 3 \times 3 P(3, 3)$

$= 1 \times \frac{5}{27} + 2 \times \frac{1}{27} + 3 \times \frac{3}{27} + 2 \times \frac{4}{27} + 4 \times \frac{3}{27} + 6 \times \frac{4}{27} + 3 \times \frac{2}{27} + 6 \times \frac{3}{27} + 9 \times \frac{2}{27} = \frac{102}{27} = 3.78$

iv. $Cov(X, Y) = E(XY) - E(X)E(Y) = 3.78 - 1.93 \times 1.93 = 0.055$

v. $E(X + Y) = \sum_{i=1}^{3} \sum_{j=1}^{3} (x_i \times y_j) P(x_i, y_j)$

$= 2 \times P(1, 1) + 3 \times P(1, 2) + 4 \times P(1, 3) + 3 \times P(2, 1) + 4 \times P(2, 2) + 5P(2, 3) + 4P(3, 1) + 5P(3, 2) + 6P(3, 3)$

$= 2 \times \frac{5}{27} + 3 \times \frac{1}{27} + 4 \times \frac{3}{27} + 3 \times \frac{4}{27} + 4 \times \frac{3}{27} + 5 \times \frac{4}{27} + 4 \times \frac{2}{27} + 5 \times \frac{2}{27} + 6 \times \frac{2}{27}$

$= \frac{104}{27} = 3.85$

vi. $E(X^2 + Y^2) = \sum_{i=1}^{3} \sum_{j=1}^{3} (x_i^2 \times + y_j^2) P(x_i, y_j)$

$= 2P(1,1) + 5P(1,2) + 10P(1,3) + 5P(2,1) + 8P(2,2) + 13P(2,3) + 10P(3,1) + 13P(3,2) + 18P(3,3)$

$= 2 \times \frac{5}{27} + 5 \times \frac{1}{27} + 10 \times \frac{3}{27} + 5 \times \frac{4}{27} + 8 \times \frac{3}{27} + 13 \times \frac{4}{27} + 10 \times \frac{2}{27} + 13 \times \frac{3}{27} + 18 \times \frac{2}{27}$

$= \frac{236}{27} = 8.74$

4. The following are the joint probability data applied to the drying time hours (X) and durability (years) Y of points. [TU 2063 Kartik/2065 Chaitra]

P(X, Y)		Y			
		5	6	7	8
X	5	0.05	0.05	0	0
	10	0.10	0.15	0.05	0
	15	0.05	0.10	0.20	0.05
	20	0	0.05	0.05	0.10

i. Find the marginal probability distribution of X and Y.
ii. Determine the conditional probability distribution of Y given X = 10 hours
iii. Find F(10, 6) and F(15, 6)
iv. Are X and Y independent?

Scanned by CamScanner

Solution 4

Do yourself

5. **The joint probability distribution of a pair (X,Y) of random variables is given by following table.**

y \ x	1	2	3
1	0.1/1.6	0.2/1.6	0.2/1.6
2	0.2/1.6	0.3/1.6	0.1/1.6
3	0.3/1.6	0.1/1.6	0.1/1.6

Find: i. the marginal probability distribution of X

ii. the conditional probability distribution of X given Y = 1

iii. E(X), E(Y), V(X) and covariance between X and Y

iv. F(1, 3), F(2, 2) and F(3, 3)

Solution 5

Please refer Exercise 11 Q.No. 7

6. **The joint probability distribution of random variables X and Y is given by**

$$\begin{cases} k(6-x-y) & 0<x<2, 2<y<4 \\ 0 & \text{otherwise} \end{cases}$$

Then find (i) the constant k; (ii) $P(X<1 \cap Y<3)$, (iii) $P(X+Y<3)$ (iv) marginal density function of X and Y, (v) E(X) and E(Y)

[PU 2011 Fall/2017 Fall/TU 2058/2065/Purbanchal 2012]

Solution 6

i. Since, the given function is joint probability distribution then we must have

$$\int_0^2\int_2^4 f(x, y)\, dy\, dx = 1$$

or, $$\int_0^2\int_2^4 k(6-x-y)\, dy\, dx = 1$$

or, $$k\int_0^2 \left[6y - xy - \frac{y^2}{2}\right]_2^4 dx = 1$$

or, $$k\int_0^2 (12-2x-6)\, dx = 1$$

or, $$k(12x - x^2 - 6x]_0^2 = 1 \Rightarrow 8k = 1 \Rightarrow k = \frac{1}{8}$$

ii. $$P(X<1 \cap Y<3) = \int_0^1\int_2^3 f(x, y)\, dy\, dx$$

Scanned by CamScanner

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

$$= \frac{1}{8}\int_0^1 [6y - xy - y^2/2]_2^3 dx$$

$$= \frac{1}{8}\int_0^1 [6 \times 1 - x \times 1 - 1/2\,(3^2 - 2^2)]\, dx$$

$$= \frac{1}{8}\int_0^1 (6 - x - 2.5)\, dx$$

$$= \frac{1}{8}[3.5x - x^2/2]_0^1 = \frac{3}{8}$$

iii. $P(X + Y < 3)$ $$= \int_0^2 \int_2^{3-x} f(x, y)\, dy\, dx$$

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

$$= \frac{1}{8}\int_0^2 [6y - xy - y^2/2]_2^{3-x} dx$$

$$= \frac{1}{8}\int_0^2 [6(3 - x - 2) - x(3 - x - 2) - \frac{1}{2}[3 - x)^2 - 2^2]]\, dx$$

$$= \frac{1}{8}\int_0^2 \frac{(x^2 - 8x + 7)}{2} dx$$

$$= \frac{1}{16}\left[\frac{x^3}{3} - 8 \cdot \frac{x^2}{2} + 7x\right]_0^2$$

$$= \frac{1}{16} \times \frac{2}{3} = \frac{1}{24}$$

iv. The marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_4^2 \frac{1}{8}(6 - x - y)\, dy$$

Scanned by CamScanner

$$= \frac{1}{8}\left[6y - xy - \frac{y^2}{2}\right]_2^4$$

$$= \frac{1}{8}[12 - 2x - 6]$$

$$= \frac{3-x}{4}, 0 < x < 2$$

The marginal density function of Y is given by

$$f_Y(y) = \int_{\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{1}{8}(6 - x - y)\,dy$$

$$= \frac{1}{8}\left[6y - \frac{y^2}{2} - xy\right]_0^2 = \frac{1}{8} \times (12 - 2 - 2y) = \frac{5-y}{4}, 2 < y < 4$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^2 x\left(\frac{3-x}{4}\right)dx = \frac{1}{4}\left[3 \times \frac{x^2}{2} - \frac{x^3}{3}\right]_0^2 = \frac{1}{4} \times \frac{10}{3} = \frac{5}{6}$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_2^4 y\left(\frac{5-y}{4}\right)dy = \frac{1}{4}\left[5 \times \frac{y^2}{2} - \frac{y^3}{3}\right]_2^4 = \frac{1}{4} \times \frac{34}{3} = \frac{17}{6}$$

Define joint and marginal probability density function of a bivariate distribution, two random variables X and Y have the following pdf. **[PU 2003 Fall]**

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

Find: (i) Marginal probability density function of x and y

(ii) V(X) and V(Y)

(iii) Covariance between X and Y

Solution 7

The marginal probability density function of x is given by

$$f_X(x) = \int_{-\infty}^{\infty} x\, f(x, y)\,dy = \int_0^1 x(2 - x - y)\,dy$$

$$= \left[2xy - x^2y - \frac{x \cdot y^2}{2}\right]_0^1 = \frac{3x - 2x^2}{2}, 0 \le x \le 1$$

The marginal probability density function of y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} x\, f(x, y)\,dx = \int_0^1 y(2 - x - y)\,dx$$

$$= \left[2yx - y\frac{x^2}{2} - y^2x\right]_0^1 = \frac{3y - 2y^2}{2}, 0 \le y \le 1$$

Scanned by CamScanner

ii. $V(X) = E(X^2) - (E(X))^2$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left[\int_{-\infty}^{\infty} x\, f_X(x)\, dx\right]^2$$

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

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

$$= \frac{1}{2} \times \frac{7}{20} - \left(\frac{1}{2} \times \frac{1}{2}\right)^2 = \frac{7}{40} - \frac{1}{16} = \frac{9}{80}$$

Similarly, variance of Y i.e $V(Y) = \frac{9}{80}$

iii. We know that,

$Cov(X, Y) = E(XY) - E(X) \times E(Y)$

$$= \int_0^1 \int_0^1 xy\, f(x, y)\, dy\, dx - \int_0^1 x\, f_X(x)\, dx \cdot \int_0^1 y \cdot f_Y(y)\, dy$$

$$= \int_0^1 \int_0^1 xy\,(2 - x - y)\, dy\, dx - \int_0^1 x \times \left(\frac{3x - 2x^2}{2}\right) dx \int_0^1 y \left(\frac{3y - 2y^2}{2}\right) dy$$

$$= \int_0^1 x\left[y^2 - x\frac{y^2}{2} - \frac{y^3}{3}\right]_0^1 dx - \frac{1}{2} \times \frac{1}{2}\left[3\frac{x^3}{3} - 2\frac{x^4}{4}\right]_0^1 \times \left[3\frac{y^3}{3} - 2\frac{y^4}{4}\right]_0^1$$

$$= \int_0^1 \frac{x(4 - 3x)}{6} dx - \frac{1}{4}\left[1 - \frac{1}{2}\right]\left[1 - \frac{1}{2}\right]$$

$$= \frac{1}{6}\left[4\frac{x^2}{2} - 3\frac{x^3}{3}\right]_0^1 - \frac{1}{4} \times \frac{1}{2} \times \frac{1}{2}$$

$$= \frac{1}{6}[2 - 1] - \frac{1}{16} = \frac{5}{48}$$

8. **The joint probability distribution function of two dimensional random variables X and Y is given by** **[PU 2013 Fall/2015 Fall]**

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

Scanned by CamScanner

Find: (i) The marginal probability density function of X and Y.

(ii) The conditional probability density function of X given Y = y and conditional probability density function of given X = x.

(iii) Check for the independent of X and Y.

Solution 8

Do yourself

9. **The joint probability distribution function of two dimensional random variables X and Y is given by** [PU 2014 Spring]

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

i. Find the marginal probability density function of X and Y.
ii. $P(X \geq 1, Y \geq 1)$
iii. Check for the independent of X and Y?

Solution 9

Do yourself

10. **Two random variables X and Y have the following probability density functions:** [PU 2001/2007/2008/2009/2012/2014]

$$f(x, y) = \begin{cases} k(4 - x - y) & 0 < x < 2;\ 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

i. The constant k
ii. Marginal probability density function of X and Y
iii. Conditional probability density function of X, given Y = y
iv. E(X) and Var(X)
v. Are X and Y are independent

Solution 10

Since, the given function is probability density function so we must have

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

or, $$\int_0^2\int_0^2 k(4 - x - y)\, dy\, dx = 1$$

or, $$k\int_0^2 \left[4y - xy - \frac{y^2}{2}\right]_0^2 dx = 1$$

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

or, $$k\,[6x - x^2]_0^2 = 1 \Rightarrow 8k = 1 \Rightarrow k = \frac{1}{8}$$

Scanned by CamScanner

ii. The marginal probability density of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad = \int_0^2 \frac{1}{8}(4 - x - y)\, dy$$

$$= \frac{1}{8}\left[4y - xy - \frac{y^2}{2}\right]_0^2$$

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

The marginal probability density function of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \quad = \int_0^2 \frac{1}{8}(4 - x - y)\, dx$$

$$= \frac{1}{8}\left[4x - \frac{x^2}{2} - xy\right]_0^2 = \frac{3 - y}{4}, \quad 0 < Y < 2$$

iii. The conditional probability of X given is Y = y is

$$f_{X/Y}\left(\frac{x}{y}\right) = \frac{f(x, y)}{f_Y(y)} = \frac{\frac{1}{8}(4 - x - y)}{\frac{1}{4}(3 - y)} = \frac{4 - x - y}{2(3 - y)}, \quad 0 < x < 2; \quad 0 < y < 2$$

Similarly, the conditional probability of Y given X = x is

$$f_{Y/X}\left(\frac{y}{x}\right) = \frac{f(x, y)}{f_X(y)} = \frac{4 - x - y}{2(3 - x)}, 0 < x < 2; 0 < y < 2$$

iv. $$E(X) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^2 x\left(\frac{3 - x}{4}\right) dx = \frac{1}{4}\left[3\frac{x^2}{2} - \frac{x^3}{3}\right]_0^2 = \frac{5}{6}$$

$$V(X) = E(X^2) - (E(X))^2 \quad = \int_0^2 x^2\left(\frac{3 - x}{4}\right) dx - \left(\frac{5}{6}\right)^2$$

$$= \frac{1}{4}\left[3 \times \frac{x^3}{3} - \frac{x^4}{4}\right]_0^2 - \frac{25}{36} = \frac{1}{4} \times 4 - \frac{25}{36} = \frac{11}{36}$$

v. For the independent or dependent of X and Y

$$f_X(x) \times f_Y(y) = \frac{3 - x}{4} \times \frac{3 - y}{4} = \frac{1}{16}[9 - 3x - 3y - xy] \neq f(x, y)$$

So, X and Y are dependent.

11. Define joint and marginal probability density function of a bivariate distribution. Two random variable X and Y have the following joint pdf

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & 0 < x < 1; 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$ **. Find:**

i. Marginal probability density function of X and Y

Scanned by CamScanner

ii. Conditional density of X given Y = y.

iii. Are X and Y independent? [PU 2017 Spring/2012 Fall]

Solution 11

If X and Y are two continuous random variables. Then, the function f(x, y) is said to be joint probability density function if

i. $f(x, y) \geq 0$ for all x, y

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

If f(x, y) be the joint probability density function of two continuous random variables X and Y. Then the marginal probability density function of X is denoted and defined by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{for, } -\infty < x < \infty$$

Similarly, the marginal probability density function Y is denoted and defined by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{for } -\infty < x < \infty$$

i. The marginal probability density function of X

$$f_X(x) = \int_0^1 \frac{2}{3}(x + 2y)\, dy = \frac{2}{3}[xy + y^2]_0^1 = \frac{2(x+1)}{3}, 0 < x < 1$$

Similarly, the marginal probability density function of Y is

$$f_Y(y) = \int_0^1 \frac{2}{3}(x + 2y)\, dx = \frac{2}{3}[x^2/2 + 2yx]_0^1 = \frac{4y+1}{3}, 0 < y < 1$$

ii. The conditional density function of X

$$f_{X/Y}(x/y) = \frac{f(x, y)}{f_Y(y)} = \frac{2/3(x + 2y)}{1/3(4y + 1)} = \frac{2x + 47}{4y + 1}$$

iii. For the dependent or independence of X and Y

$$f_X(x)\, f_Y(y) = \frac{2(x+1)}{3} \times \frac{4y+1}{3} = \frac{2}{9}(4xy + x + 4y + 1) \neq f(x, y)$$

So, X and Y are dependent to each other.

12. Joint probability distribution function of continuous random variables X and Y is given by **[PU 2006 Spring/2010 Spring]**

$$f(x, y) = 4xye^{-(x^2+y^2)}, x \geq 0, y \geq 0$$

i. Test whether X and Y are independent or not.

ii. Find the conditional density of X given Y = y.

Scanned by CamScanner

Solution 12

Do yourself

13. **Define joint probability density function. If the joint probability density of two random variables is given by $f(x_1, x_2) = 6e^{-2x_1 - 3x_2}$, $x_1 > 0, x_2 > 0$**

[PU 2005 Fall/TU 2066/2067]

i. Find the marginal density function of x_1 and x_2.

ii. Are x_1 and x_2 are independent?

Solution 13

Do yourself

14. **The probability function of two random variables X and Y is given by**

$$f(x, y) = \begin{cases} Axy & 0 < x < 1 ; 0 < y < x \\ 0 & \text{otherwise} \end{cases}$$

i. Find the constant A

ii. Find the marginal density function of X and Y

iii. Find the conditional density function of X given Y = y and conditional density function of Y given X = x.

iv. Check for the independent X and Y. [PU 2012 Fall/2016 Spring]

Solution 14

Do yourself.

15. **The probability function of two random variables X and Y is given by**

$$f(x, y) = \begin{cases} 4xy & 0 < x < 1 ; 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

i. verify the f(x, y) is a joint probability density function.

ii. Find the marginal distribution of X and Y.

iii. Find the conditional distribution of X given Y = y.

iv. Find whether X and Y are independent or not.

v. Find $P(0 \le X \le 0.5, 0.5 \le Y \le 1)$

vi. Find covariance between X and Y

[PU 2007 Spring/2009 Fall/2010 Fall/2011 Spring/2013 Spring/2015 Spring /TU 2063/TU 2067]

Solution 15

i. Since, the given function is probability density function. So,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1 \Rightarrow \int_0^1\int_0^1 4xy\, dydx = 1$$

$$\Rightarrow \int_0^1 4x \left[\frac{y^2}{2}\right]_0^1 dx = 1$$

Scanned by CamScanner

$$\Rightarrow \int_0^1 2x\,dx = 1 \Rightarrow [x^2]_0^1 = 1 \Rightarrow 1 = 1$$

Hence, the given function is joint probability function.

ii. The marginal probability density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 4xy\,dy = \left[4x\frac{y^2}{2}\right]_0^1 = 2x, 0 < x < 1$$

$\therefore$ The marginal distribution function of X is $f(x) = \begin{cases} 2x & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$

Similarly, the marginal distribution function of Y is

$$F_Y(y) = f(y) = \begin{cases} 2y & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

iii. The conditional probability density function of X given Y = y is

$$F_{X/Y}\left(\frac{x}{y}\right) = \frac{f(x, y)}{f_Y(y)} = \frac{4xy}{2y} = 2x; 0 < x < 1$$

hence, the conditional distribution function of X given Y = y

$$f_{X/Y}\left(\frac{x}{y}\right) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

iv. For the dependence or independence of X and Y we have

$$F_X(x)\, f_Y(y) = 2x \times 2y = 4xy = f(x, y)$$

Hence X and Y are independent

v. Now, $P(0 \le X \le 0.5, 0.5 \le Y \le 1) = \int_0^{0.5}\int_{0.5}^{1} 4xy\,dy\,dx$

$$= \int_0^{0.5} 2x\,dx \int_{0.5}^{1} 2y\,dy = [x^2]_0^{0.5} \times [y^2]_{0.5}^{1}$$

$$= (0.5)^2 \times [1^2 - 0.5^2]$$

$$= 0.25 \times 0.75 = 0.1875$$

vi. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

$= E(X)E(Y) - E(X)E(Y)$ [$\because$ X and Y are independent]

$= 0$

16. The joint probability function of two discrete random variables X and Y is given by; [TU 2056 Back/2068 Back]

$$\begin{cases} C(2x + y) & 0 \le x \le 2;\ 0 \le y \le 3;\ x, y \in Z \\ 0 & \text{otherwise} \end{cases}$$

i. Find the constant C

ii. Find $P(X \ge 1, Y \le 2)$

iii. Find the marginal function of X and Y

iv. Are X and Y independent?

Solution 16

Please refer solution 15

Scanned by CamScanner

17. **If two random variables have joint density function**

$$\begin{cases} 6/5(x+y)^2 & 0 \le x \le 1;\ 2 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

i. Find $P(0.2 < X < 0.5 \cap 0.4 < Y < 1)$
ii. Verify that the given function is probability density function.
iii. Find $P(0 \le X \le 0.5, 0 \le Y \le 0.5)$
iv. Find the marginal function of X and Y?
v. Find the marginal cumulative distribution for Y.

[TU 20645S/2064 P/2066M/2069 P/2070 M/2068]

Solution 17

Please refer solution 15

Short Questions

18. **Write short note on marginal density function.** **[PU 2005 Spring]**

If f(x, y) is the joint probability density function of continuous random variables X and Y. Then, the marginal probability density function of X is denoted and defined by $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ for $-\infty < y\ \infty$

Similarly, the marginal probability density function of Y is defined by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \text{ for } -\infty < x < \infty$$

19. **Write short note on marginal and joint probability** **[PU 2007 Fall]**

Let X and Y be two random variables defined over the same probabilities space S. The joint probability function of X and Y is denoted and defined by

$$f(x, y) = P(X = x_i, Y = y_j) = P(x_i, y_j) = P_{ij}$$

is called the joint probability function.

Let $P(x_i, y_j)$ be the probability function of random variables X and Y then the marginal probability function of X is denoted and defined by;

$$P_i = P_X(x_i) = P(X = x_i) = \sum_{j=1}^{n} P(x_i, y_j)$$

Also, the marginal probability function of Y is denoted and defined by

$$P_j = P_Y(y_j) = P(Y = y_j) = \sum_{j=1}^{m} P(x_i, y_j)$$

20. **Write short note on joint probability mass function** **[PU 2010 Fall]**

Let X and Y denote the discrete random variables defined on the sample space S. The joint probability function $P(x_i, y_j)$, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ is said to be probability mass function off it satisfies

i. $P(x_i, y_j) \ge 0$ for all i, j

ii. $\sum_{j=1}^{m} \sum_{j=1}^{n} P(x_i, y_j) = 1$

Scanned by CamScanner

13
Chapter

Sampling and Estimation

Sampling method in statistics is as important as salt in food. Sampling theory is a study of relationship that exists between a population and samples which are drawn from the population.

Population:

A complete set or group of objects or things or individuals or item related to the topic of the study is called s population (or universe). Thus the population is the whole aggregate things or objects in which the researcher has to study.
According to the Young "The entire group from which a sample is chosen is known as the population or universe.
The number of objects or items which is included in the population is called the population size and it is denoted by N. There are different types of population.

i. Finite population and infinite population.
ii. Real population and Hypothetical population.

Sample:

A fraction or part of a population selected in some manner from the population for any purpose of investigation is called a sample. It is the finite subset of population. The number of items in the sample is called the sample size and it is denoted by n.

"Census and Sampling"

For the investigation we need much information about the population which can be collected in two ways

i. Census method.
ii. Sampling method.

Scanned by CamScanner

Census:

A study conducted on all elements or units of the population is called census. In other words a complete enumeration of all elements of the population is called the census survey.

Sampling:

When one by one study of all elements of population is not possible due to some factors we take a small representative part from the population for the study. This small representative part selected for the study from the population is called sample. The process of selecting a sample from a population is called sampling.

Parameter and statistic

Parameter:

There are various statistical measures in statistics such as mean, median, standard deviation, proportion, correlation coefficient etc. The statistical measures computed from the population are called parameter. A parameter is denoted by $\hat{\theta}$.

Statistics

Any statistical measures such as mean, median, mode, standard deviation, variance, proportion, etc. computed from the sample data is called statistic. A statistic is denoted by $\hat{\theta}$.

Some useful symbols

S.No.	Parameters	Statistics
i	Population size = N	Sample size = n
ii	Population mean = μ	Sample mean = $\bar{X}$
iii	Population variance = σ^2	Sample variance = s^2
iv	Population s.d = σ	Sample s.d = s
v	Population proportion = p	Sample proportion = $\hat{p}$
vi	Population correlation = ρ	Sample correlation = r

Techniques or Types of Sampling:

The sampling techniques usually depend upon the nature of population under investigation. The sampling techniques commonly used are as follows,

a. Random or probability sampling.
b. Non - random or Non - probability sampling.
c. mixed sampling.

a. **Random or probability sampling:**

The method of sampling in which each unit of the population has some fixed probability of being selected in the sample is called Random or probability sampling. There are various random sampling techniques for selecting sample from a population;

i. Simple random sampling.
ii. Stratified random sampling.
iii. Systematic random sampling.
iv. Cluster random sampling.
v. Multi-stage random sampling.

Scanned by CamScanner

b. **Non - random or Non - Probability Sampling.**

The method of sampling in which each unit in a sample is selected on the basis of personal judgment is called non - random or non - probability sampling. There are various methods for selecting samples from population which are as follows;

i. Haphazard, Accidental or Convenience Sampling.
ii. Judgmental Sampling. iii. Quota Sampling.
iv. Snow - ball Sampling.

Error in statistics

The error involved in collection, processing and analysis of data in a survey may be classified in two parts,

a. Sampling error. b. Non - Sampling error.

a. **Sampling error:**

Usually, different samples selected from the same population will give different results because they obtain different elements. The results obtained from any one sample will generally be different from the result obtained from the corresponding population. The difference between these values is called sampling error.

In another words the difference between the values of a sample statistic and the value of the corresponding population parameter is called sampling error.

i.e. Sampling error $= \bar{X} - \mu$

b. **Non - Sampling error:**

The errors that occur in the collection, recording and tabulation of the data are called the non - sampling error. The non - sampling error is likely to increase with increase in sample size.

Standard Error (S.E)

The standard deviation of the sampling distribution of a statistics is known as the standard error of the statistic standard error of any statistic T is denoted and

defined by $S.E(T) = \sqrt{var(T)} = \sqrt{\frac{\Sigma(T - \bar{T}^2)}{K}}$

"Some Important results"

1. Population mean, $\mu = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{\Sigma X}{n}$
2. Population variance, $\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$
3. Population s.d $\sigma = \sqrt{var} = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$
4. Population proportion $p = \frac{X}{N}$
5. Sample mean $\bar{X} = \frac{\Sigma X_i}{n}$
6. Sample variance $s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n - 1}$
7. Sample proportion $\hat{p} = \frac{X}{n}$

Scanned by CamScanner

8. The total no. of ways of selecting n objects from N objects with replacement is given by N^n.
9. The total no. of ways of selecting n objects from N objects without replacement is given by ${}^NC_n$.
10. Sampling interval (k) $= \dfrac{\text{population size}}{\text{sample size}} = \dfrac{N}{n}$
11. If the samples are drawn from a large population with replacement ($n \geq 30$).
    a. Standard Error of mean S.E $(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$
    b. Standard Error of mean S.E $(\bar{X}) = \dfrac{S}{\sqrt{n}} = \sigma_x$
    c. Standard Error of Sample proportion S.E $(\hat{p}) = \sqrt{\dfrac{pq}{n}}$
       (when population proportion p is known), where, $q = 1 - p$
    d. Standard Error of sample proportion (when population proportion p is not known)
       $$S.E(\bar{X}) = \sigma_{\bar{X}} = \sqrt{\frac{pq}{n}},\ \hat{p} = p,\ \hat{q} = q$$
12. If the samples are drawn without replacement from a finite population.
    a. Standard Error of sample mean $\bar{X}$
       $$S.E(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$
    b. Standard Error of sample proportion p (when population proportion p is known)
       $$S.E(\hat{p}) = \sqrt{\frac{pq}{n}} \times \sqrt{\frac{N-n}{N-1}}$$
    c. Standard Error of sample proportion $\hat{p}$ (when population proportion p is unknown)
       $$S.E(\hat{p}) = \sqrt{pq \cdot \frac{N-n}{N(n-1)}}$$
13. If $\bar{X}_1$ and $\bar{X}_2$ are the means of two random samples of size n_1 and n_2 drawn from two normal populations with S.D σ_1 and σ_2 respectively then
    a. S.E of difference of two sample means is
       $$S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
    b. If the populations have same S.D i.e. $\sigma = \sigma_1 = \sigma_2$ then
       $$S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
    c. If σ_1^2 and σ_2^2 both are unknown then
       $$S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
14. If $\hat{p}_1$ and $\hat{p}_2$ are the proportions of two random samples of sizes n_1 and n_2 drawn from two proportion p_1 and p_2 respectively then
    a. S.E of difference of two sample proportion is

Scanned by CamScanner

$$S.E(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

b. If two population proportion p_1 and p_2 are equal then

$$S.E(\hat{p}_1 - \hat{p}_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

15. S.E of sample standard deviation is given by

$$S.E(s) = \frac{\sigma}{\sqrt{2n}}$$

16. S.E of difference of two Sample Standard deviation

$$S.E(s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

Estimation:

The statistical method of estimating the unknown population parameters from the sample observation drawn from the population is called estimation.
In other words estimation is a procedure by which a numeric value or values are assigned to a population parameter based on the information collected from a sample.

Estimate:

The values assigned to a population parameter based on the value of a sample statistic are called an estimate of the population parameter. There are two types of estimates.

a. Point estimate.
b. Interval estimate.

Point estimate [or point estimation]:

A single value which is used to estimate the population parameter is called the point estimate. For example the sample proportion $\hat{p}$ is the best point estimate of population proportion p. Also sample mean $\bar{X}$ is the best point estimate of the population mean μ.
Thus the procedure of estimating a single value about the value of population parameter with the help of information collected from sample is called point estimation.

Estimator:

The sample statistic that is used to estimate a population parameter is called an estimator. Thus the sampling mean ($\bar{X}$) is an estimator of the population mean (μ)

Criteria or characteristics of good estimators

A good estimator is one which is as close to the true value of the parameter as possible. A good estimator must posses the following characteristics.

i. Unbiasedness
ii. Consistency
iii. Efficiency
iv. Sufficiency

i. **Unbiasedness:**

An estimator $\hat{\theta}$ of θ is said to be unbiased if $E(\hat{\theta}) = \theta$. That is the mean value of the sampling distribution of the statistic is equal to the parameter.

Scanned by CamScanner

Example: Sample mean is am unbiased estimator of population mean.

Proof: Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n, selected from a population with mean μ and variance σ_2. Then the sample mean $\bar{X}$ is given by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The mean of the sample is given by,

$$E(\bar{X}) = E\left[\frac{1}{n}\sum_{i=1}^{n} E(X_i)\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

Since, $E(X_i) = \mu$ for all $i = 1, 2, 3, \ldots, n$

$$\therefore \quad E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

Example:

The sample proportion $\hat{p}$ is an unbiased estimator of the population p

Proof:

Let $X_1, X_2, X_3, \ldots, X_n$ be n independent Bernoulli random variables. The probability $p = X/N$ is called the proportion of success in the population and X being the total number of success in the parent population of size N i.e. $\bar{X} = \sum_{i=1}^{n} = X_i$

Here $X_1, X_2, X_3, \ldots, X_n$ constitute a random sample of size n and the total number of successes in the sample is the sum $\sum_{i=1}^{n} X_i = \bar{X}$, say which in fact follows Binomial distribution with parameter n and p. Therefore $E(\bar{X}) = np$ and $Var(\bar{X}) = npq$.

Now, let $\hat{p} = \frac{X}{n}$ be the proportion of total number of success in the sample i.e.

$$\hat{p} = \frac{\Sigma xi}{n} = \bar{X} \text{ then } E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}\Sigma X = \frac{1}{n} \times np = p$$

Hence the sample proportion **p** is an unbiased estimation of the population proportion p.

Note:

1. Sample variance S^2 is an unbiased estimation of the normal population variation σ^2.
2. However, the sample standard deviation S is biased estimator of the population standard deviation σ i.e. $E(S) \neq \sigma$. For large sample, the bias is negligible.
3. The sample variance defined by $S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_1 - \bar{X})^2$ is biased estimator of the population variance σ^2 because $E(S^2) \neq \sigma^2$.

ii. **Consistency:** A sample statistic $\hat{\theta}$ is known to be constant estimator of the population parameter θ if the difference between $\hat{\theta}$ and θ can be made smaller and smaller by taking the sample size (n) larger and larger.

That is, $\hat{\theta}$ is consistent estimator for θ if $E(\hat{\theta}) \to \theta$ and $V(\hat{\theta}) \to 0$

Scanned by CamScanner

For example: The sample mean $\overline{X}$ is a consistent estimator of the population mean μ.

iii. **Efficiency:** If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimator of a parameter θ such that $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ for all samples of size n, then $\hat{\theta}_1$ is said to be more efficient then $\hat{\theta}_2$. In other words, an estimator with lesser variability is said to be more efficient and consequently more reliable than the other.

iv. **Sufficiency:** An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

Interval Estimate [Interval Estimation]

In interval estimate instead of a single value of a population parameter, an interval is constructed around the point estimate is called the interval estimate. The procedure of estimating a minimum value and maximum value for the value of population parameter with the help of information collected from sample is called interval estimation.

Note: The interval estimate is also called the confidence interval (CI) or fiducial interval. The upper boundary and lower boundary of confidence interval are called upper confidence limit (UCL) and lower confidence limit (LCL) respectively.

Level of Significance

The probability that a calculated interval for a population parameter does not consist of the actual or true value of the population parameter is called the level of significance. It is denoted by α.

Level of Confidence:

The probability that a calculated interval for a population parameter consists of the actual or true value of population parameter is called the level of confidence. It is the complementary probability level of significance. So it is denoted by $1 - \alpha$.

Z-value obtained from standard normal table for different level of confidence.

Nature of Alternative Hypothesis	Level of Confidence (1 – α)									
	99%	98%	97%	96%	95%	94%	93%	92%	91%	90%
	Level of Significance (α)									
	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Two tailed test $Z_{\alpha/2}$	±2.57	±2.33	2.17	2.05	1.93	1.88	1.81	1.75	1.7	1.645
Right tailed test Z_α	2.33	2.05	1.88	1.75	1.645	1.555	1.48	1.41	1.35	1.28
Left tailed test $-Z_\alpha$	–2.33	–2.05	–1.88	–1.75	–1.645	–1.555	–1.48	–1.41	–1.34	–1.28

1. **Confidence interval for normal population mean (μ)**

a. For large sample ($n > 30$)

The confidence interval (C.I) for the population mean μ is given by

C.I for $\mu = \overline{X} \pm t_{\alpha/2}$ S.E. $(\overline{X})$

Where, $S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ or $\frac{S}{\sqrt{n}}$ when N is infinite

$S.E(\overline{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ or $\frac{S}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$, if N is finite.

b. For small sample ($n \leq 30$)

The confidence interval (C.I) for the population mean μ is given by,

Scanned by CamScanner

C.I. for $\mu = \bar{X} \pm t_{\alpha/2\, n-1}$ SE. $(\bar{X})$

Where, S.E $(\bar{X}) = \dfrac{S}{\sqrt{n-1}}$

$t_{\alpha, n-1}$ = t value at α% level of significance and $(n-1)$ degree of freedom.

