LECTURE SHEET ON APPLIED STATISTICS

PREPARED BY

MD. SHAH NAJMUS SHAKIB

LECTURER, DEPARTMENT OF MATHEMATICS

UNIVERSITY OF GLOBAL VILLAGE (UGV)

BASIC INFORMATION

- **Course Code: CSE 0541-3101**
- **Course Credit: 03**
- **CIE Marks: 90**
- **SEE marks: 60**
- **Semester End Exam (SEE): 3 hours**

COURSE LEARNING OUTCOMES

01 Student will acquire proficiency in fundamental concepts and apply it to solve different problems.

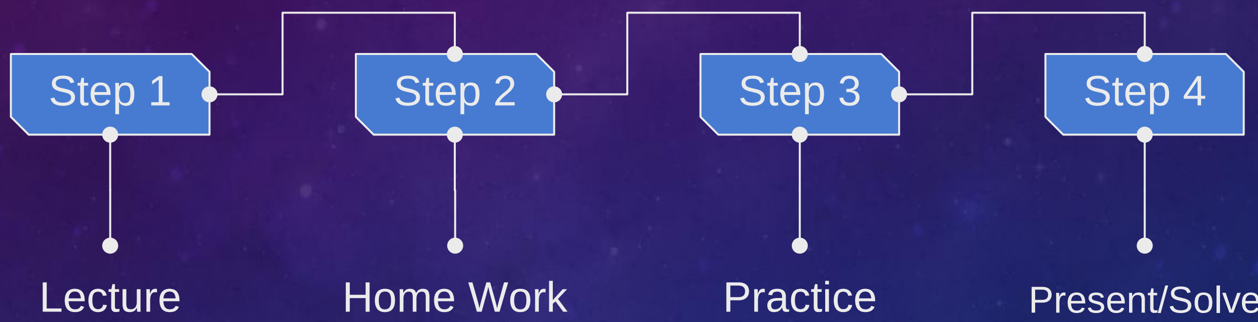
02 Develop advanced analytical and problem-solving skills essential for tackling complex computational challenges.

03 Demonstrate mastery several mathematical proofs and apply the method to solve problems.

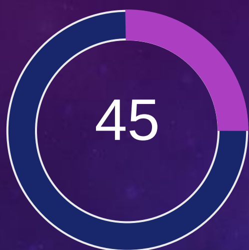
04 Apply theoretical knowledge to effectively solve practical problems.

SL.	Content of Courses	Hrs	CLO's
1	Introduction to Statistics: Definition and scope. Mean, Median, Mode, Frequency Distribution (mean, median, mode), Geometric and Harmonic Mean.	10	CLO1, CLO2
2	Standard Deviation, Variance, Mean Deviation (concept and calculation), Pie Charts, Histograms, Bar Diagrams, Frequency Polygons, and Ogives.	8	CLO1, CLO3
3	Probability: Introduction, rules, conditional probability, and practical	10	CLO2, CLO4

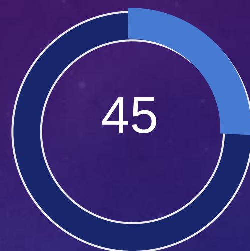
STEP-BY-STEP PROGRESSION



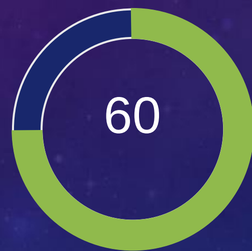
MARKS DISTRIBUTION



Assignment,
Quiz, Attendence



Mid Term



Final

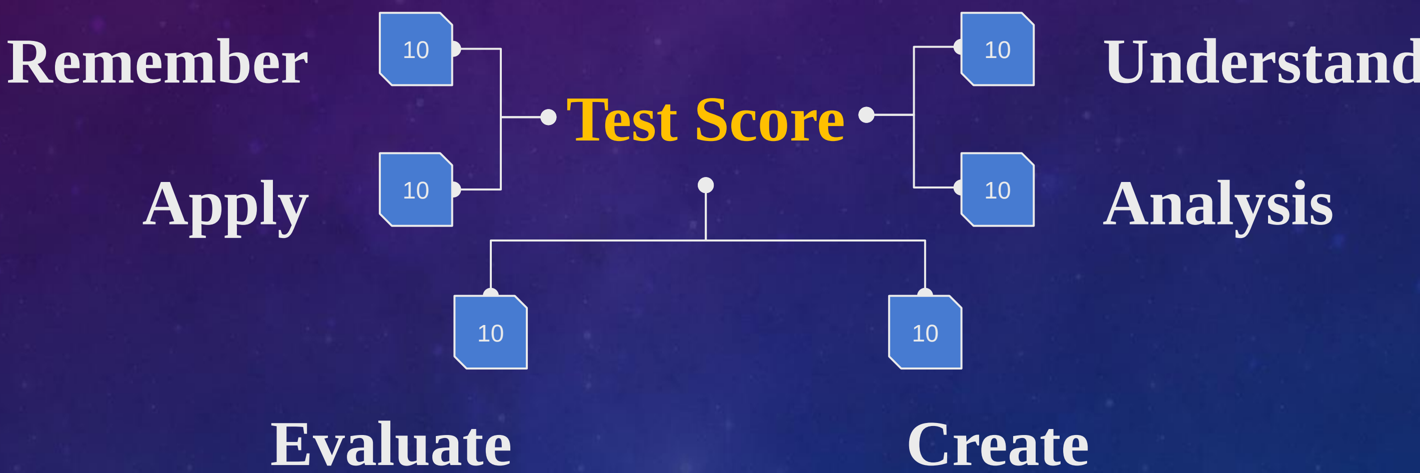
Assessment Pattern

Blooms Category	Test (Out of 45)	Assignments (15)	Quiz (15)	Co-curricular Activities (15)
Remember	05		5	Attendance 15
Understand	05			
Apply	10			
Analysis	8	7	10	
Evaluate	7	8		
Create	10			

Semester End Examination (SEE 60)

Blooms Category	Test (Out of 60)
Remember	10
Understand	10
Apply	10
Analysis	10
Evaluate	10
Create	10

ASSESSMENT PATTERN (SEE)



Assessment Pattern

Continuous Internal Evaluation (CIE 90 marks)

Blooms Category	Test (Out of 45)	Assignments (15)	Quiz (15)	Co-curricular Activities (15)
Remember	05		5	Attendance 15
Understand	05			
Apply	10			
Analysis	8	7	10	
Evaluate	7	8		
Create	10			

HOW TO GET BEST RESULT?

Stay consistent with a study routine and focus on understanding concepts instead of memorizing. Practice regularly and discuss with friends.

Attend Class

Group Study

Practice

Understan
d

Apply

WEEKLY PLAN

W S

Week 1							
Week 2							
Week 3							
... … ….							
Week 17							

Course Plan Specific Content, CLOs, Teaching Learning and Assessment Strategy mapped with CLOs.

Week	Task Heading	Topics	Teaching-Learning Strategy	Assessment Strategy	Corresponding CLO's
1	Introduction to Statistics	Definition and scope	Lecture, Discussion	Quiz	CLO1

2	Median and Mode	Concept and definition	Lecture, Practice Exercises	Written Assignment	CLO1
3	Median and Mode	Examples and Applications	Discussion, Problem Solving	Quiz	CLO1
4	Median and Mode	Examples	Lecture, Discussion	Group Assignment	CLO1, CLO2
5	Median and Mode	Examples and	Discussion,	Written Assignment	CLO1, CLO2

		Applications	Examples		
6	Standard Deviation	Formula and examples	Lecture, Practice Exercises	Quiz	CLO1, CLO2
7	Variance	Calculation and examples	Problem Solving	Quiz	CLO1, CLO2
8	Mean Deviation	Concept and calculation	Lecture, Problem Solving	Written Assignment	CLO1, CLO3

9	Pie Charts	Steps and examples	Lecture, Practice Exercises	Quiz	CLO1, CLO3
10	Histograms	Construction and uses	Lecture, Problem Solving	Written Assignment	CLO1, CLO3
11	Bar Diagrams	Steps and examples	Lecture, Problem Solving	Quiz	CLO1, CLO3
12	Frequency Polygons	How to draw and use	Lecture, Discussion	Written Assignment	CLO1, CLO3

13	Ogives	Construction and uses	Lecture, Examples	Quiz	CLO1, CLO3
14	Probability: Introduction	Definition, rules, and examples	Lecture, Discussion	Quiz	CLO2, CLO4
15	Conditional Probability	Concept and applications	Discussion, Problem Solving	Written Assignment	CLO2, CLO4
16	Practical Applications	Solving real-life problems	Problem Solving,	Oral Presentation	CLO2, CLO4

			Examples		
17	Distributions	Normal, Binomial, and Poisson Distributions	Lecture, Examples	Written Assignment, Quiz	CLO4

1^{st} Week

Topic: Introduction to Statistics

Page: 05- 06

Introduction to Statistics: A Visual Guide

Statistics is a powerful tool for understanding the world around us. It helps us make sense of data, uncover patterns, and draw meaningful conclusions. This presentation will introduce you to the fundamentals of statistics, its advantages, disadvantages, and effective ways to visualize data.

Advantages of Statistical Analysis

Data-Driven Decision-Making

Informed decisions based on objective evidence instead of intuition or assumptions.

Improved Efficiency

Identifying patterns and trends to optimize processes and reduce waste.

Uncovering Hidden Insights

Revealing relationships and correlations that may not be apparent through casual observation.

Increased Accuracy

Minimizing bias and subjectivity through rigorous analysis techniques.

Disadvantages and Limitations

Data Quality

Inaccurate, incomplete, or biased data can lead to misleading results.

Oversimplification

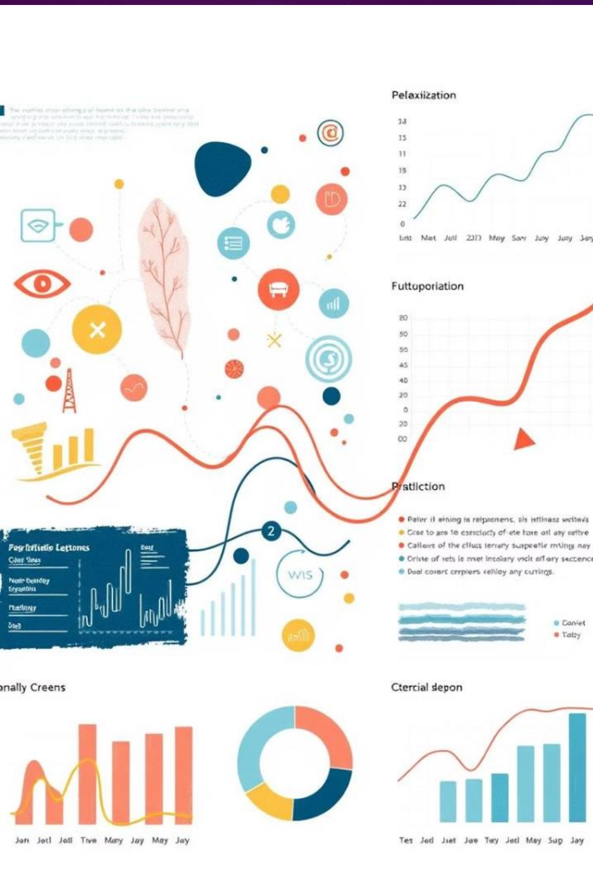
Statistical models can sometimes oversimplify complex phenomena, leading to an incomplete picture.

Misinterpretation

Statistical findings can be misinterpreted or misrepresented, leading to incorrect conclusions.

Ethical Concerns

The use of statistics can raise ethical concerns, such as privacy violations or data misuse.



Designing Effective Data Visualizations

Choose the Right Chart Type

Select a chart that effectively represents the data and communicates the desired message.

Use Color Strategically

Employ a color palette that enhances readability and highlights key trends.

Label Clearly and Concisely

Provide clear and concise labels to guide the viewer's understanding.

Prioritize Simplicity and Clari

Avoid clutter and ensure the visualization is easy to understand and interpret.

Conclusion and Key Takeaways

Statistics is a powerful tool for understanding the world around us. By mastering key concepts, leveraging its advantages, and designing effective visualizations, we can unlock valuable insights and make informed decisions.

Statistics:

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.

Statistics is the discipline that concerns the collection, organization, displaying, analysis, interpretation and presentation of data.

Limitation of statistics

1.Statistics does not deal with the individual items

2.It deals with quantitative data only.

3.It may mislead to wrong conclusion in the absence of details.

4.It laws are true only on averages.

