MODULE-03

Statistics And Probability, basic data visualization, probability, common probability distributions: common probability mass functions, Bernoulli, binomial, poisson distributions, common probability density functions, uniform, normal, student's t-distribution.

Statistical computing: The statistical computing is defined as the bond between statistics and computer science is called Statistical computing

Statistics And Probability

- Statistics is the practice of turning *data* into *information* to identify trends and understand features of populations.
- Statistics is simply defined as the study and manipulation of data.
- Statistics is a branch of mathematics that involves collecting, analyzing, interpreting, presenting, and organizing data.
- It helps in understanding patterns, making predictions, and drawing conclusions from data by using various methods, tools, and techniques.

  Example: To find the mean of the marks obtained by each student in the class whose strength is 50.

  There are two main types of statistics

  1. **Descriptive Statistics:** It involves summarizing and describing the features of a dataset. Measures such as mean, median, mode, range, variance, and standard deviation fall under this category. Descriptive statistics help in organizing and presenting data to understand its main characteristics.

     Example:

     a real-life example of descriptive statistics in action:

     Imagine you're working for a company that sells smartphones. You've collected data on the battery life (in hours) of a new model of smartphones. Your goal is to use descriptive statistics to summarize and understand this dataset.

     Using descriptive statistics, you can analyze this data:

     Measures of Central Tendency:

     Mean: Calculate the average battery life.

     Median: Find the middle value when the data is arranged in order.

     Measures of Dispersion:

     Range: Determine the difference between the maximum and minimum battery life values.

Standard Deviation: Measure the spread of the data around the mean.

Visualization:

Create a histogram or boxplot to visualize the distribution of battery life values.

2. **Inferential Statistics:** This branch involves making inferences or predictions about a larger population based on a sample of data. It includes techniques like hypothesis testing, regression analysis, and confidence intervals. Inferential statistics helps in drawing conclusions or making predictions beyond the immediate data.

   Example:

   Inferential statistics involves making inferences or predictions about a larger population based on a sample of data. Here's a real-life example:

   Example: Customer Satisfaction Survey

   Imagine you're a manager at a retail store, and you want to assess the overall satisfaction of your customers regarding the quality of service. Instead of surveying every single customer (which might be impractical), you decide to use inferential statistics.

   Population: All customers who visit your store.

   Sample: You randomly select a sample of 200 customers and ask them to rate their satisfaction on a scale of 1 to 10.

Types of Statistics

The two main branches of statistics are:

- Descriptive Statistics
- Inferential Statistics

Descriptive Statistics – Through graphs or tables, or numerical calculations, descriptive statistics uses the data to provide descriptions of the population.

Inferential Statistics – Based on the data sample taken from the population, inferential statistics makes the predictions and inferences.

Characteristics of Statistics

The important characteristics of Statistics are as follows:

- Statistics are numerically expressed.
- It has an aggregate of facts
- Data are collected in systematic order

- It should be comparable to each other
- Data are collected for a planned purpose

Importance of Statistics

The important functions of statistics are:

- Statistics helps in gathering information about the appropriate quantitative data
- It depicts the complex data in graphical form, tabular form and in diagrammatic representation to understand it easily
- It provides the exact description and a better understanding
- It helps in designing the effective and proper planning of the statistical inquiry in any field
- It gives valid inferences with the reliability measures about the population parameters from the sample data
- It helps to understand the variability pattern through the quantitative observations

Variable in statistics: A *variable* is a characteristic of an individual in a population, the value of which can differ between entities within that population.

NUMERICAL VARIBLE: A *numeric* variable is one whose observations are naturally recorded as numbers.

There are two types of numeric variables:

1. Continuous and
2. Discrete

1.Continuous variable: A *continuous* variable can be recorded as any value in some interval, up to any number of decimals (which technically gives an infinite number of possible values, even if the continuum is restricted in range).

Example: if you were observing rainfall amount, a value of 15 mm would make sense, but so would a value of 15.42135 mm.Any degree of measurement precision gives a valid observation.

2.Discrete variable: A discrete variable, on the other hand, may take on only distinct numeric values—and if the range is restricted, then the number of possible values is finite.

Example: if you were observing the number of heads in 20 flips of a coin, only whole numbers would make sense. It would not make sense to observe 15.42135 heads; the possible outcomes are restricted to the integers from 0 to 20 (inclusive).

Categorical Variables: numeric observations are common for many variables, it's also important to consider categorical variables.

- Like some discrete variables, cate-gorical variables may take only one of a finite number of possibilities.
- Unlike discrete variables, however, categorical observations are not always recorded as numeric values.

There are two types of categorical variables.

1.Nominal

2.Ordinal

1. Nominal: Those that cannot be logically ranked are called nominal.

Example: A good example of a categorical-nominal variable is sex. In most data sets, it has two fixed possible values, male and female, and the order of these categories is irrelevant.

2.Ordinal: Categorical variables that can be naturally ranked are called ordinal.

Example: ordinal variable would be the dose of a drug, with the possible values low, medium, and high. These values can be ordered in either increasing or decreasing amounts, and the ordering might be relevant to the research.

Univariate and Multivariate Data

Univariate: When discussing or analyzing data related to only one dimension, you're dealing with *univariate* data.

Example, the weight variable in the earlier example is univariate since each measurement can be expressed with one component—a single number.

Multivariate Data

To consider data with respect to variables that exist in more than one dimension (in other words, with more than one component or measurement associated with each observation), your data are considered *multivariate*.

Population: defined as the entire collection of individuals or entities of interest.

Parameter or Statistic:

Parameter: The characteristics of that population are referred to as parameters.

Statistic: They may then estimate the parameters of interest using the sample data—and those estimates are the *statistics*.

Random variable: The random variable is any function that assign the numerical value at each possible outcome is called **Random variable**.

Types of Random variable

1.Continuous random variable:

- A discrete random variable can take only a finite number of distinct values such as 0, 1, 2, 3, 4, … and so on.

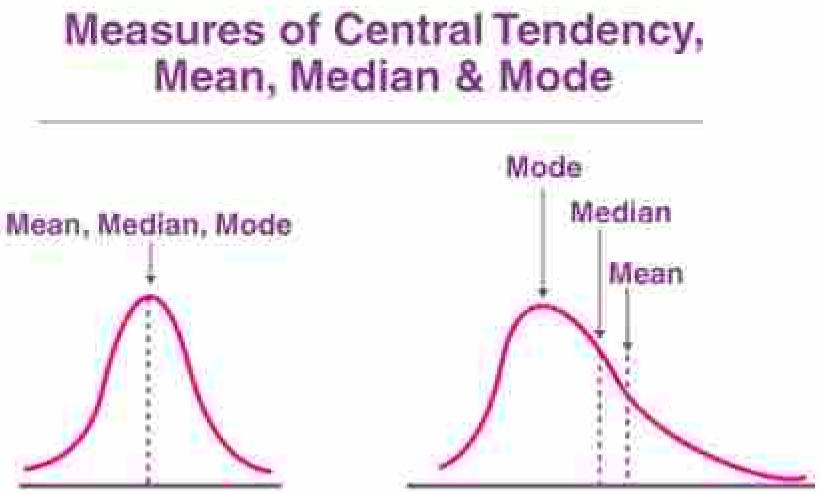
- The probability distribution of a random variable has a list of probabilities compared with each of its possible values known as probability mass function.

2.Discreate random variable:

- A numerically valued variable is said to be continuous if, in any unit of measurement, whenever it can take on the values a and b.
- If the random variable X can assume an infinite and uncountable set of values, it is said to be a continuous random variable.
- When X takes any value in a given interval (a, b), it is said to be a continuous random variable in that interval.

Example: For instance, the height of individuals, the time it takes for a computer program to execute, or the temperature measured at a specific time.