Remarks: The factor $\sqrt{\dfrac{N-n}{N-1}}$ is called the finite population correction factor or finite population multiplier. The finite population correction factor can be ignored if sampling fraction $\dfrac{n}{N}$ is less than 5% i.e. $n < 5\%\ N$.

2. **Confidence interval for population proportion (p)**

The confidence interval (C.I) for the population proportion p is given by

C.I for $p = p \pm Z_{\alpha/2}$ S.E (p)

Where, S.E. $(p) = \sqrt{\dfrac{pq}{n}}$ or $\sqrt{\dfrac{pq}{n}} \times \sqrt{\dfrac{N-n}{N-1}}$

3. The sample size for estimating mean μ is, $n = \left(\dfrac{Z_\alpha \times \sigma}{E}\right)^2$

Where, Z_α = value of Z at α% level of significance

σ = standard deviation

$E = |\bar{X} - \mu|$ = maximum permissible error

4. The sample size for estimation of population proportion (p) is given by

$$n = \left(\frac{Z_\alpha}{E}\right)^2 \times pq$$

Methods of Estimation

One of the basic problem is how to find an estimator of population parameter θ. There are several methods for finding an estimator of θ which are given below:

i. Method of maximum livelihood estimation
ii. Method of least square
iii. Method of moments
iv. Method of minimum variance
v. Method of inverse probability theory
vi. Method of minimum Chi-square

Method of Maximum Likelihood Estimation

The most general method of estimation is the method of maximum likelihood estimator (MLE).

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population density function $f(x, \theta)$. Then the likelihood function of the sample values $x_1, x_2, \ldots, x_n$ usually denoted by,

$L = L(\theta)$ is their joint density function given by,

$$L = f(x_1, \theta) \times f(x_2, \theta) \times \ldots \times f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Note: The principle of maximum likelihood is selecting that values as the estimator of θ which maximizes $L(\theta)$. If $\hat{\theta}$ is called MLE of θ. In this case, we must have

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L(\theta)}{\partial \theta^2} < 0$$

Scanned by CamScanner

Exam Asked Question

Long Question

1. Hotel's manger in Kathmandu wants to know the hotels average daily registrations. The following table presents the number of guest registered each of 27 randomly selected days. Calculate the sample mean and 95% confidence limits of population mean. PU 2016 Spring]

61	57	53	60	64	57	54	58	63
61	50	59	50	60	57	58	62	63
60	54	54	61	51	53	62	57	60

Solution 1

We know that the above data can be written as .

x	f	fx	fx^2
50	2	100	5000
51	1	51	2601
53	2	106	5618
54	3	162	8748
57	4	228	12996
58	2	116	6728
59	1	59	3481
60	4	240	14400
61	3	183	11163
62	2	124	7688
63	2	126	7936
64	1	64	4096
	N = 27	$\Sigma fx = 1359$	$\Sigma fx^2 = 90457$

We know that,

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{1359}{27} = 50.33$$

Also, $S = \sqrt{\frac{\Sigma fx^2 - n(\bar{X}^2)}{n-1}} = \sqrt{\frac{90457 - 27\,(50.33)^2}{27.1}} = 29.13$

Now, the standard error of mean

$$S.E(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{29.13}{\sqrt{27}} = 5.6$$

Since $\alpha = 5\% = 0.05$

Now the 95% confidence limits for the population mean,

$= \bar{X} \pm t_{(\alpha/2,\, n-1)} \cdot S.E(\bar{x})$

$= 50.332 \pm t_{(0.025,\, 26)} \times 5.6$

$= 50.33 \pm 2.565 \times 5.6$

$= 50.33 \pm 2.056 \times 5.6$

Lower confidence interval = 50.33 – 11.51 = 38.82

Upper confidence interval = 50.33 + 11.51 = 61.84

2. A random sample of 8 envelopes is taken from letter box of a post office and their weight in grams are found to be 12.1, 11.9, 12.3, 11.9, 12.1, 12.4, 12.1 and

Scanned by CamScanner

11.5. Find 99% fiducial limits for the mean weight of envelops received at that post office. [P.U. 2017 Spring]

Solution 2

Here given the weights (in gms) of envelops is 12.1, 11.9, 12.3, 11.9, 12.1, 12.4, 12.1, and 11.5.

Also, $1 - \alpha = 99\% = 1 - \alpha = 0.99 \quad = \alpha = 0.01$

Now, calculation of mean and standard deviation are as follows:

x_i:	11.5	11.9	11.9	12.1	12.1	12.1	12.3	12.4	$\Sigma x_i = 963.3$
x_i^2:	132.25	141.61	141.61	146.41	146.41	146.41	151.29	153.76	$\Sigma X^2 = 1159.75$

Now, $\bar{X} = \frac{\Sigma x_i}{n} = \frac{96.3}{8} = 12.0375$

$$S = \sqrt{\frac{\Sigma x_i - n(\bar{x})^2}{n-1}} = \sqrt{\frac{1159.75 - 8 \times (12.375)^2}{8-1}} = 0.077$$

Now for 99% fiducial limits for the mean weight,

Lower limit of the confidence intervals is

$$\bar{x} - t_{(\alpha/2, n-1)} \times \frac{S}{\sqrt{n}}$$

$$= 12.0375 - t_{(0.01/2, 7)} \times \frac{0.077}{\sqrt{8}} = 12.0375 - \frac{3.355 \times 0.77}{2.83} = 12.0375 - 0.091 = 11.95$$

Upper limit of the confidence interval is

$$\bar{x} + t_{(\alpha/2, n-1)} \times \frac{S}{\sqrt{n}} = 12.0375 + 0.091 = 12.13$$

Hence the required limit is [11.95, 12.13]

3. **From a population of 540, a sample of 60 individuals is taken. From this sample, the mean is found to be 6.2 and the standard deviation is 1.368.**
   (i) Find the estimated error of the mean
   (ii) Construct 96% confidence interval for mean. [PU 2016 Fall/PU 2015 Spring]

Solution 3

Here we have given

Sample size (n) = 60 Population size (N) = 540

Sample mean ($\bar{x}$) = 6.2 Sample sd. (s) = 1.368

i. The estimated standard error of the mean is given by,

$$S.E(\bar{x}) = \frac{S}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{1.368}{\sqrt{60}} \times \sqrt{\frac{540-60}{540-1}} = \frac{1.368}{7.75} \times 0.94 = 0.17$$

ii. Level of confidence $(1 - \alpha) = 96\% = 0.96$

$\therefore \quad \alpha = 1 - 0.96 = 0.04$

Now, $\frac{\alpha}{2} = 0.02$

Now, the 96% confidence interval for mean is given by

$$= Z_{\alpha/2} \times S.E(\bar{X}) = 6.2 \pm 2.05 \times 0.17 = 6.2 \pm 0.35$$

Scanned by CamScanner

Lower limit = 6.2 – 0.35 = 5.85

Upper limit = 6.2 + 0.35 = 6.55

Therefore, the 96% confidence interval for the average or mean population is 5.85 to 6.55.

4. **A population consist of the value 7, 6, 8, 4, 10,. Prove that sample mean of size 3 is unbiased estimator of the population mean.** **[P.U 2015 Fall]**

Solution 4

In population there are 5 units. The population mean is

$$\mu = \frac{\Sigma x_i}{n} = \frac{7+6+8+4+10}{5} = 7$$

We have, 3 units can be selected from 5 units in C(5, 3) = 10 ways

Sample No.	Observations	Sample mean $\bar{X}_i$	$P(\bar{x}_i)$
1	7, 6, 8	(7 + 6 + 8)/3 = 7	1/10
2	7, 6, 4	17/3	1/10
3	7, 6, 10	23/3	1/10
4	7, 8, 4	19/3	1/10
5	7, 8, 10	25/3	1/10
6	7, 4, 10	7	1/10
7	6, 8, 4	6	1/10
8	6, 4, 10	8	1/10
9	6, 8, 10	20/3	1/10
10	8, 4, 10	22/3	1/10
		$\Sigma X_i = 70$	

Now, $E(X) = \sum_{i=1}^{10} \bar{x}_i P(\bar{x}_i) = \frac{1}{10} \times 70 = 7$

$\therefore$ $E(\bar{X}) = 7 = \mu$

This gives sample mean is unbiased estimator of population mean.

5. **400 orange are taken from a large consignment and 50 are found to be bad. Estimate the percentage bad orange in the consignment and assign 95% confidence limit.** **[P.U 2014 Fall]**

Solution 5

Here, we have given

Sample size (n) = 400 Number of bad oranges (x) = 50

Sample proportion $(P) = \frac{x}{n} = \frac{50}{400} = 0.125$

The percentage of bad orange in the consignment is estimated by $\hat{P}$ in percentage because of $E(\hat{P}) = P$

Therefore the required percentage = 0.125 × 100% = 12.5%

Scanned by CamScanner

For the 95% confidence level

Now, the standard error of the sample proportion is

$$S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.125 \times 0.875}{400}} = 0.017$$

Now, the level of confidence $(1 - \alpha) = 95\% = 0.95$

$\therefore \quad \alpha = 0.05$

Now, the 95% confidence interval for the population proportion is

C.I for p $= p \pm Z_\alpha \times S.E(p) = 0.125 \pm 1.96 \times 0.017 = 0.125 \pm 0.03$

Lower limit $= 0.125 - 0.03 = 0.095$

Upper limit $= 0.125 + 0.03 = 0.155$

Hence, the percentage of bad oranges in the consignment lies almost certainly between 9.5 and 15.5.

6. **A sample poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favor of a particular candidate. Find the 95% confidence limits for the proportion of all the voter in the favor of this candidate.** **[P.U 2014 Fall]**

Solution 6

We have given sample size $(n) = 100$

Sample proportion $(p) = 0.55$

Then, $q = 1 - p = 1 - 0.55 = 0.45$

Now, standard error of sample proportions is

$$S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.55 \times 0.45}{100}} = 0.0497$$

For level of confidence $(1 - \alpha) = 95\% = 0.95$

$\therefore \quad \alpha = 0.05$

Also, the Z value at $\alpha = 0.005$ is

$Z_\alpha = Z_{0.05} = 1.96$

The 95% confidence interval for the population proportion is

C.I for p $= p \pm Z_\alpha \times S.P(p) = 0.55 \pm 1.96 \times 0.0497 = 0.55 \pm 0.097$

Lower limit $= 0.55 + 0.097 = 0.453$

Upper limit $= 0.55 + 0.097 = 0.647$

Therefore, the 95% confidence interval for the population proportion of all voters is 0.453 to 0.647.

7. **What is estimation? What are the criterias to be a good estimator? A researcher wishes to estimate the mean of a population by using sufficiency large sample. The probability is 0.95 that the sample mean will not differ from the true mean by more than 25%. How large a sample should be taken?** **[P.U 2011 Spring]**

Solution 7

Estimation: The statistical method of estimating the unknown population parameters from the sample observations drawn from the population is called estimation.

Scanned by CamScanner

E.g. The civil engineer estimates the strength of rod, brick, etc. from the sample; a businessman estimates his future sales from the past records. The estimation of population parameters can be done either of the two ways:

i. Point estimation ii. Interval estimation

Second part

Using a sufficiently large sample, the mean of a population can be estimated by interval estimation. By question we have given the probability = 0.95

Also, the sample mean will not differ from the true mean by more than 25% of σ.

That is, $P(|\overline{X} - \mu|) \leq 25\%$ of $\sigma = 0.95$ ...(i)

But, we know that, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Therefore, $P[|\overline{X} - \mu| \leq 1.96 \frac{\sigma}{\sqrt{n}}] = 0.95$ ...(ii)

From equation (i) and (ii), we have

$$1.96 \times \frac{\sigma}{\sqrt{n}} = 25\% \text{ of } \sigma$$

or, $$\frac{1.96 \times \sigma}{\sqrt{n}} = \frac{25}{100} \times 6$$

$$\therefore \quad \sqrt{n} = \frac{196}{25}$$

$$\Rightarrow \quad n = 61.45 = 62$$

Hence, in order to estimate the population means μ a sample of size 62 should be taken.

8 **In a random sample 136 of 400 person given a flue vaccine experienced some discomfort; construct a 98% confidence interval for true proportion of persons who will experiences some discomfort from the vaccine.** **[P.U 2013 Fall]**

Solution 8

Here, given total no. of persons in sample (n) = 400

No. of flue vaccine experience some discomfort (X) = 136

Sample proportion $(p) = \frac{X}{n} = \frac{136}{400} = 0.34$

$q = 1 - p = 0.66$

Now, the standard error of sample proportion is

$$S.E\ (p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.34 \times 0.66}{400}} = 0.024$$

Now, for 98% of confidence

$1 - \alpha = 98\% = 0.98 \Rightarrow \alpha = 0.02$

Then, $Z_\alpha = 2.33$

Now, confidence interval for population proportions is given by

$$P \pm Z_\alpha\ S.E\ (P) = 0.34 \pm 2.33 \times 0.024 = 0.34 \pm 0.056$$

Scanned by CamScanner

Lower limit = 0.34 – 0.056 = 0.284
Upper limit = 0.34 + 0.056 = 0.396
Hence, the required confidence interval is (0.284, 0.396)

9. Define point estimation. Name the properties to be followed by a good estimator. A population consists of the values 3,6,8,9,12 and 15. Draw the sample of size 3 and show that the sample mean is an unbiased estimator of the population mean. **[PU 2003 Fall]**

Solution 9

Point estimation: The procedure of estimating a single value about the value of population parameter with the help of information collected from sample is called the point estimation. For example, suppose the census bureau takes a sample of 1,000 households and determines that the mean of housing expenditure per month ($\bar{X}$) for this sample is Rs.10,000. Then, using sample mean ($\bar{X}$) as a point estimate for population mean (μ) the bureau can state that the mean housing expenditure per month. This process is called point estimation.

Properties of good estimator

A good estimator is one which is one which is as close to the value of the parameter as possible. The properties of good estimator are mentioned below:

i. Unbiasedness ii. Consistency
iii. Efficiency iv. Sufficiency

Second part

Here, population size (N) = 6
Sample size (n) = 3
The number of possible samples of size 3 that can be drawn from the no. of possible sample = ${}^{N}C_n = {}^{6}C_3 = 20$

Now, the population mean (μ) = $\frac{3+6+8+9+12+15}{6} = 8.83$

S.N.	Observations	Sample mean ($\bar{X}_i$)	$P(\bar{x}_i)$
1	(3, 6, 8)	17/3	1/20
2	(3, 6, 9)	6	1/20
3	(3, 6, 12)	7	1/20
4	(3, 6, 15)	8	1/20
5	(3, 8, 9)	20/3	1/20
6	(3, 8, 12)	23/3	1/20
7	(3, 8, 15)	26/3	1/20
8	(3, 9, 12)	8	1/20
9	(3, 9, 15)	9	1/20
10	(3, 12, 15)	10	1/20
11	(6, 8, 9)	23/3	1/20
12	(6, 8, 12)	26/3	1/20
13	(6, 8, 15)	29/3	1/20
14	(6, 9, 12)	9	1/20
15	(6, 9, 12)	10	1/20
16	(6, 12, 15)	11	1/20
17	(8, 9, 12)	29/3	1/20
18	(8, 9, 15)	32/3	1/20
19	(8, 12, 15)	35/3	1/20
20	(9, 12, 15)	12	1/20

Scanned by CamScanner

New, $E(\bar{X}) = \sum_{i=1}^{20} \bar{x}_i P(\bar{x}_i) = \frac{1}{20} \times 176.67 = 8.83$

Since, $E(\bar{x}) = \mu = 8.83$

Thus, the sample mean is unbiased estimator of population mean.

10. **Dr. Sharma, a noted social psychologist, surveyed 150 top executives and found that 42 percent of them were unable to add fractions correctly.**

i. Estimate the standard error of the proportion

ii. Construct a 99 percent confidence interval for the true proportion of top executives who cannot correctly add fractions. [PU 2007 Fall]

Solution 10

We have given that,

Sample size (n) = 150

Proportion of top executives who can not correctly add fractions p = 42% = 0.42

Then q = 1 – p = 1 – 0.42 = 0.58

i. The standard error of the proportion is $S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.42 \times 0.58}{150}} = 0.04$

ii. For the 99% of confidence

$1 - \alpha = 99\% = 0.99 \quad \Rightarrow \alpha = 0.01$

Then, $Z_\alpha = 2.58$

Now, the confidence interval for the true proportion of the executives who can not correctly add fractions is

CI for P $= P \pm Z_\alpha S.E(P) = 0.42 \pm 2.58 \times 0.04 = 0.42 \pm 0.1032$

Lower limit = 0.423 – 0.1032 = 0.3168

Upper limit = 0.42 + 0.1032 = 0.5232

Hence, the required interval is [0.3168, 0.5232]

11. **What do you mean by unbiased estimator? Consider a population of 5 units with 4, 2, 10, 6 and 8. Show that sample mean of size 2 is unbiased estimator of population mean.** [PU 2006 Spring]

Solution 11

Unbiased estimator: A statistic said to be an unbiased estimator of the population parameter θ if $E(\hat{\theta}) = \theta$.

i.e. If the mean value of the sampling distribution of the test statistic θ is equal to the population parameter θ. Whatever the value of θ, Such estimator $\hat{\theta}$ is called unbiased estimator.

2nd part:

Here, given population size (N) = 5

Sample size (n) = 2

Then the population mean $(\mu) = \frac{4+2+10+6+8}{5} = 6$

Scanned by CamScanner

The number of possible sample of size 2 that can be drawn from the population size 5 is given by C (N,n) = C (5,2) = 10

S.N	Observations	Sample mean ($\bar{X}_i$)	$P(x_i)$
1.	(4,2)	3	1/10
2.	(4,10)	7	1/10
3.	(4,6)	5	1/10
4.	(4,8)	6	1/10
5.	(2,10)	6	1/10
6.	(2,6)	4	1/10
7.	(2,8)	5	1/10
8.	(10,6)	8	1/10
9.	(10,8)	9	1/10
10.	(6,8)	7	1/10
		$\Sigma(\bar{X}_i) = 60$	

$$E(\bar{X}_i) = \sum_{i=1}^{20} \bar{x}_i P(\bar{x}_i) = 1/10 \times 60 = 6$$

Since $E(\bar{X}_i) = \mu = 6$ so it is an unbiased estimator of population mean.

12. **Define interval estimation. A random sample of 100 patients suffering from a certain disease was given serum treatment. It was observed that 75% were released of the disease. Find 95% confidence limit for the % of patient cured.** **[PU 2005 FALL]**

Solution 12

The process is similar to Q.No. 4

13. **What do you mean by unbiased estimator? Consider a population of 5 units with 4, 2, 10, 6 and 8. Show that sample mean of size 2 is unbiased estimator of population mean.** **[PU 2006 SPRING]**

Solution 13

The process is similar to Q.No. 4

14. **A population consists of value 1, 2, 3, 6, 8. Prove that the sample mean of size 3 is unbiased estimator of population mean.** **[PU 2007 FALL]**

Solution 14

The process is similar to Q.No. 4

Short Answer Questions

15. **Write short note on Population and Sample** **PU 2009 Spring**

Solution 15

See in the theoretical part of this chapter.

16. **Write short note on Point and interval estimate** **PU 2005Fall/2006Spring**

Solution 16

See in the theoretical part of this chapter.

17. **Write short note on Sample space** **PU 2011Spring/2005 Spring/2017Spring**

Solution 17

See in the theoretical part of this chapter.

18. **Write short note on Central Limit Theorem** **PU 2012 Fall/2010Spring/2013Fall**

Solution 18

See in the theoretical part of this chapter.

Scanned by CamScanner

19. **Write short note on Criteria/Chrematistics for good estimator**
PU 2013 Spring /2015 Fall/2015S/pring2016Fall/2016Spring /2017Spring /2010Spring

Solution 19

See in the theoretical part of this chapter.

20. **Write short note on Probability Sampling** **PU 2014 Spring**

Solution 20

See in the theoretical part of this chapter.

21. **Write short note on Probability and non-probability sampling** **PU 2017 Fall**

Solution 21

See in the theoretical part of this chapter.

22. **Write short note on Confidence interval** **PU 2017Spring /2009Fall**

Solution 22

See in the theoretical part of this chapter.

23. Write short note on Confidence level PU 2005 Fall

Solution 23

See in the theoretical part of this chapter.

24. Write short note on Parameter and statistic PU2003 Fall/2006Spring /2010Fall

Solution 24

See in the theoretical part of this chapter.

Additional Questions for Practice

1. A certain cinema hall sampled 55 people who viewed "DHOOM 3" and asked them whether they planned to see it again. Only 10 of them believed the film was worthy of second look.
   (i) Estimate the standard error of the proportion of movie viewer who will viewed the film a second time.
   (ii) Construct a 90% C.I for this proportion. [P.U 2013 Spring]
2. General Cinema sampled 55 people who viewed Ghost hunter 8 and asked them whether they planned to see it again. Only 10 of them believed the film was worthy of a second look.
   i) Estimate the standard error of the proportion of moviegoers who will view the film a second time.
   ii) Construct 95 percent confidence intervals for this proportion.
   PU 2009 SPRING
3. What do you mean by efficient estimator? A sample of 900 members has a mean 3.4 cm and standard deviation 2.61. If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of population mean.
   PU 2009 FALL
4. Ten objects are chosen on random from the large population and their weights are found to be in gms 63, 64, 65, 66, 69, 69, 70, 70 and 71. In the light above that, discuss the suggestion that the mean weight in the universe is 65 gms. Also construct 95% fudicial limits for the population mean. PU 2010 FALL
5. A random sample of 900 items is drawn from a normal population whose mean is 3.25 cm. and whose standard deviation is 2.61 cm. If the sample mean is 3.4 cm.

Scanned by CamScanner

Can the sample be regarded as a true random sample at 5% level of significance? Also construct the 95% confidence interval for the true mean in the population.

PU 2003 FALL

6. Define interval estimation. A random sample of 100 patients suffering from a certain disease was given serum treatment. It was observed that 75% were released of the disease. Find 95% confidence limit for the % of patient cured.

PU 2005 Fall

7. Dr. Sharma, a noted social psychologist, surveyed 150 top executives and found that 42 percent of them were unable to add fractions correctly.

   i. Estimate the standard error of the proportion

   ii. Construct a 99 percent confidence interval for the true proportion of top executives who cannot correctly add fractions. PU 2007 FALL

8. The quality control manager at a light bulb factory needs to estimate the mean life of a large shipment of life bulbs. The process standard deviation is known to be 100 hours. A random sample of 64 light bulbs indicated a sample mean life of 350 hours.

   i. Obtain the standard error of the mean.

   ii. Set up 95% confidence interval of the true population mean life of bulbs in this shipment. [PU2010 SPRING]

Do yourself

9. Define point estimation. Name the properties to be followed by a good estimator. A population consists of the values 3,6,8,9,12 and 15. Draw the sample of size 3 and show that the sample mean is an unbiased estimator of the population mean.

PU 2003 FALL

Scanned by CamScanner

14
Chapter

Test of Hypothesis

Hypothesis:

A Hypothesis is a tentative assertion or idea or assumption about the parameters of a population.

Once hypothesis is formed, we must test it to decide the validity of the hypothesis whether it has to be accepted or rejected. The only information which helps us to take decision is contained in the sample. Based on sample information at hand, we make a decision about the population parameter.

Testing of Hypothesis:

The Testing of Hypothesis is a statistical procedure that involves the formulating hypotheses and testing of validity of these hypotheses on the basis of sample information.

Types of hypothesis: There are two types of hypothesis.

1. Null Hypothesis
2. Alternative Hypothesis
3. Null hypothesis

The assumption no significant difference between actual value and claimed value of a population parameter is called null hypothesis. Thus null hypothesis is a hypothesis of no difference. It represents the default possibility that we will accept unless we have convincing evidence to the contrary. The null hypothesis is denoted by H_0. Suppose a population has a specified mean value (say μ_0) and actual mean value (μ) then the null hypothesis is set up as

H_0: $\mu = \mu_0$ i.e. there is no significant difference between actual mean (μ) and specified mean (μ_0). In other words

Alternative hypothesis

A complementary statement of null hypothesis is called alternative hypothesis. In other words, the assumption of significant difference between actual/ true value and claimed/ specified value of a population parameter is called alternative hypothesis. It represents the conclusion or decision taken if the null hypothesis is rejected. The alternative hypothesis is denoted by H_1. Suppose a population has a specified mean value (say μ_0) and actual mean value (μ) then the null hypothesis may be

Scanned by CamScanner

$H_1: \mu \neq \mu_0$ i.e. there is significant difference between actual mean (μ) and specified mean (μ_0).

$H_1: \mu < \mu_0$ i.e. population mean is less than the specified value μ_0.

$H_1: \mu > \mu_0$ i.e. population mean is more than the specified value μ_0.

Risks or errors in hypothesis testing

There are two ways that we could be wrong when we perform a hypothesis test, which is presented in the following table

Two Types of errors

Statistical Decision	Actual Situation	
	H_0 is True	H_0 is False
Accept H_0	Correct Decision	Type II error
Reject H_0	Type I error	Correct Decision

Thus, there are two types of errors in decision-making process with testing of hypothesis. i) Type I error ii) Type II error

Type I error

The error committed in rejecting true null hypothesis is known as Type I error. A Type I error is made when we reject the null hypothesis when the null hypothesis is actually true. The probability of making Type I error is denoted by Greek letter alpha (α). That is, α = P[Reject H_0 when H_0 is true] = P[Reject H_0 | H_0 is true]

Since producers have to bear this kind of risk, the Type I error is also known as *Producer's Risk.*

Type II error

The error committed in accepting false null hypothesis is known as Type II error. A Type II error is made when we fail to reject the null hypothesis when the null hypothesis is false. The probability of making Type II error is denoted by Greek letter beta (β).That is, β = P[Accept H_0 when H_0 is false] = P[Accept H_0 | H_0 is false]

Since consumers have to bear this kind of risk, the Type II error is also known as *Consumer's Risk.*

Level of significance

The probability of committing a Type I error is referred to as the level of significance of the statistical test. It is denoted by α. The probability of making a correct decision is $(1 - \alpha)$.The most commonly used level of significance in practice is 5%. If we are using level of significance as 5%, it means that probability of committing Type I error is 0.05. It also means that we are 95% confident that a correct decision has been made.

The complimentary probability of level of significance is called *level of confidence.* In other words, the complimentary probability of Type I error is called level of confidence. More precisely, it is the probability that the null hypothesis H_0 is accepted when in fact it is true. The level of confidence of a hypothesis test is denoted by $(1 - \alpha)$.

Scanned by CamScanner

Test Statistic

A test statistic is a value which is calculated from the sample data. The value of test statistic is used to decide whether the null hypothesis should be accepted or rejected in our hypothesis test.

The value of test statistic is calculated by using the following formula

$$\text{Test Statistic} = \frac{\text{Sample Statistic} - \text{Population Paramter}}{\text{Standard Error of sample statistic}}$$

The choice of a test statistic is guided by the sample size and the nature of the hypothesis testing.

Critical Region

The sampling distribution of the test statistic is divided into two regions – a region of rejection called critical region and a region of acceptance. If the test statistic falls into the region of acceptance, a null hypothesis is accepted, If the value of test statistic lies in the region of rejection, then null hypothesis is rejected and so alternative hypothesis is accepted. The value that divides the region of rejection and the region of acceptance is called critical value. The critical value can be obtained from the corresponding standard table of the test statistic.

Tails of a test in testing of hypothesis

A test having region of rejection on both sides of the distribution curve of a test statistic is called *Two Tailed Test*.

A test having region of rejection on either (left or right) side of the distribution curve of a test statistic is called *One Tailed Test*. A one tailed test is called *Left Tailed Test* if the region of rejection lies in the left side of distribution curve of a test statistic. Similarly a one tailed test is called *Right tailed* if the region of rejection lies in the right side of the distribution curve of a test statistic.

Remark:

i. Some most commonly used words which are useful to detect one tailed test and two tailed test are as follows:

ii. **For two tailed test:** change, difference, differ, equal to, same, exactly, neither less nor more, unbiased, biased etc.

iii. **For one tailed test:** less than, more than, greater, smaller, taller, shorter, lower, higher, bigger, only, silently claim, over, under, above, below, increase, decrease, at least, at most, superior, inferior, majority, minority, better, effective, enhance, improve, positively biased, negatively biased etc.

Procedure in testing of hypothesis

The following five steps are involved in hypothesis testing problems.

Setting up hypotheses: Set null hypothesis and alternative hypothesis as

Null hypothesis: H_0: $\theta = \theta_0$ i.e. the actual value of population parameter is equal to specified value θ_0.

Alternative hypothesis: Set any one of the following hypothesis according to nature of the problem of hypothesis testing.

Scanned by CamScanner

H_1: $\theta \neq \theta_0$ i.e. the actual value of population parameter is not equal to specified value θ_0. [Two tailed test]

H_1: $\theta < \theta_0$ i.e. the actual value of population parameter is less than specified value θ_0. [Left tailed test]

H_1: $\theta > \theta_0$ i.e. the actual value of population parameter is more than specified value θ_0. [Right tailed test]

Level of significance (α): The level of significance (α) is fixed in advance. The most commonly used level of significance is 5%.

Test Statistic: Under the null hypothesis, calculate the value of appropriate test statistic by using sample information as

$$\text{Test Statistic} = \frac{\text{Sample Statistic} - \text{Population Paramter}}{\text{Standard Error of Sample Statistic}}$$

It is called calculated value.

Critical Value: The critical or tabulated value of the test statistic is obtained from the corresponding standard table according to pre determined level of significance and tail of the test.

Decision: The last step in the testing of hypothesis is to make a decision about the null hypothesis, i.e., whether to accept it or to reject it. The decision is made by comparing the calculated value of test statistic with the critical value of the test statistic and decides as under:

If calculated value of test statistic lies in the region of acceptance, null hypothesis is accepted. In other words, if the calculated value of the test statistic is less or equal to ($\leq$) tabulated value, then accept the null hypothesis.

If calculated value of test statistic lies in the region of rejection, null hypothesis is rejected and alternative hypothesis is accepted. In other words, if the calculated value of the test statistic is greater than (>) tabulated value, then accept the alternative hypothesis.

Test of Significance for large samples (Z-test)

The sample is considered as large if $n > 30$. The large sample is generally desirable when the units in the population under study are not homogeneous or uniform. To get more reliable results about the population parameter, small sample is not sufficient for heterogeneous population, so large sample test has to be carried out.

Assumptions of Z test

The Z test is used under the following assumptions:

The sample size is large ($n > 30$).

The population from which the samples are drawn is normally distributed.

The population standard deviation is known.

The samples are independent.

Applications of Z test

The Z test has a large number of applications in management science. Some of the applications are:

Scanned by CamScanner

a. Z test for mean of variables
b. Test of significance of a single mean
c. Test of significance of difference between two means
d. Z test for proportion of attributes
e. Test of significance of single proportion
f. Test of significance of difference between two proportions

Test of significance of a single mean

Suppose that the sample of size n drawn from a normal population with mean μ and variance σ^2. The sample mean $\bar{X}$ is also a normal variate with mean μ and variance $\frac{\sigma^2}{n}$. Hence, the standardized normal variate Z corresponding to the statistic $\bar{X}$ is given by $Z = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

It is used to test whether the population mean has a specified value or not. The procedure for testing the significance of a single mean is as follows:

Step 1: Setting up Hypotheses

Null hypothesis: H_0: $\mu = \mu_0$ i.e. the actual value of population mean is equal to specified value μ_0 or there is no significant difference between sample mean ($\bar{X}$) and population mean (μ) or there is no significant difference between actual value (μ) and specified value (μ_0) of population mean or sample has been drawn from a population with mean μ_0.

Alternative hypothesis: Set any one of the following hypothesis according to nature of the problem of hypothesis testing.

H_1: $\mu \neq \mu_0$ i.e. the actual value of population mean is not equal to specified value μ_0 or there is significant difference between actual value (μ) and specified value (μ_0) of population mean or sample has not been drawn from the population with mean μ_0. [Two tailed test]

H_1: $\mu < \mu_0$ i.e. the actual value (μ) of population mean is less than specified value μ_0. [Left tailed test]

H_1: $\mu > \mu_0$ i.e. the actual value (μ) of population mean is more than specified value μ_0. [Right tailed test]

Step 2: Level of significance (α)

Choose the appropriate level of significance in advance. The most commonly used level of significance is $\alpha = 5\%$ unless otherwise stated.

Step 3: Test Statistic The test statistic under H_0

$$Z = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

However if σ is unknown, then we use estimate of population standard deviation provided by the sample standard deviation, i.e., $\hat{\sigma} = s$. Then,

$$Z = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

Scanned by CamScanner

Step 4: Critical Value

The critical or tabulated value of the test statistic Z at the pre-specified level of significance is obtained from the area under normal table.

Table of the Normal Distribution

(Critical Values Z_α of Z)

Nature of Alternative Hypothesis	Level of confidence (1 - α)				
	99%	98%	96%	95%	90%
	Level of significance (α)				
	1%	2%	4%	5%	10%
Two tailed test	± 2.57	± 2.33	± 2.05	± 1.96	± 1.64
Right tailed test	2.33	2.05	1.75	1.64	1.28
Left tailed test	– 2.33	– 2.05	– 1.75	– 1.64	– 1.28

Step 5: Decision

If the calculated value of Z (Z_{cal}) is less or equal to (≤) the tabulated value of Z (Z_{tab}), H_0 is accepted which means the population mean has a specified value μ_0. In other words, there is no significant difference between sample mean and the population mean, or, the sample has been drawn from a normal population with population mean μ_0.