5.It does not reveal/depict the entire story of phenomena.

6.It data should be uniform and homogeneous.

7.It laws are not exact.

8. It is liable to be misused.
9. It needs too many methods to study problems.

Distrust of statistics:

1. Statistics is an unreliable science.
2. It can prove anything.
3. It is a tissue of falsehood.
4. There are three type of lies-lies, damned lies and statistics.
5. Statistics are lies of the first order.
6. Figures do not lie, liars figure.

7. If figures say so it cannot be otherwise.

Statistics is the mathematical study of data collection, analysis, interpretation, and presentation. It's a branch of mathematics that's used to draw conclusions from data, and is considered the analytical foundation of science.

Here are some things to know about statistics:

- **Purpose**

Statistics is used to help make decisions in the face of uncertainty. It's concerned with how data can be used to extract meaningful conclusions.

- **Mathematical techniques**

Mathematical statistics uses mathematical analysis, linear algebra, stochastic analysis, differential equations, and measure-theoretic probability theory.

•Populations and samples

When applying statistics to a problem, you usually start by studying a population or process. If a census isn't possible, you can study a representative subset of the population, called a sample.

•Uncertainty and variation

Uncertainty and variation are two basic ideas in statistics. Statistical analysis is used to determine these uncertainties in different fields.

2nd Week

Topic: Mean, Median, Mode Basic

Page: 6-7

Mean: Mean is a value which is typical or representative of a set of data.

Mean is a single number describing some features of a set of data.

Median:

"The median is that value of the variable which divides the group into two equal parts, one part comprising all value greater, and the other, all values less than median."- **L.R Cannor.**

"Median of a series is the value of the item actual or estimated when a series is arranged in order of magnitude which divides the distribution into two parts." - **Secrist**

Mode: Mode is the value which occurs most densely in a set of observations and around which the other items of the set cluster densely. In other words, mode is the value of the variable which is predominant in the series.

Arithmetic Mean: Arithmetic mean of a set of observations is their sum divided by the number of observations. The arithmetic mean $\bar{x}$ of n observations $x_1, x_2, \cdots\cdots, x_n$ is given by

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

In case of frequency distributions x_i , f_i , $i = 1, 2, \cdots\cdots, n$, where f_i is the frequency of the variable x_i ,

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots\cdots + f_n x_n}{f_1 + f_2 + \cdots\cdots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \quad \left[\sum_{i=1}^{n} f_i = N\right]$$

Geometric Mean: Geometric mean of a set of observations is their n^th root of the product of that

$\frac{1}{n}$

Harmonic Mean: Harmonic mean of a set of observations, none of which is zero, is the reciprocal of the arithmetic mean of the reciprocals of the given values. The Harmonic mean H of n observations $x_1, x_2, \cdots, x_n$ is given by

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

Formulae of **Median** and **Mode** for continuous frequency distribution are as follows:

Median $= L + \frac{\frac{N}{2} - F_c}{f_m} \times C$ where

$L =$ LOWER LIMIT OF THE MEDIAN CLASS

$F_M =$ FREQUENCY OF THE MEDIAN CLASS

$C =$ Class interval

$F_c =$ Cumulative frequency of the class preceding median class.

Mode $= L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$ where

$L =$ Lower limit of the modal class

$f_1 =$ Frequency of the modal class

$f_0 =$ Frequency of the class preceding the modal class

$f_2 =$ Frequency of the class succeeding of the modal class

$C =$ Class interval

Week: 3-5

Topic: Mean, Median, Mode Examples

Page: 8-12

Question: Calculate the mean for the following frequency distributions:

Class interval	0-6	6-12	12-18	18-24	24-30	30-36
Frequency	5	10	15	20	14	8

Solution:

Class interval	Frequency (f_i)	Midpoint (x_i)	$f_i x_i$
0-6	5	3	15
6-12	10	9	90
12-18	15	15	225

18-24	20	21	420
24-30	14	27	378
30-36	8	33	264
	$N = 72$		$\sum f_i x_i = 1392$

According to the formula we have, mean

$$x = \frac{\sum f_i x_i}{N} = \frac{1392}{72} = 19.33$$

Question: Obtain the Median for the following frequency distributions:

Class interval:	10-20	20-30	30-40	40-50	50-60	60-70
Frequency:	4	6	10	15	12	7

Solution:

Class interval	Frequency (f_i)	Cumulative frequency (F)
10-20	4	4
20-30	6	10

30-40	10	20
40-50	15	35
50-60	12	47
60-70	7	54
	$N = 54$	

Here $\frac{N}{2} = 27$. Therefore, the median lies on the 40-50 class.

Median is calculated as follows: Median $= L + \frac{\frac{N}{2} - F_c}{f_m} \times C$.

We have $L = 40$, $f_m = 15$, $F_c = 20$ and $C = 10$.

Median $= 40 + \frac{27 - 20}{15} \times 10 = 40 + 4.667 = 44.667$

<u>Question:</u> Find the mode for the following frequency distributions:

Class interval:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency:	5	8	7	12	28	20	11	9

SOLUTION:

Class interval	Frequency (f_i)	Cumulative Frequency(F)
0-10	5	5
10-20	8	13
20-30	7	20
30-40	12	32
40-50	28	60
50-60	20	80
60-70	11	91
70-80	9	100

We have, Mode $= L + \dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$

Here $L = 40$, $f_1 = 28$, $f_0 = 12$, $f_2 = 20$ and $C = 10$

Mode $= 40 + \dfrac{28 - 12}{2 \times 28 - 20 - 12} \times 10 = 40 + \dfrac{16 \times 10}{24} = 40 + 6.667 = 46.667$

Question: Given the following data:

Daily wages	Number of workers
100-120	8
120-140	12

140-160	20
160-180	K
180-200	30
200-220	J
220-240	18
	N=160

If the Median is 186, then find the values of k and j.

SOLUTION:

Daily wages	Number of workers (frequency (f_i))	Cumulative Frequency (F)
100-120	8	8
120-140	12	20
140-160	20	40
160-180	k	40+k
180-200	30	70+k
200-220	j	70+k+j

220-240	18	88+k+j
	N=160	

According to the question $160 = 88 + k + j$ implies

$k + j = 72 \quad \cdots\cdots \quad (1)$

Median is calculated as follows: Median $= L + \dfrac{\frac{N}{2} - F_c}{f_m} \times C$.

Since Median is given to be 186, the class 180-200 is the Median class.

We have $L = 180$, $f_m = 30$, $F_c = 40 + k$, $C = 20$ and Median $= 186$.

Therefore

$$186 = 180 + \frac{80 - (40 + k)}{30} \times 20$$

$$\Rightarrow 186 - 180 = \frac{2(40-k)}{3} \Rightarrow 3\times 6 = 80 - 2k \Rightarrow 2k = 62 \therefore k = 31$$

Putting the value of k in (1) we get $31 + j = 72 \quad \therefore j = 41$

Question: The Median and Mode of the following wage distribution are known to be tk.335 and tk.340 respectively. Find the values of *f*, *g* and *h*.

Daily wages	Number of workers
0-100	4
100-200	16
200-300	*f*

300-400	*g*
400-500	*h*
500-600	6
600-700	4
	N=230

Solution:

Daily wages	Number of workers(frequency (f_i))	Cumulative Frequency(F)
0-100	4	4
100-200	16	20

200-300	f	20+f
300-400	g	20+f+g
400-500	h	20+f+g+h
500-600	6	26+f+g+h
600-700	4	30+f+g+h
	N=230	

ACCORDING TO THE QUESTION

$230 = 30 + f + g + h$ implies

$f + g + h = 200 \quad \cdots\cdots \quad (1)$

Median and Mode are calculated as follows:

Median $= L + \dfrac{\frac{N}{2} - F_c}{f_m} \times C$ and Mode $= L + \dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$

Since Median and Mode are given to be 335 and 340, the class 300-400 is both the Median and Modal class.

We have $L = 300$, $f_m = g$, $F_c = 20 + f$, $C = 100$, $f_1 = g$, $f_0 = f$, $f_2 = h$,

Median $= 335$ and Mode $= 340$.

Therefore and

$$335 = 300 + \frac{115 - (20 + f)}{g} \times 100$$

$$\Rightarrow 335 - 300 = \frac{100\,(95 - f)}{g}$$

$$\Rightarrow 35g = 9500 - 100f$$

$$\therefore \quad 20f + 7g = 1900 \cdots\cdots \quad (2)$$

$$340 = 300 + \frac{g - f}{2g - f - h} \times 100$$

$$\Rightarrow 340 - 300 = \frac{100\,(g - f)}{2g - f - h}$$

$$\Rightarrow 40\,(2g - f - h) = 100\,(g - f)$$

$$\Rightarrow 4g - 2f - 2h = 5g - 5f$$

$$\therefore \quad 3f - g - 2h = 0 \cdots\cdots \quad (3)$$

Now we get the following system of linear equations using (1), (2) and (3)

$$\left.\begin{aligned} f + g + h &= 200 \\ 20f + 7g &= 1900 \\ 3f - g - 2h &= 0 \end{aligned}\right\}$$

Solving the above system we obtain $f = 60, \quad g = 100, \quad h = 40$.

QUESTION: THE AVERAGE SALARY OF MALE EMPLOYEES IN A

firm was 520 and that of female was 420. The mean salary of all employees was 500. Find the percentage of male and female employees.

Solution: Let n_1 and n_2 denote respectively the number of male and female employees. Also let $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}$ denote respectively the average salary of male, female and all employees.

Therefore $\bar{x}_1 = 520$, $\bar{x}_2 = 420$ and $\bar{x} = 500$.

We have $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$\Rightarrow 500 = \frac{520n_1 + 420n_2}{n_1 + n_2}$$

$$\Rightarrow 500n_1 + 500n_2 = 520n_1 + 420n_2$$

$$\Rightarrow 80n_2 = 20n_1 \quad \Rightarrow \frac{n_2}{n_1} = \frac{20}{80} = \frac{1}{4} \quad \Rightarrow n_2 : n_1 = 1 : 4$$

Thus the percentage of male employees is $= \frac{4}{4+1} \times 100\% = 80\%$

and female employees is $= \frac{1}{4+1} \times 100\% = 20\%$

Question: Obtain the Median for the following frequency distributions:

x	6	7	8	9	10	11	12	13	14
f	8	10	11	16	20	25	15	9	6

SOLUTION :

x	F	F_c
6	8	8
7	10	18
8	11	29
9	16	45
10	20	65
11	25	90
12	15	105
13	9	114
14	6	120
Total N=120		

HERE $N = 120 \Rightarrow \frac{N}{2} = 60$. The cumulative frequency just greater than $\frac{N}{2}$ is 65 and the value of x corresponding to 65 is 10. Therefore Median is 10.

Week: 6-8

Topic: Standard Deviation, Mean deviation, Variance

Page: 13-15

DEVIATION IS DEFINED AS A STATISTICAL MEASURE THAT IS USED TO CALCULATE THE AVERAGE DEVIATION FROM THE MEAN VALUE OF THE GIVEN DATA SET. IF

$f_i, \; i = 1, \ldots\ldots, n$ is the frequency distribution, then mean deviation from the average is given by:

Mean deviation= $\frac{\sum \; f_i(|x - \bar{x}|)}{\sum f_i}$

Variance: The variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as 'σ^2'.

The population variance formula is given by:

Variance, $\sigma^2 = = \frac{\sum \; f_i(|x - \bar{x}|)}{\sum f_i}$.

Standard deviation: The square root of the variance.

✓Differences Between Standard Deviation and Variance

	Standard Deviation	**Variance**
What Is It?	The square root of the variance	The average of the squared differences from the mean
What Does It Indicate?	The spread between numbers in a data set	The average degree to which each point differs from the mean

✓Differences Between Standard Deviation and Variance

How Is It Expressed?	The same as the units in the data set	In squared units or as a percentage
What Does It Mean?	A low standard deviation (spread) means low volatility, while a high standard deviation (spread) means higher volatility	The degree to which returns vary or change over time

Question: The weights recorded by the students of a class are given. Calculate the mean, mean deviation, variance, and standard deviation.