Centrality: Mean, Median, Mode



Measures of centrality are commonly used to explain large collections of data by describing where numeric observations are centered. One of the most common measures of centrality is of course the arithmetic *mean*. It's consid- ered to be the central "balance point" of a collection of observations.

Centrality measures in statistics describe the central tendency or typical value within a dataset. **The primary measures of centrality are the mean, median, and mode:**

1.Mean: The mean, often referred to as the average, is calculated by summing all values in a dataset and dividing the total by the number of values. It's sensitive to extreme values (outliers) and is suitable for symmetrically distributed data.

Mean= $\frac{\textit{sum of the observations}}{\textit{Number of observations}}$ $\frac{\sum_{i=1}^{n} x}{n}$

Example, if you observe the data 2, 4.4, 3, 3, 2, 2.2, 2, 4, the mean is calculated like this:

$$\frac{2 + 4.4 + 3 + 3 + 2 + 2.2 + 2 + 4}{8} = 2.825$$

2. Median: median represents the mid-value of the given set of data when arranged in a particular order.

- Given that the data collection is arranged in ascending or descending order, the following method is applied:
- If number of values or observations in the given data is odd, then the median is given by [(n+1)/2]th observation.
- If in the given data set, the number of values or observations is even, then the median is given by the average of (n/2)th and [(n/2) +1]th observation.

  Median=$\frac{(n/2)^{th} \text{ and } [(n/2)+1]^{th}}{2}$

Mode: The *mode* is simply the "most common" observation. This statistic is more often used with numeric-discrete data than with numeric-continuous

Example:

Consider the following dataset:

{5,8,12,8,5,6,8,5,9,8} {5,8,12,8,5,6,8,5,9,8}

To find the mode, we'll determine the value that appears most frequently in this dataset.

Step 1: Arrange the data in ascending order to see the pattern more clearly: {5,5,5,6,8,8,8,8,9,12} {5,5,5,6,8,8,8,8,9,12}

Step 2: Count how many times each value appears:

5 appears 3 times

6 appears 1 time

8 appears 4 times

9 appears 1 time

12 appears 1 time

Step 3: Identify the value(s) with the highest frequency:

In this case, the value 8 appears most frequently (4 times), so the mode of this dataset is 8.

Therefore, for the dataset {5, 8, 12, 8, 5, 6, 8, 5, 9, 8}, the mode is 8.

Prog1: Write R Program to summary descriptive statistics.

1. **Mean, Median, and Mode:**

Example:

```
# Generate a sample data
data <- c(12, 15, 18, 20, 22)
# Mean
mean(data)
```

Output: 17.4

```
# Median
median(data)
```

Output: 18

```
# Mode (using the 'DescTools' package)
install.packages("DescTools")
library(DescTools)
 Mode(data)
```

Output: [1] NA

```
attr(,"freq")
```

Output: [1] 1

Standard Deviation and Variance:

Example:

```
# Standard Deviation
sd(data)
```

output: 3.974921

```
# Variance
var(data)
```

output: 15.8

Summary Statistics:

Example:

```
summary(data)
```

Output:

```
Min. 1st Qu. Median    Mean 3rd Qu.    Max
  12.0    15.0    18.0   17.4    20.0    22.0
```

Counts, Percentages, and Proportions

Counts:

Counts simply refer to the number of occurrences or frequency of a particular event or observation in a dataset.

Example: the count of heads in ten coin flips, the count of people in different age groups, or the count of cars passing through a toll booth in an hour.

Percentages:

Percentages express a part of a whole as a fraction of 100. They are calculated by taking a part of the whole and expressing it as a proportion of 100.

Example: if 30 out of 100 students like chocolate, the percentage of students who like chocolate is$\frac{30}{100}$ *100 which equals 30%.

Proportions:

Proportions are ratios expressing the size or frequency of one group relative to the size of the entire dataset or another group.

They are calculated by dividing the count of a specific event by the total count.

Example: the proportion of red cars out of the total number of cars observed.

Quantiles, Percentiles, and the Five-Number Summary

Quantiles and percentiles are measures used to divide a dataset into equal parts. They help in understanding the distribution and spread of values within a dataset. The five-number summary is a concise statistical summary that includes the minimum, first quartile (Q1), median (second quartile, Q2), third quartile (Q3), and maximum of a dataset.

Quantiles: Quantiles divide a dataset into equal parts.

Example:

- **Median (Second Quartile, Q2):** Divides the data into two equal halves; 50% of the data falls below and 50% above this value.
- **Quartiles (Q1, Q2, Q3):** Divides the data into four equal parts, where Q1 represents the value below which 25% of the data falls, Q2 is the median, and Q3 is the value below which 75% of the data falls.

Percentiles:

Percentiles are a type of quantile that divides the data into 100 equal parts. For example:

- **Median (50th percentile):** Same as the median, dividing the data into two equal parts.
- **25th percentile (First Quartile, Q1):** Divides the data into four parts with 25% of the data falling below.
- **75th percentile (Third Quartile, Q3):** Divides the data into four parts with 75% of the data falling below.

Five-Number Summary

The five-number summary includes the minimum, maximum, median, and quartiles (Q1 and Q3) of a dataset. It provides a quick overview of the data's center, spread, and outliers.

- **Minimum:** The smallest value in the dataset.
- **Q1 (First Quartile):** The value below which 25% of the data falls.
- **Median (Second Quartile):** The middle value of the dataset.
- **Q3 (Third Quartile):** The value below which 75% of the data falls.
- **Maximum:** The largest value in the dataset.

The five-number summary is often used in box plots to visually represent the distribution of the data.

Variance, Standard Deviation, and the Interquartile Range

Variance:

- Variance measures the spread or dispersion of a dataset around its mean. It quantifies how much the values in a dataset differ from the mean value.
- It's calculated by averaging the squared differences between each data point and the mean.
- According to layman's words, 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'

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

Standard deviations:

Standard Deviation=$\sqrt{Variance}$

- **Formula for Variance (Population variance):**

	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

Variance Formula:

The **population variance** formula is given by:

σ^2 = Population variance

N = Number of observations in population

X_i = ith observation in the population

μ = Population mean

The **sample variance** formula is given as:

s^2 = Sample variance

n = Number of observations in sample

x_i = ith observation in the sample

$\bar{X}$= Sample mean

Standard Deviation Formula

The population standard deviation formula is given as:

σ = Population standard deviation

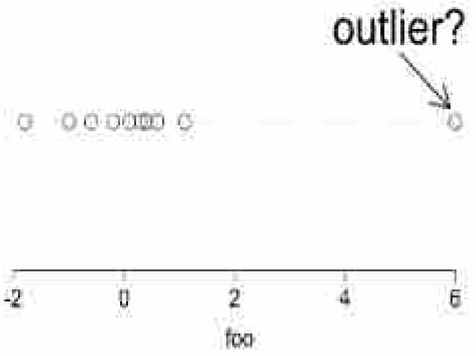
Similarly, the sample standard deviation formula is:

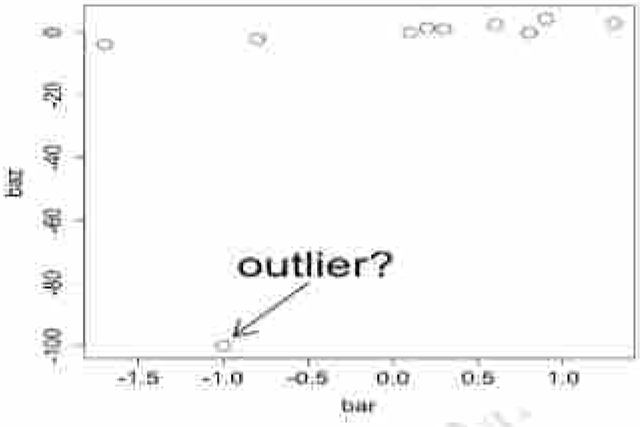
s = Sample standard deviation

Outliers

An outlier is an observation that does not appear to "fit" with the rest of the data. It is a noticeably extreme value when compared with the bulk of the data, in other words, an anomaly.