If the calculated value of Z (Z_{cal}) is greater than (>) the tabulated value of Z (Z_{tab}), H_1 is accepted which means the mean of population has not a specified value μ_0. In other words, there is significant difference between sample mean and the population mean, or the sample has not been drawn from a normal population with population mean μ_0.

Remarks:

Confidence limits for population mean (μ): The (1 – α)% confidence limits for the population mean μ is given by

$$\text{C. I. for } \mu = \bar{X} \pm Z_\alpha \text{ S.E.}(\bar{X}) = \bar{X} \pm Z_\alpha \frac{\sigma}{\sqrt{n}} = \bar{X} \pm Z_\alpha \frac{s}{\sqrt{n}}.$$

Test of significance of difference between two means

Suppose that two independent random samples of size n_1 and n_2 drawn from two populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. Let $\bar{X}_1$ and $\bar{X}_2$ be the corresponding sample means. Then under the null hypothesis

H_0: $\mu_1 = \mu_2$, the formula becomes

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

If the population variances σ_1^2 and σ_2^2 are unknown, then sample variances s_1^2 and s_2^2 are used instead of population variances σ_1^2 and σ_2^2 respectively. The test statistics under null hypothesis becomes

Scanned by CamScanner

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$$

Remarks: If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then the test statistic under null hypothesis becomes

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

If σ^2 is not known, then its estimate is based on both sample variances and is given by $\hat{\sigma}^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$

Remarks:

The test statistic under H_0 is $Z = \frac{p - E(p)}{S.E.(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$ for large samples

Where, S.E.(p) = Standard error of the sample proportion $= \sqrt{\frac{PQ}{n}}$

p = Sample proportion of success $= \frac{x}{n}$

x = Number of success possessing the given attribute in sample

n = Sample size (Number of items in the sample).

P = Population proportion of success possessing the given attribute

$Q = 1 - P$ = Population proportion of failure not possessing the given attribute

If we have drawn sample from finite population of size N, the standard error of sample proportion is given by

S.E. (p) $= \sqrt{\left(\frac{N-n}{N-1}\right)\frac{PQ}{n}}$, then the test statistic in testing the single population proportion is $Z = \frac{p - P}{\sqrt{\left(\frac{N-n}{N-1}\right)\frac{PQ}{n}}} \sim N(0, 1)$

As we have already familiar with the estimation of population proportion in estimation – chapter 4, the confidence interval for population proportion (P) is given by

C.I. for $P = p \pm Z_\alpha \times S.E.(p)$

Test of significance of difference between two proportions

Let us consider two independent populations each having population proportion P_1 and P_2 of the certain attributes or characteristics. Now, we want to test whether there is any difference between the proportions of these populations with respect to this attribute..

Scanned by CamScanner

The procedure for testing the significance of difference between two proportions is as follows:

Under H_0, the test statistic is $Z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

In general P, common population proportion is not known and we use its unbiased estimate based on two sample proportions given by

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \therefore Z = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}, \text{where, } \hat{Q} = 1 - \hat{P}$$

The t- Distribution

Let $X_1, X_2, \ldots, X_n$ is a random sample of size $n \le 30$ drawn from a normal population with mean μ and variance σ^2 (not known) then the student's t statistic is defined as

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}, \text{ where, } \bar{X} = \frac{\Sigma X}{n} = \text{sample mean}$$

$= \text{Population mean} \quad n = \text{Sample size}$

$$S^2 = \frac{1}{n-1}\Sigma(X - \bar{X})^2 = \text{Unbiased estimate of population variance } (\sigma^2)$$

Student's t-statistic follows the student's t-distribution with degree of freedom = n – 1 and probability density function

$$f(t) = \frac{1}{\sqrt{n-1}\,\beta\left(\frac{1}{2}, \frac{n-1}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}}} \; ; -\infty < t < \infty$$

Properties of student's t- distribution

Some of the important properties of student's t-distribution are as follows:

1. The t- distribution is also a continuous distribution like standard normal distribution (Z).
2. The curve of t- distribution is also symmetrical about the mean t = 0 but flatter than the curve of standard normal distribution (Z).
3. The t- distribution is asymptotic to x-axis.
4. The shape of the curve of t- distribution varies with the degrees of freedom.
5. The t- distribution is lower at mean and higher at tails than a standard normal distribution.
6. As the sample size increases, the t- distribution tends to a normal distribution.
7. The total area under a t-distribution curve is 1 or 100%.

Assumptions of t- test

Some basic assumptions of t- test are as follows:

1. The sample size should be small i.e. $n \le 30$.
2. The population from which sample is drawn should be normal.

Scanned by CamScanner

3. The sample should be randomly and independently drawn.
4. The standard deviation of the population should be unknown.

Applications of t-test

The t- test has a wide number of applications in management. Some of them are

a. Test of significance of single mean, population variance being unknown.

b. Test of significance of difference between two independent population means, population variances being equal but unknown.

c. Paired t test for difference of means.

d. Test of significance of an observed sample correlation coefficient.

Test of Significance of a Single Mean

The procedure for testing the significance of a single mean is as follows:

Step 1: Setting up Hypotheses

Null hypothesis: H_0: $\mu = \mu_0$ i.e. the actual value of population mean is equal to specified value μ_0 or there is no significant difference between sample mean ($\bar{X}$) and population mean (μ) or there is no significant difference between actual value (μ) and specified value (μ_0) of population mean or sample has been drawn from a population with mean μ_0.

Alternative hypothesis: Set any one of the following hypothesis according to nature of the problem of hypothesis testing.

i. H_1: $\mu \neq \mu_0$ i.e. the actual value of population mean is not equal to specified value μ_0 or there is significant difference between actual value (μ) and specified value (μ_0) of population mean or sample has not been drawn from the population with mean μ_0. [*Two tailed test*]

ii. H_1: $\mu < \mu_0$ i.e. the actual value (μ) of population mean is less than specified value μ_0. [*Left tailed test*]

iii. H_1. $\mu > \mu_0$ i.e. the actual value (μ) of population mean is more than specified value μ_0. [*Right tailed test*]

Step 2: Level of significance (α)

Choose the appropriate level of significance in advance. The most commonly used level of significance is $\alpha = 5\%$ unless otherwise stated.

Step 3: Test Statistic The test statistic under H_0

$$t = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad \text{and degree of freedom} = n - 1$$

Step 4: Critical Value

The critical or tabulated value of the test statistic t at the pre-specified level of significance and degree of freedom is obtained from t table.

Step 5: Decision

i. If the calculated value of t (t_{cal}) is less or equal to ($\leq$) the tabulated value of t (t_{tab}), H_0 is accepted which means the population mean has a specified value μ_0. In other words, there is no significant difference between sample mean and the population mean, or, the sample has been drawn from a normal population with population mean μ_0.

Scanned by CamScanner

ii. If the calculated value of t (t_{cal}) is greater than (>) the tabulated value of t (t_{tab}), H_1 is accepted which means the population mean has not a specified value μ_0. In other words, there is significant difference between sample mean and the population mean, or the sample has not been drawn from a normal population with population mean μ_0.

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} \qquad \text{C.I. for } \mu = \bar{X} \pm t_{n-1,n} \times \frac{s}{\sqrt{n-1}}$$

Test of Significance of Difference between Two Independent Populations Means

Let us consider two independent populations with corresponding means μ_1 & μ_2 and unknown standard deviations σ_1 and σ_2 respectively.

Test Statistic

Under the assumption of $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and H_0: $\mu_1 = \mu_2$, the test statistic is

$$t = \frac{(\bar{X}_1 - \bar{X})_2 - E(\bar{X}_1 - \bar{X})_2}{S.E\,(\bar{X}_1 - \bar{X})_2} \quad t = \frac{\bar{X}_1 - \bar{X}_2}{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \sim t_{n_1+n_2-2} \text{ degree of freedom} = n_1 + n_2 - 2.$$

Calculations of S_p^2, an unbiased estimate of population variance (σ^2):

Actual Mean Method:

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2\right]$$

Direct Method:

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1} + \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2}\right]$$

Short-cut method

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1} + \Sigma d_2^2 n \frac{(\Sigma d_2)^2}{n_2}\right]$$

Where, $d_1 = X_1 - A$, $d_2 = X_2 - B$, A and B are assumed mean.

When biased sample variances s_1^2 and s_2^2 are given, then Sp^2 is estimated as

$$S_p2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

When unbiased sample variances S_1^2 and S_2^2 are given, then S_p^2 is estimated as

$$S_p^2 = \frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

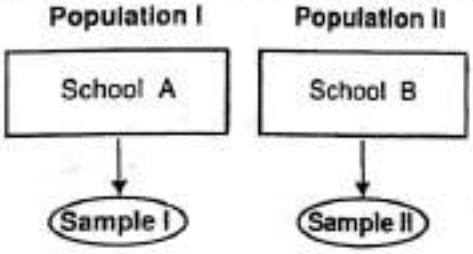
Paired t-test for difference of means

In t-test for difference of means, we performed the hypothesis testing procedure about the difference between two population means when the samples are independent of each other. However there are many situations where samples are pairly dependent to each other. *Independent Samples:* Two samples are called independent samples if the sample selected from one population is not related to the sample selected from other population. For example, suppose a random sample of 5 students from school A and another random sample of 4 students from school B are selected. The marks of these students are given below:

Scanned by CamScanner

Sample from school A	65	70	55	62	75
Sample from school B	50	72	65	68	

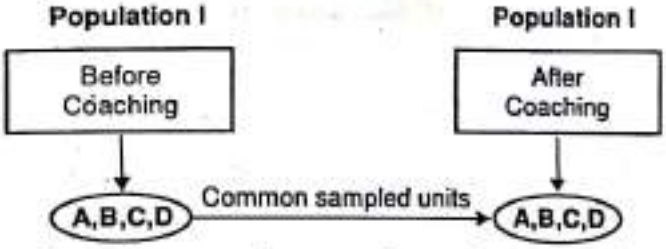
Here, these two samples are independent because sample of 5 students from school A is not related to the sample of 4 students from school B.



Dependent Samples: Two samples are called dependent samples if the selected units or items in both samples are related and same. Such samples are generally known as paired samples. For example, suppose a random sample of 5 students from a school is selected before and after a coaching class. The marks of these students are given below:

Students	A	B	C	D	E
Before coaching class	50	65	58	62	75
After coaching class	55	64	60	68	80

Here, these two samples are dependent because same 5 students have been selected before and after the coaching class. So both samples have same common students as sampled units.



In case of two dependent samples, two data values are collected before and after from the same source and hence these are also called paired samples

The sample sizes are equal, i.e., $n_1 = n_2 = n(\leq 30)$.

The two samples are dependent.

The test statistic under H_0 is

$$t = \frac{\bar{d}}{S.E.(\bar{d})} = \frac{\bar{d}}{S_d/\sqrt{n}} \sim t_{n-1} \text{ and degree of freedom} = n-1$$

Where, $\bar{d}$ the mean of the paired differences for the sample $= \frac{\Sigma d}{n}$

$d = X_2 - X_1$ = difference of observations

X_1 = observations in the sample before treatment

X_2 = observations in the sample after treatment

S_d = unbiased standard deviation of paired difference is given by

$$S_d = \sqrt{\frac{1}{n-1}\Sigma(d-\bar{d})^2} = \sqrt{\frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]}$$

Scanned by CamScanner

Test of Significance of an Observed Sample Correlation Coefficient

Let 'r' be the observed correlation coefficient in a sample of n pairs of observations from a bivariate normal distribution.

The test statistic under H_0

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2} \sim t_{n-2} \text{ and degree of freedom} = n-2$$

Where, r = sample correlation coefficient

n = sample size (number of pair of observations in the sample)

Calculations of sample correlation coefficient (r):

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma X-\bar{X}^2}\sqrt{\Sigma(Y-\bar{Y})^2}} \quad \text{Actual mean method}$$

$$r = \frac{n\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{n\Sigma X^2-(\Sigma X)^2}\sqrt{n\Sigma Y^2-(\Sigma Y)^2}} \quad \text{Direct mean method}$$

Remark: Confidence limits for population correlation coefficient (ρ)

The (1 - α)% confidence limits for estimating the population correlation coefficient (ρ) is given by

$$\text{C.I. of } p = r \pm t_{n-2,\alpha} \times \text{S.E.}(r) = r \pm t_{n-2,\alpha}\frac{1-r^2}{\sqrt{n}}$$

F -Distribution

F -Statistic

The ratio of two independent chi square variables divided by their respective degree of freedom is called F-statistic. Thus if X_1 is a chi square variable with degree of freedom $v_1 = n_1 - 1$ and X_2 is another independent chi square variable with degree of freedom $v_2 = n_2 - 1$ then ratio is defined as

$$F = \frac{\frac{X_1}{v_1}}{\frac{X_2}{v_2}} \sim F(v_1, v_2)$$

is called F-statistic with degree of freedom = (v_1, v_2) where n_1 and n_2 are number of observations in the first sample and second sample respectively. The distribution of F statistic is called Fisher's F-distribution. The sampling distribution of the F-statistic does not involve any population parameter and depends only on the degrees of freedom v_1 and v_2.

The F-distribution, also called variance ratio distribution, is skewed to the right and the F values can never be negative. The value of F-statistic ranges from 0 to infinity (∞). The F-distribution has two numbers of degree of freedom: degrees of freedom for numerator and degrees of freedom for denominator. Each combination of degrees of freedom for the numerator and for the denominator gives a different F-distribution curve as shown in the figure below

Assumptions of F Distribution

The following assumptions must hold true to use F-statistic:

Scanned by CamScanner

i. The populations from which samples are selected must be normally distributed.

ii. All the observations in the sample must be selected randomly and independently.

iii. The different samples should be independent of each others.

Applications of F-Distribution

i. F-test for equality of population variances.

ii. F-test for testing the equality of several population means.

iii. F-test for testing the significance of an observed sample multiple correlation.

iv. F-test for testing the significance of an observed sample correlation ratio.

v. F-test for testing the linearity of regression model.

F-test for Equality of Population Variances [Variance Ratio Test]

In the previous three section of hypothesis testing, we became familiar to the techniques (Z-test and t-test) used for single or two population means, for single or two population proportions, population correlation etc. In this section we are going to be familiar with the statistic (F-test) used for population variances. This test is used to

i. Test whether two random samples have been taken from two normal populations with same variance.

ii. Test whether there is significant difference between two variances of normal populations or not.

iii. Test whether two normal populations have equal variance or not.

iv. The procedure for testing the equality of population variances is as follows:

Step 1: Setting up Hypotheses

Null hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ i.e. there is no significant difference between variances of the two populations. In other words, two samples have been taken from two populations with equal variance.

Alternative hypothesis: H_1: $\sigma_1^2 \neq \sigma_2^2$ i.e. there is significant difference between variances of the two populations. In other words, two samples have not been taken from two populations with equal variance.

Step 2: Level of significance (α)

Choose the appropriate level of significance in advance. The most commonly used level of significance is $\alpha = 5\%$ unless otherwise stated.

Step 3: Test Statistic The test statistic under H_0 is

$$F = \frac{S_1^2}{S_2} \sim F(n_1 - 1, n_2 - 1) \text{ and } df = (n_1 - 1, n_2 - 1) \quad \text{if } S_1^2 > S_2^2$$

and $$F = \frac{S_2^2}{S_1^2} \sim F(n_1 - 1, n_2 - 1) \text{ and } df = (n_2 - 1, n_1 - 1) \quad \text{if } S_2^2 > S_1^2$$

Where, n_1 = size of the first sample taken from first population

n_2 = size of the second sample taken from second population

S_1^2 = unbiased estimate of variance of first population

S_2^2 = unbiased estimate of variance of second population

Computations of S_1^2 and S_2^2:

Scanned by CamScanner

Actual mean method:

$$S_1^2 = \frac{1}{n_1 - 1}\Sigma(X_1 - \bar{X}_1)^2, S_2^2 = \frac{1}{n_2 - 1}\Sigma(X_2 - \bar{X}_2)^2$$

Direct method:

$$S_1^2 = \frac{1}{n_1 - 1}\left[\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1}\right], S_2^2 = \frac{1}{n_2 - 1}\left[\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2}\right]$$

Short-cut method:

$$S_1^2 = \frac{1}{n_1 - 1}\left[\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1}\right], S_2^2 = \frac{1}{n_2 - 1}\left[\Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right]$$

Where, $d_1 = X_1 - A$, $d_2 = X_2 - B$, A and B are assumed or arbitrary means.

If the biased estimates of population variances – sample variances s_1^2 and s_2^2 – are given then unbiased sample variances S_1^2 and S_2^2 are calculated by using the following relation:

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} \text{ and } S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

Where, $S_1^2 = \frac{1}{n_1}\Sigma(X_1 - \bar{X}_1)^2$, $S_2^2 = \frac{1}{n_2}\Sigma(X_2 - \bar{X}_2)^2$

Step 4: Critical value

The critical or tabulated value of the test statistic F at the pre-specified level of significance and degree of freedom is obtained from the F-table.

Step 5: Decision

If the calculated value of F (F_{cal}) is less or equal to (≤) the tabulated value of F (F_{tab}), H_0 is accepted which means there is no significant difference between variances of the two populations. In other words, two samples have been taken from two populations with equal variance.

If the calculated value of F (F_{cal}) is greater than (>) the tabulated value of F (F_{tab}), H_1 is accepted which means there is significant difference between variances of the two populations. In other words, two samples have not been taken from two populations with equal variance.

Chi-Square Test

The test of hypothesis about the mean, difference between two means, the proportion, and difference between two proportions as well as equality of two population variances and test of equality of more than two population means were discussed in the previous sections. These tests require assumptions about population parameters so they are called *parametric tests*.

The chi-square test provides hypothesis test for qualitative data, where you have categorical data instead of numerical data. In other words, this test is appropriate for *nominal* or *ordinal* data. Nominal data are obtained from just categorization of qualitative variables such as defective and non defective, male and female, smoker and non smoker, success and failure, boy and girl etc. ordinal data are obtained by ranking the qualitative variables such as secretary, officer, clerk, peon etc. Chi square test is therefore based on frequencies that the number of items in

Scanned by CamScanner

the sample falling into each category. The distribution of chi square does not involve any parameter of the distribution, and hence it is called a *non parametric test*. Chi square test measures the difference between the observed frequency and the expected frequency. Chi-square statistic is denoted by a Greek alphabet χ^2 read as kai square. It is defined as

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$, Where, O = Observed frequency E = Expected frequency

Chi-Square Distribution

The square of a standard normal variate is known as chi-square variate with 1 degree of freedom. If X is a normal distribution with mean μ and standard deviation σ then

$$\chi^2 = \left(\frac{X-\mu}{\sigma}\right)^2$$ is a chi-square variate with 1 degree of freedom.

In general, if X_i, i = 1, 2, ..., n are n independent normal variates with means μ_i and variances σ_i^2, i = 1, 2, ..., n then

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma_i}\right)^2$$, is a chi-square variate with n degree of freedom.

The probability density function of chi-square distribution with v degree of freedom is given by

$$P(\chi^2) = \frac{1}{2^{v/2}\,\Gamma(v/2)} e^{-\chi^2} (\chi^2)^{v/2-1};\ 0 < \chi^2 < \infty$$

Properties of χ^2 distribution

The some of the properties of chi square (χ^2) distribution are as follows:

i. The mean of a chi-square distribution is equal to its degree of freedom and the standard deviation is equal to $\sqrt{2 \text{ degree of freedom}}$, i.e., for a chi-square with v degree of freedom, the Mean = v and standard deviation = $\sqrt{2v}$.

ii. The chi-square distribution assumes non-negative values only. The range of value of chi square is 0 to ∞ i.e., $0 < \chi^2 < \infty$.

iii. The shape of a chi-square distribution curve for various values of degrees of freedom is different.

iv. The shape of a chi-square distribution curve is skewed to the right for small degree of freedom and becomes symmetric for large degree of freedom.

v. The total area under each curve of χ^2 distribution is 1 or 100%.

Conditions for the validity of χ^2 test

The χ^2 test is used under the following assumptions or conditions:

i. The sample observations should be independent.

ii. The sum of observed frequency should be equal to the sum of expected frequency i.e. $\Sigma O = \Sigma E = N$.

iii. The total frequency should be reasonably large, say greater than 50.

iv. No theoretical cell frequency should be less than 5. If any theoretical frequency is less than 5, then for the applications of chi-square test, it is pooled with the preceding or succeeding frequency so that the pooled frequency is more than 5 and finally adjusts for the degree of freedom lost in pooling.

Scanned by CamScanner

Note: that all the above conditions for the validity of chi-square test are essential when the data is discrete. For continuous data, these conditions are not necessary.

Applications of Chi Square (χ^2) distribution

The χ^2-test has a number of applications in testing of hypothesis when the data available are in the form of frequency. The mostly used important applications of χ^2-test are as follows:

1. Test of goodness of fit.
2. Test of independence of attributes.
3. Test of population variance.

Chi-Square test of goodness of fit

The test of significance difference between observed frequencies and theoretical or expected frequencies is called goodness of fit. In a goodness of fit, we compare the observed frequencies for an experiment with theoretical or expected frequencies. Observed frequencies are those that the researcher obtains empirically through direct observation; theoretical or expected frequencies are developed on the basis of some hypothesis. For example, in 200 flips of a coin, one would expect 100 heads and 100 tails. This is theoretical or expected frequency. But 200 flips of coin may shows 92 heads and 108 tails. This is observed frequency.

Karl Pearson, 1900, developed a test for testing the significance of the discrepancy between experimental frequencies and theoretical frequencies and the procedure for testing the goodness of fit is summarized as follows:

Step 1: Setting up Hypotheses

Null hypothesis H_0 : There is no significant difference between observed frequencies and expected frequencies. In other words, there is good compatibility between theory and experiment, or, experiment supports the theory.

Alternative hypothesis H_1: There is significant difference between observed frequencies and expected frequencies. In other words, there is no compatibility between theory and experiment, or, experiment does not support the theory.

Step 2: Level of significance (α)

Choose the appropriate level of significance in advance. The most commonly used level of significance is $\alpha = 5\%$ unless otherwise stated.

Step 3: Test Statistic The test statistic under H_0 is

$$\chi^2 = \Sigma\frac{(O-E)^2}{E} \quad \text{and degree of freedom} = n_1 - 1 - k_1 - k_2$$

Where, O = Observed frequency

E = Expected frequency n_1 = no. of observed frequency

i. One degree of freedom is lost because of the assumption of linear condition $\Sigma O = \Sigma E = N$.

ii. k_1 is the number of parameters computed from the given data and used in estimating the theoretical frequencies of the distribution, if there is no need of computation of parameters we take k_1 as zero.

iii. k_2 is the number of degree of freedom lost in pooling of theoretical cell frequencies which are less than 5. If pooling is not necessary we take k_2 as zero.

Step 4: Critical value

The critical or tabulated value of the test statistic χ^2 at the pre-specified level of significance and degree of freedom is obtained from the χ^2-table.

Scanned by CamScanner

Step 5: Decision

If the calculated value of χ^2 (χ^2_{cal}) is less or equal to ($\leq$) the tabulated value of χ^2 (χ^2_{tab}), H_0 is accepted which means there is no significant difference between observed frequencies and expected frequencies. In other words, there is good compatibility between theory and experiment, or, experiment supports the theory.

If the calculated value of χ^2 (χ^2_{cal}) is greater than (>) the tabulated value of χ^2 (χ^2_{tab}), H_1 is accepted which means there is significant difference between observed frequencies and expected frequencies. In other words, there is no compatibility between theory and experiment, or, experiment does not support the theory.

Test of independence of attributes

The test statistic under H_0 is

$\chi^2 = \Sigma\frac{(O-E)^2}{E}$ and degree of freedom = $(r-1) \times (c-1)$; provided that no theoretical frequency is less than five. If any theoretical frequency is less than five, then degree of freedom is $df = (r-1) \times (c-1) - k_1$,where k_1 is no. of lost frequencies due to pooling.

Where, O = Observed frequency E= Expected frequency

r = number of rows c= number of columns

Calculation of expected frequency:

The expected frequencies corresponding to each cell are calculated as follows.

$$E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}} = \frac{RT \times CT}{N}$$

Chi-square test for Population Variance

Suppose we want to test whether the population variance (σ^2) has a specified value (σ_0^2), the chi-square test for population variance is used. The working procedure for testing the population variance is summarized as follows:

The test statistic under H_0 is

$\chi^2 = \frac{\Sigma(X-\bar{X})^2}{\sigma^2} = \frac{ns^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2}$ and degree of freedom = n – 1

Where, (a) Actual mean method

$s^2 = \frac{1}{n}\Sigma(X-\bar{X})^2$ and $S^2 = \frac{1}{n-1}\Sigma(X-\bar{X})^2$

(b) Direct method

$s^2 = \frac{1}{n}\left[\Sigma X^2 - \left(\frac{(\Sigma X)^2}{n}\right)\right]$ and $S^2 = \frac{1}{n}\left[\Sigma X^2 - \left(\frac{(\Sigma X)^2}{n}\right)\right]$

(c) Short-cut method

$s^2 = \frac{1}{n}\left[\Sigma d^2 - \left(\frac{(\Sigma d)^2}{n}\right)\right]$ and $S^2 = \frac{1}{n}\left[\Sigma d^2 - \left(\frac{(\Sigma d)^2}{n}\right)\right]$

$d = X - A$ and A = assumed mean

Remarks: If sample size is large (n > 30), we use Fisher's normal approximation to the chi-square distribution and then apply the normal test considering the test statistic

$Z = \sqrt{2\chi^2} - \sqrt{2n-1}$ **Error! Bookmark not defined.**

Scanned by CamScanner

Exam Asked Questions

'Long questions'

1 The quantity control manager at a light bulb factory needs to estimate the mean life of a large shipment of light bulbs. The process standard deviation is known to be 100 hours. A random sample of 64 light bulbs indicated a sample mean life of 350 hours.

(i) Obtain the standard error of the mean and interpret its meaning.

(ii) Set up 95% confidence interval of the true population mean life of bulbs in this shipment.

(iii) Do you think that manufacturer has the right to state the light bulbs last an average of 390 hours? [PU 2017 Fall/2011 Fall/2010 Spring]

Solution 1

Here we have given

Standard deviation (σ) = 100 hrs. Number of sample light bulb (n) = 64

Sample mean ($\bar{X}$) = 350 hrs

i. The standard Error of the mean S.E. $(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{64}} = 12.5$

ii. Now 95% confidence interval of the true population mean life of bulbs is given by

$$\bar{X} \pm Z_{\alpha/2} S.E.(\bar{X}) = 350 \pm 1.64 \times 12.5 = 350 \pm 20.5$$

Lower confidence interval $= 350 - 20.5 = 329.5$

Upper confidence interval $= 350 + 20.5 = 370.5$

iii.

a. Setting up hypothesis

Null hypothesis $H_0 : \mu = 390$

Alternative hypothesis $H_1 : \mu \neq 390$ [Two tailed test]

b. Level of significance : Since we do not have given the given the level of significance so suppose $\alpha = 5\% = 0.05$

c. Test of statistics : The test of statistics under H_0 is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{350 - 390}{100\sqrt{64}} = \frac{-40}{12.5} = -3.2$$

$\therefore$ $|Z_{cal}| = 3.2$

d. Critical value: The critical value of the test statistics Z at 5% level of significance for two tailed test is ± 1.96, i.e $|Z_{tab}| = 1.96$

e. Decision: Since $|Z_{cal}| = 3.2$ is greater than $|Z_{tab}| = 1.96$, H_1 is accepted which means the light bulbs last an average is not 390 hrs.

2. A social study group wants to determine of people think that there is too much violence on the television in these days. 70% of 2000 men selected at random believed so 60% of 4000 women selected also selected at random also believed that there is too much violence in the television. The group want to test if there is significant difference in the viewers of men as compared to views expressed by women. Test the hypothesis is at $\alpha = 0.05$. [PU 2016 Spring/2015 Fall]

Scanned by CamScanner

Solution 2

Here, we have given

$n_1 = 2000$ $n_2 = 4000$ $p_1 = 70\% = 0.7$ $p_2 = 60\% = 0.6$

(a) Now setting up hypothesis

Null Hypothesis H_o : $P_1 = P_2$

Alternative Hypothesis H_1: $P_1 \neq P_2$ (two failed test)

(b) Level of significance: Here we have given level of significance $\alpha = 0.05$

(c) Test of statistics: The test statistics under H_o is,

$$Z = \frac{P_1 - P_2}{\sqrt{\hat{P}\,\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad \text{Where, } \hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = 0.54$$

$$Z = \frac{0.7 - 0.6}{\sqrt{0.54 \times 0.46 \times 0.00075}} = \frac{0.1}{\sqrt{0.0136}} = 7.35$$

$\therefore$ $Z_{cal} = 7.35$

(d) Critical value: The critical value of Z at 5% level of significant for two tailed test is 1.96

$\therefore$ $Z_{tab} = 1.96$

(e) **Decision:** Since, $Z_{cal} = 7.35$ is greater than $Z_{tab} = 1.965$, H_1 is accepted which means there is difference between men and women.

3. The National Bureau of standard has previously reported the value of Selenium (Se) in NBS orchard leaves to be 0.080. A random sample of size n = 6 gives the following determinations:

0.072	0.073	0.080	0.078	0.088	0.080

Does the data contradict the preciously reported value? [P. U 2013 Spring/2017 Fall]

Solution 3

Here given the specific reported value of selenium (μ_0) = 0.08

Random sample size (n) = 6

Null hypothesis H_0: $\mu = 0.08$ there is not sufficiently different between sample mean $\bar{X}$ and reported value of selenium.

Alternative hypothesis H_1 : $\mu \neq 0.08$ (two tailed test)

Under H_o, test statistic $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

Where, $$\bar{X} = \frac{0.072 + 0.073 + 0.080 + 0.078 + 0.088 + 0.080}{6} = 0.0785$$

$$S^2 = \sum_{i=1}^{n} \frac{(xi)^2 - n\bar{X}^2}{n-1} = \frac{0.037141 - 6 \times (0.0785)^2}{6-1} = 0.00578$$

Where,

$$\sum_{i=1}^{n} x_i^2 = (0.072)^2 + (0.073)^2 + (0.080)^2 + (0.078)^2 + (0.088)^2 + (0.080)^2 = 0.037141$$

Scanned by CamScanner

$$\therefore \quad t = \frac{0.0785 - 0.08}{0.00578/\sqrt{6}} = -0.6356$$

The critical value of t_{n-1} at 5% level of significant for two tailed test is

$t_{0.025,5} = 2.571$

The calculated value of $|t|$ is less than 2.571. Therefore we accept μ_0, it means that the data does not contradict the previous reported value of Se.

4. In factory, 2% fans were found defective in a lot of 2000 fans and in another factory, 2.5% fans were found defective in a lot of 3000 fans. Do you find that the fans in the second factory are significantly inferior compared to the fans in the first factory (Use $\alpha = 5\%$) [P. U 2017 Fall 2017 Spring / 2011 Fall]

Solution 4

We have given

$n_1 = 2000$,

$n_2 = 3000$

Sample proportion of defective fans of in a factory $\hat{P}_1 = 2\% = 0.02$

Sample proportion of defective fans in another factory $\hat{P}_2 = 2.5\% = 0.025$

Let P_1 and P_2 be population's proportions of defective fans produced by a factory and another factory respectively. Therefore, we set up hypothesis as follows:

(i) Null Hypothesis: $H_0 : P_1 = P_2$

There is no significant difference in population proportion.

(ii) Alternative hypothesis: $H_1 : P_1 < P_2$ (one tailed test)

(iii) Test statistics: Under null hypothesis. $H_0: P_1 = P_2$

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

$$\text{Where, } \hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1 + n_2} = \frac{3000 \times 0.02 + 2500 \times 0.03}{3000 + 2500} = 0.023$$

$$\therefore \quad Z = \frac{0.02 - 0.025}{\sqrt{0.023 \times 0.977 \times \frac{1}{2000} + \frac{1}{3000}}} = -1.155$$

Scanned by CamScanner

5. What is P - value? In a factory 2% fans were found to be defective in a lot of 3000 fans, and in another factory 3% fan were found to be defective in lot of 2500 fans. Do you find that the fans in secondary factory are significantly inferior compared to the fans in the first factory? (Use $\alpha = 0.05$) [P. U 2017 Spring]

Solution 5

P - value: The probability that the value of test statistics is at least as extreme as the computed value of the test statistics on the basis of the sample data under H_o is called its P - value (probability or tail value). The null hypothesis is rejected if the P - value is very small such as 0.05 or less.

In other words, the smallest level of significance 'α' that would lead to rejection of null hypothesis H_o is called the P-value.

2nd Part

Let P_1 and P_2 be the population proportion of defective fans produced by a factory and another factory respectively. Then we have,

$n_1 = 3000$ $\qquad$ $n_2 = 2500$

$\hat{P}_1 = 2\% = 0.02$ $\qquad$ $\hat{P}_2 = 3\% = 0.03$

Now, we set up hypothesis as follows:

(i) Null hypothesis: $H_o : P_1 = P_2$

There is no significant difference in population proportion.

(ii) Alternative hypothesis: $H_o: P_1 < P_2$ [one tailed test]

(iii) Test statistics: Under null hypothesis $H_1: P_1 = P_2$

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0, 1)$$

$$\text{Where } \hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1 + n_2} = \frac{3000 \times 0.02 + 2500 \times 0.03}{3000 + 2500} = 0.025$$

$$Q = 1 - \hat{P} = 1 - 0.025 = 0.975$$

$$\therefore \quad Z = \frac{0.02 - 0.03}{\sqrt{0.025 \times 0.975 \times 0.0007}} = \frac{-0.01}{0.00423} = -2.364$$

(iv) Critical value: The critical value of Z at $\alpha = 0.05$ for left tailed test is

$$Z_\alpha = Z_{0.05} = -1.645$$

(v) Decision: Since the critical value of (i.e. $Z_\alpha = -1.645$) is greater than the calculated value ($Z = -2.364$) so, Z lies in the rejection region therefore the null hypothesis rejected.