Weights (Kg)	Frequency
54-57	5
58-61	7
62-65	10
66-69	12
70-73	6
74-77	5
78-81	4
82-85	1

2

Answer:

HERE, WE HAVE;

Weights (Kg)	Frequency (f_i)	x_i	$f_i x_i$	$\|x - \bar{x}\|$	$f_i(\|x - \bar{x}\|)$	$f_i(\|x - \bar{x}\|)^2$
54-57	5	55.5	277.5	11.44	57.20	654.368
58-61	7	59.5	416.5	7.44	52.08	387.4752
62-65	10	63.5	635	3.44	34.40	118.336
66-69	12	67.5	810	0.56	6.72	3.7632
70-73	6	71.5	429	4.56	27.36	124.7616

74-77	5	75.5	377.5	8.56	42.80	366.368
78-81	4	79.5	318	12.56	50.24	631.0144
82-85	1	83.5	83.5	16.56	16.56	274.2336
	50		3347		287.36	2560.32

3

Mean$= \frac{\sum f_i x_i}{\sum f_i} = \frac{3347}{50} = 66.94$

Mean deviation$= \frac{\sum f_i(|x-\bar{x}|)}{\sum f_i} = \frac{287.36}{50} = 5.7472$

Variance, $\sigma^2 = \frac{\sum f_i(|x-\bar{x}|)^2}{\sum f_i} = \frac{2560.32}{50} = 51.2064$

Standard deviation, $\sigma = \sqrt{51.2064} = 7.1559$.

Question: If a die is rolled, then find the variance and standard deviation of the possibilities.

Solution: When a die is rolled, the possible outcome will be 6. So the sample space, n = 6 and the data set = { 1;2;3;4;5;6}.

To find the variance, first, we need to calculate the mean of the data set.

Mean, $\bar{x}$ = (1+2+3+4+5+6)/6 = 3.5

We can put the value of data and mean in the formula to get;

$\sigma^2 = \Sigma (x_i - \bar{x})^2/n$

$\sigma^2 = \frac{1}{6} (6.25+2.25+0.25+0.25+2.25+6.25)$ $\sigma^2 = 2.917$

Now, the standard deviation, $\sigma = \sqrt{2.917} = 1.708$

Week: 9

Topic: Pie Diagram

Page: 15-16

PIE DIAGRAM: STEPS OF CONSTRUCTION OF PIE CHART FOR A GIVEN DATA:

1. Find the central angle for each component using the formula $\theta = \frac{Value\ of\ the\ component}{Total\ value} \times 360^{\circ} = \frac{f}{N} \times 360^{\circ}$
2. Draw a circle of any radius.
3. Draw a horizontal radius.
4. Starting with the horizontal radius, draw radii, making central angles corresponding to the values of respective components.
5. Repeat the process for all the components of the given data.

6. These radii divide the whole circle into various sectors.
7. Now, shade the sectors with different colors to denote various components.
8. Thus, we obtain the required pie chart.

Question: The number of hours spent by a school student on various activities on a working day is given below. Construct a pie chart using the angle measurement.

Daily Activity	Sleep	School	Play	Homework	Others	Total

Number of hours	6	6	3	4	5	24

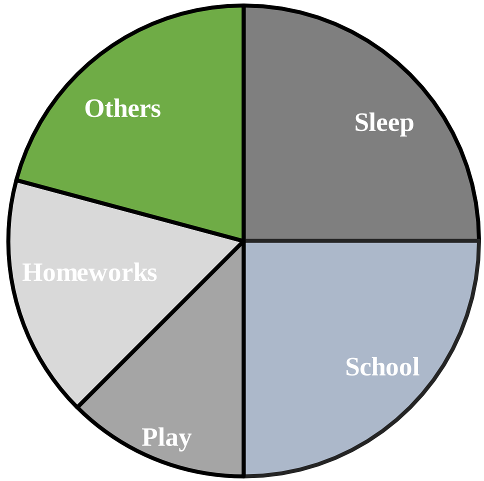
Solution:

Daily Activity	No. of Hours	Central Angle
Sleep	6	$\frac{6}{24} \times 360^{\circ} = 90^{\circ}$
School	6	$\frac{6}{24} \times 360^{\circ} = 90^{\circ}$
Play	3	$\frac{3}{24} \times 360^{\circ} = 45^{\circ}$
Homework	4	$\frac{4}{24} \times 360^{\circ} = 60^{\circ}$

Others	5	$\frac{5}{24} \times 360^\circ = 75^\circ$
Total	24	

Now, we shall represent these angles within the circle as different sectors.

Then we now make the pie chart:


Others
Sleep
Homeworks
School
Play

Week: 10

Topic: Histogram

Page: 16-17

Histogram: To draw a histogram we have to follow the following steps:

1.The magnitude of the class interval is plotted along the horizontal axis (X-axis) and the frequency on the vertical axis (Y-axis).

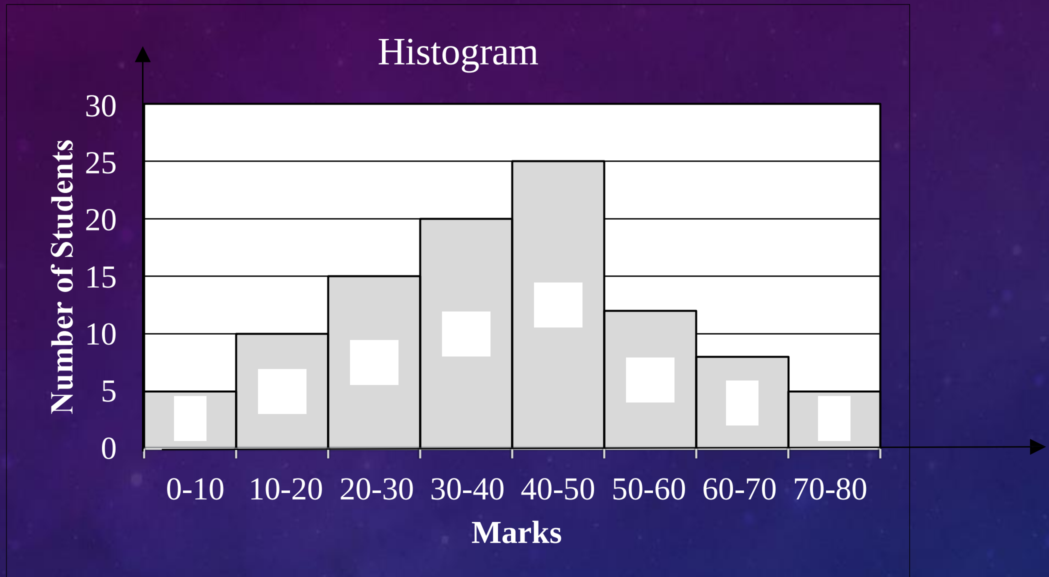
2.Each class has lower and upper values. This gives us two equal vertical line representing the frequency.

3.Upper end of the line are joined together. This process will give us rectangle, as there are classes; and the height of the rectangles are proportional to the frequencies. Histogram is also known as block diagram.

Question: Draw a histogram for the following table which represents the marks obtained by 100 students in an examination:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	5	10	15	20	25	12	8	5

Solution: The class intervals are all equal with length of 10 marks. Let us denote marks along the X-axis. Denote the number of students along the Y-axis, with appropriate scale. The histogram is given below.


Histogram
Number of Students
30
25
20
15
10
5
0
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
Marks

Week: 11

Topic: Bar Diagram

Page: 17-18

Bar Diagram: **Steps in construction of Bar Diagram**

1.On a graph, draw two lines perpendicular to each other, intersecting at 0. The horizontal line is x-axis and vertical line is y-axis.

2.Along the horizontal axis, choose the uniform width of bars and uniform gap between the bars and write the names of the data items whose values are to be marked.

3. Along the vertical axis, choose a suitable scale in order to determine the heights of the bars for the given values. (Frequency is taken along y-axis).
4. Calculate the heights of the bars according to the scale chosen and draw the bars.

Question: The percentage of total income spent under various needs by a family is given below.

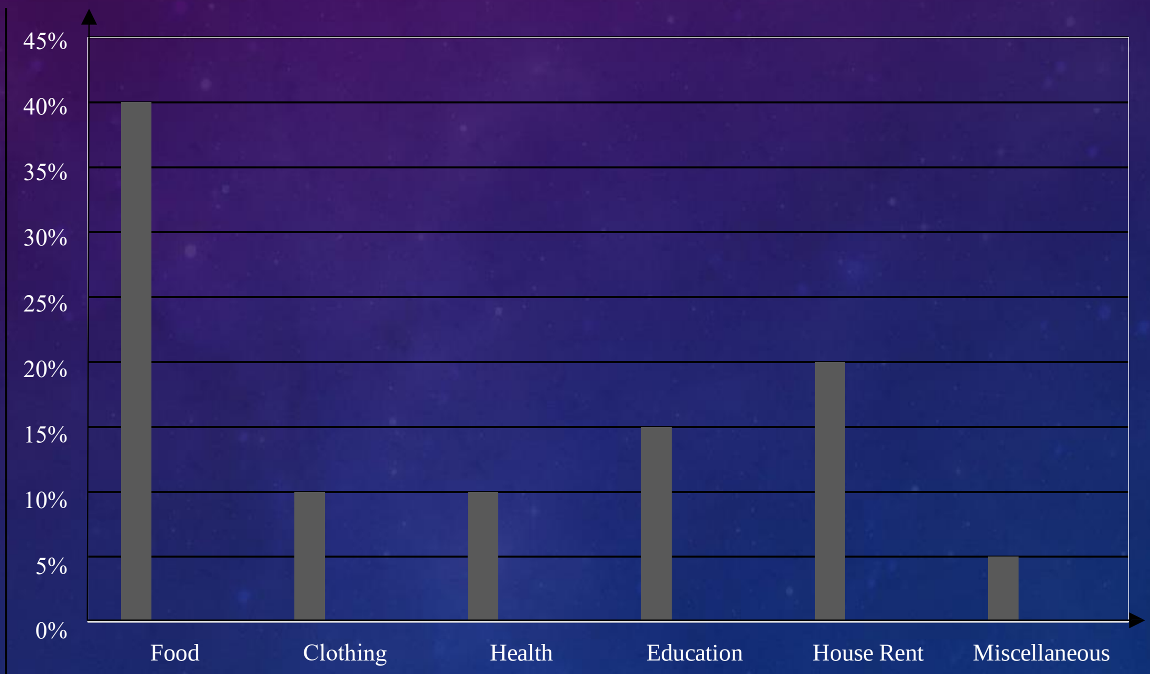
Different needs	Food	Clothing	Health	Education	House Rent	Miscellaneous
Percentage of total number	40%	10%	10%	15%	20%	5%

Solution: Let us denote these needs along the X-axis.

Denote the percentage of total

NUMBER ALONG THE Y-AXIS, WITH APPROPRIATE SCALE. THE BAR

diagram is given below.



Week: 12

Topic: Bar Diagram

Page: 18-19

<u>Frequency Polygon:</u> To draw frequency polygons, we follow the following steps:

1.Choose the class interval and mark the values on the horizontal axes.

2.Mark the mid value of each interval on the horizontal axes.

3.Mark the frequency of the class on the vertical axes.

4. Corresponding to the frequency of each class interval, mark a point at the height in the middle of the class interval.
5. Connect these points using the line segment.
6. The obtained representation is a frequency polygon.

Question: Draw a frequency polygon for the following data

Class interval : 10-20 20-30 30-40 40-50 50-60 60-70 70-80 80-90 90-100

Frequency	**:**	3	5	9	11	14			
18	12	7	4						

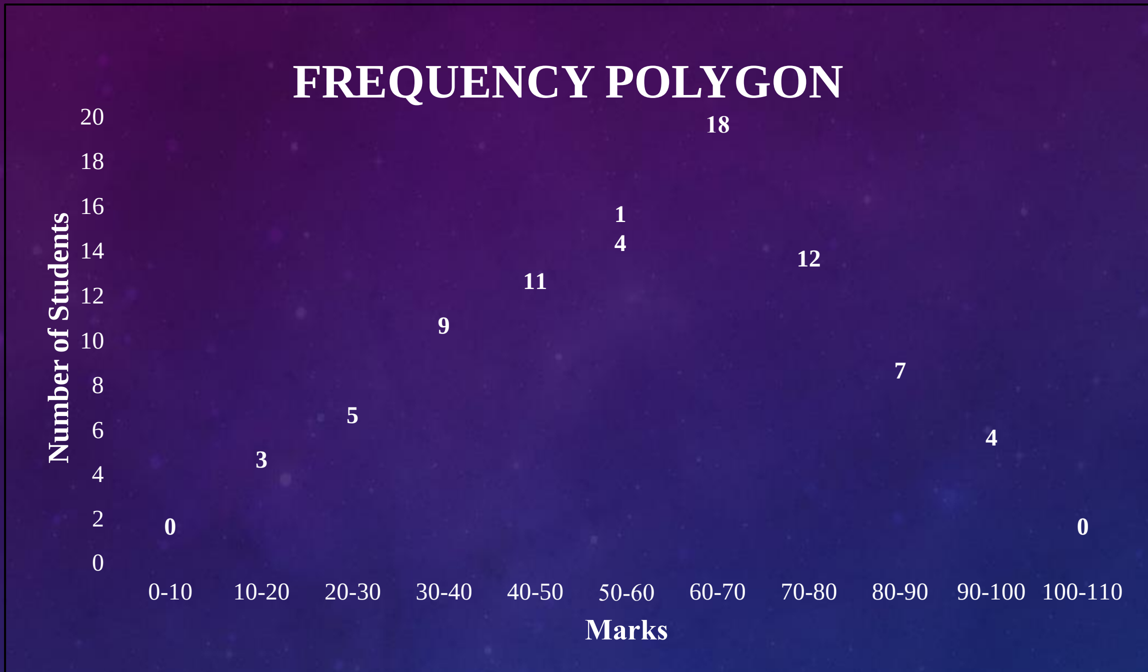
Solution: Mark the class intervals along the X-axis and the frequency along the Y-axis. We take the imagined classes 0-10 at the beginning and 100-110 at the end, each with frequency zero.