Example:





BASIC DATA VISUALIZATION

Data visualization is an important part of a statistical analysis. The visualization tools appropriate for a given data set are dependent upon the types of variables.

- The pictorial representation of the data is called data visualization.
- Data visualization is a graphical representation of quantitative information and data by using visual elements like graphs, charts, and maps.
- Data visualization convert large and small data sets into visuals, which is easy to understand and process for humans.

Bar plots and Pie charts

Bar plots and pie charts are commonly used to visualize qualitative data by category frequency

1.Bar plots:

A barplot draws either vertical or horizontal bars, typically separated by white space, to visualize frequencies according to the relevant categories.

There are two types of bar plots- horizontal and vertical which represent data points as horizontal or vertical bars of certain lengths proportional to the value of the data item.

- A bar chart is a pictorial representation in which numerical values of variables are represented by length or height of lines or rectangles of equal width.
- A bar chart is used for summarizing a set of categorical data.
- In bar chart, the data is shown through rectangular bars having the length of the bar proportional to the value of the variable.

syntax:

barplot(h,x,y,main, names.arg,col)

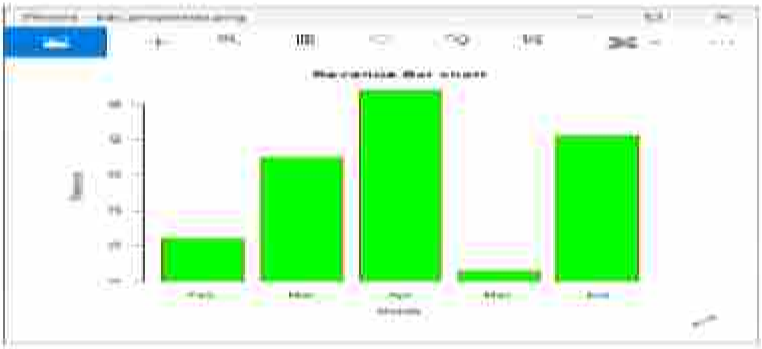
S.No	Parameter	Description
1.	H	A vector or matrix which contains numeric values used in the bar chart.
2.	xlab	A label for the x-axis.
3.	ylab	A label for the y-axis.
4.	main	A title of the bar chart.
5.	names.arg	A vector of names that appear under each bar.
6.	col	It is used to give colors to the bars in the graph.

```
# Creating the data for Bar chart
H <- c(12,35,54,3,41)
M<- c("Feb","Mar","Apr","May","Jun")
```

Example:

```
# Giving the chart file a name
png(file = "bar_properties.png")

# Plotting the bar chart
  barplot(H,names.arg=M,xlab="Month",ylab="Revenue",col="Green", main="Revenue Bar
        chart",border="red")

# Saving the file
dev.off()
```

Output:



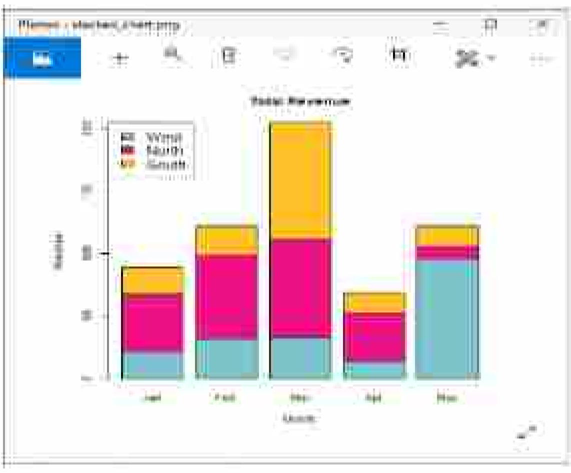
Group Bar Chart & Stacked Bar Chart

We can create bar charts with groups of bars and stacks using matrices as input values in each bar. One or more variables are represented as a matrix that is usedto construct group bar charts and stacked bar charts.

Example:

```
months <- c("Jan","Feb","Mar","Apr","May") regions
<- c("West","North","South")

# Creating the matrix of the values

Values <- matrix(c(21,32,33,14,95,46,67,78,39,11,22,23,94,15,16), nrow = 3,
ncol = 5, byrow = TRUE)

 # Giving the chart file a name
 png(file = "stacked_chart.png")

 # Creating the bar chart
barplot(Values, main = "Total Revenue", names.arg = months, xlab = "Month", ylab =
"Revenue", ccol =c("cadetblue3","deeppink2","goldenrod1"))

# Adding the legend to the chart
legend("topleft",regions,cex =1.3,fill=c("cadetblue3","deeppink2","goldenrod1"))
 # Saving the file
 dev.off()
```

Output:



R Pie Charts

A pie-chart is a representation of values in the form of slices of a circle with different colors. Slices are labeled with a description, and the numbers corresponding to each slice are also shown in the chart.

The Pie charts are created with the help of pie () function, which takes positive numbers as vector input.

Syntax:

pie(X, Labels, Radius, Main, Col, Clockwise)

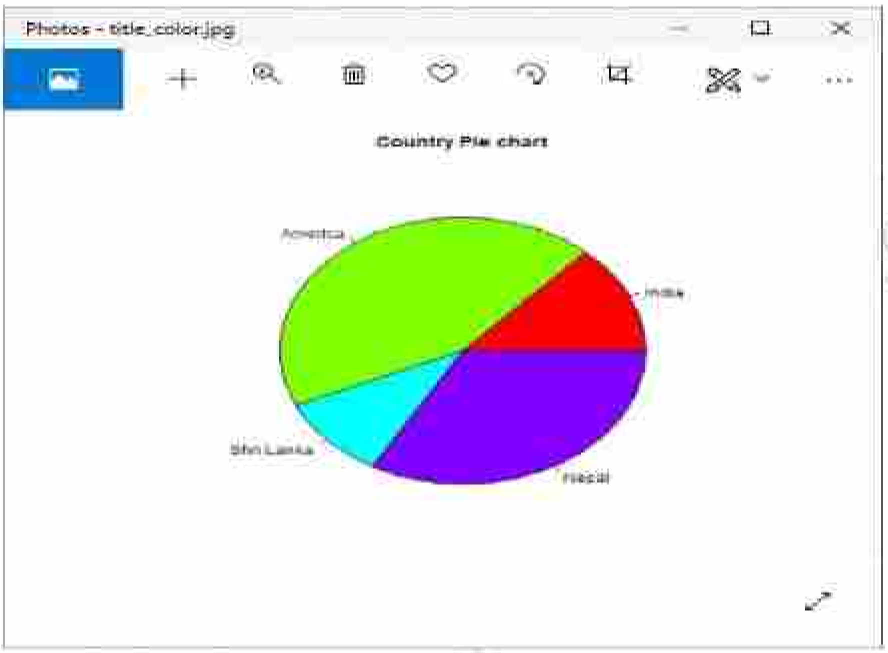
S.No	Parameter	Description
1.	**X**	is a vector that contains the numeric values used in the pie chart.
2.	**Labels**	are used to give the description to the slices.
3.	**Radius**	describes the radius of the pie chart.
4.	**Main**	describes the title of the chart.
5.	**Col**	defines the colour palette.
6.	**Clockwise**	is a logical value that indicates the clockwise or anti-clockwise direction in which slices are drawn.