We conclude that the rods in the second factory are sufficiently inferior compared to the fans in the first factory.

Scanned by CamScanner

6. An I.Q. test was administered to 5 persons before and after they were trained. The results are given below:

Candidates	1	2	3	4	5
IQ Before Training	110	120	123	132	125
IQ After Training	120	118	125	136	121

Is training effective at 5% level of significance? [P.U 2017 Spring]

Solution

Let the marks of IQ test of 5 persons before and after the training be denoted by X and Y respectively. Now the calculations of $\bar{d}$ and S_d is as follows.

Before (X) Training	After training (Y)	d = Y – X	d^2
110	120	10	100
120	118	–2	4
123	125	2	4
132	136	4	16
115	121	–4	16
		$\Sigma d = 10$	$\Sigma d^2 = 140$

$$\text{Now, } \bar{d} = \frac{\Sigma d}{n} = \frac{10}{5}$$

$$S_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{(n-1)}} = \sqrt{\frac{140-20}{4}} = 5.48$$

(i) **Setting up Hypothesis:**

Null Hypothesis: $H_o : \mu_1 = \mu_2$

i.e. The average knowledge level of students before and after the completion of training are equal. So the training is not effective.

Alternative Hypothesis: $H_1: \mu_1 < \mu_2$

i.e. The average knowledge level after the training is greater than the before training. So, the training is effective to increase the knowledge of students. [Right tailed test]

(ii) Level of significance: It is given that $\alpha = 5\% = 0.05$

(iii) Test statistics; The test statistics under the null hypothesis H_o is.

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{2}{5.48/2.24} = \frac{2}{2.45} = 0.816$$

$\therefore |t_{cal}| = 0.816$ and degree of freedom (df)= n – 1 = 4

(iv) Critical value: The critical (tabulated) value of the test statistics t at 5% level of significance and df = 4 for right tailed test is = $t_{0.05,4} = 2.132$ i.e. $t_{tab} = 2.132$

(v) Decision: Since $|t_{cal}| = 0.816$ is less than the $|t_{tab}|$, H_o is accepted. Which means that the average knowledge after and before training are equal. so the training is not effective.

7. A machine is used to cut wheels of cheese into blocks of specified weight. On the basis of long experience it has been observed that the weight of the blocks is normally distributed with a standard deviation of 0.3 kg. The machine is

Scanned by CamScanner

currently set to cut the block of wt. 12 kg. A sample of 9 blocks is found to have an average block of weight 12.25 kg set up the null hypothesis and determine whether it is a one tailed or two tailed test. Use the critical value approach to determine whether the cutting machine needs to be recalibrated. [P.U 2015]

Solution 7

Here given standard deviation (σ) = 0.3 kg

The no. of data in the sample (n) = 9

Average weight $\bar{X}$ = 12.25 kg

(a) Setting up Hypothesis

Null Hypothesis H_o: $\mu = 12$ kg

Alternative Hypothesis: H_1 $\mu \neq 12$ kg

It is a two tailed test.

(b) Level of significance: It is not given but we suppose $\alpha = 5\% = 0.05$.

(c) Test statistics: The test statistics under H_o is

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{12.25 - 12}{0.3/\sqrt{9}} = \frac{0.25}{0.1} = 2.5$$

$\therefore$ $Z_{cal} = 2.5$

(d) Critical value: The critical value or tabulated value of the test statistics Z at $\alpha = 5\%$ level of significance for two failed test is 1.96.

$\therefore$ $Z_{tab} = 1.96$

Decision: since $Z_{cal} = 2.5$ id greater than $Z_{tab} = 1.96$, H_1 is accepted.

8. The accompanying data on the cube compressive strength (MPa) concrete specimen appeared in the article "Experimental Study of Recycled Rubber Filled High Strength Concrete". 112.3, 97.0, 92.7, 86.0, 102.0, 99.2, 9*5.8, 103.5, 89.0, 86.7 Suppose the concrete will be used for the particular application unless there is strong evidence that true average strength is les than 100MPa. Should the concrete be used. [P.U 2016 Spring]

Solution 8

Here given data in ascending order is

X:	86.0	86.7	89.0	92.7	95.8	97.0	99.2	102.0	103.5	112.3
$X - \bar{X}$	−10.42	−9.72	−7.42	−3.72	−0.62	−0.58	2.78	5.58	7.08	15.88
$(X - \bar{X})^2$	108.58	94.48	55.06	13.84	0.38	0.34	7.73	31.14	50.13	252.17

We know that $\bar{X} = \frac{\Sigma x}{n} = \frac{964.2}{10} = 96.42$

and $S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{613.85}{9}} = 8.26$

(a) Setting up Hypothesis

Null Hypothesis H_o : $\mu = 100$ Mpa

Alternative Hypothesis H_1: $\mu < 100$ Mpa (Right tailed test)

Scanned by CamScanner

(b) Level of significance: Since we do not have given level significance so suppose $\alpha = 5\% = 0.05$

(c) Test statistics: The test statistics under H_o is

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{96.42 - 100}{8.26/\sqrt{10}} = \frac{-3.58}{2.61} = -1.37$$

(d) Critical value: The tabulated value of the test statistics t at 5% level of significance and degree of freedom = 9 for right tailed test is 2.262

(e) Decision: Since $|t_{cal}| = 1.37$ is less than the $|t_{tab}| = 2.062$, H_o is accepted.

9. In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 grams. With standard deviation of 12 grams, while the corresponding figures in a sample of 400 items from the other process are 124 and 14. Test whether two mean weight differ significantly or not at 10% level of significance. [P.U 2016 Fall/ 2015 Spring]

Solution 9

We have given:

For process I:

The number of items in sample (n_1) = 250

Average weight of items in sample ($\bar{X}_1$) = 120 gms

Sample standard deviation (S_1) = 12gms

For process II:

The number of items in sample (n_2) = 400

Average weight pf items in sample ($\bar{X}_2$) = 124 gms

Sample standard deviation (S_2) = 14

(i) Setting up Hypothesis:

Null Hypothesis H_o: $\mu_1 = \mu_2$

i.e. the means of two populations are equal, so the difference between weight of two process is not significant.

Alternative Hypothesis: H_1: $\mu_1 \neq \mu_2$ [i.e. two tailed test]

(The means of two populations are not equal, so the difference between mean weights of two processes is not significant)

(ii) Level of significance: It is given that $\alpha = 10\% = 0.1, \frac{\alpha}{2} = 0.05$

(iii) Test statistics: The test statistics under H_o is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{144}{250} + \frac{196}{400}}} = \frac{-4}{1.0325} = -3.87$$

$\therefore |Z_{cal}| = 3.87$

(iv) Critical value: The critical (tabulated) value of Z at $\alpha = 10\%$ level of significance for two tailed test is $Z_{0.05} = 1.645$ i.e $|Z_{tab}| = 1.645$

Scanned by CamScanner

(v) Decision: Since $|Z_{cal}| = 3.87$ is greater than $|Z_{tab}| = 1.645$, H_1 is accepted. We can conduce that the means of two populations are not equal. So the difference between means weights of two processes is significant.

10. The average hourly wage of a sample of 150 workers in plant A was Rs. 256 with a standard deviation of Rs. 108. The average wage of a sample of 200 workers in plant B was Rs. 282 with a standard deviation of Rs. 128. Can an applicant safely assume that the hourly wages paid by plant B are higher than those by plant A? Test the hypothesis at 3% level of significance. [P.U 2012 Fall]

Solution 10

We have given

For plant A: No. of workers in sample (n_1) = 150

Average hourly wages of workers in sample ($\bar{X}$) = 256

Sample standard deviation (S_1) = 108

For plant B: No. of workers in sample (n_2) = 200

Average hourly wages of workers in sample ($\bar{X}_2$) = 282

Sample standard deviation (S_2) = 128

(i) Setting up Hypothesis

Null Hypothesis H_0: $\mu_1 = \mu_2$

i.e. The average hourly wages of workers in two plants A and B are equal. So the hourly wages paid by plant is not higher than those paid by plant A.

Alternative Hypothesis H_1: $\mu_1 < \mu_2$ [Left tailed test]

i.e. The average hourly wages of workers in plant is not B higher than the workers in plant A.

(ii) Level of significance: We have given the level of significance $\alpha = 3\% = 0.03$

(iii) Test statistic : The test statistics under H_0 is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{256 - 282}{\sqrt{\frac{(108)^2}{150} + \frac{(128)^2}{200}}} = \frac{-26}{12.64} = -2.057$$

$\therefore |Z_{cal}| = 2.057$

(iv) Critical value: The critical (tabulated) value of Z at 3.

Level of significant for left tailed test is $Z_\alpha = Z_{0.03} = 1.75$

$\therefore$ $|Z_{tab}| = 1.75$

(v) Since $|Z_{cal}| = 2.057$ is greater than $|Z_{tab}| = 1.75$ H_0 is accepted which means the mean hourly wages of workers in plant B is higher than the workers in plant A.

So, the hourly wages paid by plant B is greater than plant A.

11. ABC Physical fitness claims that completion of their weight loss programmed will result in weight loss programme will results in weight loss, to tests the claim, 6 persons were selected at random and they were put through the weight loss programme and weights before and after the program recorded. Test the claim of the fitness centre at 0.05 level of significance. The weight in pounds in six persons before and after the programme are as follows: [P.U 2016 Fall/2015 Spring]

Scanned by CamScanner

Persons	1	2	3	4	5	6
Weight (before)	145	200	160	165	164	175
Weight (after)	143	190	165	183	160	176

Solution 11

Let the weights of persons before and after the physical fitness be denoted by X_1 and X_2 respectively. Now the calculations of $\bar{d}$ and S_d are as follows:

Weight (before) X_1	Weight after (X_2)	$d = X_2 - X_1$	d^2
145	143.	−2	4
200	100	−10	100
160	165	5	25
185	]183	−2	4
164	160	−4	16
175	176	1	1
N= 6		$\Sigma d = -12$	$\Sigma d^2 = 150$

Now, $\bar{d} = \frac{\Sigma d}{n} = \frac{-12}{6} = -2$ $\quad S_d = \sqrt{\frac{1}{n-1}[\Sigma d^2 - n\bar{d}^2]} = \sqrt{\frac{150 - 6 \times 4}{5}} = 5.02$

(i) Setting up hypothesis:

Null Hypothesis H_o: $\mu_1 = \mu_2$

i.e. The average weights before and after the physical fitness are equal. So the physical fitness is not effective in reducing the weights.

Alternating Hypothesis H_1: $\mu_1 > \mu_2$ [i.e. Left tailed test]

i.e. The average weights after the physical fitness is less than before the physical fitness.

(ii) Level of significance: It is given that $\alpha = 0.05$

(iii) Test statistic: The test statistic under H_o is

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{2}{5.02/2.45} = \frac{-2}{2.05} = -0.98$$

$\therefore$ $|t_{cal}| = 0.98$ and degree of freedom (df) = n− 1 = 6 − 1 = 5

(iv) Critical value: The critical (tabulated) value of the test statistic t at $\alpha = 0.05$ and df = 5 for left tailed test is $t_{\alpha, n-1} = t_{(0.05, 5)} = -2.571$

$\therefore |t_{tab}| = 2.571$

(v) Decision: Since $|t_{cal}| = 0.98$ is less than $|t_{tab}| = 2.571$ H_o is accepted which means the average weight before and after the physical fitness is remain constant. His claim is wrong.

12. An IQ test was administered to 7 persons before and after they were trained. The results are given below:

Candidates	1	2	3	4	5	6	7
IQ (before training)	130	135	120	133	142	137	126
IQ (after training)	138	131	130	135	146	146	138

Test whether the training is designed properly. [P. U 2015 Fall]

Scanned by CamScanner

Solution 12

Let X_1 and X_2 denote the marks of the persons before and after the training.

The calculations of $\bar{d}$ and S_d are as follows:

IQ before training (X_1)	IQ after training (X_2)	$d = X_2 - X_1$	d^2
130	138	8	64
135	131	–4	16
120	130	10	100
133	135	2	4
142	146	4	16
137	146	9	81
126	138	112	144
n = 7		$\Sigma d = 45$	$\Sigma d^2 = 425$

Now, $\bar{d} = \frac{\Sigma d}{n} = \frac{45}{7} = 6.429$ $\qquad S_d = \sqrt{\frac{1}{n-1}[\Sigma d^2 - n\bar{d}^2]} = 4.76$

(i) Setting of Hypothesis:

Null Hypothesis H_o: $\mu_1 = \mu_2$

i.e. The average marks of students before and after the training are equal. So, the training is not designed properly.

Alternative Hypothesis H_1: $\mu_1 < \mu_2$ [Right tailed test]

i.e. The average marks of students after training is greater than the before training. So, the training is designed properly.

(ii) Level of significant: We do not have given the level of significant let us choose $\alpha = 5\% = 0.05$

(iii) Test statistics: The test statistic under H_o is

$$t = \frac{\bar{d}}{Sd/\sqrt{n}} = \frac{6.429}{4.76/\sqrt{7}} = \frac{6.429}{1.8} = 3.57$$

$\therefore$ $|t_{cal}| = 3.57$, $\qquad df = n - 1 = 7 - 1 = 6$

(iv) Critical value : The critical value of t at $\alpha = 0.05$ and df = 6 for right tailed test id $t_{tab} = t_{\alpha, n-1} = t_{0.05,6} = 2.447$

(v) Decision: Since $t_{cal} = 3.57$ is greater than $t_{tab} = 2.447$, H_1 is accepted which means the average marks after training increases then the before training. Hence the training is designed properly.

13. A manufacturer of electronic calculator takes a random sample of 200 calculators and finds that there are eight defective units.

i. Construct a 95% confidence interval on the population proportion.

ii. Is there evidences to support a claim that the fraction of defective units produced is 1% or less? [PU 2014 Spring]

Solution

We have given, the sample size (n) = 1200

No. of defective units in sample (x) = 8

Sample proportion $(p) = \frac{x}{n} = \frac{8}{1200} = 0.0067$

Scanned by CamScanner

Then $q = 1 - p = 0.9933$

Now, the standard error of sample proportions is

$$S.E(P) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.0067 \times 0.9933}{1200}} = 0.0024$$

i. The level of confidence $(1 - \alpha) = 95\% = 0.95$

$$\therefore \alpha = 0.05$$

Now the 95% confidence interval for the population proportion is,

C. I for p $= p \pm Z_\alpha \times S.E(p) = 0.0067 \pm 1.96 \times 0.0024 \quad = 0.0067 \pm 0.0047$

Lower limit $= 0.0067 - 0.0047 = 0.002$

Upper limit $= 0.0067 + 0.0047 = 0.0114$

ii.

a. **Setting of Hypothesis:**

Null Hypothesis H_0: $P = 0.01 (= P_0)$

i.e. The proportion is less or equal to 1%.

Alternative Hypothesis H_2: $P > 0.01 = (P_0)$ [left tailed test]

i.e. The proportion is greater than 1%.

(b) Level of significant: Level of significant (α) is specified as 1% when $\alpha = 10\%$.

(c) Test statistics: The test statistics under H_0 is

$$Z = \frac{P - P_0}{\sqrt{pq/n}} = \frac{0.0067 - 0.01}{\sqrt{0.0067 \times 0.9933/1200}} = \frac{-0.0033}{0.0024} = -1.375$$

$\therefore \quad |Z_{cal}| = 1.375$

(d) Critical value: The critical value of Z at $\alpha = 0.01$ for the test is $Z_\alpha = Z_{0.01} = -2.33$

$\therefore \quad |Z_{tab}| = 2.33$

(e) Decision: Since $|Z_{cal}| = 1.375$ is less than $|Z_{tab}| = 2.33$ H_0 is accepted, which means the proportion is less or equal to 1%. It is significant.

14. The following random samples are measurements of the heat producing capacity (in millions of calories per 100 kg) of specimens of coal from two mines,

Mine 1:	82	81	83	80	83	
Mine 2:	72	78	79	81	79	78

Use the 0.01 level of significance to test whether the difference between the means of these two mines are significant. [P.U 2014 Spring]

Solution 14

Let X_1 and X_2 denote the data for mine-1 and mine-2 respectively.

For mine – 1: $n_1 = 5$

$$\bar{X} = \frac{82+81+83+80+83}{5} = \frac{406}{5} = 81.8$$

$$S_1 = \sqrt{\frac{1}{n_1 - 1}\left[\Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n_1}\right]} = \sqrt{\frac{1}{4}[33463 - 33456.2]} = 1.3$$

Where $\Sigma x_1^2 = 82^2 + 81^2 + 83^2 + 80^2 + 83 = 33463$

Scanned by CamScanner

For mine -2:

$n_2 = 6$

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{72+78+79+81+79+78}{6} = \frac{467}{6} = 77.83$$

$$S_2 = \sqrt{\frac{1}{n_2-1}\left[\Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n}\right]} = \sqrt{\frac{1}{5}[36395 - 3648.17]} = 3.06$$

Where $\Sigma x^2_2 = 72^2 + 78^2 + 79^2 + 81^2 + 79^2 + 78^2 = 36395$

Let μ_2 be the population means of mine – 1 and mine – 2 respectively.

a. **Setting of Hypothesis:**

Null Hypothesis Ho: $\mu_1 = \mu_2$

i.e. $\mu_1 - \mu_2 = 0$, the difference is not significant.

Alternative Hypothesis: H_1: $\mu_1 \neq \mu_2$ [two tailed test]

i.e. the difference is significant.

b. **Level of significance:** The level of significance $\alpha = 0.01$.

c. **Test statistic:** The test statistic under H_0 is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{81.8 - 77.83}{\sqrt{5.95 \times (0.25 + 0.2)}} = \frac{3.97}{1.64} = 2.42$$

$$\text{Where, } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{4 \times 1.69 + 5 \times 9.3636}{9} = 5.95$$

$\therefore$ $t_{cal} = 2.42$

d. **Critical value:** The critical value of t at $\alpha = 0.01$ with degree of freedom (df) = $n_1 + n_2 - 2 = 9$ for two tailed test is $t_{\alpha/2, n_1 + n_2 - 2} = t_{0.005,9} = 3.250$

$\therefore$ $t_{tab} = 3.250$.

e. **Decision:** Since, $t_{cal} = 2.42$ is less than $t_{tab} = 3.250$, so, H_0 is accepted. The difference is not significant.

15. The quantity demanded of a certain product before advertisement and after advertisement is given below:

Before advertisement	50	40	45	50	60	70	35	38	60	70
After advertisement	55	41	51	51	55	75	50	54	70	78

Is advertisement effective? Test at 5% level of significance. [P.U 2014 Spring]

Solution 15

The process is similar as Q. 12

16. The average weekly wages of a sample of 200 workers in industry A was Rs. 1150 with standard deviation of Rs. 100. The weekly wages of a sample of 300 workers in industry B was Rs. 1000 with standard deviation of Rs. 50. Can we consider the weekly wages paid by industry A is higher than those paid by industry B. (Use level of significance $\alpha = 0.05$) [PU 2014 Fall/P.U 2013 Spring]

Scanned by CamScanner

Solution 16

For industry A

No. of workers in sample (n_1) = 200

Average weekly wages of workers $\overline{X}_1$ = Rs.1150

Sample standard deviation (S_1) = Rs.100

For industry B

No. of workers in sample (n_2) = 300

Average weekly wages of workers $\overline{X}_2$ = Rs.1000

Sample standard deviation (S_2) = Rs.50

a. **Setting up hypothesis**

Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e. the average weekly wages of workers in two industries A and B are equal.

Alternative hypothesis $H_1: \mu_1 > \mu_2$ [Right tailed test]

i.e. the average weekly wages of workers in industry A is higher than the workers of industry B.

b. **Level of significance:** We have given the level of significance $\alpha = 0.05$

c. **Test statistic:** The test statistic under H_0 is

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{1150-1000}{\sqrt{\left(\frac{100^2}{200}+\frac{50^2}{300}\right)}} = \frac{150}{7.64} = 19.63$$

$\therefore$ $Z_{cal} = 19.63$

d. **Critical value:** The critical value of Z at $\alpha = 0.05$

For right tailed test is $Z_\alpha = 1.64$

$\therefore$ $Z_{tab} = 1.64$

e. **Decision:** Since, $Z_{cal} = 19.63$ is higher than $Z_{tab} = 1.64$, H_1 is accepted which means the mean of weekly wages of workers of industry A is higher than the workers of industry B

17. The sales figure of an item in eight shops before and after advertisement is given as follows:

Before	70	65	48	72	80	92	98	100
After	72	70	53	75	84	95	105	104

Test whether advertisement was effective at 5% level of significance.

[PU 2014 Fall/2013 Spring]

Solution 17

The process is similar as Q. 12

18. A company claims that its light bulbs are superior to those of a competitor. A study showed that a sample of 40 bulbs had an average life of 628 hours of continuous use with a standard deviation of 27 hours. While a sample of 30 bulbs made by the competitor has an average life of 619 hours of continues use with a standard deviation of 25 hours. Test at the 5% level of significance, whether this claim is justified?

Scanned by CamScanner

Solution 18

For company A:

No. of bulbs in sample (n_1) = 40 Average life time of bulbs $\overline{X}_1 = 628$

Sample standard deviation (S_1) = 27

For Competitor Company

No. of bulbs in sample (n_2) = 30 Average life time of bulbs $\overline{X}_2 = 619$

Sample standard deviation (S_2) = 25

a. **Setting up hypothesis**

Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e. the average life time of company A is equal to the competitor company.

Alternative hypothesis $H_1: \mu_1 > \mu_2$ **[Right tailed test]**

i.e. the average life time of company A is greater than the competitor company.

b. **Level of significance:** We have given the level of significance $\alpha = 5\% = 0.05$

c. **Test statistic:** The test statistic under H_0 is

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{628-619}{\sqrt{\left(\frac{27^2}{40}+\frac{25^2}{30}\right)}} = \frac{9}{6.25} = 1.44$$

$\therefore$ $Z_{cal} = 1.44$

d. **Critical value:** The critical value of Z at $\alpha = 0.05$

For right tailed test is $Z_\alpha = 1.64$

$\therefore$ $Z_{tab} = 1.64$

e. **Decision:** Since, $Z_{cal} = 1.44$ is less than $Z_{tab} = 1.64$, H_0 is accepted which means the mean of life time of both companies is same.

Thus, the claim is not significant.

19. Mr. A claims that TU graduates have a higher IQ score than the PU graduates. A sample of 15 TU graduates taken at random revealed an average IQ score 130 with a standard deviation of 10. A similar random sample of 12 PU graduates revealed IQ of 115 with a standard deviation of 7. At 99% confidence level, can you conclude that on average, TU graduates are more intelligent than PU graduates as measured by IQ scores? [P.U 2013 Fall]

Solution 19

The process is similar to Q.No. 18

20. The following are the weights in pounds before and after, of 10 persons who stayed on certain weight reducing diet for four weeks:

Before	After
147:0	137.9
183:5	176.2
232:1	219.0
161:6	163.8
196:5	193.5

Scanned by CamScanner

206.3	201.4
177.0	180.6
215.4	203.2
147.7	149.0
208.1	196.4

Test at the 0.05 level of significance, whether the diet is effective in term of reducing weight. [P.U 2013 Fall]

Solution 20

The process is similar to Q. 12

21. The scores of 10 candidates prior and after training are given below:

Before	80	50	36	40	54	69	83	96	90	65
After	90	57	56	50	62	81	84	86	84	75

Is training effective? (use 5% level of significance)

Solution 21

The process is similar to Q. 12

22. Two types of drugs are used for 5 and 7 patients to reduce their weight. Drug A was imported and Drug B indigenous. The decrease in the weight after using the drug for six month was as follows:

Drug A	10	12	13	11	14		
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficiency of two drugs? If not which drug should you buy? Consider $\alpha = 5\%$. [P.U 2011 Spring

Solution 22

Let X_1 and X_2 denote the drug A and drug B respectively.

Calculations of Sample means and Sp^2

Drug A			Drug B		
X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
10	−2	4	8	−3	9
12	0	0	9	−2	4
13	1	1	12	1	1
11	−1	1	14	3	9
14	2	4	15	4	16
			10	−1	1
			9	−2	4
$\Sigma X_1 = 60$		$\Sigma(X_1 - \bar{X}_1)^2 = 10$	$\Sigma X_2 = 77$		$\Sigma(X_2 - \bar{X}_2)^2 = 44$

Here,

$n_1 = 5$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{60}{5} = 12$$

$n_2 = 7$

Scanned by CamScanner

$$\overline{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{77}{7} = 11$$

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(X_1 - \overline{X}_1) + 2 + \Sigma(X_2 - \overline{X}_2)^2\right] = \frac{1}{10}[10 + 44] = 5.4$$

a. **Setting up hypothesis**

Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e. the average decrease in weight of patients treating with drug A is equal to the average decrease in weight of patients treating with drug B.

Alternative hypothesis $H_1: \mu_1 > \mu_2$ [Right tailed test]

i.e. the average decrease in weight of patients treating with drug A is not equal to the average decrease in weight of patients treating with drug B.

b. **Level of significance:** We have given the level of significance $\alpha = 5\% = 0.05$

c. **Test statistic:** The test statistic under H_0 is

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{12 - 11}{\sqrt{5.4\left(\frac{1}{5} + \frac{1}{7}\right)}} = \frac{1}{3.6} = 0.74$$

$\therefore$ $Z_{cal} = 0.74$

d. **Critical value:** The critical value of Z at $\alpha = 0.05$

and df = 10 for two tailed test $t_{\alpha/2}, n_1 + n_2 - 2 = 2.228$

e. **Decision:** Since $t_{cal} = 0.74$ is less than the $t_{tab} = 2.228$ so H_0 is accepted. Which means the average decrease in weight of patients treating with drug A is equal to the average decrease in weight of patients treating with drug B. So, there is no significance in the efficacy of two drugs A and B.

Hence, any drugs A or B can be bought.

23. A television producer of a farming show believes that the show is more popular with rural viewers than urban viewers. To test the claim, a TV station showed the program to 300 rural viewers and 100 urban viewers. It was noted that 65 of the rural viewers and 18 of the urban viewers enjoyed the program. Test the producer's belief at 5% level of significance. [P.U 2013 Fall]

Solution 23

For rural viewers

The total number of people in the sample $(n_1) = 300$

Total number of people enjoyed the programme $(x_1) = 65$

Proportion of rural viewers $P_1 = \frac{x_1}{n_1} = \frac{65}{300} = 0.22$

For urban viewers

The total number of people in the sample $(n_2) = 100$

Total number of people enjoyed the programme $(x_2) = 18$

Proportion of rural viewers $P_2 = \frac{x_2}{n_2} = \frac{18}{100} = 0.18$

Scanned by CamScanner

Let P_1 and P_2 be the population proportion of people of rural area and urban area respectively who enjoyed the programme.

a. **Setting up hypothesis**

Null hypothesis $H_0 : P_1 = P_2$

The population proportion of rural and urban area equally enjoyed the programme.

Alternative hypothesis H_1: $P_1 > P_2$ [Right tailed test]

i.e. the population proportion of rural viewers is greater than urban viewers is greater than urban viewers.

b. **Level of significance:** We have given the level of significance $\alpha = 5\% = 0.05$

c. **Test statistic:** The test statistic under H_0 is

$$t = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.22 - 0.18}{\sqrt{0.21 \times 0.79\,(0.003 + 0.01)}} = \frac{0.04}{0.046} = 0.87$$

$\therefore$ $Z_{cal} = 0.87$

d. **Critical value:** The critical value of Z at $\alpha = 0.05$ for right tailed test $Z_\alpha = 1.645$

$\therefore$ $Z_{tab} = 1.645$

e. **Decision:** Since $Z_{cal} = 0.87$ is less than $Z_{tab} = 1.645$, H_0 is accepted which means the proportion of rural viewers and urban viewers is equal. So, the producers belief is not significant.

24. A machine put out 16 defective articles in a sample of 600 articles. After machine is overhauled, it puts out 1 defective articles in sample of 100 articles. Has the machine improved? Suggest your answer at 5% level of significance?

Solution 24

We have,

Before overhauled

No. of articles in the sample $(n_1) = 600$

No. of defective articles in the sample $(x_1) = 16$

Proportion of defective articles in the sample

$$(P_1) = \frac{x_1}{n_1} = \frac{16}{600} = 0.027$$

After overhauled

No. of articles in the sample $(n_2) = 100$

No. of defective articles in the sample $(x_2) = 1$

Proportion of defective articles in the sample

$$(P_2) = \frac{x_2}{n_2} = 1/100 = 0.01$$

Since the population proportion of defective articles P is not given, It is estimated as

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{16 + 1}{600 + 100} = \frac{17}{700} = 0.024$$

Scanned by CamScanner

$\hat{Q} = 1 - \hat{P} = 1 - 0.024 = 0.976$

a. Setting up Hypotheses:

Null Hypothesis $H_0 : P_1 = P_2$

i.e. the proportion of defective articles by the machine before and after over hauling of it are same.

Alternative Hypothesis H_2: $P_1 > P_2$ [Right tailed test]

i.e. the proportion of defective articles by the machine before overhauling is greater than after overhauling.

b. level of significance : Here given level of significance $\alpha = 5\% = 0.05$

c. Test statistic : the test statistic under H_0 is

$$Z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.027 - 0.01}{\sqrt{0.024 \times 0.976\left(\frac{1}{600} + \frac{1}{100}\right)}} = 0.017$$

d. Critical value : the critical value of z at $\alpha = 0.05$ for right tailed test is 1.64.

i.e. $Z_{tab} = 1.64$

e. Decision : Since $Z_{cal} = 0.017$ is less than $Z_{tab} = 1.64$

So H_0 is accepted which means the proportion of defective articles by the machine before and after overhauling of it are same. So, the machine has not improved after overhauling.

25. The average height of 50 students who showed no athletic interest was 68.2 inches with standard deviation of 2.5 inches, while another set of 50 students who showed athletic interest was average height of 67.5 inches with a standard deviation of 2.8 inches. Test the hypothesis that the athletics makes student taller. [PU 2010 SPRING]

Solution 25

For no interest in Athletics:

No. of students in the sample $(n_1) = 50$

Average height of students in the sample $(\bar{X}_1) = 68.2$

Standard deviation in sample $(S_1) = 2.5$

For interested students in Athletics:

The no. of students in the sample $(n_2) = 50$

Average height of students $(\bar{X}_2) = 67.5$

Sample standard deviation $(S_2) = 2.8$

a. Setting up Hypothesis:

Null Hypothesis H_0: $\mu_1 = \mu_2$

i.e. the population mean height of two types of students are same.

Alternative Hypothesis: H_1: $\mu_1 < \mu_2$ [left tailed test]

i.e. the population mean height of students having interest in athletics is greater than having no interest in athletics.

b. Level of significance: Since it is not given let us take

$\alpha = 5\% = 0.05$

Scanned by CamScanner

c. Test statistic: Tne test statistic under H_0 is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{68.2 - 67.5}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}} = \frac{0.7}{0.53} = 1.32$$

$\therefore$ $|Z_{cal}| = 1.32$

d. Critical value: The critical value of Z at $\alpha = 0.05$ for left tailed test $Z_\alpha = -1.64$

e. Decision: Since $|Z_{cal}| = 1.32$ is less than $|Z_{tab}| = 1.64$, H_0 is accepted which means that the mean height of students having interest in athletics and no interest have equal height.

Hence Athletics can not make the students taller.

26. A potential buyer of light bulbs bought 50 bulbs of each of two brands upon testing these bulbs, he found that Brand A has a mean life of 1282 hours with a standard deviation of 80 hrs whereas B had a mean life of 1208 hrs with a standard deviation of 94 hrs can the buyers be quite certain that the two brands differ in quality? [PU 2005 FALL]

Solution 26

For Brand A:

No. of bulbs in the sample $(n_1) = 50$

Average life of bulbs in the sample $(\bar{X}_1) = 1282$ hrs.

Sample standard derivation $(S_1) = 80$ hrs.

For Brand B:

No. of bulbs in the sample $(n_2) = 50$

Average life of bulbs in the sample $(\bar{X}_2) = 1208$ hrs.

Sample standard derivation $(S_2) = 94$ hrs.

a. Setting up Hypothesis:

Null Hypothesis H_0: $\mu_1 = \mu_2$

i.e. the population mean life of bulbs of brand A and brand B are equal.

Alternative Hypothesis: H_1: $\mu_1 \neq \mu_2$ [left tailed test]

i.e. the population mean life of bulbs of brand A and brand B are different.

b. Level of significance: We do not have given the significance so suppose

$\alpha = 5\% = 0.05$

c. Test statistic: The test statistic under H_0 is

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1282 - 1208}{\sqrt{\frac{(80)^2}{50} + \frac{(94)^2}{50}}} = \frac{74}{17.46} = 4.24$$

d. Critical value: The critical value of Z at $\alpha = 0.05$ for two tailed test $Z_\alpha = 1.96$

e. Decision: Since $|Z_{cal}| = 4.24$ is greater than $|Z_{tab}| = 1.96$. Hence H_1 is accepted which means that the average life time of bulbs of both brands are different.