Marks	**Midpoint** (x)	**No. of Students** (f)
0-10	5	0
10-20	15	3
20-30	25	5

30-40	35	9
40-50	45	11
50-60	55	14
60-70	65	18
70-80	75	12
80-90	85	7
90-100	95	4
100-110		0

Using the adjacent table, plot the points A (5, 0), B (15, 3), C (25, 5), D (35, 9), E (45, 11), F (55, 14), G (65, 18),

H (75, 12), I (85, 7), J (95, 4) and K (105, 0). Joining the above two adjacent points wen obtain the required frequency polygon.


FREQUENCY POLYGON
Number of Students
Marks
20
18
16
14
12
10
8
6
4
2
0
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
100-110
0
3
5
9
11
1
4
18
12
7
4
0

Week: 13

Topic: Ogives

Page: 18-19

Ogives: The class limits are shown along the X-axis and cumulative frequencies are shown along the Y-axis. In drawing an Ogives, the cumulative frequency is plotted at the upper limit of the class interval. The successive points are joined together to get an Ogives curve.

There are two methods of constructing an Ogives:

1. Less than Ogives
2. More than Ogives

In less than Ogives, the less than cumulative frequencies are plotted against upper class boundaries of the respective classes. Then the points are joined by a smooth free hand curve and has the shape of an elongated S.

In more than Ogives, the more than cumulative frequencies are plotted against lower class boundaries of the respective classes. Then the points are joined by a smooth free hand curve and has the shape of an elongated S, upside downward.

From the intersection point of these Ogives, a perpendicular line touches x-axis, is the value of median.

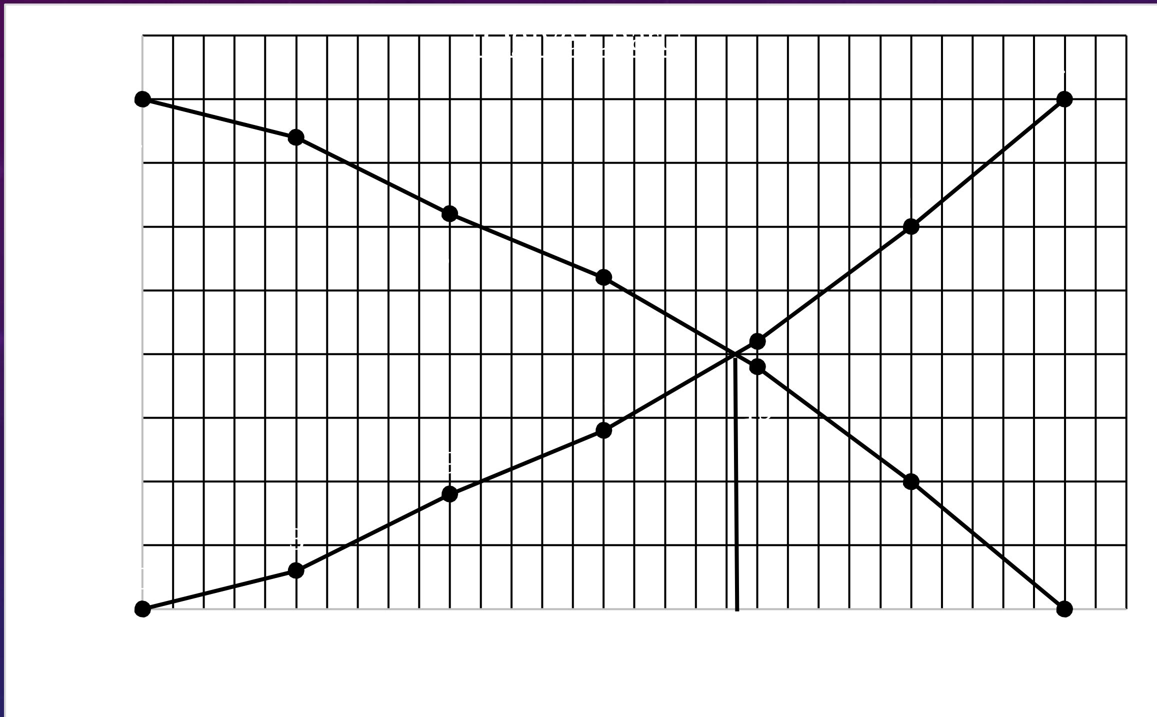
Question: Draw Ogives for the following table which represents the frequency distribution of weights of 40 students.

Weights in kg	No. of students
40-45	3
45-50	6
50-55	5
55-60	7
60-65	9
65-70	10

Solution:

Weights less than	No. of students	Weights more than	No. of students
40	0	40	40
45	3	45	37
50	9	50	31
55	14	55	26
60	21	60	19
65	30	65	10
70	40	70	0

Let us denote marks along the X-axis. Denote the number of students along the Y-axis, with appropriate scale. The graphical representation of the above tabular form is given below.



From this graph it is clear that the value of the median is 59.3.

<u>Correlation and Regression:</u> Correlation and Regression are the two analysis based on multivariate distribution. A multivariate distribution is described as a distribution of multiple variables. Correlation is described as the analysis which lets us know the association or the absence of the relationship between two variables ‘x’ and ‘y’. On the other hand, Regression analysis, predicts the value of the dependent variable based on the known value of the independent variable, assuming that average mathematical relationship between two or more variables. The difference between correlation and regression is one of the commonly asked questions in interviews. Moreover, many people suffer ambiguity in understanding these two. So, take a full read of this article to have a clear understanding on these two.

Basis for Comparison	**Correlation**	**Regression**
Meaning	Correlation is a statistical measure which determines co-relationship or association of two variables.	Regression describes how an independent variable is numerically related to the dependent variable.

Usage	To represent linear relationship between two variables.	To fit a best line and estimate one variable on the basis of another variable.
Dependency	Dependent and Independent variables are not different.	Both Dependent and Independent variables are different.
Indicates	Correlation coefficient indicates the extent	Regression indicates the impact of a unit change in

	to which two variables move together.	the known variable (x) on the estimated variable (y).

Karl Pearson (1867-1936) a British Biometrician, developed the coefficient of correlation to express the degree of linear relationship between two variables Correlation co-efficient between two random variables X and Y denoted by $r(X,Y)$, is given by $r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

Where $Cov(X,Y) = \frac{1}{n}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$ (covariance between X and Y)

$\sigma_X = \sqrt{\frac{1}{n}\sum_i (X_i - \bar{X})^2}$ (Standard deviation of X)

$$\sigma_Y = \sqrt{\frac{1}{n}\sum_i (Y_i - \bar{Y})^2}$$ (Standard deviation of Y)

Thus $$r(X,Y) = \frac{\frac{1}{n}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n}\sum_i (X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum_i (Y_i - \bar{Y})^2}} = \frac{\sum x \cdot y}{\sqrt{\sum x^2 \sum y^2}}$$

Where $x = X_i - \bar{X}$ and $y = Y_i - \bar{Y}$

Regression line

For the pair of values of (X,Y), where X is an independent variable and Y is the dependent variable. The line of regression of Y on X is given by $Y - \bar{Y} = b_{yx}(X - \bar{X})$ where b_{yx} is the regression co-efficient of Y on X and given by $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

WHERE *R* IS THE CORRELATION CO-EFFICIENT BETWEEN X AND Y

and σ_x and σ_y are the standard deviations of X and Y, respectively.

$\therefore \quad b_{yx} = \frac{\sum xy}{\sum x^2}$ where $x = X - \bar{X}$ and $y = Y - \bar{Y}$

Similarly when Y is treated as an independent variable and X as dependent variable, the line of regression of X on Y is given by $X - \bar{X} = b_{xy}(Y - \bar{Y})$ where b_{xy} is the regression co-efficient of X on Y and given by $b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\sum xy}{\sum y^2}$ where $x = X - \bar{X}$ and $y = Y - \bar{Y}$.

Question: Find if there is any significant correlation between the heights and weights given below.

Height in inches	57	59	62	63	64	65	55	58	57
Weight in lbs.	113	117	126	126	130	129	111	116	112

Solution:

eight ches	Deviations from mean(60) $x = X - \bar{X}$	Square of the deviations $x^2 = (X - \bar{X})^2$	Weight in lbs. Y	Deviations from mean(120) $y = Y - \bar{Y}$	Square of the deviations $y^2 = (Y - \bar{Y})^2$	Produc of deviati o of X an Y $x \cdot y$
57	-3	9	113	-7	49	21

59	-1	1	117	-3	9	3
62	2	4	126	6	36	12
63	3	9	126	6	36	18
64	4	16	130	10	100	40
65	5	25	129	9	81	45
55	-5	25	111	-9	81	45
58	-2	4	116	-4	16	8
57	-3	9	112	-8	64	24
540	**0**	**102**	**1080**	**0**	**472**	**216**

Here $X = \frac{\sum X}{n} = \frac{540}{9} = 60$ and $Y = \frac{\sum Y}{n} = \frac{1080}{9} = 120$

Co-efficient of Correlation $r = \frac{\sum x \cdot y}{\sqrt{\sum x^2 \sum y^2}} = \frac{216}{\sqrt{102 \times 471}} = 0.985469555$

REGRESSION EQUATION OF Y ON X IS

$Y - \bar{Y} = b_{yx}(X - \bar{X})$ where $b_{yx} = \frac{\sum x \cdot y}{\sum x^2} = \frac{216}{102} = \frac{36}{17}$ therefore

$Y - 120 = \frac{36}{17}(X - 60)$ $\quad \therefore 17Y = 36X - 120$

Regression equation of X on Y is

$X - \bar{X} = b_{xy}(Y - \bar{Y})$ where $b_{xy} = \frac{\sum x \cdot y}{\sum y^2} = \frac{216}{1080} = \frac{1}{5}$ therefore

$X - 60 = \frac{1}{5}(Y - 120)$ $\quad \therefore 5X = Y + 180$

<u>**Question:**</u> From the data given below find out

Ages of Husbands	22	23	23	24	26	27	27	28	30	30

Ages of Wives	18	20	21	20	21	22	23	24	25	26

a) Coefficient of correlation between the ages of husbands and wives.

b) The two regression line.

c) Most likely age of husband when the wife's age is 30. d)Most likely age of wife when the husband's age is 29.

Answer: $r = 0.95, \quad Y = 0.82X + 0.789474, \quad X = 1.1Y + 1.643$

<u>Question:</u> From the data given below find out

Marks in Mathematics	55	56	56	57	57	58	58	60	64	69

Marks in Statistics	65	65	66	66	67	67	68	69	70	71

a) Coefficient of correlation between the marks in Mathematics and Statistics.

b) The two regression line.

c) Most likely mark in Mathematics when in Statistics is 69.

d) Most likely mark in Statistics when in Mathematics is 66.

$r = 0.9238, \quad Y = .4412X + 41.37$

Week: 14

Topic: Probability

Page: 24-28

Experiment: An experiment is an act that can be repeated under given identical conditions.

Example: Throwing a die, tossing a coin, drawing cards from a bridge deck are the examples of experiment.

Trial: Each of the repetition in an experiment is called a trial. This means that trial is a special case of experiment. Experiment may be a trial or two or more trials.

Outcomes: The results of an experiment are known as outcomes.