Example:

```
 # Creating data for the graph.
 x <- c(20, 65, 15, 50)
labels <- c("India", "America", "Shri Lanka", "Nepal")
# Giving the chart file a name.
png(file = "title_color.jpg")
```

```
# Plotting the chart.
pie(x,labels,main="Country Pie chart",col=rainbow(length(x)))
 # Saving the file.
 dev.off()
```

Output:



R Histogram

A histogram is a type of bar chart which shows the frequency of the number ofvalues which are compared with a set of values ranges.

The histogram is used for the distribution, whereas a bar chart is used forcomparing different entities.

In the histogram, each bar represents the height of the number of values present in the given range.

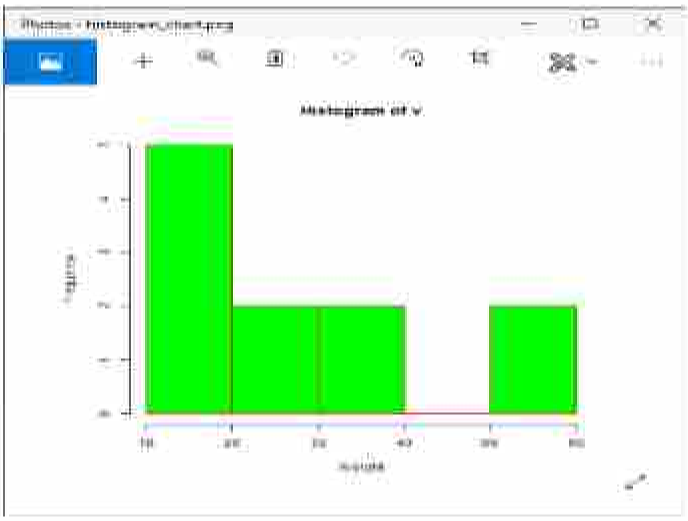
For creating a histogram, R provides hist() function, which takes a vector as aninput.

Syntax:

```
hist(v,main,xlab,ylab,xlim,ylim,breaks,col,border)
```

S.No	Parameter	Description
1.	v	It is a vector that contains numeric values.
2.	main	It indicates the title of the chart.
3.	col	It is used to set the color of the bars.
4.	border	It is used to set the border color of each bar.
5.	xlab	It is used to describe the x-axis.
6.	ylab	It is used to describe the y-axis.
7.	xlim	It is used to specify the range of values on the x-axis.
8.	ylim	It is used to specify the range of values on the y-axis.
9.	breaks	It is used to mention the width of each bar.

```
# Creating data for the graph.
v <- c(12,24,16,38,21,13,55,17,39,10,60)
# Giving a name to the chart file.
png(file = "histogram_chart.png")
# Creating the histogram.
hist(v,xlab = "Weight",ylab="Frequency",col = "green",border = "red")
# Saving the file.
dev.off()
```



R Boxplot

Boxplots are a measure of how well data is distributed across a data set. This divides the data set into three quartiles. This graph represents the minimum, maximum, average.

Boxplot is also useful in comparing the distribution of data in a data set by drawing a boxplot for each of them.

R provides a boxplot() function to create a boxplot.

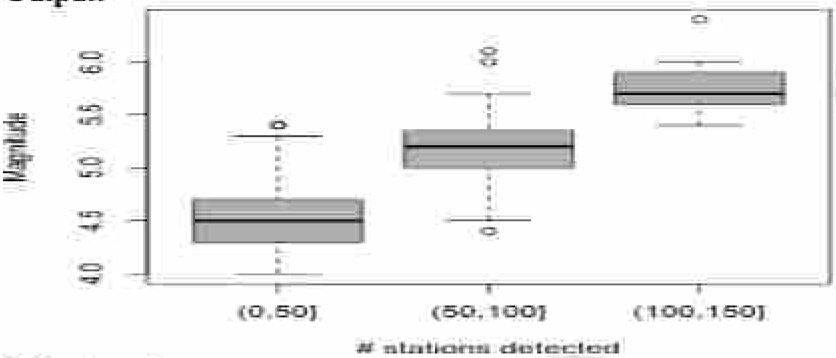
Syntax:

boxplot(x, data, notch, varwidth, names, main)

S.No	Parameter	Description
1.	x	It is a vector or a formula.
2.	data	It is the data frame.
3.	notch	It is a logical value set as true to draw a notch.
4.	varwidth	It is also a logical value set as true to draw the width of the box same as the sample size.
5.	names	It is the group of labels that will be printed under each boxplot.
6.	main	It is used to give a title to the graph.

```
# Giving a name to the chart file.

png(file = "boxplot.png")

# Plotting the chart.

  boxplot(mpg ~ cyl, data =mtcars, xlab = "Quantity of Cylinders", ylab =

          "Miles Per Gallon", main = "R Boxplot Example")

    # Save the file.

dev.off()
```

Output:



R Scatterplots

The scatter plots are used to compare variables. A comparison between variables is required when we need to define how much one variable is affected by another variable.

In a scatterplot, the data is represented as a collection of points. Each point on the scatterplot defines the values of the two variables.

One variable is selected for the vertical axis and other for the horizontal axis.

Syntax:

plot(x, y, main, xlab, ylab, xlim, ylim, axes)

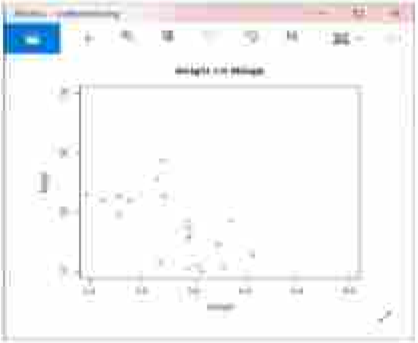
S.No	Parameters	Description
1.	x	It is the dataset whose values are the horizontal coordinates.
2.	y	It is the dataset whose values are the vertical coordinates.

3.	main	It is the title of the graph.
4.	xlab	It is the label on the horizontal axis.
5.	ylab	It is the label on the vertical axis.
6.	xlim	It is the limits of the x values which is used for plotting.
7.	ylim	It is the limits of the values of y, which is used for plotting.
8.	axes	It indicates whether both axes should be drawn on the plot.

Example: In our example, we will use the dataset "mtcars", which is thepredefined dataset available in the R environment.

```
#Fetching two columns from mtcars

data <-mtcars[,c('wt','mpg')]

# Giving a name to the chart file.

png(file = "scatterplot.png")

# Plotting the chart for cars with weight between 2.5 to 5 and mileage between 15 and 30.

plot(x =data$wt,y = data$mpg, xlab = "Weight", ylab = "Milage", xlim =c(2.5,5),ylim = c(15,30), main ="Weight v/sMilage")
# Saving the file.dev.off()
```

Output:



Advantages of Data Visualization in R:

R has the following advantages over other tools for data visualization:

- R offers a broad collection of visualization libraries along with extensive online guidance on their usage.
- R also offers data visualization in the form of 3D models and multipanel charts.
- Through R, we can easily customize our data visualization by changing axes, fonts, legends, annotations, and labels.

Disadvantages of Data Visualization in R:

R also has the following disadvantages:

- R is only preferred for data visualization when done on an individual standalone server.
- Data visualization using R is slow for large amounts of data as compared to other counterparts.

PROBABILITY:

A probability is a number that describes the "magnitude of chance" associated with making a particular observation or statement.

It's always a number between 0 and 1 (inclusive) and is often expressed as a fraction. Exactly how you calculate a probability depends on the definition of an event.

Events and Probability

An event typically refers to a specific outcome that can occur.

To describe the chance of event A actually occurring,to use a probability, denoted by Pr(A). At the extremes, Pr(A) = 0 suggests *A* cannot occur,

Pr(*A*) = 1 suggests that *A* occurs with complete certainty.

Example: Let's say you roll a six-sided, fair die. Let *A* be the event "you roll a 5 or a 6." You can assume that each outcome on a standard die has a probability of occurring 1/6 in any given roll.

$$\Pr(A)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$$

Probability: Probability means possibility. It is a branch of mathematics that deals with the occurrence of a random event.