27. Experience shows that a fixed dose of a certain drug causes an average increase of pulse rate of 10 beats per minute. A group of 9 patients given the same dose showed the following increase 13, 15, 14, 10, 8, 12, 16, 9, 20

Test at 5% level of significance whether this group is different in response to the drug? [PU 2005 FALL]

Scanned by CamScanner

Solution 27:

Here given the mean increase of pulse rate

$\mu = 10$ (beats per minute)

Also sample size (n) = 9

The calculation of samples mean and sd

X	d = X − 14	d^2
13	−1	1
15	1	1
14	0	0
10	−4	16
8	−6	36
12	−2	4
16	2	4
9	−5	25
20	6	36
	$\Sigma d = -9$	$\Sigma d^2 = 123$

$$\bar{X} = A + \frac{\Sigma d}{n} = 14 + \frac{-9}{9} = 14 - 1 = 13$$

$$S = \sqrt{\frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]} = \sqrt{\frac{1}{8}[123 - 9]} = 3.78$$

a. Setting up Hypothesis:

Null Hypothesis H_0: $\mu_1 = \mu_2$ beats per minute.

i.e. the mean increase of pulse rate is equal to 10 beats per minute.

Alternative Hypothesis: H_1: $\mu_1 \neq \mu_2$ [left tailed test]

i.e. the mean increase of pulse rate is different than 10 beats per minute.

b. Level of significance: We have given level of significance

$\alpha = 5\% = 0.05$

c. Test statistic: The test statistic under H_0 is

$$t = \frac{\bar{x}_1 - \mu}{s/\sqrt{n_1}} = \frac{13 - 10}{3.78/\sqrt{9}} = \frac{3}{1.26} = 2.38$$

$\therefore$ $t_{cal} = 2.38$ and degree of freedom d.f = 9 − 1 = 8

d. Critical value: The critical value of t at $\alpha = 0.05$ for d.f = 8 two tailed test is

$t_{tab} = 2.306$

e. Decision: Since $|t_{cal}| = 2.38$ is greater than $|t_{tab}| = 2.306$. Hence H_1 is accepted which means that the two drugs are different.

28.. 1000 articles from factory A are examined and found to contain 3% defective. 1500 similar articles from factory B are found to contain only 2% defectives. Would you conclude that the products of factory B are superior to the factory A?

[PU 2005 FALL]

Solution 28

For Factory A:

$n_1 = 1000$ $\qquad$ $p_1 = 3\% = 0.03$

Scanned by CamScanner

For Factory B:

$n_2 = 1500$ $\qquad$ $p_2 = 2\% = 0.02$

Since, the population percentage of particles which are found to be defective, P is estimated as.

$$\hat{P} = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{1000 \times 0.03 + 1500 \times 0.02}{1000 + 1500} = 0.024$$

$$\hat{Q} = 1 - \hat{P} = 1 - 0.024 = 0.976$$

a. Setting up Hypothesis:

Null Hypothesis H_0: $P_1 = P_2$

i.e. the proportion of defective particles in factory A and B are same.

Alternative Hypothesis: H_1: $P_1 > P_2$ [right tailed test]

i.e. the proportion of defective particles in factory A is greater than that in factory B.

b. Level of significance: Suppose that

$\alpha = 5\% = 0.05$

c. Test statistic: The test statistic under H_0 is

$$z = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{0.03 - 0.02}{\sqrt{0.024 \times 0.976\left(\frac{1}{1000} + \frac{1}{1500}\right)}} = \frac{-0.01}{0.006} = -1.67$$

d. Critical value: The critical value of Z at $\alpha = 0.05$ for right tailed test $Z_{0.05} = 1.64$

e. Decision: Since $|Z_{cal}| = 1.67$ is greater than $|Z_{tab}| = 1.64$, H_1 is accepted which means that the proportion of defective particles in factory A is greater than factory B. Hence factory B is superior to the factory A.

29. Memory capacity of a student was tested before and after training. Is the difference between the memory capacities of the students significant? [PU 2005 Fall]

Roll No.	1	2	3	4	5	6	7	8	9
Before Training	10	15	9	3	7	12	16	17	4
After training	12	17	8	5	6	11	18	20	3

Solution 29

Let X_1, X_2 denote the memory capacity of students before and after the training.

The calculations of $\bar{d}$ and s.d is as follows

Memory capacity before training (X_1)	Memory capacity after training (X_2)	$d = X_2 - X_1$	d^2
10	12	2	4
15	17	2	4
9	8	−1	1
3	5	2	4
7	6	−1	1
12	1	−1	1
16	18	2	4
17	20	3	9
4	3	−1	1
n = 9		$\Sigma d = 7$	$\Sigma d^2 = 29$

Scanned by CamScanner

Now

$$\bar{d} = \frac{\Sigma d}{n} = \frac{7}{9} = 0.77$$

$$Sd = \sqrt{\frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]} = \sqrt{\frac{1}{8}\left[29 - \frac{(7)^2}{9}\right]} = \sqrt{\frac{23.5}{8}} = 1.71$$

a. Setting up Hypothesis:

Null Hypothesis H_0: $\mu_1 = \mu_2$

i.e. the average memory capacity of students before and after the training remains constant.

Alternative Hypothesis: H_1: $\mu_1 < \mu_2$ [$\because$ right tailed test]

i.e. the average memory capacity of students after the training is greater than before training.

b. Level of significance: Since we do not have given the level of significance let us choose $\alpha = 5\% = 0.05$

c. Test statistic: The test statistic under H_0 is

$$t = \frac{\bar{d}}{sd/\sqrt{9}} = \frac{0.77}{1.71/\sqrt{9}} = \frac{0.77}{0.57} = 1.35$$

$\therefore$ $t_{cal} = 0.35$

d. Critical value: The critical value of t at $\alpha = 0.05$ for degree of freedom (d.f) = (9 – 1) = 8 is

$\therefore$ $t_{tab} = 1.86$

e. Decision: Since, $|t_{cal}| = 1.35$ is less than $|t_{tab}| = 1.86$, H_0 is accepted which means that the memory capacity of students before and after the training remains constant. That is the training is not significant at 5% level of significance.

Exam asked Short Questions

30. Write short note on Paired T –test PU 2010 Spring

Solution 30

See in the theoretical part of this chapter.

31. Write short note on Level of Significance and power test PU 2011Spring /2006 Spring

Solution 31

See in the theoretical part of this chapter.

32. Write short note on Coefficient of determination
PU 2011Spring /2003Fall/2006Spring /2010 Fall/2013Fall/2017 Fall

Solution 32

See in the theoretical part of this chapter.

33. Write short note on Null and alternative Hypothesis
PU 2003 Fall /2010Spring /2009Spring /2007Fall

Solution 33

See in the theoretical part of this chapter.

Scanned by CamScanner

34. Write short note on P-Value and Critical Value PU 2014 Spring

Solution 34

See in the theoretical part of this chapter.

35. Write short note on Types of error in hypothesis

PU 2011Spring /2005Spring /2010 Spring /2009 Spring /2009Fall/2012Fall/2013 Fall/2013Spring /2015Spring /2016Fall/2016Spring /2017 Fall/2017Spring

Solution 35

See in the theoretical part of this chapter.

Additional Questions for Practice

1. A television producer of a farming show believes that the show is more popular with rural viewers than urban viewers. To test the claim, a TV station showed the program to 300 rural viewers and 100 urban viewers. It was noted that 65 of the rural viewers and 18 of the urban viewers enjoyed the program. Test the producer's belief at 5% level of significance. [P.U 2013 Fall]

2. The controller of examination of certain university claimed that 60% of the students have passed in the university. An examination is conducted to a random sample of 200 students and it is found that 140 students were passed in the examination. Test whether the controller's claim valid or not at the 5% level of significance. [P.U 2012 Fall]

3. Discuss the importance of central limit theorem in statistics. A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Obtain the 95% and 99% confidence limits for the good apples. [P.U 2011 Spring]

4. What is the test of significance? A social worker believes that fewer than 14% of the couples in a certain area ever used any form of birth control. A random sample 120 couples was contacted. 25 of them said they had used some method of birth control. Comment on the social worker's belief at $\alpha = 2\%$ level of significance. [P.U 2011 Spring]

5. A study on expenditure behavior of tourist in Nepal revealed that the average expenditure of 100 European tourist per day is Rs 9200 with standard deviation of Rs. 600. Also, the average expenditure of 100 American tourists per day is Rs. 10,000 with standard deviation of Rs. 500. Can you conclude that the expenditure behavior of European and American tourists are same? [P.U 2011 Spring]

6. A study was made of a sample of 25 records of patients seen at a chronic diseases hospital on an out patient basis. The mean number of outpatient visits per patient was 4.8 and the sample standard deviation was 2.0. Can it be touched from these data that the population mean is greater than 4 visits per patients. Let the probability of committing type I error be 0.05. [P.U 2011 Fall]

7. A store keeper wanted to but a large quantity of light bulbs from two brands labeled A and B. He bought 100 bulbs from each brand and found by testing that6 brand A has mean lifetime of 1120 hours and standard deviation 75 hours and brand B had mean lifetime 1062 hours and the standard deviation 82 hours. Examine whether the difference of means is significance or not at 5% level of significance. [P.U 2011 Fall/ 2010 Spring]

Scanned by CamScanner

9. A random sample of 10 boys had the following I.Q's: 70, 120, 110, 83, 88, 95, 98, 107, 99 and 100. Do these data support the assumption of a population mean I.Q. of 100 at 1% level of significance?

10. The average height of 50 students who showed athletic interest was 68.2 inches with standard deviation of 2.5 inches, while another set of 50 students who showed no athletic interest had the average height of 67.5 inches with a standard deviation of 2.8 inches. Test the hypothesis that athletics interest makes a student taller. PU 2009 SPRING

10. The heights of six randomly chosen sailors are in inches: 63, 65, 68, 69, 71 and 72. Those of 10 randomly chosen heights of soldiers are 61, 62, 65, 66, 69, 70, 71, 72, 60 and 73 inches. Test whether sailors are on the average taller than soldiers or not. Use 5% level of significance. PU 2009 FALL

11. Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase in duty, 900 people were tea drinkers in a sample of 1150 people. State whether there is a significant decrease in the consumption of tea after the increase in excise duty. (Use 1 % level of significance) PU 2006 SPRING

12. Price of shares of a company on a different days in a year were found to be in Rs. 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68.

Test the mean price of the shares in the year is more than 65 at 5% level of significance. [Given $P(t>1.833) = 0.05$ PU 2006 SPRING

13. Two research laboratories have independently produced drugs that provide relief to arthritis suffers. The first drug was tested on a group of 90 arthritis sufferers and produced an average of 8.5 hours of relief and a sample standard deviation of 1.8 hours. The second drug has tested on 80 arthritis sufferers, producing an average of 7.9 hours of relief and a sample standard deviation of 2.1 hours. At 0.05 level of significance; does the second drug provide a significantly shorter period of relief? PU 2007 FALL

14. The following data gives the experience of machine operators in years and their performance as given by the number of good parts turned out per 100 pieces.

Operator:	1	2	3	4	5	6	7	8
Experience:	16	12	18	4	3	10	5	12
Performance:	87	88	89	66	78	80	75	83

Calculate the regression equation of performance ratings on experience and estimate the probable performance if an operator has 8 year experience.

PU 2007 FALL

15. Explain the types of errors of testing of hypothesis. Two horses A and B were tested according to the times (in seconds) to run a particular track with the following results:

Horse A:	28	30	32	33	33	29	34
Horse B:	29	30	30	24	27	27	32

Analyze the above data and report whether, or not, you can discriminate between the two horses at 5% level of significance. PU 2003 FALL

16. Explain the meaning of confidence interval. One thousand articles from a factory are examined and found to be 2.5% defective. Fifteen hundred similar articles from another factory are found to have only 2% defective. Can it be reasonably concluded that the products of the first factory are inferior to those of the second? PU 2003 FALL

Scanned by CamScanner

17. Eleven students were given a test in statistics. They were given a month's tuition and a second test was taken at the end of it. Do the marks give evidence that the students have benefited by the extra coaching? Test at 5% level of significance.

Students	1	2	3	4	5	6	7	8	9	10	11
Marks in 1st test:	21	20	19	21	18	20	18	17	23	10	19
Marks in 2nd test	24	19	22	23	20	22	20	20	23	20	18

PU 2003 FALL

18. The manufacturer of Shilpa electric bulbs claims that mean life of their bulb is 25 months. A random sample of 9 such bulbs gave the following values: Life in months:24, 26, 32, 28, 20, 20, 23, 27 and 34
Can you regard the manufacturer's claim to be valid at 5 percent level of significance? PU 2010 Spring

19. At a certain date in a large city 400 out of a random sample of 500 men were found to be smokers. After the tax on tobacco had been heavily increased another random sample of 600 men in the same city included 400 smokers. Was the observed decrease in the proportion of smokers significant? PU 2009 Spring

20. In a random sample of 1000 houses in a certain city it was found that 318 own color T.V. In the neighboring city it was found that 450 out of 1200 houses own color T.V. Test whether there is significant difference between the two samples of housing with respect to using color T.V. (Use α =5%) PU 2009 Fall

21. Nine observations of surface-soil PH were made at each of two different locations at the Central Soil Salinity Research Institute experimental farm, and the resulting data appeared in the articles" Sodium Calcium Exchange Equilibrium in Soils as Affected by Calcium Carbonate and Organic Matter" (Soil Science). Do the data suggest that the true mean soil PH values differ for the two locations? Test the appropriate hypothesis using 0.05 level of significance. PU 2010 Fall

PH-VALUE
Location A: 8.53, 8.52, 8.01, 7.99, 7.93, 7.89, 7.85, 7.82, 7.80
Location B: 7.85, 7.73, 7.58, 7.40, 7.35, 7.30, 7.27, 7.27, 7.23

22. The manufacturer of Shilpa electric bulbs claims that mean life of their bulb is 25 months. A random sample of 9 such bulbs gave the following values:

Life in months:	24,	26,	32,	28,	20,	20,	23,	27	34

Can you regard the manufacturer's claim to be valid at 5 percent level of significance?[PU 2010 SPRING]

23. At a certain date in a large city 400 out of a random sample of 500 men were found to be smokers. After the tax on tobacco had been heavily increased another random sample of 600 men in the same city included 400 smokers. Was the observed decrease in the proportion of smokers significant? PU 2009 SPRING

24. In a random sample of 1000 houses in a certain city it was found that 318 own color T.V. In the neighbouring city it was found that 450 out of 1200 houses own color T.V. Test whether there is significant difference between the two samples of housing with respect to using color T.V. (Use α=5%) PU 2009 FALL

25. A store keeper wanted to buy a large quantity of light bulbs from two brands labeled A and B. He bought 100 bulbs from each brand and found by testing that brand A had mean life time 1120 hours and standard deviation 75 hours and brand B had mean life time 1062 hours and standard deviation 82 hours. Examine whether the difference of means is significant or not at 5% level of significance. PU 2010 Spring

Scanned by CamScanner

15
Chapter

Simple Linear Regression and Correlation

Correlation Analysis

A statistical (tool) measure which is used to study the degree of relationship between two or more variables is called correlation. The measure of correlation is called the correlation coefficient and it is simply denoted by r.

Types of correlation

Correlation can be classified into various types. The most important ones which we discuss here are

(a) Simple correlation (b) Partial correlation

(c) Multiple correlation

Simple Correlation:- The measure of simple linear relationship between only two variables is called the simple correlation. For example the relationship between the age and blood pressure of a man, the relationship between height and weight of a women etc. Simple correlation is also classified as

(i) Positive and Negative Correlations

(ii) Linear and Non-linear correlations

Partial Correlation:- The correlation between two or more variables considered but the correlation is studied only between two variables by keeping other variables constant is called partial correlation.

Multiple Correlation:- The relationship between three and more variables is called the multiple correlation.

Methods of Studying Simple Correlation

There are variable methods of studying correlations. The important methods are given below:

(i) Scatter Diagram Method (Graphical/Qualitative method)

(ii) Karl Pearson's correlation coefficient (For Quantitative Date)

(iii) Spearman's Rank Correlation coefficient (For ordinal Data)

Scanned by CamScanner

Scatter Diagram Method:- Independent variable (X) is plotted along x-axis and dependent variable (Y) is plotted along y-axis. The diagram of dots of obtained is called scatter diagram.

Note:- The scatter diagram gives only the nature or direction of correlation between two variable, it fails to give degree of relationship between two variables.

Karl Person's Correlation Coefficient

It is the method of finding the relationship between two variables whose values are given and gives the linear relation between then which is denoted by r.

1. $$r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}.\sqrt{\text{Var}(Y)}} = \frac{\sigma_{xy}}{\sigma_x.\sigma_y}$$

   Where, $\text{Cov}(X,Y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n}$;

   $\text{Var}(X) = \frac{\Sigma(X-\bar{X})^2}{n}$, $\text{Var}(Y)\frac{\Sigma(Y-\bar{Y})}{n}$

2. $$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}} = \frac{n.\Sigma(X-\bar{X})(Y-\bar{Y})}{\sigma_x.\sigma_y}$$

3. $$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}, \text{ Where } x = X-\bar{X}, y = Y-\bar{Y}$$

4. $$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}.\sqrt{n.\Sigma Y^2 - (\Sigma Y)^2}}$$

5. $$r = \frac{\frac{\Sigma XY}{n} - \bar{X}\bar{Y}}{\sqrt{\frac{\Sigma X^2}{n} - (\bar{X})^2}.\sqrt{\frac{\Sigma Y^2}{n} - (\bar{Y})^2}} = \frac{\Sigma XY - n.\bar{X}\bar{Y}}{\sqrt{\Sigma X^2 - n(\bar{X})^2} \times \sqrt{\Sigma Y^2 - n(\bar{Y})^2}}$$

6. $$r = \frac{N\Sigma fXY - (\Sigma fX)(\Sigma fY)}{\sqrt{N\Sigma fX^2 - (\Sigma fX)^2}\sqrt{N.\Sigma fY^2 - (\Sigma fY)^2}} = \frac{\Sigma fXY - N\bar{X}\bar{Y}}{\sqrt{\Sigma fX^2 - N(\bar{X})^2}.\sqrt{\Sigma fY^2 - N(\bar{Y})^2}}$$

7. $$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n.\Sigma u^2 - (\Sigma u)^2}\sqrt{n.\Sigma v^2 - (\Sigma v^2)}}, \quad u = \frac{x-a}{h}, \quad v = \frac{y-b}{k}$$

8. $$r = \frac{N.\Sigma fuv - (\Sigma fu)(\Sigma fv)}{\sqrt{N\Sigma fu^2 - (\Sigma fu)^2}\sqrt{N\Sigma fv^2 - (\Sigma f u)^2}} \quad \text{Where, } u = \frac{x-a}{h}, \quad v = \frac{y-b}{k}$$

Spearman's Rank Correlation Coefficients

Karl Pearson's coefficient of correlation measures the degree of linear relationship between two variables whose quantitative date is given. But some characteristics such as beauty. intelligence honesty, knowledge, taste etc cannot measured quantitatively but they can be measured by assigning ranks. Such case we use the spearman's rank correlation coefficient to find the relation between two variables.

1. $$R = 1 - \frac{6 \times \Sigma d^2}{n(n^2-1)}, \text{ where, } d = |R_1 - R_2|$$

   R_1 (rank of first variable), R_2 (rank of second variable)

   n – no. of observation

Scanned by CamScanner

2 **For Repeated Ranks:**

$$R = 1 - \frac{6\left[\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \ldots\right]}{n(n^2-1)}$$

Where, $m_1 \rightarrow$ no. of times first value repeated.
$m_2 \rightarrow$ no. of times 2nd value repeated.

Probability of Correlation coefficients

(i) The correlation coefficient between two variables is symmetrical i.e. $r_{xy} = r_{yx}$.

(ii) The correlation coefficient between two variables is in dependent of change of origin and scale.

(iii) The correlation coefficient lies between –1 and 1. i.e. $-1 \le r \le 1$.

(iv) Correlation coefficient is geometric mean between two regression coefficients. i.e. $r = \sqrt{b_{xy} \cdot b_{yx}}$

(v) It two variables are independent then correlation coefficient ($r = 0$), but the converse may not true.

(vi) If $r = 0$, then the correlation between two variables X and Y is not linear.

Interpretation of Correlation Coefficient:

(i) If $r = +1$, perfect positive correlation between two variables.

(ii) If $r = -1$, perfect negative correlation.

(iii) If $r > \pm 0.75 \rightarrow$ High degree of (Positive/negative) Correlation.

(iv) If $r < \pm 0.25$ low degree of (Positive/negative) correlation

(v) If $r = 0$, there is no correlation.

Probable Error of the correlation coefficient:

It is used to test the reliability of the calculated value of correlation coefficient (r). Probability error of correlation coefficient r is given by

$$P.E(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

Notes:

(i) If $r < P.E.(r)$, the value of r is not significant.

(ii) If $r > 6 \times P.E.(r)$, the value of r is definitely significant.

(iii) If $P.E.(r) \le 6 \times P.E.(r)$, nothing can be concluded with certainty.

(iv) Where $S.E(r) = \frac{1-r^2}{\sqrt{n}}$.

Coefficient of Determination [Explained Variation]

The square of the correlation coefficient r is known as the coefficient of determination. It gives the percentage variation. (Change) in the dependent variable that is accounted for by the independent variable.

i.e. Coefficient of determination $(r^2) = \frac{\text{Explained Variance}}{\text{Total variance}}$

Partial Correlation

The degree of relationship between two variable at a time by keeping the effect of other variables constant then the variable is called the partial correlation. For example if we are studying the relationship between production and the seed quality by keeping fertilizer constant then the correlation is said to be partial correlation.

Scanned by CamScanner

Let X_1, X_2 and X_3 be the three variables then the partial correlation coefficient between them is denoted and defined as follow:

(i) The partial correlation coefficient between X_1 and X_2 keeping X_3 constant is given by $r_{12.3} = \dfrac{r_{12} - r_{13}\, r_{23}}{\sqrt{1 - r_{13}^2}\,.\,\sqrt{1 - r_{23}^2}}$

(ii) The partial correlation coefficient between X_1 and X_3 keeping X_2 constant is given by $r_{13.2} = \dfrac{r_{13} - r_{12}\,.\,r_{32}}{\sqrt{1 - r_{12}^2}\,.\,\sqrt{1 - r_{32}^2}}$

(iii) The partial correlation coefficient between X_2 and X_3 keeping X_1 constant is given by $r_{23.1} = \dfrac{r_{23} - r_{21}\,.\,r_{31}}{\sqrt{1 - r_{21}^2}\,.\,\sqrt{1 - r_{31}^2}}$

Multiple Correlation

The degree of relationship between more than two variables simultaneously at a time. The correlation is said to be multiple correlation. For example if the relationship between production seed quality and fertilizer at a time then the correlation is said to be multiple correlation.

Let X_1, X_2 and X_3 denote the tri-variable distribution then the multiple correlation between them is denoted and defined as follows:

(i) The multiple correlation coefficient between dependent variable X_1 and joint effect of independent variables X_2 and X_3 is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

(ii) The multiple correlation coefficient between dependent variable X_2 and joint effect of independent variables X_1 and X_3 is given by

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{13}^2}}$$

(iii) The multiple correlation coefficient between dependent variable X_3 and joint effect of independent variables X_1 and X_2 is given by

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{23}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{12}^2}}$$

Regression Analysis

The estimation of unknown values of one variable form known values of another variable is called regression.

The equation of straight line that represents the relation between two variables also estimates the value of one variable with respect to another variable such line is called the regression line and its equation is called the regression equation.

Types of Regression:-

Here we discuss their types of regression.

(a) Simple and multiple regression

(b) Linear and non - linear regression

(c) Total and partial regression.

Lines of regressions/ Equations of regression

(i) (a) The regression line (equation) of y on x is $y = a + bx$

Where $a \rightarrow$ intercept of y-axis,

$b \rightarrow$ slope of the regression line (regression coefficient of y on x i.e. b_{yx}

Whose normal equations are

$\sum y = na + b\sum x$, $\sum xy = a\sum x + b\sum x^2$

Scanned by CamScanner

(b) The regression equation of y on x is given by $y - \bar{y} = b_{yx}(x - \bar{x})$

Where, $\bar{X} = \frac{\Sigma x}{n}$ $\bar{Y}, = \frac{\Sigma y}{n}$,

$$b_{yx} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2} = \frac{n\Sigma xy - (\Sigma x)\Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = r\frac{6_y}{6_x}$$

byx – regression coefficient of y on x.

(ii) (a) The regression line (Equation) of X on Y is given by $x = a + by$

Where, a → intercept of x-axis

b → slope of the regression line

(Regression coefficient of x on y i.e b_{xy})

Whose normál equations are

$\Sigma x = na + b\Sigma y$, $\Sigma xy = a\Sigma y + b\Sigma y^2$

(b) The regression equation of x on y is given by $x - \bar{x} = b_{xy}(y - \bar{y})$

Where b_{xy} is regression coefficient of x on y.

$$b_{xy} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (y - \bar{y})^2} = \frac{n\Sigma xy - \Sigma x.\Sigma y}{n\Sigma y^2 - (\Sigma y)^2} = r\frac{6_x}{6_y}.$$

Properties of Regression Coefficients

(i) Correlation coefficient r is the geometric mean of the regression coefficients.

$r = \pm\sqrt{b_{xy}.b_{yx}}$

(ii) Both the regression coefficient have the same sign.

(iii) If one regression coefficient is greater than one than other is less than one.

(iv) The arithmetic mean of regression coefficients is greater than the correlation coefficient r.

(v) Regression coefficients are independent of change of origin but not of scale.

Angle between two lines of Regression:-

If θ is the angle between two regression line then, $\tan\theta = \left(\frac{1 - r^2}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$.

(i) If r = 0, then the regression lines are perpendicular to each other

(ii) If r = ±1, then the regression line are either parallel or coincide.

(iii) If tan θ > 0 then θ is acute.

(iv) If tan θ < 0 then θ is obtuse.

Coefficient of Determination and Interpretation

The product of the regression coefficient of y on x and x on y is known as the coefficient of determination and is represented by r^2

i.e $r^2 = b_{xy} \times b_{yx}$

or $$r^2 = \frac{a\Sigma y + b\Sigma xy - n(\bar{y})^2}{\Sigma y^2 - n(\bar{y})^2}.$$

Multiple Regression

Let $y = a + bx_1 + cx_2$ be the linear multiple regression equation of dependent variable y on x_1 and x_2 then its normal equations are.

$\Sigma y = na + b\Sigma x_1 + c\Sigma x_2$

$\Sigma x_1 y = a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2$

$\Sigma x_2 y = a\Sigma x_2 + b\Sigma x_1 x_2 + c\Sigma x_2^2$

Scanned by CamScanner

Exercise 15-1

1. Fit a straight line to the following data:

x:	1	2	3	4	6	8
y:	2.4	3	3.6	4	5	6

Solution 1

Let the best fitted line to the given data be

$$y = a + bx \qquad \ldots(i)$$

Where a and b satisfy the normal equations

$\Sigma y = na + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$

Which are obtained as follow:-

x	y	x^2	y^2	xy
1	2.4	1	5.76	2.4
2	3	4	9	6
3	3.6	9	12.96	10.8
4	4	16	16	16
6	5	36	25	30
8	6	64	36	48
$\Sigma x = 24$	$\Sigma y = 24$	$\Sigma x^2 = 130$	$\Sigma y^2 = 104.72$	$\Sigma xy = 113.2$

Substituting the values the normal equation reduces to

$6a + 24b = 24$ and $24a + 130b = 113.2$

Solving we get, $b = 0.506$ and $a = 1.976$

Hence the best fitted line is $y = 1.976 + 0.506x$.

2.(a) Find the line of best fit to the following data:

x:	0	5	10	15	20	25
y:	12	15	17	22	24	30

Solution 2a

Let the best fitted line to the given data be

$$y = a + bx \qquad \ldots(i)$$

Where a and b satisfy the normal equations

$\Sigma y = na + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$

x	y	x^2	y^2	xy
0	12	0	144	0
5	15	25	225	75
10	17	100	289	170
15	22	225	484	330
20	24	400	576	480
25	30	625	900	750

We have,

$n = 6, \Sigma x = 75\ \Sigma y = 120$

$\Sigma x^2 = 1375\ \Sigma y^2 = 2618\ \Sigma xy = 1805$

Substituting these value the normal equation reduces to,

Scanned by CamScanner

$6a + 75b = 120 \Rightarrow 2a + 25b = 40$

$75a + 1375b = 1805 \Rightarrow 15a + 275b = 361$

Solving we get,

$a = 1.29$ $\quad$ $b = 0.7$

Hence the best fitted line is $y = 11.29 + 0.7x$. #

b. In some determination of the volume V of carbon dioxide dissolved in a given volume of water at different temperatures θ the following pairs of value were abstained:

θ:	0	5	10	15
V:	1.80	1.45	1.18	1.00

Obtained by the method of least square, a relation of the form $v = a + b\theta$ which best fits to these observations.

Solution 2b

Let $v = a + b\,\theta$ be the best fitted line to the given data,

Where the values of a and b are given by the normal equations

$\Sigma v = na + b\Sigma\theta$ and

$\Sigma v\theta = a\Sigma\theta + b\Sigma\theta^2$

Which are given by the following data

θ	ϖ	θ^2	ϖ^2	θ.ϖ
0	1.80	0	3.27	0
5	1.45	25	2.10	7.25
10	1.18	100	1.39	11.8
15	1.00	225	1	15
$\Sigma\theta = 30$	$\Sigma\varpi = 5.43$	$\Sigma\theta^2 = 350$	$\Sigma\varpi^2 = 7.73$	$\Sigma\varpi\theta = 34.05$.

Substituting these values in the normal equations we get,

$4a + 30b = 5.43$ $\quad$ and $30a + 350b = 34.05$

Solving we get, $a = 1.758$ $\quad$ $b = 0.0534$

Hence the best fitted line is

$v = 1.758 - 0.0534\theta$

3. Fit a straight line to the following data taking x as the dependent variable:

x:	1	3	4	6	8	9	11	14
y:	1	2	4	4	5	7	8	9

Solution 3

The best fitted line with x as the dependent variable is given by

$$x = a + by \qquad \ldots(i)$$

Where a and b satisfy the normal equations

$$\Sigma x = na + b\Sigma y \quad \text{and} \quad \Sigma xy = a\Sigma y + b\Sigma y^2$$

x	y	x^2	y^2	xy
1	1	1	1	1
3	2	9	4	6
4	4	16	16	16
6	4	36	16	24
8	5	64	25	40
9	7	81	49	63
11	8	121	64	88
14	9	196	81	126
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma x^2 = 524$	$\Sigma y^2 = 256$	$\Sigma xy = 374$

Substituting these values in the normal equations we get.

Scanned by CamScanner

$8a + 40b = 56$ ...(A)

$40a + 256b = 374$ ...(B)

Solving (A) and (B) we get,

$a = -6.4$ and $b = 2.68$

Hence the best fitted line is, $x = -6.4 + 2.68y$.

4. a. A simply supported beam carries a concentrated load P (kg) at its mid point. The following table gives maximum deflection y (cm) corresponding to various value of p.

p :	100	120	140	160	180	200
y:	0.45	0.55	0.60	0.70	0.80	0.85

Find a law of the form $y = a + b\,p$.

Also find value of maximum deflection when $P = 150$ kg.

Solution 4a

Let the best fitted regression line the maximum deflection in cm by concentrated load b in kg is

$y = a + bp$ ..... (1)

Where a and b satisfy the normal equations

$\Sigma y = na + b\Sigma p$ and $\Sigma py = a\Sigma p + b\Sigma p^2$

p (kg)	y (cm)	p^2	y^2	py
100	0.45	10,000	0.2	45
120	0.55	14400	0.3	66
140	0.60	19600	0.36	84
160	0.70	25600	0.49	112
180	0.80	32400	0.64	144
200	0.85	40000	0.72	170
$\Sigma P = 900$	$\Sigma y = 3.95$	$\Sigma P^2 = 142000$	$\Sigma y^2 = 2.71$	$\Sigma py = 591$

Substituting these values, the normal equation reduces to,

$6a + 900b = 3.95$ and $900a + 42000b = 591$

Solving we get,

$a = 0.69$ $b = -0.0002$

Hence line (1) becomes.

$y = 0.69 - 0.0002p$ .....(2)

Also the maximum deflection when

$P = 150$ kg is

$y = 0.69 - 0.0002 \times 150 = 0.66$ cm.

b. If P is the pull required to lift a loan W means of a pulley block, find a liner law of the form $P = m\,W + c$ connecting P and W, using the following data: [T. U. 2073]

P:	12	15	21	25
y:	50	70	100	120

Where P and W are taken in kg. wt. Computer P when W = 150 kg . wt.

(Bangalore, 1987)

Scanned by CamScanner

Solution 4b

Let $P = mW + C$ be the best fitted regression line where m and c satisfy the equation $\Sigma P = nc + m\Sigma W$ and $\Sigma WP = C\Sigma W + m\Sigma W^2$

P	W	P^2	W^2	PW
12	50	144	2500	600
15	70	225	4900	1050
21	100	441	10,000	2100
25	120	625	14400	3000
$\Sigma P = 73$	$\Sigma W = 340$	$\Sigma P^2 = 1435$	$\Sigma W^2 = 31800$	$\Sigma PW = 6750$

Substituting these values the normal equation reduces to,

$4c + 340m = 75$ and $340c + 31800\,m = 6750$

Solving we get

$C = 2.1$ $\quad$ $m = 0.19$.

Hence the best fitted linear line is.

$P = 2.1 + 0.19W$ .....(1)

Also the value of P when W = 150 kg is

$P = 2.1 + 0.19 \times 150 = 30.6$ kg.

5. In a study between the amount of rain fall and the quantity of air pollution removed the following data were collected:

Daily rainfall:(X) (in .01 cm)	4.3	4.5	5.9	5.6	6.1	5.2	3.8	2.1
Pollution removed (Y)(mg/m3)	12.6	12.1	11.6	11.8	11.4	11.8	13.2	14.1

Find the regression line of y on x.