Examples:

- If we throw a die we get 1 or 2 or 3 or 4 or 5 or 6. So that individually 1 is an outcome, 2 is an outcome.
- If we toss a coin we get head or tail. Individually head and trial are two outcomes.

Sample space: A sample space is the set of all outcomes.

Example:

- If we throw a die the outcomes are 1, 2, 3, 4, 5 and 6. Then $S = \{1, 2, 3, 4, 5, 6\}$ is a sample space.
- If we toss a coin then the outcomes are head (H) and tail (T). Then $S = \{H, T\}$ is a sample space.

Event: An event is the collection of one or more outcomes of an experiment.

Example: If we throw a die the outcomes are 1, 2, 3, 4, 5, and 6. Then the outcomes of even numbers are 2, 4, 6. Then $A = \{2, 4, 6\}$ is called an event of even numbers.

Sure Event: An event is called sure event when it always happens.

The probability of a sure event is 1.

Example: Today the sun rises in the east is a sure event and its probability is 1.

Impossible Event: An event is called impossible event when it never happens.

OF AN IMPOSSIBLE

The probability

event is 0.

Example: A river does not contain any fish is an impossible event and its probability is 0.

Equally Likely Outcomes: Outcomes of an experiment are said to be equally likely if we have no reason to expect any one rather than the other.

Example: i) In tossing a fair coin, the outcomes ‘head’ and ‘tail’ are equally likely.

ii)In throwing a balanced die, all the six faces are equally likely.

iii)Drawing a card from a bridge deck, all the 52 cards are equally likely to come.

Mutually Exclusive event: When an event occurs and none of the other events will occur at the same time, then the event is called mutually exclusive event.

Example:

If we toss a coin two outcomes head (H) and tail (T) are mutually exclusive event. Because if it appears head (H) or tail (T) not both head and tail at the same time.

Formally, two events A and B are mutually exclusive if and only if $A \cap B = \phi$.

Favorable Outcomes: The outcomes of an experiment are said to be favourable to an event if they entail the happening of the event.

Example: i) In throwing a die, the favourable outcomes of the even numbers on the faces of the die will be 2, 4 and 6.

ii) In drawing a card from a bridge deck, the favourable outcomes of the event heart will be 13 because it may be any one of these 13 hearts.

Exhaustive events: The total number of possible outcomes in any trial is known as exhaustive events or exhaustive cases.

Example:

- In tossing of a coin there are two exhaustive cases viz, head and tail.
- In throwing of a die, there are six exhaustive cases since any one of the 6 faces 1, 2, 3, 4, 5, 6 may come upper most.

Probability:

If there are ‘n’ mutually exclusive, equally likely and exhaustive outcomes of an experiment and if ‘m’ of these outcomes are favourable to a event A, then the probability of the event A is denoted by p (A) and defined as

$$p = p(A) = \frac{\text{Favourable Outcomes of an event A}}{\text{Total Number of Outcomes of the Experiment}} = \frac{m}{n}$$

Properties of probability

Let E be an experiment. Also let S be a sample space associated with E, with each event A we associate a real number, designed by $P(A)$ and called the probability A of satisfying the following properties:

- ❖ $0 \leq p(A) \leq 1$.
- ❖ $p(S) = 1$.

❖If A and B are mutually exclusive events,

$p(A \cup B) = p(A) + p(B)$.

Example: A bag contains 4 white and 6 black balls. If one ball is drawn at random from the bag, what is the probability that the ball is (a) black (b) white (c) white or black and (d) red.

SOLUTION:

Hence the total numbers of balls are 10. Since one ball is drawn from the bag, there are 10 mutually exclusive, equally likely and exhaustive outcomes of this experiment.

(a) Let A be the event that the ball is black, then the number of favorable outcomes to A is 6. So that

$$p(A) = \frac{\text{number of black balls}}{\text{total number of balls}} = \frac{6}{10}.$$

(b) Let B be the event that the ball is white, then the favorable outcomes corresponding to B are 4. Therefore

$$P(B)=\frac{4}{10}.$$

(c) Let C be the event that the ball is white or black, then the favorable outcomes corresponding to C are 10. Therefore

$$p(C)=\frac{10}{10}=1.$$

(d) Let D be the event that the drawn ball is red, then the favorable outcomes corresponding to D is zero. Therefore

$$p(D)=\frac{0}{10}=0$$

Example: Three contractor, A, B and C are bidding for the construction of a new cinema hall. Some expert in

THIS INDUSTRY BELIEVES THAT A HAS EXACTLY HALF THE CHANCE

that B has, B, in turn is $\frac{4}{5}$ the as like as C to win the contract. What is the probability for each to win the contract if the expert's estimates are accurate?

Solution: Let the probability of C's wining the contract is x. Then $p(C)=x$, $p(B)=\frac{4}{5}x$ and $p(A)=\frac{4}{5}.\frac{1}{2}x$. We know that total probability is one.

Therefore $x+\frac{4}{5}x+\frac{2}{5}x=1 \Rightarrow \frac{5x+4x+2x}{5}=1 \Rightarrow \frac{11x}{5}=1 \quad \therefore x=\frac{5}{11}$

Hence, $p(C)=\frac{5}{11}$, $p(B)=\frac{4}{11}$ and $p(A)=\frac{2}{11}$.

Example: Tickets are numbered from 1 to 100. They are well shuffled and a ticket is drawn at random. What is the probability that the drawn ticket has

1)An odd number

2)A number 4 or multiple of 4

3)A number which is greater than 70 and 4)A number which is square?

Solution: Since there are 100 tickets, the total number of exclusive mutually exclusive and equally likely case is 100.

1)Let A denote the event that the ticket drawn an odd number. Since there are 50 odd number tickets so the number of cases favorable to the event A is 50.

$$\therefore p(A) = \frac{50}{100} = 0.5.$$

2)Let B denote the event that the ticket drawn has a number 4 or multiple of 4. The numbers favorable to event B are $4,\ 8, 12, 16,\ 20,..., 92,\ 96, 100$. The total number of

cases will be $\frac{100}{4} = 25$. $\therefore p(B) = \frac{25}{100} = 0.25$.

3) Let C denote the event that the drawn ticket has a number greater than 70. Since the number greater than 70 are $71,\ 72,\ 73, \ldots,\ 100$. Therefore, 30 cases are favorable to the event C. $\therefore\ p(C) = \frac{30}{100} = 0.3$.

4) Let D denote the event that the drawn ticket has a number which is a square. Since the squares between 1 and 100 are $1,\ 4,\ 9,\ 16,\ 25,\ 36,\ 49,\ 64,\ 81$ and 100. So the cases favorable to event D are 10 in number. Hence,

$$\therefore\ p(D) = \frac{10}{100} = 0.1.$$

Week: 15

Topic: Conditional probability

Page: 29-31

Conditional probability: Let A and B be two events. The conditional probability of event A given that B has occurred, is defined by the symbol $p(A|B)$ and is found to be:

$$p(A|B)=\frac{p(A\cap B)}{p(B)}; \text{ provided } p(B)>0.$$

Similarly, $$p(B|A)=\frac{p(A\cap B)}{p(A)}; \text{ provided } p(A)>0.$$

Example: A hamburger chain found that 75% of all customers use mustard, 80% use ketchup and 65% use

BOTH. WHAT ARE THE PROBABILITIES THAT A KETCHUP USER USES MUSTARD AND THAT A MUSTARD USER USES KETCHUP?

Solution: Let A be the event "customer uses mustard" and B be the event "customer uses ketchup". Thus, we have, $p(A) = 0.75$, $p(B) = 0.80$ and $p(A \cap B) = 0.65$.

The probability that a ketchup user uses mustard is the conditional probability of event A, given event B is

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{0.65}{0.80} = 0.8125.$$

SIMILARLY, THE PROBABILITY THAT A MUSTARD USER USE KETCHUP IS

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{0.65}{0.75} = 0.8667$$

<u>Rules of Addition:</u>

Special rule of Addition: If two events A and B are mutually exclusive, the special rule of addition states that the probability of one or the other events occurring equals the sum of their probabilities i., e.,

$$p(A \text{ or } B) = p(A) + p(B) \text{ or, } \quad p(A \cup B) = p(A) + p(B).$$

For three mutually exclusive events designated A, B and C the rule is written as

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

Or, $p(A \cup B \cup C) = p(A) + p(B) + P(C)$.

Example: If we toss a coin then what is the probability of head or tail?

SOLUTION: HERE THERE ARE TWO EVENTS, NAMELY EVENT $A = H$ AND EVENT $B = T$. SO THAT

$$p(A \text{ or } B) = p(A) + p(B) = \frac{1}{2} + \frac{1}{2} = 1$$

- **The general rules for addition:** When two or more events are not mutually exclusive then we use the general rule for addition. The rule is
- $p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B)$

or $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Week: 16

Topic: Examples

Page: 31-32

EXAMPLE: MR. X FEELS THAT THE PROBABILITY THAT HE WILL

pass Mathematics is $\frac{2}{3}$ and Statistics is $\frac{5}{6}$. If the probability that he will pass both the course is $\frac{3}{5}$. What is the probability that he will pass at least one of the courses?

Solution: Let M and S be the events that he will pass the courses Mathematics and Statistics respectively. The event $M \cup S$ means that at least one of M or S occurs. Therefore

$$p(M \cup S) = p(M \text{ or } S) = p(\text{he pass at least one of the course})$$

$$= p(M) + p(S) - p(M \text{ or } S) = \frac{2}{3} + \frac{5}{6} - \frac{3}{5} = \frac{9}{10} .$$

EXAMPLE: MR. Y FEELS THAT THE PROBABILITY THAT HE WILL

get A in Calculus is $\frac{3}{4}$, A in Statistics is $\frac{4}{5}$ and A in both the courses is $\frac{3}{5}$. What is the probability that Mr. Y will get

a) At least one A
b) No A's?

Solution: Let C be the event that Mr. Y will get A in Calculus and S be the event that he will get A in Statistics. We have, $p(C) = \frac{3}{4}$, $p(S) = \frac{4}{5}$ and $p(C \text{ and } S) = p(C \cap S) = \frac{3}{5}$

.

a) $p(\text{at least one } A) = p(C \cup S)$

$$= p(C) + p(S) - p(C \cup S) = \frac{3}{4} + \frac{4}{5} - \frac{3}{5} = \frac{19}{20}$$

b) $p(\text{no } A's) = p(\overline{CS}) = 1 - p(C \cup S) = 1 - \frac{19}{20} = \frac{1}{20}$

Complement rule

The complement rule is used to determine the probability of an event occurring by subtracting the probability of the event not occurring from 1 i., e. $p(A) = 1 - p(\overline{A})$.

Example:

Weight	Event	Probability
Underweight	A	0.025
Satisfactory	B	??
Overweight	C	0.075

Find $p(B)$.

Solution: We know that $p(B) = 1 - p(\overline{B}) = 1 - (0.025 + 0.075) = 0.90$.

Rules of Multiplication: Special rule of multiplication requires that two events A and B are independent (two events are independent if the occurrence of one event does not alter the probability of the occurrence of the other event).

For two independent events A and B, the probability that A and B will both occur is found by multiplying the two probabilities i., e., $p(A \text{ and } B) = p(A) * p(B)$.

For three events A, B and C the special rule of multiplication used to determine the probability that all the events will occur is:

$$p(A \text{ and } B \text{ and } C) = p(A) * p(B) * p(C)$$

Example: A company has two large computers. The probability that the newer one will breakdown on any particular month is 0.05, the probability that the older one will breakdown on any particular month is 0.1. What is the probability that they will both breakdowns in a particular month?

Solution: Let, Event A is the newer one will breakdown and Event B is the older one will breakdown. So that

$p(A) = 0.05$ and $p(B) = 0.1$.

$$\therefore P(A \text{ AND } B) = P(A) * P(B) = 0.05 * 0.1 = 0.005 .$$

General rule of Multiplication: The general rule of multiplication states that for two events A and B, the joint probability that both events will happen is found by multiplying the probability of event A will happen by the conditional probability of event B occurring given that event A has occurred. Symbolically, the joint probability is

$$p(A \text{ and } B) = p(A) * p(B \mid A) .$$

Example: There are 10 rolls of film in a box, 3 of which are defective. Two rolls are to be selected one after another. What is the probability of selecting a defective roll followed by another defective roll?

Solution: The first roll of film selected from the box being found defective is event D_1. Therefore, $p(D_1) = \frac{3}{10}$. The second roll selected being found defective is event D_2. Therefore, $p(D_2 | D_1) = \frac{2}{9}$. Since, after the first selection was found to be defective, only 2 defective rolls of film remained in the box containing 9 rolls.