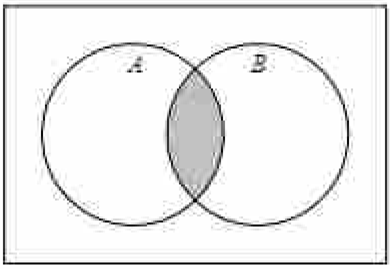
For example, when we toss a coin, either we get Head OR Tail, only two possible outcomes are possible (H, T). But when two coins are tossed then there will be four possible outcomes, i.e {(H, H), (H, T), (T, H), (T, T)}.

Conditional Probability

The conditional probability of an event B, assuming that the event A has happened.

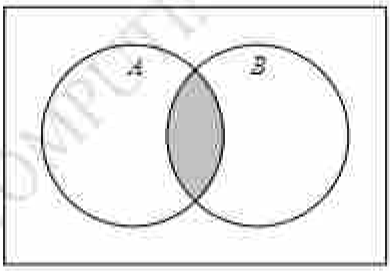
$P(B/A) = \frac{P(A \cap B)}{P(A)}, provided P(A) \neq 0$

$P(A/B) = \frac{P(A \cap B)}{P(B)}, provided P(B) \neq 0$



Intersection

The intersection of two events is written as Pr(A∩B) and is read as "the probability that both A and B occur simultaneously." It is common to represent this as a Venn diagram, as shown here:



Here, the disc labeled A represents the outcome (or outcomes) that satisfies A, and disc B represents the outcomes for B.

$Pr(A \cap B) = Pr(A|B) \cdot Pr(B)$ or $Pr(B|A) \cdot Pr(A)$

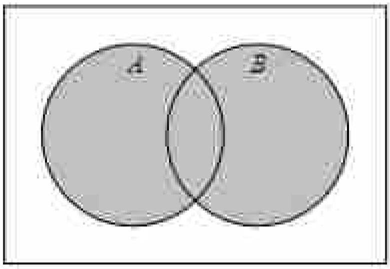
to the die example, what is the probability that on a single toss you roll an even number *and* it's a 4 or more Using the fact that $Pr(A|B) = 2/3$ and that $Pr(B) = 1/2$it is easy to compute

$Pr(A \cap B) = 2/3*1/2=1/3$

```
R> (2/3)*(1/2)
[1] 0.3333333
```

Union

The union of two events is written as Pr(AUB) and is read as "the probability that A or B occurs." Here is the representation of a union as a Venn diagram:



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

If two sets A and B are given, then the union of A and B is equal to the set that contains all the elements present in set A and set B. This operation can be represented as;

A ∪ B = {x: x ∈ A or x ∈ B}

Where x is the elements present in both sets A and B.

Example: If set A = {1,2,3,4} and B {6,7}

Then, Union of sets, A ∪ B = {1,2,3,4,6,7}

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

To subtract the intersection in this diagram is that in summing Pr(A) and Pr(B) alone, you'd be incorrectly counting Pr($A \cup B$) twice.

to find that $\Pr(A \cup B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3}$

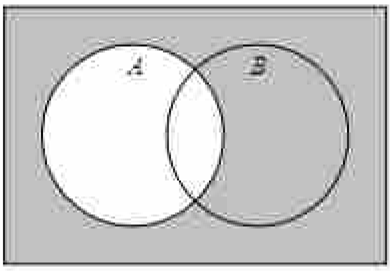
In R progam: (1/2)+(1/2)-(1/3)

```
[1] 0.6666667
```

Complement

The probability of the *complement* of an event is written as Pr($\bar{A}$) and is read as "the probability that A does *not* occur."

Here it is as a Venn diagram:



From this diagram, you can see the following:

$\Pr(\bar{A}) = 1 - \Pr(A)$

Wrapping up the running example, it's straightforward to find the probability that you do not roll a 4 or greater: $\Pr(\bar{A})$ = 1-1/2=1/2

Random Variables and Probability Distributions

In probability, a random variable is a real valued function whose domain is the sample space of the random experiment.

- A *random variable* is a variable whose specific outcomes are assumed to ariseby chance or according to some random or *stochastic* mechanism
- It means that each outcome of a random experiment is associated with a single real number, and the single real number may vary with the different outcomes of a random experiment. Hence, it is called a random variable and it is generally represented by the letter "X".

For example, let us consider an experiment for tossing a coin two times.

Hence, the sample space for this experiment is S = {HH, HT, TH, TT}

If X is a random variable and it denotes the number of heads obtained, then the values are represented as follows:

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

Similarly, we can define the number of tails obtained using another variable, say Y.

(i.e) Y(HH) = 0, Y(HT) = 1, Y(TH) = 1, Y(TT)= 2.

Two types of random variables.

1. discrete random variable
2. Continuous random variable

1. Discrete random variable: A discrete random variable can be defined as a type of variable whose value depends upon the numerical outcomes of a certain random phenomenon. It is also known as a stochastic variable.

Discrete random variables are always whole numbers, which are easily countable. A probability mass function is used to describe the probability distribution of a discrete random variable.

Example: Suppose 2 dice are rolled and the random variable, X, is used to represent the sum of the numbers. Then, the smallest value of X will be equal to 2, which is a result of the outcomes 1 + 1 = 2, and the highest value would be 12, which is resulting from the outcomes

6 + 6 = 12. Thus, X could take on any value between 2 to 12 (inclusive).

Mean Of Discrete Random Variable

The average value of a random variable is called the mean of a random variable. The mean is also known as the expected value. It is generally denoted by E[X], where X is the random variable.

Mean of a Discrete Random Variable: $E[X] = \sum xP(X=x)$

Here $P(X = x)$ is the probability mass function.

Variance Of Discrete Random Variable

The variance of a random variable can be defined as the expected value of the square of the difference of the random variable from the mean. The variance of a random variable is given by Var[X] or σ22. If μ is the mean then the formula for the variance is given as follows:

- Variance of a Discrete Random Variable: $Var[X] = \sum(x - \mu)^2P(X=x)$

Discrete Random Variable – Types

A discrete random variable is a variable that can take on a finite number of distinct values. For example, the number of children in a family can be represented using a discrete random variable.

1. Bernoulli Random Variable

A Bernoulli random variable is the simplest type of random variable. It can take only two possible values, i.e., 1 to represent a success and 0 to represent a failure.

A Bernoulli random variable is given by X~Bernoulli(p), where p represents the success probability.

Probability mass function: $P(X = x) = \begin{cases} p \ if \ x = 1 \\ 1 - p \ if \ x = 0 \end{cases}$

2. Binomial Random Variable

A random variable that represents the number of successes in a binomial experiment is known as a binomial random variable.

A binomial experiment has a fixed number of repeated Bernoulli trials and can only have two outcomes, i.e., success or failure.

The number of trials is given by n and the success probability is represented by p.

A binomial random variable, X, is written as X~Bin(n,p)

The probability mass function is given as P(X=x)=$\binom{n}{x}P^{x}q^{n-x}$

3. Poisson Random Variable

A Poisson random variable is used to show how many times an event will occur within a given time period. These events occur independently and at a constant rate. The parameter of a Poisson distribution is given by λ which is always greater than 0.

A Poisson random variable is represented as X~Poisson(λ)

The probability mass function is given by $P(X = x) = \frac{\lambda^{X} e^{x}}{x!}$

Continuous Random Variable

- Continuous random variable is a random variable that can take on a continuum of values. In other words, a random variable is said to be continuous if it assumes a value that falls between a particular interval.
- Continuous random variables are used to denote measurements such as height, weight, time, etc.

According to the definition A continuous random variable can be defined as a random variable that can take on an infinite number of possible values.

the probability that a continuous random variable will take on an exact value is 0.