Solution 5

Let $y = a + bx$ be the regression line of y on x where a and b satisfy the condition $\Sigma y = na + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$

Which by given

X (0.1 cm)	Y (mg/m³)	x^2	xy
4.3	12.6	18.49	54.18
4.5	12.1	20.25	54.45
5.9	11.6	34.81	68.44
5.6	11.8	31.36	66.08
6.1	11.4	37.21	69.54
5.2	11.8	27.04	61.36
3.8	13.2	14.44	50.16
2.1	14.1	4.41	29.61
$\Sigma P = 37.5$	$\Sigma y = 98.6$	$\Sigma x^2 = 188.01$	$\Sigma xy = 453.82$

Substituting these value in normal equations we get,

We get, $8a + 37.5b = 98.6$

$37.5a + 188.1b = 453.82$

Solving we get, $a = -15.51$ $\quad$ and $\quad$ $b = 0.68$

Hence the regression equation of y on x is, $Y = 15.51 - 0.68x$

Scanned by CamScanner

6. The following table gives age (x) in years of cars and annual maintenance cost (y) in hundred rupees.

X	1	3	5	7	9
Y	15	18	21	23	22

Estimate the maintenance cost for a 4 years old car after finding the regression equation and also find coefficient of determination and interpret it.

Solution 6

Let X denote the age of car in years and Y denote the annual maintenance cost in hundred rupees. Now the regression equation of y on x is given by

$$y = a + bx \qquad \text{..... (1)}$$

Where a and b satisfy the normal equations

$$\Sigma y = na + b\Sigma x \quad \text{and} \quad \Sigma xy = a\Sigma x + b\Sigma x^2.$$

Form calculator we get,

$$n = 5, \Sigma x = 25, \Sigma y = 99, \Sigma x^2 = 165\ \Sigma y^2 = 2003\ \Sigma xy. = 533$$

Substituting these values the normal equation reduces to,

$$5a + 25b = 99$$
$$25a + 1656 = 533$$

Solving we get a = 16, b = 0.76

Hence regression equation is,

$$y = 16 + 0.76x. \qquad \text{..... (2)}$$

The estimate maintenance cost for a 4 year old car is given by

$$y = 16 + 0.76 \times 4 = 18.96$$

i.e. Rs. 1896

Also the coefficient of determination is given by

$$r^2 = \frac{a\Sigma y + b\Sigma xy - n\bar{y}^2}{\Sigma y^2 - n(\bar{y})^2} = \frac{16 \times 99 + 0.76 \times 533 - 5 \times 392.04}{2003 - 5 \times 392.04} = \frac{28.89}{42.8} = 0.675$$

Hence the one variable accounts for 67.5% of the total variation in the other variable.

7. The following data gives the experience of machine operations in years and their performance as given by the number of good parts turned out per 100 pieces,

Operator	1	2	3	4	5	6	7	8
Experience	15	14	18	5	3	10	7	13
Performance	88	89	90	78	68	75	83	87

i. Estimate the performance of an operator has 10 years of experience.

ii How much variation in performance is explained by the machine's experience?

Based on this value, determine the reliability of the fitted regression equation?

[2011 spring]

Solution 7

Let X denote the experience of the operator in years and Y denote the performance as given by the number of good parts turned out per 100 pieces.

Since we have to estimate the performance, these the regression equation of y on x is given by

$$y = a + bx \qquad \text{..... (I)}$$

Where a and b satisfy the normal equations.

Scanned by CamScanner

$\Sigma y = na + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$.

Now by calculator we get.

$n = 8$, $\Sigma x = 85$ $\Sigma y = 658$ $\Sigma x^2 = 1097$

$\Sigma y^2 = 54556$ $\Sigma xy = 7242$

Substituting these values the normal equation reduces to,

$8a + 85b = 658$

$85a + 107b = 7242$

Solving we get $a = 68.5$ $b = 1.29$.

Hence the required regression equation is

$y = 68.5 + 1.29x$ ..... (2)

(i) The performance of an operator has 10 year experience is

$= 71.92 + 1.29 \times 10 = 81.44$ good parts out of 100.

(ii) Also the coefficient of determination is given by

$$r^2 = \frac{a\Sigma y + b\Sigma xy - n(\bar{y})^2}{\Sigma y^2 - n(\bar{y})^2} = \frac{68.5 \times 658 + 1.29 \times 7242 - 8 \times 6765.06}{54556 - 8 \times 6765.06} = \frac{294.7}{435.52} = 0.68$$

8. a. calculate the correlation co-efficient for the following height in inches of father (X) and their sons (Y). [2006 spring]

X:	65	66	67	68	68	69	70	72
Y:	67	68	65	68	72	72	69	71

Solution 8a

Let X denote the height of father in inches, and Y denote the height of sons in inches. Then the correlation between them are as follows:

X	Y	$u = X - 68$	$v = Y - 68$	u^2	v^2	uv
65	67	−3	−1	9	1	3
66	68	−2	0	4	0	0
67	65	−1	−3	1	9	3
67	68	−1	0	1	0	0
68	72	0	4	0	16	0
69	72	1	4	1	16	4
70	69	2	1	4	1	2
72	71	4	3	16	9	12
		$\Sigma u = 0$	$\Sigma v = 8$	$\Sigma u^2 = 36$	$\Sigma v^2 = 52$	$\Sigma uv = 24$

We know that

$$r = \frac{n\Sigma uv - \Sigma u.\Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}} = \frac{8 \times 24 - 0 \times 8}{\sqrt{8 \times 36 - 0^2}\sqrt{8 \times 52 - (8)^2}} = \frac{192}{\sqrt{288}.\sqrt{252}} = 0.603.$$

b. Find the correlation co-efficient between x and y from the given data:

X:	78	89	97	69	59	79	68	57
Y:	125	137	156	112	107	138	123	108

Scanned by CamScanner

Solution 8b

Calculation of correlation coefficient

x	y	u = X – 77	v = Y – 130	u^2	v^2	uv
78	125	1	– 5	1	25	5
89	137	12	7	144	49	84
97	156	20	– 4	400	16	– 80
69	112	– 8	– 18	64	324	144
59	107	– 18	– 23	324	529	414
79	138	2	8	4	64	16
68	123	– 9	– 7	81	49	63
57	106	– 20	– 22	400	484	440
		$\Sigma u = 20$	$\Sigma v = 64$	$\Sigma u^2 = 1418$	$\Sigma v^2 = 1540$	$\Sigma uv = 1086$

We know that

$$r = \frac{n\Sigma uv - \Sigma u.\Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}} = \frac{8 \times 1086 - (-20)(-64)}{\sqrt{8 \times 1418 - (20)^2}\sqrt{8 \times 1540 - (64)^2}} = 0.78$$

c. Find the correlation co-efficient between x and y from the given data:

X:	92	89	87	86	83	77	71	63	53	50
Y:	86	88	91	77	68	85	52	82	37	57

Solution 8c

Calculation of correlation coefficients

x	y	u = X – 72	v = Y – 64	u^2	v^2	uv
92	86	20	22	400	484	440
89	88	17	24	289	576	408
87	91	15	27	225	729	405
86	77	14	13	196	169	182
83	68	11	4	121	16	44
77	85	5	21	25	441	105
71	52	– 1	– 12	1	144	12
63	82	– 9	18	81	324	– 162
53	37	– 19	– 27	361	729	513
50	57	– 22	– 7	484	49	154
		$\Sigma u = 31$	$\Sigma u = 83$	$\Sigma u^2 = 2183$	$\Sigma v^2 = 3661$	$\Sigma uv = 2101$

The correlation coefficient

$$r = \frac{n\Sigma uv - \Sigma u \times \Sigma u}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma u^2 - (\Sigma u)^2}} = \frac{10 \times 2101 - 31 \times 83}{\sqrt{10 \times 2183 - 31^2}} = \frac{18437}{\sqrt{20869} \times \sqrt{29721}}$$

$= 0.74$

9. **Calculate the co-efficient of correlation for the following ages of husbands and wives:**

Husband's age x:	23	27	28	28	29	3	31	33	35	36
Wife's age y:	18	20	22	27	21	29	27	29	285	29

Solution 9

Let X denote the age of husbands and Y denote the age of wife's in years The correlation between them is given as follow:

Scanned by CamScanner

x	Y	U = X − 29	v = y − 23	u^2	v^2	uv
23	18	−6	−5	36	25	30
27	20	−2	−3	4	9	6
28	22	1	−1	1	1	−1
28	27	1	4	1	16	4
29	21	0	−2	0	4	0
30	29	1	6	1	36	6
31	27	2	4	4	16	8
33	29	4	6	16	36	24
35	28	6	5	36	25	30
36	29	7	6	49	36	42
		$\Sigma u = 14$	$\Sigma u = 20$	$\Sigma u^2 = 148$	$\Sigma v^2 = 204$	$\Sigma uv = 112$

The correlation coefficient is given by

$$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{10 \times 112 - 14 \times 20}{\sqrt{10 \times 148 - 14^2}\sqrt{10 \times 204 - 20^2}} = \frac{840}{\sqrt{1284}\sqrt{1640}} = 0.58$$

10. Establish the formula $\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y - 2r\sigma_x\sigma_y$, where r is the correlation between x and y.

Solution 10

Let $Z = X - Y$

Taking expectation we get,

$E(Z) = E(X - Y)\ [E(X) - E(Y)]$

$Z - E(Z) = \{X - E(X)\} - \{Y - E(Y)\}$

Squaring both sides we get,

$[Z - E(Z)]^2 = [X - E(X)]^2 - 2[X - E(X)][Y - E(Y)] + [Y - E(Y)]^2$

Taking expectation we get,

$E[Z - E(Z)]^2 = E[(X - E(X)]^2 - 2[X - E(X)][Y - E(Y)] + [Y - E(Y)]^2]$

$\sigma_z^2 = E[X - E(X)]^2 - 2E(X - E(X))[Y - E(Y)] + E[Y - E(Y)]^2$

$= \sigma_X^2 - 2\,Cov(X,Y) + \sigma_Y^2$

$\sigma_z^2 = \sigma_X^2 - 2r\,\sigma_X\sigma_Y + \sigma_Y^2$ $\left[\because r = \frac{cov(X,Y)}{\sigma_X.\sigma_Y}\right]$

Hence $\sigma(_{x-y})^2 = \sigma_X^2 - 2r\,\sigma_X\sigma_Y + \sigma_Y^2$

11. a. Calculate the co-efficient of correlation for the following table:

y \ x	16–18	18–20	20–22	22–24
10–20	2	1	1	
20–30	3	2	3	2
30–40	3	4	5	6
40–50	2	2	3	4
50–60		1	2	2
60–90		1	2	1

Scanned by CamScanner

Solution 11a

Calculation of correlation coefficient of bivariate distribution;

		X	16 – 18	18 – 20	20 – 22	22 – 24				
		Mid value (x)	17	19	21	23				
Y	Mid value (y)	$u = x - 20$ / $v = \frac{x-45}{10}$	– 3	– 1	1	3	f	fv	fv^2	fuv
10 – 20	15	– 3	2 | 18	1 | 3	1 | – 3		4	– 9	36	18
20 – 30	25	– 2	3 | 16	2 | 4	3 | – 6	2 | – 12	10	– 20	40	4
30 – 40	35	– 1	3 | 9	4 | 4	5 | – 5	6 | – 18	18	– 18	18	–18
40 – 50	45	0	2 | 0	2 | 0	3 | 0	4 | 0	11	0	0	0
50 – 60	55	1		1 | – 1	2 | 2	2 | 2	5	5	5	7
60 – 80	75	3		1 | – 3	2 | 6	2 | 6	4	12	36	12
		f	10	11	16	15	52	– 33	135	31
		fu	– 30	– 11	16	45	20			
		fu^2	90	11	16	135	252			
		fuv	45	7	– 6	– 15	31			

Form the above table we get.

$N = \Sigma f = 52 \qquad \Sigma fu = 20 \qquad \Sigma fu^2 = 252$

$\Sigma fv = -33 \quad \Sigma fv^2 = 135 \quad \Sigma fuv = 31$

$$\because \quad r = \frac{N.\Sigma fuv - (\Sigma fu).(\Sigma fv)}{\sqrt{N.\Sigma fu^2 - (\Sigma fu)^2}\sqrt{N.\Sigma fv^2 - (\Sigma fv)^2}}$$

$$= \frac{52 \times 31 - 20 \times (-33)}{\sqrt{52 \times 252 - (20)^2}\sqrt{52 \times 125 - (-33)^2}} = \frac{2272}{\sqrt{1274 \times 5931}} = 0.262$$

b. Find the correlation between x(marks in mathematics) and y (marks in engineering Drawing) given in the following data.

x / y	10–40	40–70	70-100	total
0–30	5	20	-	25
30–60	-	28	2	30
60–90	-	32	13	45
Total	5	80	15	100

Solution 11b

b. Calculation of correlation coefficient of bivariate distribution

		X	10–40	10–70	70–100				
		Mid value (X)	25	55	85				
Y	Mid value (Y)	$v = \frac{x-45}{40}$ v $= \frac{X-55}{10}$	– 3	0	3	f	fv	fv^2	fuv
0 – 30	15	– 3	5 | 45	20 | 0		25	– 75	225	45
30 – 60	45	0		28 | 0	2 | 0	30	0	0	0
60 – 90	75	3		32 | 0	13 | 117	45	135	405	117
		f	5	80	15	100	60	630	162
		fu	– 15	0	45	30			
		fu^2	45	0	135	180			
		fuv	45	0	117	162			

Scanned by CamScanner

$$\because \quad r = \frac{N\Sigma fuv - (\Sigma fu).(\Sigma fv)}{\sqrt{N.\Sigma fu^2 - (\Sigma fu)^2}\sqrt{N.\Sigma fv^2 - (\Sigma fv)^2}}$$

$$= \frac{100 \times 163 - 30 \times 60}{\sqrt{100 \times 180 - 30^2}\sqrt{100 \times 630 - (60)^2}} = \frac{14400}{\sqrt{17100 \times 59400}} = 0.452$$

12. Ten students got the following % og marks in chemistry and physics.

Students:	1	2	3	4	5	6	7	8	9	10
Marks in chemistry:	78	36	98	25	75	82	90	62	65	39
Marks in physics:	84	51	91	60	68	62	86	58	63	47

Calculate the rank correlation co-efficient.

Solution 12

Calculation of Spearman's Rank Correlation

Marks in chemistry	Marks in physics	R_1	R_2	$d = R_1 - R_2$	d^2
78	84	4	3	−1	1
36	51	9	9	0	0
98	91	1	1	0	0
25	60	10	7	3	9
75	68	5	4	1	1
82	62	3	6	−3	9
90	86	2	2	0	0
62	58	7	8	−1	1
65	63	6	5	1	1
39	47	8	10	−2	4
					$\Sigma d^2 = 26$

Now the spearman's Rank correlation coefficient is given by

$$R = 1 - \frac{6\Sigma d^2}{n^3 - n} = 1 - \frac{6 \times 26}{10^3 - 10} = \frac{744}{900} = 0.827$$

14. A sample of 12 fathers and their eldest sons gave the following data about their heights in inches:

Father:	65	63	67	64	68	62	70	66	68	67	69	71
sons:	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the co-efficiency of rank correlation. [T. U. 2068]

Solution 14

Calculation of rank correlation coefficients

Height of fathers	Height of sons	R_1	R_2	$d = R_1 - R_2$	d^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	−1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	−3.5	12.25
66	65	8	11.5	−3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	−1.5	2.22
69	68	3	5.5	−2.5	6.25
71	70	1	2	−1	1
					$\Sigma d^2 = 72.5$

In X, the value 68 repeat twice with rank 4 and 5 so the common rank is $\frac{4+5}{2} = 4.5$

Scanned by CamScanner

Also the value 67 repeat twice with ranks 6 and 7, thus common rank is $\frac{6+7}{2} = 6.5$

Also In Y, the value 68 repeat 4 times with ranks 4, 5, 6, 7, the common rank is $\frac{4+5+6+7}{4} = 5.5$.

The value 66 repeats 2 times and 65 also repeats two times.

Hence the rank correlation is given by

$$R = 1 - 6\left[\frac{f\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12} + \frac{m_4(m_4^2-1)}{12} + \frac{m_5(m_5^2-1)}{12}}{n^3 - n}\right]$$

$$= 1 - \frac{6\left[72.5 + \frac{2\times3}{12} + \frac{2\times3}{12} + \frac{4\times15}{12} + \frac{2\times3}{12} + \frac{2\times3}{12}\right]}{12^3 - 12} = \frac{1716 - 6\times79.5}{1716} = 0.722$$

15. If $r = 0$, show that the two line of regression are parallel to the axes.

Solution 15

Since r is the correlation coefficient between two variables x and y then

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

If $r = 0$ then we get,

$$n\Sigma xy = \Sigma x . \Sigma y \qquad \text{.....} (1)$$

Let b_{yx} and b_{xy} be the regression coefficient of y on x and x on y. given by

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x^2)} = 0 = b_{xy}.$$

Now the regression line of Y on X is given by $y - \bar{y} = b_{yx}(x - \bar{x})$

$\Rightarrow y - \bar{y} = 0 \Rightarrow y = \bar{y}$ Which is a line parallel to x-axis.

Also the regression line of X on Y is given by $x - \bar{x} = b_{xy}(y - \bar{y})$

$\Rightarrow x - \bar{x} = 0 \qquad \Rightarrow x = \bar{x}$ which is a line parallel to y-axis.

Hence if $r = 0$ the regression lines are parallel to axis.

16. If the two regressions co-efficient were 0.8 and 0.2, what would be the value of co-efficient correlation?

Solution 16

We have given two regression coefficients $b_{xy} = 0.8$ and $b_{yx} = 0.2$

Then $r = \pm\sqrt{b_{yx}.\, b_{xy}} = \sqrt{0.8\times0.2} = \sqrt{0.16} = 0.4$.

17. a. Find the correlation co-efficient and the equations of regres sion lines for the following values of x and y:

x:	1	2	3	4	5
y:	2	5	3	8	7

Scanned by CamScanner

Solution 17a

Calculation of correlation and regression between x and y.

x	y	x^2	y^2	xy
1	2	1	4	2
2	5	4	25	10
3	3	9	9	9
4	8	16	64	32
5	7	25	49	35
$\Sigma x = 15$	$\Sigma y = 25$	$\Sigma x^2 = 55$	$\Sigma y = 151$	$\Sigma xy = 88$

The correlation coefficient between x and y is given by

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{5 \times 88 - 15 \times 25}{\sqrt{5 \times 55 - 15^2}\sqrt{5 \times 151 - 25^2}} = \frac{65}{\sqrt{50}\sqrt{130}} = 0.806$$

Also, $\bar{X} = \frac{\Sigma x}{n} = \frac{15}{5} = 3, \quad \bar{Y} = \frac{\Sigma y}{x} = \frac{25}{5} = 5$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{5 \times 88 - 15 \times 25}{5 \times 55 - 15^2} = \frac{65}{50} = 1.3$$

$$b_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2} = \frac{65}{130} = 0.5$$

The regression equation of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - 5 = 1.3(x - 3) \Rightarrow y = 1.3x + 1.1$$

Also the regression equation of x on y is given by

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - 3 = 0.5(y - 5) \quad \Rightarrow x = 0.5y + 0.5$$

b. Find the correlation co-efficient between x and y for the given values. Find also the two regression lines. [P. U. 2010 S]

x:	1	2	3	4	5	6	7	8	9	10
y:	10	12	16	28	25	36	41	49	40	50

Solution 17b

The calculation of correlation and regression between x and y.

x	y	x^2	y^2	xy
1	10	1	100	10
2	12	4	144	24
3	16	9	256	48
4	28	16	784	112
5	25	25	625	125
6	36	36	1296	216
7	41	49	1681	287
8	49	64	2401	292
9	40	81	1600	360
10	50	100	2500	500
$\Sigma x = 55$	$\Sigma y = 307$	$\Sigma x^2 = 385$	$\Sigma y = 11387$	$\Sigma xy = 2074$

Scanned by CamScanner

We know that $\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7$$

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x)\,\Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{10 \times 2074 - 55 \times 307}{10 \times 385 - 55^2} = \frac{3855}{825} = 4.67$$

$$b_{xy} = \frac{n.\,\Sigma xy - \Sigma x.\,\Sigma y}{n.\,\Sigma y^2 - (\Sigma y)^2} = \frac{3855}{19621} = 0.196$$

$$\because \quad r = \sqrt{b_{xy} \times b_{xy}} = \sqrt{0.196 \times 4.67} = 0.96$$

The regression equation of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - 30.7 = 4.67\,(x - 5.5) \qquad \Rightarrow y = 4.67x + 5.015$$

Also the regression equation of x on y is given by

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - 5.5 = 0.196\,(y - 30.7)$$

$$x = 0.196y - 0.52$$

18. The two regression equations of the variables x and y are $x = 19.13 - 0.87y = y = 11.64 - 0.50x$. Find (i) mean of x' s, (ii) mean of y' s and (iii) the correlation of coefficient between x and y.

Solution 18

The regression equation of x on y is $x = 19.13 - 0.87y$ ..... (1)

The regression equation of y on x is $y = 11.64 - 0.50x$ .....(2)

Since the regression lines passes though their means i.e. $(\bar{x},\bar{y})$ then

$\bar{x} = 19.13 - 0.87\bar{y}$ and

$\bar{y} = 11.64 - 0.50\,\bar{x}$

Solving we get, $\bar{x} = 15.935$ and $\bar{y} = 3.67$

Form equation (1) $b_{xy} = -0.87$

Form equation (2) $b_{yx} = -0.50$

$\because$ Correlation coefficient between a and y

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{(-0.87) \times (-0.50)} = 0.66$$

19. Two random variables have the regression lines with equations $3x + 2y = 26$ and $6x + y = 31$. Find the mean values and the correlation coefficient between x and y.

Solution 19

Here given line is,

$$6x + y = 31 \Rightarrow x = -\frac{y + 31}{6} \Rightarrow x = -0.167y + 5.167$$

Which is the regression equation of x on y with $b_{xy} = -0.167$.

Also, $3x + 2y = 26 \Rightarrow y = \frac{-3x + 26}{2} \Rightarrow y = -1.5x + 13$

Which is the regression equation of y on x with $b_{yx} = -1.5$

Scanned by CamScanner

$\because$ Correlation coefficient $r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{(-0.167) \times (-1.5)} = 0.2505$.

Since the regression equations passes though their means $(\bar{x}, \bar{y})$ then we have,

$3\bar{x} + 2\bar{y} = 26$ and $6\bar{x} + \bar{y} = 31$

Solving we get $\bar{x} = 4$ and $y = 7$.

20. In a partially destroyed laboratory data, only the equations giving the two lines of regression y on x and x and yare available and are respectively, $7x - 16y + 9 = 0$, $5y - 4x - 3 = 0$.

Calculate the coefficient of correlation, $\bar{x}$ and $\bar{y}$.

Solution 20

Here given line is,

$7x - 16y + 9 = 0 \Rightarrow 16y = 7x + 9 \quad \Rightarrow y = 0.44x + 0.56$

$\because b_{yx} = 0.44$

Again, $5y - 4x - 3 = 0$

$4x = 5y + 3 \Rightarrow x = 1.25\,y + 0.75$

$\because b_{xy} = 1.25$

$\because r = \sqrt{b_{xy} + b_{yx}} = \sqrt{0.44 \times 1.25} = 0.74$

Also since the regression line passes through their means $(\bar{x}, \bar{y})$ so, the have, $7\bar{x} - 16\bar{y} + 9 = 0$ and $5\bar{y} - 4\bar{x} - 3 = 0$

Solving the equations we get $\bar{x} = -0.1$ and $\bar{y} = 0.517$.

22. The following table gives the ages and blood pressure fo 10 women.

Age X:	56	42	36	47	49	42	60	72	63	55
Blood Pressure Y:	147	125	118	128	145	140	155	160	149	150

i Find the correlation coefficient between X and Y.

ii Determine the least square regression equation of Y on X.

iii Estimate the blood pressure of a women whose age is 45 years.

[P. U 2008 F, 2009 S, 2013 S, 2015 F, 2017 S, 2011 S]

Solution 22

Let X denote the age and Y denote the blood pressure of 10 women's. Then the correlation and regression coefficients are given as follows:

X	Y	u = X – 50	v = Y – 136	u^2	v^2	uv
56	147	6	11	36	121	66
42	125	– 8	– 11	64	121	88
36	118	– 14	– 18	196	324	252
47	128	– 3	– 8	9	64	24
49	145	– 1	9	1	81	– 9
42	140	– 8	4	64	16	– 32
60	155	10	19	100	261	190
72	160	22	24	484	576	528
63	149	13	13	169	169	169
55	150	5	14	25	196	70
		$\Sigma u = 22$	$\Sigma v = 57$	$\Sigma u^2 = 1148$	$\Sigma v^2 = 2029$	$\Sigma uv = 1346$

Scanned by CamScanner

(i) We know that $r = \dfrac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n.\Sigma v^2 - (\Sigma v)^2}}$

$$= \frac{10 \times 1346 - 22 \times 57}{\sqrt{10 \times 1148 - 10^2}\sqrt{10 \times 2029 - 57^2}} = \frac{12206}{\sqrt{10996}\sqrt{17041}} = 0.89$$

(ii) We know that:

$$\bar{x} = A + \frac{\Sigma u}{n} = 50 + \frac{22}{10} = 52.2$$

$$\bar{y} = B + \frac{\Sigma v}{n} = 136 + \frac{57}{10} = 141.7$$

Also $b_{yx} = \dfrac{n\Sigma uv - (\Sigma u).\Sigma v}{n.\Sigma u^2 - (\Sigma u)^2} = \dfrac{10 \times 1346 - 22 \times 57}{10 \times 1148 - 22^2} = \dfrac{12206}{10996} = 1.11$

The regression line of y on x is given by $y - \bar{y} = b_{yx}(x - \bar{x})$

$y - 141.7 = 1.11(x - 52.2)$

$y = 1.11x + 83.785$

(iii) When $x = 45$ then

$y = 1.11 \times 45 + 83.758 = 133.708 \approx 134$

23. Two random variables have the least square regression lines with equation $3X + 2Y - 26 = 0$ and $6X + Y - 31 = 0$.

Find:

i the mean of X and Y.

ii the regression coefficient of X and Y.

iii the coefficient of correlation between Y on X and X on Y.

Solution 23

Repeated question (19)

24. Following are the data pertaining to the production and export of sugar in lakh tones.

Production (X)	Export(Y)
37.4	3.90
31.1	1.33
38.7	1.1
39.5	4.39
47.9	9.41
42.6	9.67
48.6	3.41
64.6	2.51
58.4	8.62
38.6	9.9
51.4	6.64
84	6.5

a. Find the regression of Y on X.

b. Test whether the regression line in the population pass through the origin.

c. What export of sugar can be expected when the production is 50 lakh tones?

Scanned by CamScanner

Solution 24

Let X denote the production and Y denote the export of sugar in lakh tons. Then by calculator we have,

$n = 12, \Sigma x = 582.8\ \Sigma y = 67.38$

$\Sigma x^2 = 30666.68 \qquad \Sigma y^2 = 496.1\ \Sigma xy = 3366.595$

$\bar{x} = \frac{\Sigma x}{n} = 48.57, \qquad \bar{y} = \frac{\Sigma y}{n} = 5.615$

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n.\ \Sigma x^2 - (\Sigma x)^2} = \frac{12 \times 3366.595 - 582.8 \times 67.38}{12 \times 30666.68 \times (582.8)^2} = \frac{1130}{28344.32} = 0.04$$

(a) Now the regression line of y on x is given by $y - \bar{y} = b_{yx}(x - \bar{x})$

or $\quad y - 5.615 = 0.04\ (x - 48.57)$

or $\quad y = 0.04x + 3.67$

(c) When the production is 50 lakh tons then the export is,

$y = 0.04 \times 50 + 3.67 = 5.67$

25. The coefficient of rank correlation of the marks obtained by 10 students in statistics and accountancy was found to be 0.2. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 9 instead of 7. Find the correct value of coefficient of rank correlation.

Solution 25

Let X denote the marks in statististic and Y denote the marks in Accountancy.

Here given $n = 10, \qquad R = 0.2$

Wrong $|R_1 - R_2| = 9 \qquad$ Correct $|R_2 - R_2| = 7.$

Correct rank correlation = ?

We know that:

$$R = 1 - \frac{\Sigma d^2}{n^3 - n}$$

$$\text{or,} \quad 0.2 = 1 - \frac{\Sigma d^2}{990}$$

$$\text{or,} \quad \frac{\Sigma d^2}{990} = 0.8 \quad \Rightarrow \Sigma d^2 = 792$$

$$\text{Correct Rank Correlation (R)} = 1 - \frac{\text{correct } \Sigma d^2}{n^2 - n} = 1 - \frac{792 - 9^2 + 7^2}{990} = 1 - \frac{760}{990} = 0.23$$

26. A student calculates the value of n is 10 and concludes that r is highly significant. Is his conclusion corrects? Determine the limits of population correlation coefficient.

Solution 26

Here $r = 0.81$ and $n = 10$

$$\text{Also, P.E (r)} = 0.6745\left(\frac{1 - r^2}{\sqrt{n}}\right) = \frac{0.6745.\ [1 - (0.81)^2]}{\sqrt{10}} = 0.0734$$

Scanned by CamScanner

Now, $6 \times P.E(r) = 6 \times 0.0734 = 0.44$

Since, $r = 0.81 > 6.\ P.E(r)$

So r is significant.

Yes his conclusion is correct.

The limits of population correlation coefficient is given by

$[r \pm P.E(r)] = (0.81 \pm 0.0734) = (0.81 - 0.0734, 0.81 + 0.0734) = (0.7366, 0.8834)$

27. The following table gives the distribution of the total population and those who are totally or partially blind among them. find out if there is any relation between age and blindness.

Age(years):	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
No. of persons	100	60	40	36	24	11	6	3
No. of blind	55	40	40	40	36	22	18	15

Solution 27

We are going to find the correlation between age and blindness of the people. First we calculate % number of b blindness to the total No. of person.

Age (in years)	Mid values (x)	% of blindness people (y)
0 – 10	5	55/10000× 100 = 0.055
10 – 20	15	40/60,000× 100 = 0.067
20 – 30	25	40/40,000× 100 = 0.1
30 – 40	35	40/3600 × 100 = 0.111
40 – 50	45	= 0.15
50 – 60	55	= 0.2
60 – 70	65	= 0.3
70 – 80	75	0.5

By calculator we get $n = 8$, $\Sigma x = 320$ $\Sigma y = 1.483$

$\Sigma x^2 = 17000$ $\Sigma y^2 = 0.43$

$\Sigma xy = 82.415$

$$\because \quad r = \frac{n.\Sigma xy - \Sigma x.\Sigma y}{\sqrt{n.\Sigma x^2 - (\Sigma x)^2} \times \sqrt{n.\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{8 \times 82.415 - 320 \times 1.483^2}{\sqrt{8 \times 17000 - 320^2} \times \sqrt{8 \times 0.43 - 1.483^2}} = \frac{184.76}{\sqrt{33600} \times \sqrt{1.24}} = 0.9$$

28. A computer while calculating the correlation coefficient between the variable X and Y obtained the following results.

$n = 30$, $\Sigma x = 120$, $\Sigma X^2 = 600$, $\Sigma Y = 90$, $\Sigma Y^2 = 250$, $\Sigma XY = 356$

It was, however, later discovered at the time of checking that it had copied down two pairs of observation as:

X	Y
8	10
12	7

While the correct value were

X	Y
8	12
10	8

Obtain the correct value of the correlation coefficient between X and Y.

Scanned by CamScanner

Solution 28

Here given

$n = 30, \Sigma x = 120, \Sigma x^2 = 600, \Sigma y = 90, \Sigma y^2 = 250$

The wrong input are (8, 10) and (12, 7)

The correct input are, (8, 12) and (10, 8)

Correct $\Sigma x = 120 - 12 + 8 - 8 + 10 = 118$

Correct $\Sigma y = 90 - 10 - 7 + 12 + 8 = 93$

Correct $\Sigma y^2 = 600 - 8^2 - 12^2 + 8^2 + 10^2 = 556$

Correct $\Sigma y^2 = 250 - 10^2 - 7^2 + 12^2 + 8^2 = 309$

Correct $\Sigma xy = 356 - 8 \times 10 - 12 \times 7 + 8 \times 12 + 10 \times 8 = 368.$

$$\because \quad \text{Correct } r = \frac{n.\Sigma xy - \Sigma x.\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n.\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{36 \times 368 - 118 \times 93}{\sqrt{30 \times 556 - (118)^2}\sqrt{30 \times 309 - 93^2}} = \frac{66}{\sqrt{2756}.\sqrt{621}} = 0.05$$

29. What is correlation and regression? Find correlation, coefficient and the standard error, probable error, confidence interval, and significant test of following data.

Solution 29

Calculation of correlation coefficient between the variables x and y.

x	y	x^2	y^2	xy
8	1	64	1	8
7	2	49	4	14
1	3	1	9	3
3	4	9	16	12
5	5	25	25	25
6	6	36	36	36
9	7	81	49	63
12	8	144	64	96
51	36	409	204	257

Form know that the correlation coefficient between two variables x and y are given by

$$r = \frac{n.\Sigma xy - \Sigma x.\Sigma y}{\sqrt{n.\Sigma x^2 - (\Sigma x)^2}\sqrt{n.\Sigma y^2 - (\Sigma y)^2}}$$

$$r = \frac{8 \times 257 - 51 \times 36}{\sqrt{8 \times 409 - 51^2}.\sqrt{8 \times 204 - 36^2}} = \frac{220}{\sqrt{671 \times 336}} = 0.46$$

(b) Now standard Error S.E (r) $= \frac{1 - r^2}{\sqrt{n}} = \frac{1 - (0.46)^2}{\sqrt{8}} = 0.28$

(c) Also Probable Error P.E (r) = 0.6745 × S.E (r) = 0.6745 × 0.28 = 0.19.

(d) Now 6.P.E (r) = 6 × 0.19 = 1.14 > r So

Hence the r is not significant.