SO THE PROBABILITY OF TWO DEFECTIVES IS

$$= p(D_1 \text{ and } D_2) = p(D_1) \times p\left(D_2 \middle| D_1\right) = \frac{3}{10} \times \frac{2}{9} = \frac{6}{90} = 0.07$$

Bayes Theorem:

If $E_1, E_2, \ldots, E_n$ are mutually disjoint events with $p(E_i) \neq 0\ ;\ i = 1, 2, \ldots, n$ then for any arbitrary event A which is a subset of $\bigcup_{i=1}^{n} E_i$ such that $p(A) > 0$, we have

$$p\left(E_i \middle| A\right) = \frac{p(E_i)\, p\left(A \middle| E_i\right)}{\sum_{i=1}^{n} p(E_i)\, p\left(A \middle| E_i\right)}, \quad i = 1, 2, \ldots, n.$$

Example: In a bolt factory machines A, B and C manufacturer respectively 25%, 35% and 40% of the total. Of their output 5, 4 and 2 percent are defective bolts. A bolt is drawn at randomly from the product and is found to be defective. What are the probabilities that it was manufactured by machines A, B and C?

Solution: Let E_1, E_2 and E_3 denote the events that a bolt selected at random is manufactured by the machines A, B and C respectively and let E denote the event of its being

DEFECTIVE. THEN WE HAVE,

$p(E_1)=0.25$, $p(E_2)=0.35$ and $p(E_3)=0.40$

.

The probability of drawing a defective bolt manufactured by machine A is $p(E|E_1)=0.05$. Similarly, we have $p(E|E_2)=0.04$ and $p(E|E_3)=0.02$.

Hence, the probability that a defective bolt selected at random is manufactured by machine A is given by

$$p(E_1 \mid E) = \frac{p(E_1)\, p(E \mid E_1)}{\sum_{i=1}^{3} p(E_i)\, p(E \mid E_i)} = \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = 0.362$$

Similarly, $$p(E_2 \mid E) = \frac{p(E_2)\, p(E \mid E_2)}{\sum_{i=1}^{3} p(E_i)\, p(E \mid E_i)} = \frac{0.35 \times 0.04}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= 0.406$$

and $$p(E_3 \mid E) = \frac{p(E_3)\, p(E \mid E_3)}{\sum_{i=1}^{3} p(E_i)\, p(E \mid E_i)} = \frac{0.40 \times 0.02}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = 0.232$$

<u>Question:</u> In the submission of an income tax return, a tax-payer has applied for the allowance of certain expanses, for which the legitimacy is subject to verification. A three-member review board operating on a rotation schedule receives applications for consideration. The first board, consisting mainly of members have vast experience, approves only 25% of such claims without verification. The second board, consisting mainly of midterm individuals, approves 50% of such claims. The third board with relatively

newer members dubious about refuting claims approves 70% of such cases. There is a 20% chance that the application will be considered by the first board, a 50% chance that it will be considered by the second board, and a 30% chance that it will be considered by the third board.

(i) What is the probability that the application will be approved?

APPROVED, WHAT IS THE PROBABILITY THAT IT WAS CONSIDERED BY (1) THE FIRST BOARD (2) THE SECOND BOARD (3) THE THIRD BOARD?

Solution: The approval may be granted by any one of three boards. Hence if A denotes the event that the application will be approved and B_1, B_2 and B_3 denote the considerations by the first, second and third boards respectively, the required probabilities are:

$$(i)\ p(A) = p(B_1 \cap A) + p(B_2 \cap A) + p(B_3 \cap A)$$

$$= p(B_1)\,p(B_1/A) + p(B_2)\,p(B_2/A) + p(B_3)\,p(B_3/A)$$

$$= 0.20 \times 0.25 + 0.50 \times 0.5 + 0.30 \times 0.7 = 0.51$$

(ii)

$$p(B_1/A) = \frac{p(B_1)\,p(A/B_1)}{p(A)} = \frac{0.20 \times 0.25}{0.51} = 0.098$$

$$p(B_2/A) = \frac{p(B_2)\,p(A/B_2)}{p(A)} = \frac{0.50 \times 0.5}{0.51} = 0.49$$

$$p(B_3/A) = \frac{p(B_3)\,p(A/B_3)}{p(A)} = \frac{0.30 \times 0.7}{0.51} = 0.412$$

Week: 17

Topic: Conditional probability

Page: 36-54

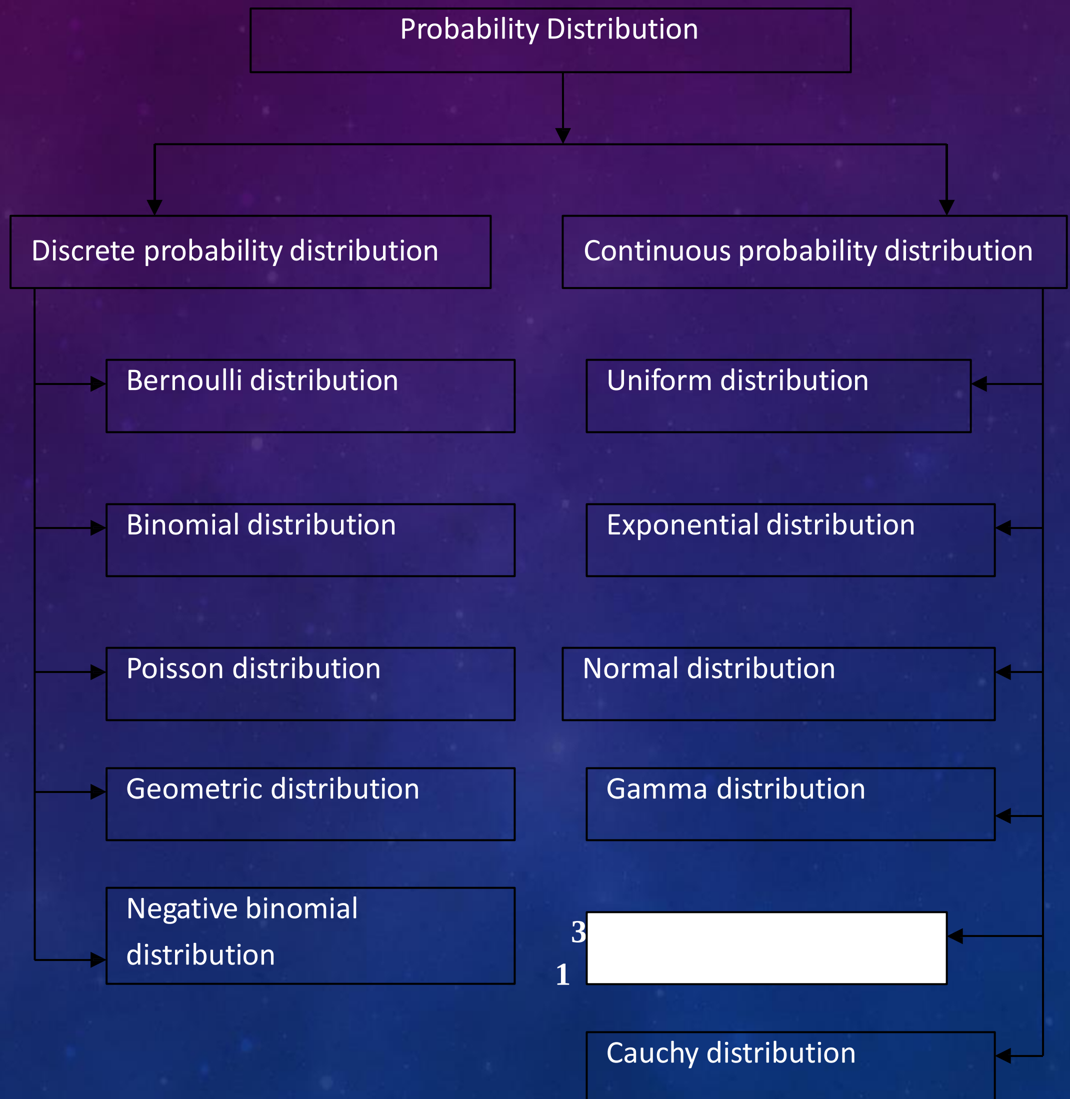
Probability Distribution

Probability Distribution: A probability distribution shows the possible outcomes of an experiment and the probability of each of these outcomes.

Or,

A listing of all the outcomes of an experiment and the probability associated with each outcome.

Types of probability distribution



Probability Distribution
Discrete probability distribution
Continuous probability distribution
Bernoulli distribution
Binomial distribution
Poisson distribution
Geometric distribution
Negative binomial distribution
Uniform distribution
Exponential distribution
Normal distribution
Gamma distribution
3
1
Cauchy distribution

Discrete probability distribution:

A discrete probability can take on only a limited number of values which can be listed.

Example: The probability that you were born in a given month is also discrete because there are 12 possible values.

Continuous probability distribution:

In a continuous probability distribution the variable under consideration is allowed to take on any within a given range. So we cannot list all the possible values.

Example: Suppose we were examining the level of effluent in a variety of streams and we measured the level of effluent by parts of effluent per million parts of water. We would expect quite a continuous range of parts per million (ppm), all the way from very low levels is clear mountains streams of extremely high levels in polluted streams. We would call the distribution of this variable (ppm) a continuous distribution.

Bernoulli distribution

Bernoulli trial: A random experiment whose outcomes have been classified into two categories namely “success” and “failure” represented by letters s and F respectively is called a Bernoulli trail.

Bernoulli distribution: A discrete random variable X is said to have a Bernoulli distribution if its probability function is given by

$$f(x, p) = \begin{cases} p^x q^{1-x} & \text{for } x = 0,1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter of the distribution satisfying $0 \leq p \leq 1$ and $p + q = 1$.

Example: A coin is tossed in which the outcome "head" is a success and the probability of head is p. Then $q = 1 - p$

IS THE PROBABILITY OF FAILURE OR TAIL. IF THE NUMBER OF HEADS OR SUCCESS IS A RANDOM VARIABLE , THE CAN TAKE VALUES

0 or 1 according to the outcome is tail (failure) or head (success). Then the probability function of x is

$$f(x,p) = \begin{cases} p^x q^{1-x} & \text{for } x = 0,1 \\ 0, & \text{otherwise} \end{cases}$$

Binomial distribution

Definition: A discrete random variable x is said to have a binomial distribution if its probability function is defined by

$$f(x;n,p)=\begin{cases}\binom{n}{x}p^{x}q^{n-x} & \text{for } x=0,1,2,\ldots,n\\ 0; & \text{otherwise}\end{cases}$$

where the two parameters n and p SATISFY $0 \le P \le 1$ AND $p+q=1$, also n is positive integer.

- **Conditions for Binomial distribution:**
- ❖Each trial results in two outcomes, termed as success and failure.
- ❖The number of trials n is finite.
- ❖The trials are independent of each other.

❖THE PROBABILITY OF SUCCESS p IS CONSTANT FOR EACH

- trial.
- **Mean of Binomial distribution: Mean** $= np$, where $n =$
- number of trials and $p =$ probability of success.
- **Variance of Binomial distribution: Variance** $= npq$, where, n $=$ number of trials, $p =$ probability of success and
- $q =$ probability of failure $= 1 - p$.

born child will be boy $\frac{2}{5}$. Among the 4 newly born

- **Example:** In a community, the probability that a newly

children in that community, what is the probability that

(A) ALL THE FOUR BOYS; (B) NO BOYS; (C) EXACTLY ONE BOY.

Solution: Let us consider the event that a newly born child is a boy as success in Bernoulli trial with probability of success $\frac{2}{5}$. Let the number of boys be a random variable x. Then x can take values 0, 1, 2, 3,
and 4.

According to binomial law, the probability function of x is

$$f\left(x,4,\frac{2}{5}\right)=\binom{4}{x}\left(\frac{2}{5}\right)^{x}\left(\frac{3}{5}\right)^{4-x}$$ FOR $X=0,1,2,3,4$.

a) $p(\text{all boys}) = p(x=4) = \binom{4}{4}\left(\frac{2}{5}\right)^{4}\left(\frac{3}{5}\right)^{4-4} = 0.0256$.

b) $p(\text{no boys}) = p(x=0) = \binom{4}{0}\left(\frac{2}{5}\right)^{0}\left(\frac{3}{5}\right)^{4-0} = 0.1296$.

c) $p(\text{exactly one boy}) = p(x=1) = \binom{4}{1}\left(\frac{2}{5}\right)^{1}\left(\frac{3}{5}\right)^{4-1} = 0.3456$.