- The cumulative distribution function and the probability density function are used to describe the characteristics of a continuous random variable.

Example

Suppose the probability density function of a continuous random variable, X, is given by $4x^3$, where x ∈ [0, 1]. The probability that X takes on a value between 1/2 and 1 needs to be determined. This can be done by integrating $4x^3$ between 1/2 and 1. Thus, the required probability is 15/16.

PDF of Continuous Random Variable

The probability density function of a continuous random variable can be defined as a function that gives the probability that the value of the random variable will fall between a range of values.

Let X be the continuous random variable, then the formula for the pdf, f(x), is given as follows:

$$f(x) = \frac{dF[x]}{dy} = F^1(x)$$

where, F(x) is the cumulative distribution function.

For the pdf of a continuous random variable to be valid, it must satisfy the following conditions:

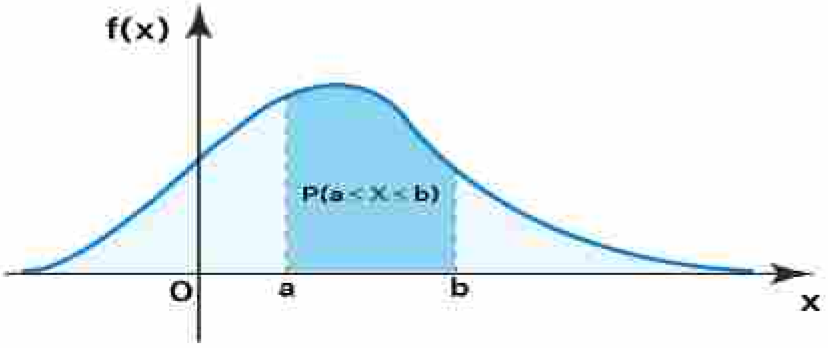
- ∫∞−∞f(x)dx=1 This means that the total area under the graph of the pdf must be equal to 1.
- f(x) $\geq$ 0. This implies that the probability density function of a continuous random variable cannot be negative.

CDF of Continuous Random Variable

The cumulative distribution function of a continuous random variable can be determined by integrating the probability density function.

- It can be defined as the probability that the random variable, X, will take on a value that is lesser than or equal to a particular value, x.
- The formula for the cdf of a continuous random variable, evaluated between two points a and b, is given below:

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\,dx$$



Mean of Continuous Random Variable

The mean of a continuous random variable can be defined as the weighted average value of the random variable, X. It is also known as the expectation of the continuous random variable. The formula is given as follows:

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

Variance of Continuous Random Variable

The variance of a continuous random variable can be defined as the expectation of the squared differences from the mean.

It helps to determine the dispersion in the distribution of the continuous random variable with respect to the mean.

The formula is given as follows:

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

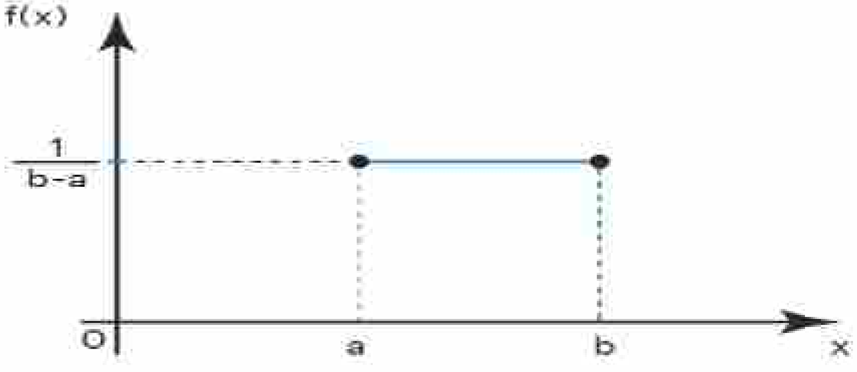
Continuous Random Variable Types

A continuous random variable is usually used to model situations that involve measurements. For example, the possible values of the temperature on any given day.

Uniform Random Variable/distribution

A continuous random variable that is used to describe a uniform distribution is known as a uniform random variable. Such a distribution describes events that are equally likely to occur. The pdf of a uniform random variable is as follows:

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



Normal Random Variable

A continuous random variable that is used to model a normal distribution is known as a normal random variable. If the parameters of a normal distribution are given as X~N(μ,σ2) then the formula for the pdf is given as follows:

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

where,

μ = mean

σ = standard deviation

σ^2 = variance.

Continuous Random Variable	**Discrete Random Variable**
The value of a continuous random variable falls between a range of values.	The value of a discrete random variable is an exact value.
The probability density function is associated with a continuous random variable.	The probability mass function is used to describe a discrete random variable.
A continuous random variable can take on an infinite number of values.	Such a variable can take on a finite number of distinct values.
Mean of a continuous random variable is E[X] = ∫∞−∞xf(x)dx∫	The mean of a discrete random variable is E[X] = ∑ x P(X = x), where P(X = x) is the probability mass function.
The variance of a continuous random variable is Var(X) = ∫∞−∞(x−μ)2f(x)dx	The variance of a discrete random variable is Var[X] = $\sum(x - \mu)^2 P(X = x)$

Continuous Random Variable	Discrete Random Variable
The examples of a continuous random variable are uniform random variable, exponential random variable, normal random variable, and standard normal random variable.	The examples of a discrete random variable are binomial random variable, geometric random variable, Bernoulli random variable, and Poisson random variable.

Common probability distributions:

A probability distribution specifies the probabilities of the possible outcomes of a random variable. The two basic types of random variables are discrete random variables and continuous random variables. Discrete random variables take on at most a countable number of possible outcomes

Two types of **common probability distribution**

1. Probability mass function(PMF)

2. Probability density function(PDF)

1. Probability mass function(PMF):

- Probability mass function gives the probability that a discrete random variable will be exactly equal to a specific value.
- The probability mass function is only used for discrete random variables.
- For continuous random variables, the probability density function is used which is analogous to the probability mass function.
- The probability mass function provides all possible values of a discrete random variable as well as the probabilities associated with it.
- Let X be the discrete random variable. Then the formula for the probability mass function, f(x), evaluated at x, is given as follows:

  $f(x) = P(X = x)$

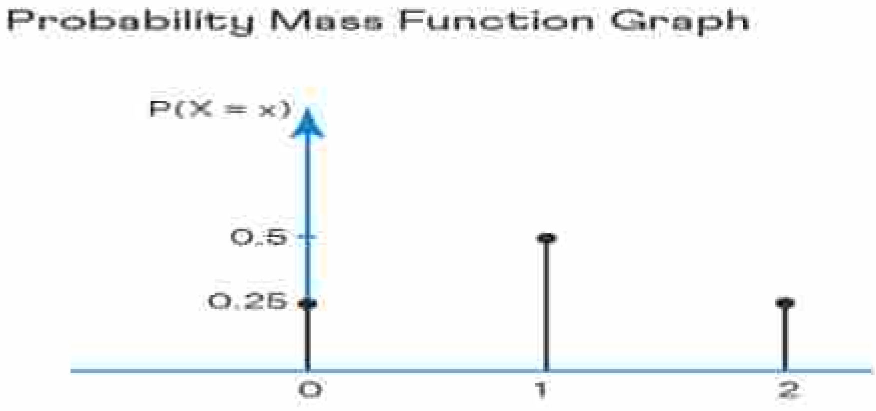
Probability Mass Function Properties

- $P(X = x) = f(x) > 0$. This implies that for every element x associated with a sample space, all probabilities must be positive.
- $\sum x\epsilon Sf(x)=1$. The sum of all probabilities associated with x values of a discrete random variable will be equal to 1.