(e) Also the confidence interval is given by

(r – P.E (r), r + P.E.(r) = (0.46 – 0.19, 0.46 + 0.19) = (0.27 0.65).

Scanned by CamScanner

30. Calculate the rank, correlation coefficient and Karl Pearson's correlation coefficient from the following data.

X	80	78	75	75	68	67	68	59
Y	12	13	14	14	14	16	15	17

Compare these correlation coefficients.

Solution 30

Calculation of Karl Pearson's coefficient of correlation and Rank correlations;

X	Y	u = X – 70	v=Y – 14	u^2	v^2	uv	R_1	R_2	$d^2=(R_1-R_2)^2$
80	12	10	–2	100	4	–20	1	8	49
78	13	8	–1	64	1	–8	2	7	25
75	14	5	0	25	0	0	3.5	5	2.25
75	14	5	0	25	0	0	3.5	5	2.25
68	14	–2	0	4	0	0	5.5	5	0.25
67	16	–3	2	9	4	–6	7	2	25
68	15	–2	1	4	1	–2	5.5	3	6.25
59	17	–11	3	121	9	–33	8	1	49
		$\Sigma u = 10$	$\Sigma v = 3$	$\Sigma u^2 = 352$	$\Sigma v^2 = 19$	$\Sigma uv = -69$			$\Sigma d^2 = 159$.

(a) The Karl Pearson's correlation coefficient between x and y is given by

$$r = \frac{n.\Sigma uv - \Sigma u.\Sigma v}{\sqrt{n.\Sigma u^2 - (\Sigma u)^2}\sqrt{n.\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{8 \times (-69) - 10 \times 3}{\sqrt{8 \times 352 - 10^2} \times \sqrt{8 \times 19 - 3^2}} = \frac{-582}{\sqrt{2716} \times \sqrt{143}} = 0.93$$

(b) Also the spearman's rank correlation between x and y given by

$$R = 1 - \frac{6\left[\Sigma d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}\right]}{n^3 - n}$$

$$= 1 - \frac{6\left[159 + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{3 \times 8}{12}\right]}{8^3 - 8} = 1 - \frac{972}{504} = -\frac{468}{504} = -0.929$$

Hence both are approximately equal and both represents the perfect negative correlation between x and y.

31. For 10 observations on price (x) and supply (Y), the following data were obtained (in appropriate units):

$\Sigma x = 130$, $\Sigma y = 220$, $\Sigma x^2 = 2288$, $\Sigma y^2 = 5506$ and $\Sigma xy = 3467$

i. Obtain the two lines of regression and estimate the supply when the price is 16 units.

ii. Find correlation coefficient and interpret the result in practical manner.

iii. Determine coefficient of determination and interpret the result.

Solution 31

Here we have given,

Scanned by CamScanner

$n = 10, \Sigma x = 130, \Sigma y = 220, \Sigma x^2 = 2288$

$\Sigma y^2 = 5506$ and $\Sigma xy = 3467$.

Now, $\bar{x} = \frac{\Sigma x}{n} = \frac{130}{10} = 13$, $\bar{y} = \frac{\Sigma y}{n} = \frac{220}{10} = 22$

$$b_{yx} = \frac{n.\Sigma xy - \Sigma x.\Sigma y}{n.\Sigma x^2 - (\Sigma x)^2} = \frac{6070}{5980} = 1.02$$

$$b_{xy} = \frac{n.\Sigma xy - \Sigma x.\Sigma y}{n.\Sigma y^2 - (\Sigma y)^2} = \frac{6070}{6660} = 0.91$$

(a) The regression line of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$y - 22 = 1.02\,(x - 13) \Rightarrow y = 1.02x + 8.74.$ .....(1)

Also the regression line of x on y is given by

$x - \bar{x} = b_{xy}\,(y - \bar{y})$

or $x - 13 = 0.91\,(y - 22) \Rightarrow x = 0.91y - 7.02.$ ..... (2)

(i) Name the supply (y) when price (x) = 16 units

is, $y = 1.02 \times 16 + 8.74 = 25.06.$

(ii) The correlation coefficient between x and y is given by

$r = \pm\sqrt{b_{xy} \times b_{yx}} = \sqrt{1.02 \times 0.91} = 0.963$

Now, $P.E\,(r) = 0.6745\left(\frac{1 - r^2}{\sqrt{n}}\right) = 0.6745 \times 0.023 = 0.016.$

Also $6 \times P.E\,(r) = 6 \times 0.016 = 0.093 < r$

So the value of r is significant.

Also there is high degree of positive correlation between price and supply.

(iii) Also the coefficient of determination

$r^2 = b_{xy} \times b_{yx} = 1.02 \times 0.91 = 0.9282$

This shows that approximately 93%

One variable accounts of the total variation.

Only 7% is unaccounted. So the value of r is statistically significant.

Exercise 15.2

1. The following are data on the ages and income of five executive working for the same company and the number of years they went to college:

Age x_1	years college x_2	Income (dollars) y
37	4	71200
45	0	66800
38	5	70300
42	2	65400
31	4	63500

Fit an equation of the form $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ to the given data, and use it to estimate how much on the average working for this company will make if he is 40 years old and has had 4 years of college.

Scanned by CamScanner

Solution 1

Let x_1 denote the age, x_2 denote the Years and y denote the income in dollars. Then the regression equation is given by

$$y = a + bx_1 + cx_2 \text{ ..... (i)}$$

Then,

x_1	x_2	y	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
37	4	71200	1369	16	148	2634400	284800
45	0	66,800	2025	0	0	3006000	0
38	5	70,300	1444	25	190	2671400	351500
42	2	65,400	1764	4	84	2746800	130800
31	4	64,500	961	16	124	1998500	254000
$\sum x_1 = 193$	$\sum x_2 = 15$	$\sum y = 337200$	$\sum x_1^2 = 7563$	$\sum x_2^2 = 61$	$\sum x_1x_2 = 546$	$\sum x_1y =$ 106228400	$\sum x_2y = 1021100$

(i) We know that (i) satisfy the normal equations

$\sum y = na + b\sum x_1 + c\sum x_2 \Rightarrow 5a + 193b + 15c = 377200$

$\sum x_1y = a\sum x_1 + b\sum x_1^2 + c\sum x_1x_2 \Rightarrow 193a + 7563\,b + 546c = 10628400$

$\sum x_2y = a\sum x_2 + b\sum x_1x_2 + c\sum x_2^2 \Rightarrow 15a + 546b + 61c = 1021100$

Solving we get,

$a = 35123.622,\ b = 681.778,\ c = 1999.917$

Then (1) becomes,

(ii) The value of y when $x_1 = 40$ and $x_2 = 4$ is

$$y = 35123.622 + 681.778 \times 40 + 1999.917 \times 4 = 70394.71$$

2. The following sample data were collected to determine the relationship between two processing variables and the current gain of a certain kind of transistor:

Diffusion time (hours) x_1	1.5	2.5	0.5	1.2	2.6	0.3	2.4	2.0	07	1.6
Sheet resistance (Ω cm) x_2	66	87	69	141	93	105	111	78	66	123
Current gain y	5.3	7.8	7.4	9.8	10.8	9.1	8.1	7.2	6.5	12.6

Fit a regression plane and use its equation to estimate the expected current gain when the diffusion time is 2.2 hours and the sheet resistance is 90 Ωcm.

Solution 2

Let x_1 denote the diffusion time in hours x_2 denote the sheet resistance and y denote the current gain. Then the regression equation of y on x_1 and x_2 is given by

$$y = a + bx_1 + cx_2 \text{ ..... (1)}$$

Which satisfy the normal equations

$$\sum y = na + b\sum x_1 + c\sum x_2 \text{ ..... (2)}$$

$$\sum x_1y = a\sum x_1 + b\sum x_1^2 + c\sum x_1x_2 \text{ ..... (3)}$$

$$\sum x_2y = a\sum x_2 + b\sum x_1x_2 + c\sum x_2^2 \text{ ..... (4)}$$

By calculator we get, n = 10

$\sum x_1 = 15.3$ $\quad \sum x_2 = 939$ $\quad \sum y = 84.6$

$\sum x_1^2 = 29.85$ $\quad \sum x_2^2 = 94131$

$\sum x_1x_2 = 1458.9$ $\quad \sum x_1y = 132.27$ $\quad \sum x_2y = 8320.2$

Scanned by CamScanner

Substituting these value (2), (3), (4) becomes.

$$10a + 15.3b + 939c = 84.6$$

$$15.3a + 29.85b + 1458.9c = 132.27$$

$$939a + 1458.9b + 94131c = 8320.2$$

Solving these equations we get.

$$a = -12.77 \quad b = 0.225 \quad c = 0.0623$$

Hence the regression line is

$$y = -12.76 + 0.225x_1 + 0.0623\,x_2$$

Also the expected current gain (y) when $x_1 = 2.2$ and $x_2 = 90$ is

$y = -12.76 + 0.225 \times 2.2 + 0.0623 \times 90 = 6.658.$

3. Fit a regression line to the following data:

Yield of crop (in '000' tones)	4	6	7	9	13	15
Amount of rainfall (in inches)	3	4	6	8	12	15
Amount of fertilizers (in kg)	4	10	14	20	24	30

Solution 3

Let Y denote the yield of crop in '000' tons x_1 denote the amount of rainfall in inches and x_2 denote the amount of fertilizers in kg.

Then the regression line is given by

$$y = a + bx_1 + c\Sigma x_2 \qquad \text{..... (1)}$$

Whose normal equations are,

$\Sigma y = na + b\Sigma x_1 + c\Sigma x_2$, $\Sigma x_1 y = a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2$

and $\Sigma x_2 y = a\Sigma x_2 + b\Sigma x_1 x_2 + cx_2^2$.

Now by calculator we get,

$n = 6$, $\Sigma x_1 = 36$, $\Sigma x_2 = 102$, $\Sigma y = 54$, $\Sigma x_1^2 = 494$

$\Sigma x_2^2 = 2188$, $\Sigma x_1 x_2 = 1034$, $\Sigma x_1 y = 531$, $\Sigma x_2 y = 1116$

Substituting the values the normal equations becomes:

$6a + 36b + 102c = 54,$

$36a + 494b + 1034c = 531$

$102a + 1034b + 2188c = 1116$

Solving we get, $a = -0.385$, $b = -0.201$, $c = 0.623$

Hence the regression equation is

$y = -0.385 - 0.201x_1 + 0.6823x_2$

4. A household survey on monthly expenditure on food yield following data:

Monthly expenditure (Rs '000')	10	15	20	25	30	35	40
Monthly income (Rs '000')	2	4	5	7	6	6	5
Size of family (number)	4	5	7	10	8	11	4

a. Estimate the line of best fit.

b. Obtain the multiple correlation coefficient and coefficient of determination.

[TU 2072/2067/2063/2059/2057]

Scanned by CamScanner

Solution 4

a. Let y denote the monthly expenditure in thousand, x_1 denote the monthly income in thousand and x_2 denote the size of family then the best fit regression line is,

$$y = a + bx_1 + cx_2 \text{ ..... (1)}$$

Where a, b, c satisfy the normal equations.

$\Sigma y = na + b\Sigma x_1 + c\Sigma x_2$,

$\Sigma x_1 y = a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2$ and

$\Sigma x_2 y = a\Sigma x_2 + b\Sigma x_1 x_2 + c\Sigma x_2^2$.

Now by the calculator we get,

$n = 7$,	$\Sigma x_1 = 35$	$\Sigma x_2 = 49$.	$\Sigma y = 175$
$\Sigma x_1^2 = 191$	$\Sigma x_2^2 = 391$	$\Sigma y^2 = 5075$	
$\Sigma x_1 x_2 = 267$	$\Sigma x_1 y = 945$	$\Sigma x_2 y = 1290$	

Now the normal equations reduces to

$$7a + 35b + 49c = 175$$
$$35a + 191b + 267c = 945$$
$$49a + 267b + 391c = 1290$$

Solving get,

$a = 3.345 \quad b = 6.795 \quad c = -1.76$

Hence the best fit regression line is

$$y = 3.345 + 6.795x_1 - 1.76x_2$$

b. Let x_1 denote the monthly expenditure x_2 denote the monthly income and x_3 denote the size of the family. Since x_1 is dependent upon x_1 and x_2.

$\because$ The multiple correlation of coefficient of x_2 and joint effect of x_2 and x_3 and x_3 is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{13}\, r_{12}\, r_{23}}{1 - r_{23}^2}} \quad \text{..... (A)}$$

Where, $$r_{12} = \frac{n.\Sigma x_1 x_2 - \Sigma x_1 . \Sigma x_2}{\sqrt{n.\Sigma x_1^2 - (\Sigma x_1)^2} \times \sqrt{n.\Sigma x_2^2 - (\Sigma x_2)^2}}$$

$$= \frac{7 \times 945 - 175 \times 35}{\sqrt{7 \times 5075 - (175)^2} . \sqrt{7 \times 191 - 35^2}} = \frac{490}{\sqrt{4900} \times \sqrt{42}}$$

$$= 0.66$$

Similarly $$r_{13} = \frac{7 \times 1290 - 175 \times 49}{\sqrt{7 \times 5075 - (175)^2} . \sqrt{7 \times 391 - (49)^2}} = \frac{455}{\sqrt{4900} \times \sqrt{336}} = 0.35$$

Also, $$r_{23} = \frac{7 \times 267 - 35 \times 49}{\sqrt{7 \times 191 - 35^2} . \sqrt{7 \times 391 - 49^2}} = \frac{154}{\sqrt{112}\sqrt{336}} = 0.79$$

Now equation reduces to,

$$R_{1.23} = \sqrt{\frac{(0.66)^2 + (0.35)^2 - 2 \times 0.35 \times 0.66 \times 0.79}{1 - (0.89)^2}} = \sqrt{\frac{0.193}{0.3759}} = 0.716$$

Also the coefficient of determination is given by $= (R_{1.23})^2 = (0.716)^2 = 0.513$

Scanned by CamScanner

5. Calculate the (s) Multiple correlation coefficient $R_{1.23}$ and interpret the result. (b) Partial correlation coefficient $r_{12.3}$; given that the simple correlation coefficient between sales in units (X_1), advertisement (X_2) and price in rupees (X_3) are $r_{12} = 0.70$, $r_{13} = 0.61$ and $r_{23} = 0.40$.

Solution 5

Let x_1 denote the sales in units, x_2 denote the advertisement and x_3 denote price in rupees.

Then $r_{12} = 0.7$, $r_{13} = 0.61$, $r_{23} = 0.40$

(a) The multiple correlation coefficient $r_{1.23}$ is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2.r_{12}.r_{13}.r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.7)^2 + (0.61)^2 - 2 \times 0.7 \times 0.61 \times 0.4}{1 - (0.4)^2}} = \sqrt{\frac{0.5205}{0.84}} = 0.787$$

(b) The partial correlation coefficient $r_{12.3}$ is given by

$$r_{12.3} = \frac{r_{12} - r_{13}.r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{0.7 - 0.61 \times 0.40}{\sqrt{1 - (0.61)^2}\sqrt{1 - (0.4)^2}} = \frac{0.456}{\sqrt{0.63}\sqrt{0.84}} = 0.63$$

6. Past experience shows that following result of productivity per hectare with the respective uses of chemical fertilizers and seeds. Calculate the coefficient of multiple correlation and coefficient of multiple determination. Also interpret the result. [TU 2065/2073/2068]

year	1976	1977	1978	1979	1980	1981
Fertilizer (kgs): (X_1)	45	30	70	75	65	80
Seeds (kgs): (X_2)	2	1.8	3	2.5	2	3
Productivity (X_3)	2000	2100	1800	1900	2400	2500

Solution 6

Calculation of multiple correlation between x_1, x_2 and x_3 as follows:

x_1	x_2	x_3	$d_1=x_1-55$	$d_2=x_2$	$d_3=x_3-\frac{2100}{100}$	d_1^2	d_2^2	d_3^2	d_1d_2	d_1d_3	d_2d_3
45	2	2000	−10	2	−1	100	4	1	−20	10	−0
30	1.8	2100	−25	1.8	0	625	3.24	0	−45	0	0
70	3	1800	15	3	−3	225	9	9	45	−45	−9
75	2.5	1900	20	2.5	−2	400	6.25	4	50	−40	−5
65	2	2400	5	2	3	25	4	9	10	15	6
80	3	2500	25	3	4	625	9	16	75	100	12
			$\Sigma d_1 = 30$	$\Sigma d_2 = 14.3$	$\Sigma d_3 = 1$	$\Sigma d_1^2 = 2000$	$\Sigma d_2^2 = 36.49$	$\Sigma d_3^2 = 39$	$\Sigma d_2 = 115$	$\Sigma d_2d_3 = 40$	Σd_2d_3

Now the multiple correlation between x_1, x_2 and x_3 is given by

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2.r_{12}.r_{13}.r_{23}}{1 - r_{12}^2}} \quad \text{..... (A)}$$

$$\text{Where, } r_{12} = \frac{n\Sigma d_1d_2 - \Sigma d_2.\Sigma d_2}{\sqrt{n.\Sigma d_1^2 - (\Sigma d_1)^2}\sqrt{n.\Sigma d_2^2 (\Sigma d_2)^2}}$$

Scanned by CamScanner

$$= \frac{6 \times 115 - 30 \times 14.3}{\sqrt{6 \times 2000 - 30^2} \times \sqrt{6 \times 35.49 - 14.3^2}} = \frac{261}{\sqrt{11100} \times \sqrt{8.45}} = 0.852$$

Similarly $r_{13} = \dfrac{6 \times 40 - 30 \times 1}{\sqrt{6 \times 2000 - 30^2}\sqrt{6 \times 29 - 12}} = 0.136$

Also, $r_{23} = \dfrac{6 \times 2 - 14.3 \times 1}{\sqrt{8.45} \times \sqrt{215}} = -0.054$

Now equation (A) reduces to

$$R_{3.12} = \sqrt{\frac{(-0.054)^2 + (0.136)^2 - 2 \times 0.852 \times 0.136 \times (0.054)}{1 - (0.852)^2}} = \sqrt{\frac{0.0339}{0.274}} = 0.35$$

Now the coefficient of determination is given by

$(R_{3.12})^2 = (0.35)^2 = 0.123$

7. The three variables production of rice (X_1), irrigation (X_2) and amount of fertilizer use (X_3) are correlated with each other. The following simple correlation coefficient were obtained from the past experience. $r_{12} = 0.92$, $r_{13} = 0.82$, $r_{23} = 0.93$. Find the correlation between production of rice and amount of fertilizer used, keeping the irrigation constant. Also calculate the correlation between production of rice and irrigation, keeping the amount of fertilizer and used constant.

Solution 7

Let x_1 denote the production of rice, x_2 denote the irrigation and x_3 denote the amount of fertilizer used. (Since the production of rice is depend upon irrigation and amount of fertilizer)

Also, $r_{12} = 0.92$, $r_{13} = 0.82$ and $r_{23} = 0.93$

The correlation between production of rice and amount of fertilizer using keeping the irrigation constant is given by

$$R_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} = \frac{0.82 - 0.92 \times 0.93}{\sqrt{1 - (0.92)^2}\sqrt{1 - (0.93)^2}} = \frac{-0.0356}{\sqrt{0.1536} \times \sqrt{0.1351}} = -$$

0.247.

Also the correlation between production of rice and irrigation keeping the amount of fertilizer used constant is given by

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{0.92 - 0.82 \times 0.92}{\sqrt{1 - 0.82^2}\sqrt{1 - 0.93^2}} = \frac{0.1574}{\sqrt{0.3276} \times \sqrt{0.1351}} = 0.748.$$

8. A sample of 10 values of three variables X_1, X_2 and X_3 were obtained as;

$X_1 = 10$	$\Sigma_2 = 20$	$\Sigma_3 = 30$
$\Sigma X_1^2 = 20$	$\Sigma X_2^2 = 68$	$X_3^2 = 170$
$\Sigma X_1 \cdot X_2 = 10$	$\Sigma X_1 X_3 = 15$	$\Sigma X_2 X^3 = 64$

Find:

a. Partial correlation between X_1 and X_3 elimination the effect of X_2.

b. Multiple correlation between X_1, X_2 and X_3 assuming X_1 as independence.

[TU 2070/2065]

Scanned by CamScanner

Solution 8

Here given

$n = 10$, $\Sigma x_1 = 10$ $\Sigma x_2 = 20$ $\Sigma x_3 = 30$

$\Sigma x_1^2 = 20$ Σx_2^2 $\Sigma x_3^2 = 170$

$\Sigma x_1x_2 = 10$ $\Sigma x_1x_3 = 15$ $\Sigma x_2x_3 = 64$

(a) The partial correlation between x_1 and x_3 eliminating the effect of x_2 is given by

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} \quad \text{..... (A)}$$

$$\text{Where, } r_{12} = \frac{n\Sigma x_1x_2 - \Sigma x_1 . \Sigma x_2}{\sqrt{n\Sigma x_1^2 - (\Sigma x_1)^2} . \sqrt{n. \Sigma x_2^2 - (\Sigma x_2)^2}}$$

$$= \frac{10 \times 10 - 10 \times 20}{\sqrt{10 \times 20 - 10^2}\sqrt{10 \times 68 - 20^2}} = \frac{-100}{\sqrt{100} \times \sqrt{280}} = -0.6$$

$$r_{13} = \frac{10 \times 15 - 10 \times 30}{\sqrt{10 \times 20 \times 10^2}\sqrt{10 \times 170 - 30^2}} = \frac{-150}{\sqrt{100} \times \sqrt{800}} = -0.53$$

$$\text{Also } r_{23} = \frac{10 \times 64 - 20 \times 30}{\sqrt{10 \times 68 - 20^2} . \sqrt{10 \times 170 - 30^2}} = \frac{40}{\sqrt{280}\sqrt{800}} = 0.085$$

Now equation (A) becomes

$$r_{13.2} = \frac{-0.53 - (-0.6) . 0.085}{\sqrt{1 - (-0.6)^2}\sqrt{1 - (0.085)^2}} = \frac{-0.479}{0.8 \times 0.995} = 0.602$$

(b) Multiple correlation between x_1, x_2 and x_3 assuming x_1 as independence is given by

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(-0.6)^2 + (-0.53)^2 - 2 \times (-0.6) \times (-0.53) \times 0.85}{1 - (0.085)^2}} = \sqrt{\frac{0.59}{0.993}} = 0.594$$

Exam Asked Questions

1. The following are the heights in centimeter and weights in kilogram of 10 men ;

Heights	162	168	174	176	180	180	182	184	186	186
Weights	65	65	84	63	75	76	82	65	80	81

i. Develop the estimating equation of weight and height iii) Estimate the weight of men whose height is 175 cm.

ii. Find the coefficient of determination and interpret it. Also find the standard error of the estimate. **[P.U 2011 Fall, 2017 Fall]**

Solution 1

Here, we have to estimate the weight (Y) when height (X) is given so we are going to find the regression of Y and X.

Calculation of Regression equation of Y on X

Heights (X)	Weights (Y)	u = X – 174	v = Y – 75	u^2	v^2	uv
162	65	–12	–10	144	100	120

Scanned by CamScanner

168	65	−6	−10	36	100	60
174	84	0	9	0	81	0
176	63	2	−12	4	144	−24
180	75	6	0	36	0	0
180	76	6	1	36	1	6
182	82	8	7	64	49	56
184	65	10	−10	100	100	−100
186	80	12	5	144	25	60
186	81	12	6	144	36	72
n = 10		Σu = 38	Σv = −14	Σu² = 708	Σv² = 636	Σuv = 250

We know that,

$$\bar{x} = A + \frac{\Sigma u}{n} = 174 + \frac{38}{10} = 177.8$$

$$\bar{y} = B + \frac{\Sigma v}{n} = 75 + \frac{(-14)}{10} = 73.6$$

$$bxy = \frac{n \times \Sigma uv - (\Sigma u) \times \Sigma v}{n\Sigma v^2 - (\Sigma v)^2} = \frac{10 \times 250 - 38 \times (-14)}{10 \times 636 - (-14)^2} = \frac{3032}{6164} = 0.49$$

$$byx = \frac{n \times \Sigma uv - (\Sigma u) \times \Sigma v}{n\Sigma u^2 - (\Sigma u)^2} = \frac{3032}{10 \times 708 - (38)^2} = \frac{3032}{5636} = 0.54$$

i. The estimating equation of height and weight is given by

$y - \bar{y} = byx(x - \bar{x})$

or, $y - 73.6 = 0.54(x - 177.8)$

or, $y = 0.54x - 22.412$

ii. The weight of men whose height is 175 cm is

$y = 0.54 \times 175 - 22.412 = 72.088 \approx 72$ kg

iii. The coefficient of determination is given by

$r^2 = b_{yx} \times b_{xy} = 0.54 \times 0.49 = 0.2646$

There is positive correlation between height and weight of men. Also there is 27% variation between heights and weights of men.

It means that 26.46% change in weights occurs due to the change in heights.

iv. Also the standard error of the estimate is given by,

$$S_e = \sqrt{\frac{\Sigma y^2 - a\Sigma xy - b\Sigma xy}{n-2}}$$

$$= \sqrt{\frac{636 - (-22.412) \times (-14) - 0.54 \times 250}{10 - 2}} = \sqrt{\frac{187.232}{8}} = 4.84$$

2. The following table gives the ages and blood pressures of 10 women:

Age X:	56	42	36	47	49	42	60	72	63	55
B.P. Y:	147	125	118	128	145	140	155	160	149	150

Then find

i) the correlation coefficient between age and blood pressure. Also find the coefficient of determination and interpret it.

ii) fit a regression equation that best describe the above data.

[P.U. 2008 Fall / 2009 Spring/2013 Spring/ 2015 Fall/ 2017 Spring]

Scanned by CamScanner

Solution 2

Calculation of correlation and regression coefficient

Age (X)	B.P. (Y)	u = X – 50	v = Y – 136	u^2	v^2	uv
56	147	6	11	36	121	66
42	125	–8	–11	64	121	88
36	118	–14	–18	196	324	252
47	128	–3	–8	9	64	24
49	145	–1	9	1	81	–9
42	140	–8	4	64	16	–32
60	155	10	19	100	361	190
72	160	22	24	484	576	528
63	149	13	13	169	169	169
55	150	5	14	25	196	70
n = 10		Σu = 22	Σv = 57	$\Sigma u^2 = 1148$	$\Sigma v^2 = 2029$	Σuv = 1346

i. We know that the correlation coefficient between age and blood pressure is given by

$$r = \frac{n \cdot \Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n \cdot \Sigma u^2 - (\Sigma u)^2}\sqrt{n \cdot \Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{10 \times 1346 - 22 \times 57}{\sqrt{10 \times 1148 - 22^2}\sqrt{10 \times 2029 - 57^2}} = \frac{12206}{\sqrt{10996}\sqrt{17041}} = 0.89$$

Now, the coefficient of determination = $r^2 = 0.7921$.

Hence, there is high degree positive correlation between age and blood pressure. Also, there is 79% of the total variation in the B.P. with respect to age.

It means that, 79.21% change in blood pressure occurs due to the change of age.

ii. Let us suppose blood pressure depends upon the age then the best fitted regression equation is y on x.

$$\bar{x} = A + \frac{\Sigma u}{n} = 50 + \frac{22}{10} = 52.2$$

$$\bar{y} = B + \frac{\Sigma v}{n} = 136 + \frac{57}{10} = 141.7$$

$$b_{yx} = \frac{n \cdot \Sigma uv - \Sigma u \cdot \Sigma v}{n \cdot \Sigma u^2 - (\Sigma u)^2} = \frac{10 \times 1346 - 22 \times 57}{10 \times 1148 - 22^2} = \frac{12206}{10996} = 1.11$$

Now, the regression equation of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

or, $y - 141.7 = 1.11(x - 52.2)$

$\Rightarrow$ $y = 1.11x + 83.758$

3. The following table gives the ages and blood pressures of 10 women:

Age X:	56	42	36	47	49	42	60	72	63	55
B.P. Y:	147	125	118	128	145	140	155	160	149	150

i. Estimate the blood pressure of a women whose age is 45 years.

ii. How much variation in blood pressure of a women is explained by the women's age? Based on this value, determine the reliability of the fitted regression model. [P.U. 2008 Fall, OR 2011 Spring]

Scanned by CamScanner

Solution 3

Please refer Q.No. 22 exercise

4. In the United States, during the fall harvest season pumpkin are sold in the large quantities at the farm stands. Often instead of weighting the pumpkin prior to sale, the farm stand operators will just place the pumpkins in the appropriate circular cutout on the counter. When asked why this was done one farmer replied," I can tell the weight of the pumpkin from its circumferences." To determine whether this was really true a sample of 8 pumpkin's circumference and weights were measured:

Circumferences(cm)	50	55	54	52	37	52	53	47
Weights(00 gram)	12	20	15	17	50	10	15	14

i. Develop the estimating linear regression equation to predict the weight of the pumpkin that is 60 cm in circumference.

ii. Calculate correlation coefficient between circumference and weight of the pumpkin and interpret it.**[P.U 2016 Fall]**

Solution 4

We have to estimate the weight of pumpkin (x) when circumference (y) is given so the regression equation of y on x is calculated as follows:

x (in cm)	y (in 00 gm)	u = x – 46	v = y – 15	u^2	v^2	uv
50	12	4	–3	16	9	–12
55	20	9	5	81	25	45
54	15	8	0	64	0	0
52	17	6	2	36	4	12
37	50	–9	35	81	1225	–315
52	10	6	–5	36	25	–30
53	15	7	0	49	0	0
47	14	1	–1	1	1	–1
n = 8		$\Sigma u = 32$	$\Sigma v = 33$	$\Sigma u^2 = 364$	$\Sigma v^2 = 1289$	$\Sigma uv = -301$

Now,

$$\bar{x} = A + \frac{\Sigma u}{n} = 46 + \frac{32}{8} = 50$$

$$\bar{y} = B + \frac{\Sigma v}{n} = 15 + \frac{33}{8} = 19.125$$

$$b_{xy} = \frac{n \cdot \Sigma uv - \Sigma u \cdot \Sigma v}{n \cdot \Sigma u^2 - (\Sigma u)^2} = \frac{8 \times (-301) - 32 \times 33}{8 \times 1289 - 33^2} = \frac{-3464}{9223} = -0.38$$

$$b_{yx} = \frac{n \cdot \Sigma uv - \Sigma u \cdot \Sigma v}{n \cdot \Sigma u^2 - (\Sigma u)^2} = -\frac{3464}{1888} = -1.83$$

i. The regression equation off y on x is given by

$y - \bar{y} - b_{yx}(x - \bar{x})$

or, $y - 19.125 = -1.83(x - 50)$

or, $y = 110.625 - 1.83x$

When circumference x = 60 cm than weight of pumpkin y = 110.625 – 1.83 × 60 = 0.825 = 82.5 gm.

Scanned by CamScanner

ii. $r = \pm\sqrt{b_{xy} \times b_{yx}} = \pm\sqrt{(-0.38) \times (-1.83)} = -0.834$

Since, $r = -0.834 \approx -1$.

Hence, there is high degree of negative correlation between them.

5. An instructor is interested in finding out the relationship between the number of students absent in school (y) and the day temperature (x). A random sample of 10 students was used for the study. The following table gives the record of the study.

Temperature (x)	10	20	25	30	40	45	50	55	59	60
No. of absent students (y)	8	7	5	4	2	3	5	6	8	9

Find the regression equation to estimate the number of students absent in the school when the temperature is 65. Also find the standard error of the estimate.

[P.U 2014 spring]

Solution 5

Here, we have to estimate the number of absent students (y). So, the regression equation of y on x can be calculated as follows:

Temperature X (in °C)	No. of absent students (y)	u = X – 35	v = y	u^2	v^2	uv
10	8	–25	8	625	64	–200
20	7	–15	7	225	49	–75
25	5	–10	5	100	25	–50
30	4	–5	4	25	16	–20
40	2	5	2	25	4	10
45	3	10	3	100	9	30
50	5	15	5	225	25	75
55	6	20	6	400	36	120
59	8	24	8	576	64	192
60	9	25	9	625	81	225
n = 10		Σu = 44	Σv = 57	$\Sigma u^2 = 2926$	$\Sigma v^2 = 373$	Σuv = 307

i. The regression equation of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x}) \qquad \ldots(i)$$

Where, $\bar{x} = A + \frac{\Sigma u}{n} = 35 + \frac{44}{10} = 39.4$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{57}{10} = 5.7$$

$$b_{yx} = \frac{n \cdot \Sigma uv - \Sigma u \cdot \Sigma v}{n \cdot \Sigma u^2 - (\Sigma u)^2} = \frac{10 \times 307 - 44 \times 57}{10 \times 2926 - 44^2} = \frac{562}{27324} = 0.021$$

Now, (i) reduces to,

$y - 5.7 = 0.021(x - 39.4)$

$\Rightarrow$ $y = 0.021x + 4.8$

The absent number of students when temperature is 65

$\hat{y} = 0.021 \times 65 + 4.87 = 6.235 \approx 6$ students

ii. The standard error or the estimate

Scanned by CamScanner

$$S_e = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{n-2}}$$

$$= \sqrt{\frac{373 - 4.87 \times 57 - 0.021 \times 307}{10-2}} = \sqrt{\frac{89.963}{8}} = 3.34$$

6. The following table gives age (x) in years of a cars and annual maintenance cost (y) in hundred rupees. Estimate the maintenance cost for a 4 years old car after finding the regression equation. Also find the coefficient of determination, standard error and interpret it. [P.U 2013 Fall]

X	1	3	5	7	9
Y	15	18	21	23	22

Solution 6

Please refer solution 5

7. The city council of Pokhara has gathered data on number of minor traffic accidents and the number of youth football games that occurred in town over the weekend.