Example: A fair coin is tossed 5 times. Find the probability of

(a) Exactly two heads (b) No head.

SOLUTION: LET THE NUMBER OF HEADS BE A RANDOM VARIABLE x WHICH CAN TAKE VALUES 0, 1, 2, 3, 4 AND 5. THEN x IS

binomial variate with $p = \frac{1}{2}$ and $n = 5$.

The probability function of x is

$$f\left(x, 5, \frac{1}{2}\right) = \binom{5}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{5-x} \quad \text{for } x = 0, 1, 2, 3, 4, 5$$

a) $p\,(\text{exactly two heads}) = p\,(x = 2) = \binom{5}{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{5-2} = 0.3125$.

b) $p\,(\text{no heads}) = p\,(x = 0) = \binom{5}{2}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{5-0} = 0.03125$.

Example: Determine the binomial distribution for which mean is 4 and variance is 3.

Solution: Let x be a binomial variate with parameters n and p. Here, we have, $np = 4$ and $npq = 3$. Thus $\frac{npq}{np} = \frac{3}{4} \Rightarrow q = \frac{3}{4}$ and $p = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$. Then $n = \frac{4}{p} = \frac{4}{\frac{1}{4}} = 16$.

Hence, the binomial distribution is

$$f\left(x;16,\frac{1}{4}\right) = \begin{cases} \binom{16}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{16-x} & \text{for } x = 0,1,2,\ldots,16. \\ 0; & \text{otherwise} \end{cases}$$

<u>QUESTION:</u> A TRAFFIC CONTROL OFFICER REPORT THAT 75% OF THE

trucks passing through a check post are from Dhaka city. What is the probability that at least three of the next five trucks are from out of the city?

<u>Solution:</u> Let X be the number of trucks that pass through are from out of Dhaka city. The probability of such an event is then $p = 1 - 0.75 = 0.25 = \frac{1}{4}$.

$$p(X \geq 3) = \sum_{x=3}^{5} f\left(x;5,\frac{1}{4}\right) = f\left(3;5,\frac{1}{4}\right) + f\left(4;5,\frac{1}{4}\right) + f\left(5;5,\frac{1}{4}\right)$$

$$= {}^{5}C_{3}\left(\frac{1}{4}\right)^{3}\left(\frac{3}{4}\right)^{2} + {}^{5}C_{4}\left(\frac{1}{4}\right)^{4}\left(\frac{3}{4}\right)^{1} + {}^{5}C_{5}\left(\frac{1}{4}\right)^{5}\left(\frac{3}{4}\right)^{0}$$

$$= \frac{90}{1024} + \frac{15}{1024} + \frac{1}{1024} = 0.1035$$

Question: It is known that 75% of the mice inoculated with a serum are protected from a certain disease. If three mice are inoculated, what is the probability that at most two of the mice will be protected from the disease? Exactly 2 will contact the disease?

Solution: Let x be the number mice inoculated. Assuming that the inoculation of one mouse is independent of inoculation of the mice, $p = \frac{3}{4}$. Since $n = 3$

$$p(X \leq 2) = \sum_{x=0}^{2} f\left(x;3,\frac{3}{4}\right) = f\left(0;3,\frac{3}{4}\right) + f\left(1;3,\frac{3}{4}\right) + f\left(2;3,\frac{3}{4}\right)$$

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

$$= \frac{1}{64} + \frac{9}{64} + \frac{27}{64} = \frac{37}{64}$$

For the second part of the problem, $p = \frac{1}{4}$

$$p(X = 2) = {}^{3}C_{2}\left(\frac{3}{4}\right)^{2}\left(\frac{1}{4}\right)^{1} = \frac{9}{64}$$

<u>QUESTION:</u> THE PROBABILITY THAT A PERSON RECOVERS FROM A

rare blood disease is 0.4. If 15 people are known to have contacted this disease, what is the probability that: (i) At least 10 people survive; (ii) From 3 to 8 survive; (iii) Exactly 5 survives.

<u>Solution:</u> Let x be the number people surviving. Then, $p = 0.4, \quad q = 0.6$. Since $n = 15$

$$(i) \quad p(X \geq 10) = 1 - p(X < 10) = 1 - \sum_{x=0}^{9} f(x;15,0.4) = 1 - 0.9662 = 0.0338$$

(ii) $P(3 \le X \le 8) = \sum_{x=0}^{8} F(X;15,0.4) - \sum_{x=0}^{2} F(X;15,0.4) = 0.9050 - 0.0271 = 0.8779$

(iii) $p(X=5) = f(5;15,0.4) = {}^{15}C_5 (0.4)^5 (0.6)^{10} = 0.1859$

POISSON DISTRIBUTION

Definition: A discrete random variable x is said to have a Poisson distribution if its probability function is given by

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \\ 0; & \text{otherwise} \end{cases}$$

where, $e = 2.71828$ and λ is the parameter of the distribution which is the mean number of success and $\lambda = np$.

Theorem: If x is a poisson variate with parameter λ, then mean $=\lambda$ and variance $=\lambda$. Hence, mean and variance of poisson distribution are equal.

Examples:

- The number of cars passing a certain street in time t.
- The number of suicide reported in a particular day.
- The number of printing mistakes at each page of a book.
- The number of air accidents in some unit of time.

- The number of deaths from a disease such as heart attack or cancer or due to snake bites.
- The number of telephone calls received at a particular telephone exchange in certain time.
- The number of defective materials in a packing manufactured by a good concern.
- The number of fishes caught in a day in a certain city.
- The number of robbers caught on a given day in a certain city.

EXAMPLE: SUPPOSE THAT THE NUMBER OF EMERGENCY

patients in a given day at a certain hospital is a Poisson variable x with parameter $\lambda = 20$. What is the probability that in a given day there will be **(a)** 15 emergency patients; **(b)** At least 3 emergency patients; **(c)** More than 20 but less than 25 patients.

Solution: We know that, $f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$

Here, $\lambda = 20$, $\therefore f(x;20) = \begin{cases} \dfrac{e^{-20}(20)^{x}}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$

a) $p(15 \text{ emergency patients}) = p(x = 15) = \dfrac{e^{-20}(20)^{15}}{15!} = 0.0516$.

B) $P\,(\text{AT LEAST 3 PATIENTS}) = P\,(X \geq 3) = 1 - P\,(X < 3)$

$= 1 - p\,(x = 0) - p\,(x = 1) - p\,(x = 2)$

$$= 1 - \frac{e^{-20}\,(20)^{0}}{0!} - \frac{e^{-20}\,(20)^{1}}{1!} - \frac{e^{-20}\,(20)^{2}}{2!} = 1.$$

C) $p\,(20 < x < 25) = p\,(x = 21) + p\,(x = 22) + p\,(x = 23) + p\,(x = 24)$

$$= \frac{e^{-20}\,(20)^{21}}{21!} + \frac{e^{-20}\,(20)^{22}}{22!} + \frac{e^{-20}\,(20)^{23}}{23!} + \frac{e^{-20}\,(20)^{24}}{24!} = 0.2841.$$

EXAMPLE: IF THE PROBABILITY THAT A CAR ACCIDENT HAPPENS

is a very busy road in on hour is 0.001. If 2000 cars passed in one hour by the road, what is the probability that:

(a) Exactly 3; (b) more than 2 car accidents happened on that hour of the road.

Solution: We know that,

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$$

HERE,

$p = 0.001$, $n = 2000$. $\therefore \lambda = np = 2000 * 0.001 = 2$.

$$\therefore f(x; 2) = \begin{cases} \dfrac{e^{-2}(2)^x}{x!} & \text{for } x = 0, 1, 2, \ldots, \infty. \end{cases}$$

a) $p(\text{exactly 3 accidents}) = p(x = 3) = \dfrac{e^{-2}(2)^3}{3!} = 0.18$.

b) $p(\text{more than 2 accidents}) = p(x > 2) = 1 - p(x \le 2)$

$= 1 - p(x = 0) - p(x = 1) - p(x = 2)$

$$= 1 - \frac{e^{-2}(2)^0}{0!} - \frac{e^{-2}(2)^1}{1!} - \frac{e^{-2}(2)^2}{2!} = 0.323.$$

EXAMPLE: A FACTORY PRODUCES BLADES IN A PACKET OF 10.

The probability of a blade to be defective is 0.2%. Find the number of packets having two defective blades in a consignment of 10,000 packets.

Solution: We know that, $f(x;\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0,1,2,\ldots,\infty. \end{cases}$

Here, $p = 0.2\% = 0.002$, $n = 10$. $\therefore \lambda = np = 10 * 0.002 = 0.02$.

$$\therefore p(2 \text{ defective blades}) = p(x = 2) = \frac{e^{-0.02}(0.02)^2}{2!} = 0.000196.$$

Therefore, the total number of packets having two defective blades in a consignment of 10,000 packet is

$$10000 \times 0.000196 = 1.96 \quad 2.$$

Example: What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model, what is the probability that a page observed at random will contain at least three misprints?

SOLUTION: WE KNOW THAT,

$$f(x;\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x=0,1,2,\dots,\infty.\end{cases}$$

we have, $p=\dfrac{1}{100}=0.01$ (because there is only one mistake on the average in a page),

$n=100$. $\therefore \lambda=np=100\times0.01=1$.

$\therefore p(\text{at least 3 misprints})=p(x\geq3)=1-p(x<3)$

$=1-p(x=0)-p(x=1)-p(x=2)$

$$=1-\frac{e^{-1}(1)^{0}}{0!}-\frac{e^{-1}(1)^{1}}{1!}-\frac{e^{-1}(1)^{2}}{2!}=0.0803$$

Question: The average number of emergence patients in a given day in a private clinic is 10. Find the probability

THAT IN A SPECIFIED DAY , THE CLINIC WILL RECEIVE (I) 5 EMERGENCE PATIENTS (II) AT LEAST 3 EMERGENCE PATIENTS (III)

- between 5 to 10 emergence patients, if the number of patients is assumed to follow the Poisson distribution?

- **Solution:** Let x denote the number of patients. Thus x is a Poisson variate with mean $\lambda = 10$

(i) Here $X = 5$, so that, $p(X = 5) = f(5,10) = \frac{e^{-10}10^5}{5!} = 0.038$

(ii) For at least three cases,

$$p(X \geq 3) = 1 - p(X \leq 2) = 1 - \sum_{x=0}^{2} f(x,10) = 1 - 0.003 = .0997$$

(*iii*) HERE X lies between 5 and 10 conclusive, so that,

$$p(5 \le X \le 10) = \sum_{x=5}^{10} f(x,10) = \sum_{x=0}^{10} f(x,10) - \sum_{x=0}^{4} f(x,10) = 0.583 - 0.029 = 0.554$$

Question: Suppose that there are, on the average, four vehicles accidents per day on the Asian Highway running from Dhaka to Manikganj. What is the probability that on a given day in the highways:

(i) There is no vehicle accident? (ii)There are 3 or fewer accidents (iii) there are more than 3 accidents?

Solution: Let X denote the number of accidents. Thus X is a Poisson variate with mean $\lambda = 4$

(i) HERE $X=0$, SO THAT, $p(X=0)=f(0,4)=\frac{e^{-4}4^{0}}{0!}=0.018$

(ii) For at least three cases, $p(X\leq 3)=\sum_{x=0}^{3}f(x,4)=0.433$

(iii) For more than three, $p(X\geq 3)=1-\sum_{x=0}^{2}f(x,4)=1-0.238=0.762$

Question: Suppose that 300 misprints are distributed randomly throughout a book of 500 pages. Find the probability that a given page contains:

(i) no misprints;

(ii) exactly 2 misprints; (iii) 2 or more misprints; (iv) between 3 and 5 misprints inclusive.