- P(X ∈ T) = ∑x∈Tf(x) The probability associated with an event T can be determined by adding all the probabilities of the x values in T. This property is used to find the CDF of the discrete random variable.
  - the number of heads in the coin tosses. The sample space created is [HH, TH, HT, TT]. This shows that X can take the values 0 (no heads), 1 (1 head), and 2 (2 heads). The probabilities of each outcome can be calculated by dividing the number of favorable outcomes by the total number of outcomes. This gives us the following probabilities.
    P(X = 0) = 1 / 4 = 0.25
    P(X = 1) = 2 / 4 = 0.5
    P(X = 2) = 1 / 4 = 0.25
  - These values can be presented as given below.
  - Probability Mass Function Table
  - A probability mass function table displays the various values that can be taken up by the discrete random variable as well as the associated probabilities. The pmf table of the coin toss example can be written as follows:

x	P(X = x)
0	0.25
1	0.5
2	0.25

  - Thus, probability mass function P(X = 0) gives the probability of X being equal to 0 as 0.25



The cumulative distribution:

In **cumulative distribution** function, the probability function value of a continuous random variable is less than or equal to the argument of the function.

The cumulative distribution function of a discrete random variable is given by the formula F(x) = P(X ≤ x).

4. **Bernoulli distribution**

   Probability mass function: P(X = x) = $\begin{cases} p \; if \; x = 1 \\ 1 - p \; if \; x = 0 \end{cases}$

5. **Binomial distribution**

   Probability mass function P(X=x)=$\binom{n}{x}P^{x}q^{n-x}$

6. **Poisson distribution**

   The probability mass function is given by P(X = x) = $\frac{\lambda^{x}e^{x}}{x!}$

- **Example 2:** The probability mass function table for a random variable X is given as follows:

x	0	1	2	3	4
P(X = x)	0	0.1	0.2	0.3	0.4

- Find the value of the CDF, P(X ≤ 2).

  Solution: P(X ≤ 2), can be computed by using the pmf property P(X ∈ T) = ∑x∈Tf(x)

  P(X ≤ 2) = P(X = 0) + P(X = 1) + P(X = 2)

  = 0 + 0.1 + 0.2

  = 0.3

  Answer: P(X ≤ 2) = 0.3

Probability Density Function

Probability density function defines the density of the probability that a continuous random variable will lie within a particular range of values.

To determine this probability, we integrate the probability density function between two specified points.

Example

Say we have a continuous random variable whose probability density function is given by $f(x) = x + 2$, when $0 < x \leq 2$.

We want to find $P(0.5 < X < 1)$. Then we integrate $x + 2$ within the limits 0.5 and 1. This gives us 1.375.

Thus, the probability that the continuous random variable lies between 0.5 and 1 is 1.375.

Probability Density Function of Continuous Random Variable

Let X be the continuous random variable, then the formula for the pdf, f(x), is given as follows:

$$f(x) = \frac{dF[x]}{dy} = F^{1}(x)$$

where, F(x) is the cumulative distribution function.

For the pdf of a continuous random variable to be valid, it must satisfy the following conditions:

- $\int_{-\infty}^{\infty} f(x)$dx=1 This means that the total area under the graph of the pdf must be equal to 1.
- $f(x) \geq 0$. This implies that the probability density function of a continuous random variable cannot be negative.

The cumulative distribution:

- In **cumulative distribution** function, the probability function value of a continuous random variable is less than or equal to the argument of the function.
- to find the probability that X lies between lower limit 'a' and upper limit 'b' then using the probability density function this can be given as:

  $P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$

  Here, F(b) and F(a) represent the cumulative distribution function at b and a respectively.

7. **Normal distribution**

   Probabilitymassfunction:$f(x) = \frac{1}{\sigma 2\pi} e^{\frac{-1}{2}\frac{(x-\mu)^{2}}{\sigma}}$

2. **Uniform distribution**

   Probability mass function: $f(x)=\begin{cases} \frac{1}{b-a} \; if \; a < x < b \\ 0 \; otherwise \end{cases}$

Probability Mass Function	Probability Density Function
Probability mass function denotes the probability that a discrete random variable will take on a particular value.	Probability density function gives the probability that a continuous random variable will lie between a certain specified interval.
It is used for discrete random variables.	It is used for continuous random variables.
It is evaluated at an exact point.	It is evaluated between a range of values.
The formula for pmf is f(x) = P(X = x)	The formula for pdf is given as p(x) = dF(x)/dx= F'(x), where F(x) is the cumulative distribution function.
To determine the CDF, P(X ≤ x), the probability mass function needs to be summed up to x values.	To determine the CDF, P(X ≤ x), the probability density function needs to be integrated from -∞ to x.

Common Probability Mass Functions

1.Bernoulli Distribution

Bernoulli Distribution is a special case of Binomial distribution where only a single trial is performed. It is a discrete probability distribution for a Bernoulli trial (a trial that has only two outcomes i.e. either success or failure).

For example, In R it can be represented as a coin toss where the probability of getting the head is 0.5 and getting a tail is 0.5. It is a probability distribution of arandom variable that takes value 1 with probability p and the value 0 with probability q=1-p.

The probability mass function f of this distribution, over possible outcomes k, is given by :

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ q = 1 - p & \text{if } k = 0. \end{cases}$$

The mean and variance are defined as follows, respectively:

$$\mu_X = p \qquad \text{and} \qquad \sigma_X^2 = p(1-p)$$

Mean and Variance of Bernoulli Distribution

The arithmetic mean of a large number of independent realizations of the random variable X gives us the expected value or mean. The expected value can also be thought of as the weighted average. Given below is the proof and formula for the mean of a Bernoulli distribution.

Mean of Bernoulli Distribution Proof:

We know that for X,

P(X = 1) = p

P(X = 0) = q

E[X] = P(X = 1) . 1 + P(X = 0) . 0

E[X] = p . 1 + q . 0

E[X] = p

Thus, the mean or expected value of a Bernoulli distribution is given by E[X] = p.

Variance of Bernoulli Distribution Proof:

The variance can be defined as the difference of the mean of X^2 and the square of the mean of X. Mathematically this statement can be written as follows:

$Var[X] = E[X^2] - (E[X])^2$

Using the properties of $E[X^2]$, we get,

$E[X^2] = \sum x2P(X=x)$

$E[X^2] = 1^2 . p + 0^2 . q = p$

Substituting this value in $Var[X] = E[X^2] - (E[X])^2$ we have

$Var[X] = p - p^2$

= p(1 - p)

= p . q

Hence, the variance of a Bernoulli distribution is Var[X] = p(1 - p) = p . q

2) Binomial Distribution

Binomial distribution in R is a probability distribution used in statistics.

The binomial distribution is a discrete distribution and has only two outcomes i.e. success or failure. All its trials are independent, the probability of success remains the same and the previous outcome does not affect the next outcome.

The outcomes from different trials are independent. Binomial distribution helps us to find the individual probabilities as well as cumulative probabilities over a certain range.