X (football Games)	20	30	10	12	15	25	34
Y (minor Accidents)	6	9	4	5	7	8	9

i. Develop the estimating linear equation o predict minor accident from football games.

ii. Predict the number of minor traffic accidents that will occur at weekends during which 32 games will takes place in Pokhara. [P.U 2012 Fall]

Solution 7

Please refer solution 5

8. Calculate the coefficient of correlation from the following marks obtained by 10 students in mathematics (x) and Statistics (y). [P.U.2005 Fall]

X	40	65	60	25	85	35	45	70	80	55
Y	30	85	65	35	90	35	55	75	75	45

Solution 8

Please refer solution 5

9. Calculate the coefficient of correlation between X and Y for the following: [P.U 2009 Fall]

X:	1	3	4	5	7	8	9	11	12
Y:	2	6	8	10	14	16	20	24	29

Interpret your answer. Also find a regression equation of Y on X.

Solution 9

Please refer solution 5

10. Obtain the regression of Y on X from the following:

Age in Years (X):	56	42	72	36	63	47	55
Blood Pressure (Y):	147	125	160	118	149	128	150

Scanned by CamScanner

Estimate blood pressure when Age is 75.

P.U. 2007 Fall\2009 Fall\2010 Spring/2011Fall /2014 Fall/TU 2065

Solution 10

Please refer solution 5

11. From the following table showing age of cars of a certain make and annual maintenance costs, obtain the regression equation for cost related to age.

Age of cars (In Yrs)	2	4	6	8	10	12	14
Annual maintenance cost: (Rs. 000)	2	5	6	7	10	11	15

Also calculate the maintenance cost for the car 20 years old.

[P.U 2003 Fall/TU 2062/TU 2068]

Solution 11

Please refer solution 5

12. Find the value of co-efficient of determination between x and y for the given data and interpret the result. Find also regression equation of Y on X. Test whether there is any significant relationship between X and Y [P.U. 2010 spring]

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

Solution 12

Calculation of correlation and regression coefficient

x	y	x^2	y^2	xy
1	10	1	100	10
2	12	4	144	24
3	16	9	256	48
4	28	16	784	112
5	25	25	625	125
6	36	36	1296	216
7	41	49	1681	287
8	49	64	2401	392
9	40	81	1600	360
10	50	100	2500	500
$\Sigma x = 55$	$\Sigma y = 307$	$\Sigma x^2 = 385$	$\Sigma y^2 = 11387$	$\Sigma xy = 2074$

We know that,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7$$

$$b_{yx} = \frac{n \cdot \Sigma xy - \Sigma X \cdot \Sigma y}{n \cdot \Sigma x^2 - (\Sigma x)^2} = \frac{20740 - 55 \times 307}{3850 - 55^2} = \frac{3855}{825} = 4.67$$

Similarly, $b_{xy} = 0.196$

$$\therefore \quad r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{4.67 \times 0.196} = 0.96$$

Now, the coefficient of determination is given by

$$r^2 = (0.96)^2 = 0.9216$$

Scanned by CamScanner

Hence, the change in y variable is 92% with respect to x The variable x and y has high degree of positive relation.

It means that 92.16% change in y occurs due to the change in x variable.

Also, the regression equation of y on x is given by

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

or, $y - 30.7 = 4.67(x - 5.5)$

or, $y = 4.67x + 5.015$

Also, for the significance of the correlation coefficient

$$\text{P.E. } (r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}} = 0.6745 \times \frac{1 - (0.96)^2}{\sqrt{10}} = 0.017$$

Also, $\sigma \times \text{P.E. } (r) = 6 \times 0.017 = 0.1 < 0.96 = r$

Since $r > 6$ P.E (r) so r is significant.

13. The following data gives the experience of machine operators in 7 years and their performance as given by the number of good parts turned out per 100 pieces. [P.U 2011 Spring]

Operator;	1	2	3	4	5	6	7	8
Experience;	16	12	18	4	3	10	5	12
Performance;	88	89	68	78	80	75	83	87

i. Estimate the performance if an operator has 8 years of experience.

ii. How much variation in performance is explained by the machine's experience? Based on this value, determine the reliability of the fitted regression equation?

Solution 13

Let x denote the experience of machine operators in years and y denote their performance in number of good parts turned out per 100 pieces.

Calculation of regression equation of y on x

x	y	u = x − 10	v = y − 78	u^2	v^2	uv
16	88	6	10	36	100	60
12	89	2	11	4	121	22
18	68	8	−10	64	100	−80
4	78	−6	0	36	0	0
3	80	−7	2	49	4	−14
10	75	0	−3	0	9	0
5	83	−5	5	25	25	−25
12	87	2	9	4	81	18
n = 8		Σu = 0	Σv = 24	$\Sigma u^2 = 218$	$\Sigma v^2 = 440$	Σuv = −19

We know that

$$\bar{x} = A + \frac{\Sigma u}{n} = 10 + \frac{0}{8} = 10$$

$$\bar{y} = B + \frac{\Sigma v}{n} = 78 + \frac{24}{8} = 81$$

$$b_{xy} = \frac{n \cdot \Sigma uv - (\Sigma u)(\Sigma v)}{n \cdot \Sigma v^2 - (\Sigma v)^2} = \frac{8 \times (-19) - 0 \times 24}{8 \times 440 - 24^2} = \frac{-152}{2944} = -0.052$$

Scanned by CamScanner

$$b_{yx} = \frac{n \cdot \Sigma uv - (\Sigma u) \cdot (\Sigma v)}{n \cdot \Sigma u^2 - (\Sigma u)^2} = \frac{8 \times (-19) - 0 \times 24}{8 \times 218 - 0^2} = -\frac{152}{1744} = -0.087$$

i. Now, the regression equation of y on x is given by

$y - \bar{y} = b_{yx}(x - \bar{x})$

or, $y - 81 = -0.087(x - 10)$

or, $y = 81.87 - 0.087x$

The estimated performance of an operator has 8 year experience is

$\hat{y} = 81.87 - 0.087 \times 8 = 81.174$ (per 100 good parts)

ii. Now, the coefficient of determination is given by

$r^2 = b_{xy} \times b_{yx} = (-0.052) \times (-0.087) = 0.004524$

It means 0.45% of variation in performance is explained by the operator's experience. Also,

$$P.E.(r) = 0.6745 \times \frac{1 - r^2}{\sqrt{n}} = 0.6745 \times \frac{1 - 0.004524}{\sqrt{8}} = 0.237$$

$6 \times P.E.(r) = 6 \times 0.237 = 1.424 > r$

So, r is not significant.

14. Shears tress (y) and rate of shear (x) can be measured for a liquid in a viscometer. For 12 pairs of values the data can be summarized as $\Sigma x = 132$. $\Sigma y = 151.7$, $\Sigma x^2 = 1944$, $\Sigma y^2 = 2570.40$, $\Sigma xy = 2233.2$

i. Find the linear regression equation of y on x

ii. What is the correlation coefficient?

iii. What fraction of the variance of y is explained by the correlation?

[P.U 2016 Spring]

Solution 14

Here, we have given

$n = 12$, $\Sigma x = 132$, $\Sigma y = 151$, $\Sigma x^2 = 1944$,

$\Sigma y^2 = 2570.48$, $\Sigma xy = 2233.2$

Then, $\bar{x} = \frac{\Sigma x}{n} = \frac{132}{12} = 11$

$\bar{y} = \frac{\Sigma y}{n} = \frac{151}{12} = 12.58$

$$b_{xy} = \frac{n \cdot \Sigma xy - (\Sigma x) \cdot (\Sigma y)}{n \cdot \Sigma y^2 - (\Sigma y)^2} = \frac{12 \times 2233.2 - 132 \times 151}{12 \times 2570.48 - (151)^2} = \frac{6866.4}{8044.76} = 0.85$$

$$b_{yx} = \frac{n \cdot \Sigma xy - (\Sigma x) \cdot (\Sigma y)}{n \cdot \Sigma x^2 - (\Sigma x)^2} = \frac{12 \times 2233.2 - 132 \times 151}{12 \times 1944 - (132)^2} = \frac{6866.4}{5904} = 1.16$$

i. Now, by the regression equation of y on x is given by

$y - \bar{y} = b_{yx}(x - \bar{x})$

or, $y - 12.58 = 1.16(x - 11)$

$\Rightarrow$ $y = 1.16x - 0.18$

ii. The correlation coefficient between x and y is given by

Scanned by CamScanner

$r = \pm\sqrt{0.85 \times 1.16} = 0.99$

iii. The coefficient of determination $(r^2) = (0.99)^2 = 0.98$.

It means 98% of variation of y variable occurs due to the variation in x variable. Also 98% fraction off the variance of y is explained by the correlation.

15. A study reported in the article *"the effect of water vapour concentrations on the rate of combustion of an artificial in humid air flow"* give data on x = temperature of nitrogen oxygen mixture (1000 degree F) under the specified conditions; y = oxygen diffusivity. Summary quantities are

$n = 6$. $\Sigma x = 12.6$, $\Sigma y = 27.68$, $\Sigma x^2 = 18.24$, $\Sigma y^2 = 93.344$, $\Sigma xy = 40.968$ then:

i. Calculate correlation coefficient between x and y and interpret your answer.

ii. Fit a regression equation of y on x.

iii. Find the coefficient of determination and interpret it. [P.U. 2015 Spring]

Solution 15

Here, we have given

$n = 9$, $\Sigma x = 12.6$, $\Sigma y = 27.68$

$\Sigma x^2 = 18.24$ $\Sigma y^2 = 93.344$ $\Sigma xy = 40.968$

i. The correlation coefficient between x and y is given by

$$r = \frac{n \cdot \Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n \cdot \Sigma x^2 - (\Sigma x)^2}\sqrt{n \cdot \Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{9 \times 40.968 - 12.6 \times 27.68}{\sqrt{9 \times 18.24 - (12.6)^2}\sqrt{9 \times 93.344 - (27)^2}} = \frac{19.944}{\sqrt{5.4} \times \sqrt{73.9136}} = 0.998$$

ii. We know that,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{12.6}{9} = 1.4; \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{27.68}{9} = 3.08$$

$$b_{yx} = \frac{n \cdot \Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{9 \times 40.968 - 12.6 \times 27.68}{9 \times 18.24 - (12.6)^2} = \frac{19.944}{5.4} = 3.69$$

Now, the regression equation of y on x is given by

$y - \bar{y} == b_{yx}(x - \bar{x})$

or, $y - 3.08 = 3.69\,(x - 1.4)$

or, $y = 3.69x - 2.086$

iii. The coefficient of determination $= r^2 = (0.998)^2 = 0.996$

It means 99.6% change in oxygen diffusivity (y) occurs due to the change in temperature of nitrogen, oxygen mixture. Also, there is high degree of positive correlation between x and y.

16. A study reported in the article" The Effects of Water Vapor Concentrations on the Rate of Combustion of an Artificial in Humid Air Flow"(Combustion and Flame) give data on x = temperature of nitrogen oxygen mixture(1000 degree F) under the specified conditions and y = oxygen diffusivity. Summary quantities are $n = 9$, $\sum x = 12.6$, $\sum y = 27.68$, $\sum x^2 = 18.24$, $\sum y^2 = 93.344$, $\sum xy = 40.968$.Calculate correlation coefficient between x and y, and interpret you answer and also fit a regression of y on x. [P.U 2010 Fall]

Scanned by CamScanner

Solution 16

Please refer solution 15

17. Calculate the rank correlation coefficient between x and y and also find the value of coefficient of determination. Interpret the results. [P.U 2012 Spring]

X	12	13	20	22	16	17	28	22	19	15	27	28	12
Y	112	134	123	126	122	134	145	140	135	122	122	135	123

Solution 17

Calculation of rank correlation coefficient x and y

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
12	112	12.5	13	−0.5	0.25
13	134	11	5.5	5.5	30.25
20	123	6	8.5	−2.5	6.25
22	126	4.5	7	−2.5	6.25
16	122	9	11	−2	4
17	134	8	5.5	2.5	6.25
28	145	1.5	1	0.5	0.25
22	140	4.5	2	2.5	6.25
19	135	7	3.5	3.5	12.25
15	122	10	3.5	−1	1
27	122	3	3.5	7	49
28	135	1.5	3.5	−2	4
12	123	12.5	8.5	4	16
					$\Sigma d^2 = 142$

We know that, the rank correlation of repeated ranks is given by

$$R = \frac{1 - 6\left[\Sigma d^2 + \frac{m_1(m_1^2 - 1)}{12} + \frac{m_2(m_2^2 - 1)}{12} + \frac{m_3(m_3^2 - 1)}{12}\right]}{n^3 - n}$$

$$= \frac{1 - 6\left[142 + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{2 \times 3}{12} + \frac{3 \times 18}{12}\right]}{13^3 - 13}$$

$$= \frac{1 - 6\,[142 + 5]}{2184} = 0.596$$

Also, the coefficient of determination = $R^2 = 0.355$. It means that 35.5% variance between X and Y. Also, the direction of relationship is positive and degree of relationship is 10.5961.

18. Calculate the rank correlation coefficient for the following height in inches of fathers(X) and their respective sons(Y). [P.U. 2006 Spring]

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Solution 18

Please refer solution 17

Scanned by CamScanner

19. Calculate the Spearman's rank correlation coefficients if the marks obtained by 10 students ,marks in Mathematics (x) and marks in Statistics (y) is given as follows: PU 2001/2002

X	78	36	98	55	75	82	90	62	65	39
Y	84	51	91	60	68	62	86	58	53	47

Solution 19

Solution of rank correlation between x and y

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
60	75	3	1	2	4
34	32	11	10	1	1
40	34	10	8	2	4
50	40	4	6	−2	4
45	45	6	4	2	4
41	33	9	9	0	0
22	12	12	12	0	0
43	30	7	11	−3	9
42	36	8	7	1	1
66	72	1	2	−1	1
64	41	2	5	−3	9
46	57	5	3	2	4
n = 12					$\Sigma d^2 = 48$

Now,

The rank correlation between x and y is

$$R = 1 - \frac{6\Sigma d^2}{n^3 - n} = 1 - \frac{6 \times 48}{12^3 - 12} = 1 - \frac{288}{1716} = 0.832$$

Also, the coefficient of determination = $R^2 = 0.69$.

Hence, there is direction of relationship is positive and degree of relationship is 0.832.

20. Calculate the rank correlation coefficients from the following data:

X	60	34	40	50	45	41	22	43	42	66	64	46
Y	75	32	34	40	45	33	12	20	36	72	41	57

Also find the value of coefficient of determination. Interpret its result. PU 2012

Solution 20

Please refer solution 19

21. Marks obtained by the 10 students in Mathematics (x) and Statistics (Y) are given below.

x	22	28	31	23	29	31	27	22	31	18
y	38	25	25	37	31	35	31	29	28	20

Obtain the rank correlation coefficient to determine the relationship between x and y. [PU 2002 D]

Solution 21

Please refer solution 17

Scanned by CamScanner

22. The following data gives the experience of a machine operators in years and their performance as given by the number of good parts turned out per 100 pieces. Fit the regression equation of performance rating on experience and estimate the probable performance if the operator has 8 years of experience. TU 2072 A /2070 B

Experience:	16	12	18	4	3	10	5	12
Performance:	87	88	89	68	78	80	75	83

Solution 22

Please refer solution 13

23. Calculate Karl Person's correlation coefficient for the following data and interpret the result. [Purbanchal 2010]

x	10	12	15	20	28
y	50	70	86	100	124

Solution 23

Calculation of Karl Pearson's Coefficient of Correlation

x	y	x^2	y^2	xy
10	50	100	2500	500
12	70	144	4900	840
15	86	225	7396	1290
20	100	400	10000	2000
28	124	784	15376	3472
$\Sigma x = 85$	$\Sigma y = 430$	$\Sigma x^2 = 1653$	$\Sigma y^2 = 40172$	$\Sigma xy = 8102$

Karl Pearson's coefficient of correlation coefficient is given bty

$$r = \frac{n\,\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\,\Sigma x^2 - (\Sigma x)^2}\sqrt{n\,\Sigma y^2 - (\Sigma y^2)}}$$

$$= \frac{5 \times 8102 - 85 \times 430}{\sqrt{5 \times 453 - 85^2}\sqrt{5 \times 40172 - (430)^2}}$$

$$= \frac{3960}{\sqrt{1040}\sqrt{15960}}$$

$$= 0.972$$

Since, $r = 0.972 \approx 1$. So, there is high degree of positive correlation between x and y.

24. The following are the number of minute it took 10 mechanics to assemble a piece of machinery in the morning (x) and in the late afternoon (y):

x	11.1	10.3	12.0	15.1	13.7	18.5	17.3	14.2	14.8	15.3
y	10.9	14.2	13.8	21.5	13.2	21.1	16.4	19.3	17.4	19.0

i. Calculate the Karl Pearson's correlation coefficient.

ii. Find the proportion of variation in y, the afternoon assembly times that can be explained by a straight line fit to x, the morning assembly time. TU 2067

Scanned by CamScanner

Solution 24

Calculation of Karl Pearson's Coefficient of Correlation

x	y	x^2	y^2	xy
11.1	10.9	123.21	118.81	120.99
10.3	14.2	106.09	201.64	146.26
12.0	13.8	144	190.44	165.6
15.1	21.5	228.01	462.25	324.65
13.7	13.2	187.69	174.24	180.84
18.5	21.1	342.25	445.21	390.35
17.3	16.4	299.29	268.96	283.72
14.2	19.3	201.64	372.49	274.06
14.8	17.4	219.64	302.76	257.52
15.3	19.0	234.09	361	290.7
$\Sigma x = 142.3$	$\Sigma y = 166.8$	$\Sigma x^2 = 2085.31$	$\Sigma y^2 = 2897.80$	$\Sigma xy = 2434.69$

i. We know that, Karl Pearson's coefficient of correlation coefficient is given by,

$$r = \frac{n\,\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\,\Sigma x^2 - (\Sigma x)^2}\sqrt{n\,\Sigma y^2 - (\Sigma y^2)}}$$

$$= \frac{10 \times 2434.69 - 142.3 \times 166.8}{\sqrt{10 \times 2085.31 - (142.3)^2}\sqrt{10 \times 2897.80 - (166.8)^2}}$$

$$= \frac{611.26}{\sqrt{603.81}\sqrt{1155.76}}$$

$$= 0.732$$

We see that there is high degree of positive correlation between them.

ii. The proportion of variation in y attributed to x is $r^2 = (0.732)^2 = 0.5358$. It means 53.58% of the variation among the afternoon times is explained by the corresponding differences among the morning times.

25. Find the Karl Pearson's coefficient of correlation between the age and the playing habit of the people from the following data. Also mention what your calculated r indicates. TU 2062

Age Group	15-20	20-25	25-30	30-35	35-40	40-45
No. of people	200	270	340	360	400	300
No. of players	150	162	170	180	180	120

Solution 25

Calculation of Karl Pearson's Coefficient of Correlation

Age-group	Mid-value (x)	% of players (y)	$u = \frac{x - 27.5}{5}$	$v = \frac{y - 55}{5}$	u^2	v^2	uv
15-20	17.5	150/200 × 100 = 75	−2	4	4	16	−8
20-25	22.5	162/270 × 100 = 60	−1	1	1	1	−1
25-30	27.5	170/340 × 100 = 50	0	−1	0	1	0
30-35	32.5	180/360 × 100 = 50	1	−1	1	1	−1
35-40	37.5	180/400 × 100 = 45	2	−2	4	4	−4
40-50	42.5	120/300 × 100 = 40	3	−3	9	9	−9
n = 6	$\Sigma y = 166.8$		$\Sigma u = 3$	Σv −2	$\Sigma u^2 = -19$	$\Sigma v^2 = 32$	$\Sigma uv = -23$

We know that, Karl Pearson's coefficient of correlation coefficient is given by,

Scanned by CamScanner

$$r = \frac{n\,\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\,\Sigma u^2 - (\Sigma u)^2}\sqrt{n\,\Sigma v^2 - (\Sigma v^2)}}$$

$$= \frac{6 \times (-23) - 3 \times (-2)}{\sqrt{6 \times 19 - (3)^2}\sqrt{6 \times 32 - (2)^2}}$$

$$= \frac{-132}{\sqrt{105}\sqrt{188}}$$

$$= -0.94$$

Hence, the Karl Pearson's coefficient of correlation between the age and playing habit of the people is $r = -0.94$.

It means they have high degree of negative correlation. We can say that when age increases the playing habit of the people decreases.

26. The following table gives the age and percentage of the blindness in respective age interval:

Age Group	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
% of blindness	70	63	21	26	45	31	46	80

i. Find out if there any correlation between age and blindness and interpret?

ii. Find the coefficient of determination. **TU 2067**

Solution 26

Calculation of Karl Pearson's Coefficient of Correlation

Age-group	Mid-value (x)	% of players (y)	$u = \frac{x - 27.5}{5}$	$v = \frac{y - 55}{5}$	u^2	v^2	uv
0-10	5	70	−3	25	9	625	−75
10-20	15	63	−2	18	4	324	−36
20-30	25	21	−1	−24	1	576	24
30-40	35	26	0	−19	0	361	0
40-50	45	45	1	0	1	0	0
50-60	55	31	2	−14	4	196	−28
60-70	65	46	3	1	9	1	3
70-80	75	80	4	35	16	1225	140
n = 8			$\Sigma u = 4$	$\Sigma v = 22$	$\Sigma u^2 = 44$	$\Sigma v^2 = 3308$	$\Sigma uv = 28$

Karl Pearson's coefficient of correlation coefficient is given by

$$r = \frac{n\,\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{n\,\Sigma u^2 - (\Sigma u)^2}\sqrt{n\,\Sigma v^2 - (\Sigma v^2)}}$$

$$= \frac{8 \times 28 - 4 \times 22}{\sqrt{8 \times 44 - (4)^2}\sqrt{8 \times 3308 - (22)^2}} = \frac{136}{\sqrt{336}\sqrt{25980}} = 0.046$$

Since, $r = 0.046$ so there is very low degree of positive correlation between age and blindness.

ii. The coefficient of determination $= r^2 = (0.046)^2 = 0.002116$

Scanned by CamScanner

31. A simply supported beam carries a concentrated load P (kg) at its mid point. The following table gives maximum deflection y (cm) corresponding to various value of p.

p :	100	120	140	160	180	200
y:	0.45	0.55	0.60	0.70	0.80	0.85

Find a law of the form y = a + b p.

Solution 31

Book Ex. 15.1 Q.No. 4a

32. A sample of 12 fathers and their eldest sons gave the following data about their heights in inches:

Father :	65	63	67	64	68	62	70	66	68	67	69	71
sons:	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the co-efficiency of rank correlation. [T. U. 2068]

Solution 32

Book Ex. 15.1 Q.No. 14

33. For 10 observations on price (x) and supply (Y), the following data were obtained (in appropriate units):

$\Sigma x = 130$, $\Sigma y = 220$, $\Sigma x^2 = 2288$, $\Sigma y^2 = 5506$ and $\Sigma xy = 3467$

i. Obtain the two lines of regression and estimate the supply when the price is 16 units.

ii. Find correlation coefficient and interpret the result in practical manner.

iii. Determine coefficient of determination and interpret the result.

Solution 33

Book Ex. 15.1 Q.No. 31

34. A household survey on monthly expenditure on food yield following data:

Monthly expenditure (Rs.'000')	10	15	20	25	30	35	40
Monthly income (Rs.'000')	2	4	5	7	6	6	5
Size of family (number)	4	5	7	10	8	11	4

a. Estimate the line of best fit.

b. Obtain the multiple correlation coefficient and coefficient of determination.

[TU 2072/2067/2063/2059/2057]

Solution 34

Book Ex. 15.2 Q.No. 4

6. Past experience shows that following result of productivity per hectare with the respective uses of chemical fertilizers and seeds. Calculate the coefficient of multiple correlation and coefficient of multiple determination. Also interpret the result. [TU 2065/2073/2068]

Scanned by CamScanner

year	1976	1977	1978	1979	1980	1981
Fertilizer (kgs): (X1)	45	30	70	75	65	80
Seeds (kgs): (X2)	2	1.8	3	2.5	2	3
Productivity (X3)	2000	2100	1800	1900	2400	2500

Solution 35

Book Ex. 15.2 Q.No. 6

8. A sample of 10 values of three variables X_1, X_2 and X_3 were obtained as;

$X_1 = 10$ $\quad$ $\Sigma_2 = 20$ $\quad$ $\Sigma_3 = 30$

$\Sigma X_1^2 = 20$ $\quad$ $\Sigma X_2^2 = 68$ $\quad$ $X_3^2 = 170$

$\Sigma X_1 \cdot X_2 = 10$ $\quad$ $\Sigma X_1 X_3 = 15$ $\quad$ $\Sigma X_2 X^3 = 64$

Find:

a. Partial correlation between X_1 and X_3 elimination the effect of X_2.

b. Multiple correlation between X_1, X_2 and X_3 assuming X_1 as independence. [TU 2070/2065]

Solution 36

Book Ex. 15.2 Q.No. 8

Short Questions

Rank Correlation **[PU 2003 Fall/2011 Fall]**

The qualitative characteristics such as beauty, intelligence, honesty, ability, knowledge, taste, etc. cannot be measured quantitatively but those types of variables can be measured by assigning ranks or some sort of ratings. The correlation which is obtained by assigning ranks is called the rank correlation. Also rank correlation is the degree of association that exists between two given variables when the data are arranged in order of ranks.

There are many methods to measure rank correlation. One of these methods is Spearman's Rank Correlation Coefficient.

The Spearman's rank correlation coefficient is denoted and defined by

$$R = 1 - \frac{6\,\Sigma d^2}{n^3 - n},$$

where $d = R_1 - R_2$

n = no. of pair of observation

When the ranks are repeated then Spearman's rank correlation coefficient is given by

$$R = 1 - \frac{6\left[\Sigma d^2 + \frac{m_1(m_1^2 - 1)}{12} + \frac{m_2(m_2^2 - 1)}{12} + \dots\right]}{n^3 - n}$$

Scanned by CamScanner

Where, m_1 = no. of times 1st value is repeated. m_2 = no. of times 2nd value is repeated and so on.

Coefficient of Determination

[PU 2003 Fall/2006 Spring/2010 Spring/2011 Spring/2013 Fall/2017 Fall]

If r is the correlation coefficient between the variables the square of correlation coefficient between the variables is called the coefficient of determination. In other words, the tool which measures the percentage variation or change in the dependent variable due to the change in the independent variable and denoted by r^2.

Also, we can say that coefficient of determination is the proportion of the variation in the dependent variable that is explained by the linear relationship between them.

Coefficient of determination can be calculated in the simple correlation, rank correlation, partial correlation, multiple correlation as well as the regression coefficients among the variables. In every types it gives the percentage variation or change of proportion of dependent variable due to the change in independent variable.

Additional Questions for Practice

Solution 27

Find the Karl Pearson's Coefficient of correlation between age and playing habit from the data. [TU 2062 Back]

Age	20	21	22	23	24	25
No. of Students	500	400	300	240	200	160
Regular players	400	300	180	96	60	24

Ans: r = –0.97

Solution 28

On 13 April 1994, the following concentrations of pollutants were recorded at eight stations of the monitoring system for air pollution control located in the downtown area of Milan, Italy. [TU 2067/2068]

Station	I	I	III	IV	V	VI	VII	VIII
NO_2 mg/m²	130	130	115	120	135	142	90	116
co_2 mg/m³	2.9	4.40	3.6	4.1	3.3	5.7	4.8	7.3

i. Compute the correlation coefficient between NO_2 and CO_2.

ii. Explain the relationship between NO_2 and CO_2.

iii. Determine the coefficient of determination between the pollutants and interpret the results.

Ans: (i) r = –0.1522; (ii) negative and low degree, (iii) r^2 = 0.2332(2.32% change)

Scanned by CamScanner

Solution 29

Observations of the yield of a chemical reaction taken at various temperatures were recorded as follows: [TU 2067/2068]

X (in °C)	150	150	200	250	250	300	150
Y (in %)	75.4	81.2	85.5	89	90.3	96.7	75.4

Fit a sample linear regression and estimate the value of yield ay 200°C.

Ans: y = 58.5806 + 0.1267x, 83.98

Solution 30

Suppose a statistic professor is interested in predicting final exam score (y) from SAT mathematics score (x) using the following data students. [TU 2062/2069]

Student	1	2	3	4	5	6	7	8	9
SAT Score (x)	440	465	282	521	535	552	572	590	607
Final score (y)	40	47	43	54	64	52	59	68	44

i. Determine the regression equation for predicting scores on the final (y) from SAT score (x).

ii. From a SAT score of 500, predict the score at final. [TU 2062/069]

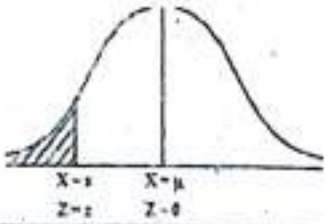
Ans: (I) y = 24.88 + 0.054 x, (II) $\hat{y}$ = 51.88

Scanned by CamScanner

Appendix

Table 3 : Standard Normal Curve Areas N_z

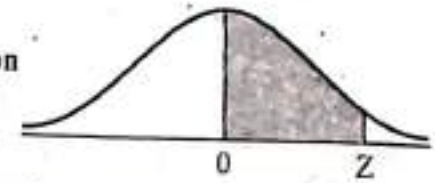
$\phi(z) = P(Z < z)$



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247

Scanned by CamScanner

Standard Normal Distribution



(The entries in this table give area under standard normal curve from 0 to z.)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4046	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49897	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

Level of significance (α)	1%	2%	4%	5%	10%
Two tailed test	± 2.57	± 2.33	± 2.05	± 1.96	± 1.64
Right tailed test	2.33	2.05	1.75	1.64	1.28

Scanned by CamScanner

Table – 2
Critical Value of t – Distribution

Degree of Freedom (df)	Level of significance for two tailed test.					
	20%	10%	5%	2%	1%	0.1%
	Level of significance for one (Left or Right) tailed test.					
	10%	5%	2.5%	1%	0.5%	0.05%
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Scanned by CamScanner

Degree of freedom for the denominator

d.f	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.14	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Scanned by CamScanner

Table – 4

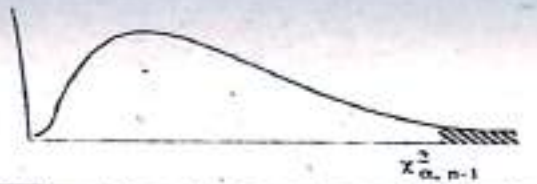
Critical Value of F – distribution at 1 % level of significance

Degree of freedom for the numerator

f	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Scanned by CamScanner

Table – 5 (Chi-square Table)



99%	97.5%	95%	90%	80%	20%	10%	5%	2.5%	1%
0.00016	0.00098	0.00398	0.0158	0.0642	1.642	2.706	3.841	5.024	6.635
0.0201	0.0506	0.103	0.211	0.446	3.219	4.605	5.991	7.378	9.201
0.115	0.216	0.352	0.584	1.005	4.642	6.251	7.815	9.348	11.345
0.297	0.484	0.711	1.064	1.649	5.989	7.779	9.488	1.143	13.277
0.554	0.831	1.145	1.610	2.343	7.289	9.236	11.070	12.833	15.086
0.872	1.237	1.625	2.204	3.070	8.558	10.645	12.592	14.449	16.812
1.239	1.690	2.167	2.833	3.822	9.803	12.017	14.067	16.018	18.475
1.646	2.180	2.733	3.490	4.594	11.030	13.362	15.507	17.535	20.090
2.088	2.700	3.325	4.168	5.380	12.242	14.684	16.919	19.023	21.666
2.558	3.247	3.940	4.865	6.179	13.442	15.987	18.307	20.483	23.209
3.053	3.816	4.575	5.578	6.989	14.631	17.275	19.675	21.920	24.725
3.571	4.404	5.226	6.304	7.807	15.812	18.549	21.026	23.337	26.217
4.107	5.009	5.892	7.042	8.634	16.985	19.812	22.362	24.736	27.688
4.660	5.629	6.571	7.790	9.467	18.151	21.064	23.685	26.119	29.141
5.229	6.262	7.261	8.547	10.307	19.311	22.307	24.996	27.488	30.578
5.812	6.908	7.962	9.312	11.152	20.465	23.542	26.296	28.845	32.000
6.408	7.564	8.672	10.085	12.002	21.615	24.769	27.587	30.191	33.409
7.015	8.231	9.390	10.865	12.857	22.760	25.989	28.869	31.526	34.805
7.633	8.907	10.117	11.651	13.716	23.900	27.204	30.144	32.852	36.191
8.260	9.591	10.851	12.443	14.578	25.038	28.412	31.410	34.170	37.566
8.897	10.283	11.591	13.240	15.445	26.171	29.615	32.671	35.479	38.932
9.542	10.982	12.338	14.041	16.314	27.301	30.813	33.924	36.781	40.289
10.196	11.689	13.091	14.848	17.187	28.429	32.007	35.172	38.076	41.638
10.856	12.401	13.848	15.658	18.062	29.553	33.196	36.415	39.364	42.980
11.524	13.120	14.611	16.473	18.940	30.675	34.382	37.652	40.647	44.314
12.198	13.844	15.379	17.292	19.820	31.795	35.563	38.885	41.923	45.642
12.879	14.573	16.151	18.114	20.703	32.912	36.741	40.113	43.194	46.963
13.565	15.308	16.928	18.939	21.588	34.027	37.916	41.337	44.461	48.278
14.256	16.047	17.708	19.768	22.475	35.139	39.087	42.557	45.722	49.588
14.953	16.791	18.493	20.599	23.364	36.250	40.256	43.773	46.979	50.892

Scanned by CamScanner

Table 7: Percentage points of the t-distribution

v	one tail α	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	two tail α	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.593
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.889
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1314	1.703	2.052	2.473	2.771	3.050	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Scanned by CamScanner