Solution: Here $n = 300 \; and \; p = \frac{1}{500}$. Thus the mean of the distribution is $np = \frac{300}{500} = 0.6$

Let X denote the number of misprints. Thus X is a Poisson variate with mean $\lambda = 4$

(i) Here $X = 0$, so that, $p(X = 0) = f(0, 0.6) = \frac{e^{-0.6}(0.6)^0}{0!} = 0.5488$

(ii) Here $X = 2$, so that, $p(X = 2) = f(2, 0.6) = \frac{e^{-0.6}(0.6)^2}{2!} = 0.0988$

(iii) For 2 or more cases,

$$p(X \geq 2) = 1 - p(X \leq 1) = 1 - \sum_{x=0}^{1} f(x, 0.6) = 1 - 0.878 = 0.122$$

(IV) HERE X LIES BETWEEN 3 AND 5 CONCLUSIVE, SO THAT,

$$p(3 \le X \le 5) = \sum_{x=3}^{5} f(x,0.6) = \sum_{x=0}^{5} f(x,0.6) - \sum_{x=0}^{2} f(x,10) = 1 - 0.977 = 0.023$$

Normal Distribution

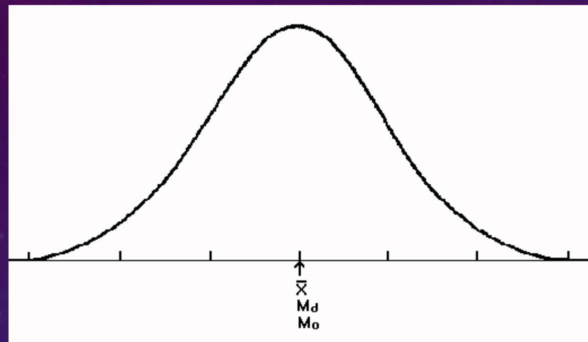
Definition: A continuous random variable X is said to have a normal distribution if its density function is given by

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

where, the parameters μ and σ^2 satisfy $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

The variable X whose density function given in (1) is called normal variate with parameters μ and σ^2 and is denoted by $N(\mu,\sigma^2)$. The parameters μ and σ^2 are actually

the mean and variance of the normal variate X. The graph of the normal curve is



Standard Normal variate: If X is a normal variate with parameters μ and σ^2, then $Z = \frac{X-\mu}{\sigma}$ is a standard normal variate with mean zero and variance unity. The density function of Z is

$$f(z, 0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}; \qquad -\infty < z < \infty$$

PROPERTIES OF NORMAL DISTRIBUTION:

- ❖The normal probability curve is symmetric about the ordinate at $x = \mu$, where μ is the mean of the distribution.
- ❖The maximum probability occurring at $x = \mu$ and is equal to $\sigma\sqrt{2\pi}$.
- ❖The total are under the curve and above the horizontal axis is equal to 1.
- ❖The parameters μ and σ^2 are respectively mean and
- standard deviation of the distribution.
- ❖The value of β_1 and β_2 are 0 and 3 respectively.
- ❖All odd moments of the distribution about the mean vanish.

❖THE MEAN, MEDIAN AND MODE OF THE DISTRIBUTION ARE ALL EQUAL TO μ.

❖The proportions of area lying between $\mu \pm \sigma,\ \mu \pm 2\sigma\ and\ \mu \pm 3\sigma$ are respectively 68.27%, 95.45% and 99.73%.

❖The mean deviation of normal curve is approximate $\frac{4}{5}\sigma$

Note: Let x be a continuous random variable with a cumulative distribution function $F(x)$ and let a and b be two possible values of x, with $a < b$. The probability that x lies between a and b is

$$p(a < x < b) = F(b) - F(a)$$

QUESTION: VERIFY THAT THE AREA UNDER THE NORMAL CURVE AND ABOVE THE X-AXIS IS EQUAL TO 1.

Proof: We have to show that $\int_{-\infty}^{\infty} f\left(x;\mu,\sigma^2\right)=1$

Now

$$\int_{-\infty}^{\infty} f\left(x;\mu,\sigma^2\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Let $y=\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$, so that $dy=\left(\frac{x-\mu}{\sigma^2}\right)dx \Rightarrow dx=\frac{\sigma}{\sqrt{2y}}dy$

Substituting in the previous equation we get,

$$\int_{-\infty}^{\infty} f\left(x;\mu,\sigma^2\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-y} \times \frac{\sigma}{\sqrt{2y}} dy$$

$$= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} y^{-\frac{1}{2}} e^{-y} dy = \frac{2}{2\sqrt{\pi}} \int_{0}^{\infty} y^{-\frac{1}{2}} e^{-y} dy = \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} y^{\frac{1}{2}-1} e^{-y} dy$$

This function $\int_{0}^{\infty} y^{\frac{1}{2}-1} e^{-y} dy$ is well-known gamma function,

whose values is $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Hence $\int_{-\infty}^{\infty} f\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{\pi}} \times \sqrt{\pi} = 1$. (Proved)

Example: A company produces light bulbs whose life time’s follow a normal distribution with mean 1200 hours and standard deviation 250 hours. If a light bulb is chosen randomly from the company’s output, what is the probability that its life time will be between 900 and 1300 hours?

SOLUTION: LET x REPRESENT LIFE TIME IN HOURS. THEN

$$p\,(900 < x < 1300) = p\left(\frac{900-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1300-\mu}{\sigma}\right)$$

$$= p\left(\frac{900-1200}{250} < z < \frac{1300-1200}{250}\right)$$

$$= p\,(-1.2 < z < 0.4)$$

$$= p\,(-\infty < z < 0.4) - p\,(-\infty < z < -1.2)$$

$$= 0.65542 - 0.11507$$ (By using Normal table)

$$= 0.54035$$

Hence, the probability is approximately 0.54 that a light bulb will last between 900 and 1300 hours.

EXAMPLE: A VERY LARGE GROUP OF STUDENTS OBTAINS TEST

scores that are normally distributed with mean 60 and standard deviation 15. What proportion of students obtained scores:

(a) Less than 85; (b) More than 90; (c) Between 85 and 95?

Solution: Let x denote the test score. Then

a) $p\left(x<80\right)=p\left(\frac{X-\mu}{\sigma}<\frac{85-\mu}{\sigma}\right)=p\left(z<\frac{85-60}{15}\right)$

$$=p\left(z<1.67\right)=p\left(-\infty<z<1.67\right)$$

$=0.9525$. (By using Normal table).

THAT IS 95.25% OF THE STUDENTS OBTAINED SCORES LESS THAN 80.

b) $p(x > 90) = p\left(\frac{X-\mu}{\sigma} > \frac{90-\mu}{\sigma}\right) = p\left(z > \frac{90-60}{15}\right)$

$= p(z > 2) = 1 - p(z < 2) = 1 - p(-\infty < z < 2)$

$= 1 - 0.9772 = 0.0228$. (By using Normal table).

That is 2.28% of the students obtained scores more than 90.

c) $p(85 < x < 95) = p\left(\frac{85-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{95-\mu}{\sigma}\right) == p\left(\frac{85-60}{15} < z < \frac{95-60}{15}\right)$

(BY USING NORMAL TABLE)

$= p(1.67 < z < 2.33) = p(-\infty < z < 2.33) - p(-\infty < z < 1.67)$

$= 0.9901 - 0.95254$

$= 0.03756$

That is 3.76% of the students obtained scores in the range 85 to 95.

Example: The average daily sales of 500 branch office were Tk. 150 thousands and the standard deviation Tk.

15 THOUSANDS. ASSUMING THE DISTRIBUTION TO BE NORMAL

indicate how many branches have sales between

(a) Tk. 120 thousands and Tk. 145 thousands. **(b)** Tk. 140 thousands and Tk. 165 thousands.

Solution: Let X be the average daily sales of 500 branch office.

a) $p(120 < x < 145) = p\left(\frac{120-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{145-\mu}{\sigma}\right)$

$$= p\left(\frac{120-150}{15} < z < \frac{145-150}{15}\right)$$

$$= p(-2 < z < -0.33)$$

$$= p(-\infty < z < -0.33) - p(-\infty < z < -2)$$

$$= 0.3707 - 0.02275 = 0.34795$$

(BY USING NORMAL TABLE)

Hence, the expected number of branches having sales between Tk. 120 thousands and Tk. 145 thousands are

$0.3479 \times 500 = 173.95 \;\; 174$.

b) $p(140 < x < 165) = p\left(\frac{140-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{165-\mu}{\sigma}\right)$

$$= p\left(\frac{140-150}{15} < z < \frac{165-150}{15}\right)$$

$= p\left(-0.67 < z < 1\right)$

$= p\left(-\infty < z < 1\right) - p\left(-\infty < z < -0.67\right)$

$= 0.84134 - 0.25143 = 0.58991$ (By using Normal table)

Hence, the expected number of branches having sales between Tk. 140 thousands and Tk. 165 thousands are

$0.58991 \times 500 = 294.955 \quad 295$.

Exponential distribution: The random variable x is said to have an exponential distribution with

IF X has a continuous distribution for which the $\lambda(\lambda > 0)$ probability density function is given by

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

<u>Question:</u> The time t second between the arrivals of successive vehicles at a particular road crossing has a probability density function

$$f(t) = 0.025e^{-0.025t}, \quad t \geq 0$$

A person, who take 20 seconds to cross the road, sets off as one vehicles passes. Find the probability that the

PERSON WILL COMPLETE THE CROSSING BEFORE THE NEXT VEHICLE ARRIVES.

Solution: Exponential distribution $f(x;\lambda)=\frac{1}{\lambda}e^{-\frac{x}{\lambda}}, x>0$. Here

$f(t)=0.025e^{-0.025t}, \quad t\geq 0$.

This is an exponential distribution with the parameter $\lambda=40$. The person will complete the crossing before the next vehicle arrive if $p(t>20)=0.025e^{-0.025\times 20}=0.6065$

Therefore the probability that the person completes the crossing before the next vehicles arrives is 0.6065.

Gamma distribution: The random variable x is said to have a gamma distribution with parameter $\lambda > 0$, if its probability density function is given by

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

Question: The daily consumption of milk in a city in excess of 20000 liters, is approximately distributed as a gamma variate with parameter $a = \dfrac{1}{1000}$ and $\lambda = 2$. The city

HAS A DAILY STOCK OF 30000 LITERS. WHAT IS THE PROBABILITY THAT THE STOCK IS INSUFFICIENT ON A PARTICULAR DAY?

Solution: Let the random variable X denote the daily consumption of milk (in liters) in a city. Then $Y = X - 20000$ has a gamma distribution with the probability density function

$$g(y) = \frac{1}{(10000)^2 \Gamma(2)} y^{2-1} e^{-\frac{y}{10000}} = \frac{y e^{-\frac{y}{10000}}}{(10000)^2}, \quad 0 < y < \infty$$

Since the daily stock of the city is 30000 liters, the probability that the stock is insufficient on a particular day is given by

$$p = p(X > 30000) = p(Y > 10000) = \int_{10000}^{\infty} g(y)dy = \int_{10000}^{\infty} \frac{ye^{-\frac{y}{10000}}}{(10000)^2} dy = \int_1^{\infty} ze^{-z}dz, \quad [z = y/10000]$$

Integrating by part we $p = \left[-ze^{-z}\right]_1^{\infty} + \int_1^{\infty} e^{-z}dz = e^{-1} - \left[e^{-z}\right]_1^{\infty} = e^{-1} + e^{-1} = \frac{2}{e}$

<u>Uniform distribution:</u> The random variable X is said to have a continuous uniform distribution over an interval

(a, b), i.e., $(-\infty < a < b < \infty)$, if its probability density function is given by

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & if \quad a < x < b \\ 0, & otherwise \end{cases}$$

Question: Subway trains on a certain line run every half hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes?

Solution: Let the random variable X denote the waiting time (in minutes) for the next train. Under the assumption that a man arrives at the station at random, X is distributed uniformly on $(0,30)$, with probability density function

$$f(x)=\begin{cases}\frac{1}{30}, & if \quad 0<x<30\\ 0, & otherwise\end{cases}$$

The probability that he has to wait at least 20 minutes is given by

$$p(X\geq 20)=\int_{20}^{30} f(x)\,dx=\int_{20}^{30}\frac{1}{30}\,dx=\frac{1}{30}[x]_{20}^{30}=\frac{30-20}{30}=\frac{1}{3}$$

Question: Arrivals of customers at a checkpoint follow uniform distribution. It is known that during a given 30-minutes period, one customer arrived at the counter. What is the probability that the customer arrived during the last 5 minutes of the 30-minutes period?

SOLUTION: WE ARE GIVEN THAT THE ACTUAL TIME OF ARRIVAL FOLLOWS A UNIFORM DISTRIBUTION OVER THE INTERVAL $(0,30)$.

Hence, if x denote the arrival time, then x is distributed uniformly with probability density function

$$f(x)=\begin{cases} \frac{1}{30}, & if \quad 0<x<30 \\ 0, & otherwise \end{cases}$$

The probability that the customer during the last arrived 5 minutes is

$$p(25 < X < 30) = \int_{25}^{30} f(x)\,dx = \int_{25}^{30} \frac{1}{30}\,dx = \frac{1}{30}[x]_{25}^{30} = \frac{30-25}{30} = \frac{1}{6}$$