In mathematical terms, for a discrete random variable X=x, the binomial mass function is

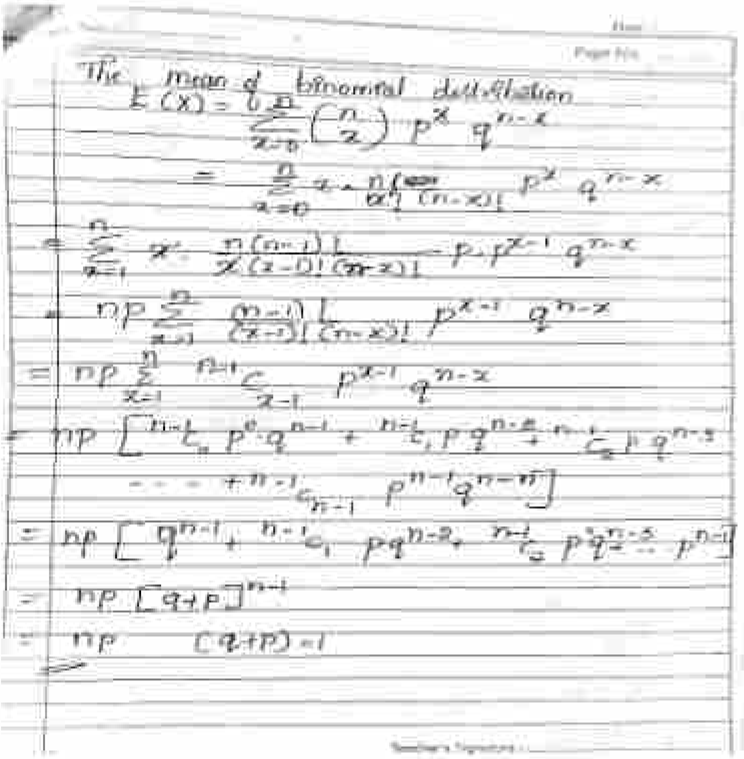
$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = \{0, 1, \ldots, n\}$$

where

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

The mean and variance are defined as follows:

$$\mu_X = np \quad \text{and} \quad \sigma_X^2 = np(1-p)$$



Date

Page No.

The mean of binomial distribution

$E(X) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$

$= \sum_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$= \sum_{x=1}^{n} x \cdot \frac{n(n-1)!}{x(x-1)!(n-x)!} p \cdot p^{x-1} q^{n-x}$

$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$

$= np \sum_{x=1}^{n} {}^{n-1}C_{x-1} p^{x-1} q^{n-x}$

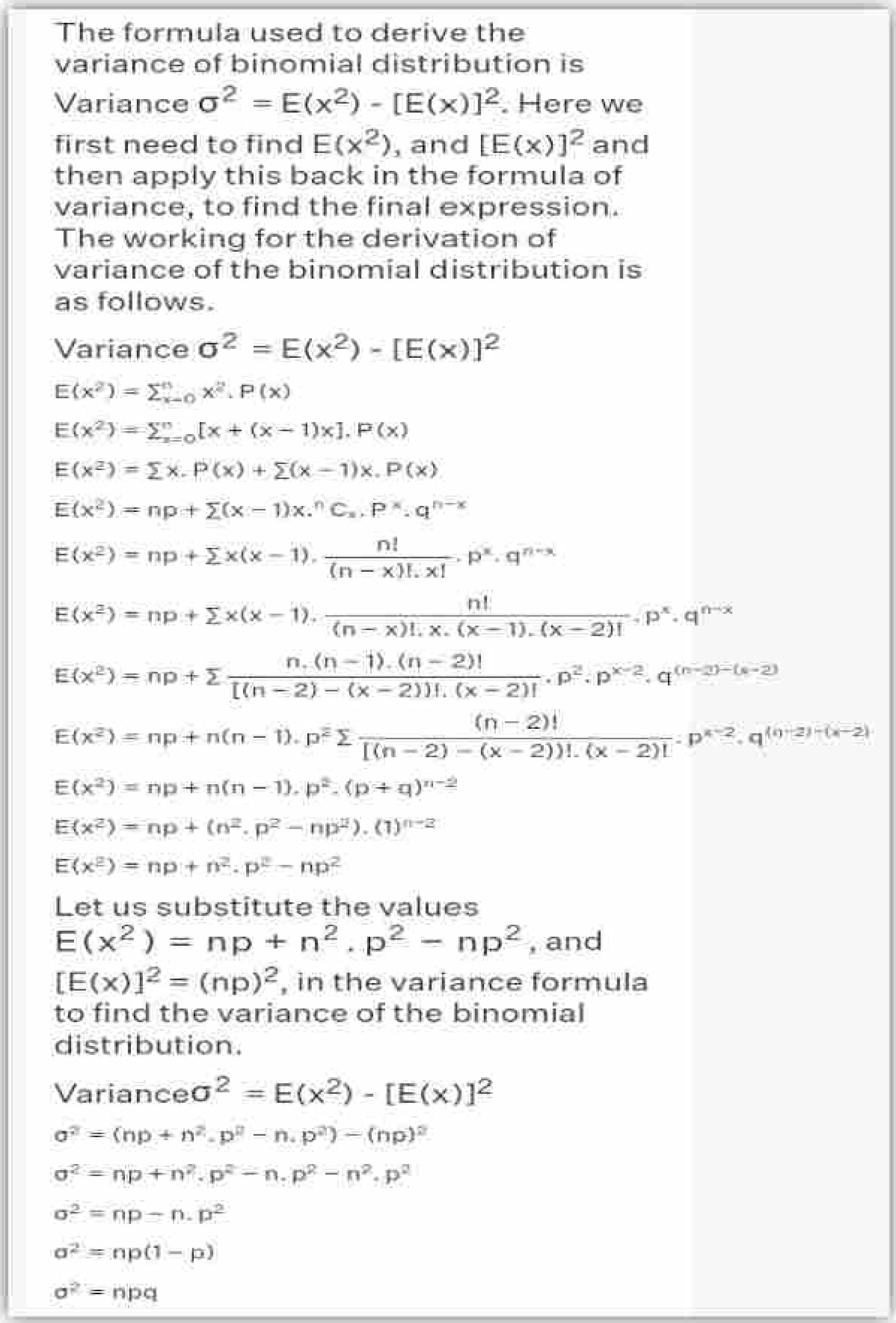
$= np \left[{}^{n-1}C_0 p^0 q^{n-1} + {}^{n-1}C_1 p q^{n-2} + {}^{n-1}C_2 p^2 q^{n-3} \right.$
$\left. \cdots + {}^{n-1}C_{n-1} p^{n-1} q^{n-n} \right]$

$= np \left[q^{n-1} + {}^{n-1}C_1 pq^{n-2} + {}^{n-1}C_2 p^2 q^{n-3} \cdots p^{n-1} \right]$

$= np [q+p]^{n-1}$

$= np \quad (q+p)=1$

Teacher's Signature



The formula used to derive the variance of binomial distribution is Variance $\sigma^2 = E(x^2) - [E(x)]^2$. Here we first need to find $E(x^2)$, and $[E(x)]^2$ and then apply this back in the formula of variance, to find the final expression. The working for the derivation of variance of the binomial distribution is as follows.

Variance $\sigma^2 = E(x^2) - [E(x)]^2$

$$E(x^2) = \sum_{x=0}^{n} x^2 . P(x)$$

$$E(x^2) = \sum_{x=0}^{n} [x + (x-1)x] . P(x)$$

$$E(x^2) = \sum x . P(x) + \sum (x-1)x . P(x)$$

$$E(x^2) = np + \sum (x-1)x . {}^{n}C_{x} . P^{x} . q^{n-x}$$

$$E(x^2) = np + \sum x(x-1) . \frac{n!}{(n-x)! . x!} . p^{x} . q^{n-x}$$

$$E(x^2) = np + \sum x(x-1) . \frac{n!}{(n-x)! . x . (x-1) . (x-2)!} . p^{x} . q^{n-x}$$

$$E(x^2) = np + \sum \frac{n . (n-1) . (n-2)!}{[(n-2)-(x-2)]! . (x-2)!} . p^2 . p^{x-2} . q^{(n-2)-(x-2)}$$

$$E(x^2) = np + n(n-1) . p^2 \sum \frac{(n-2)!}{[(n-2)-(x-2)]! . (x-2)!} . p^{x-2} . q^{(n-2)-(x-2)}$$

$$E(x^2) = np + n(n-1) . p^2 . (p+q)^{n-2}$$

$$E(x^2) = np + (n^2 . p^2 - np^2) . (1)^{n-2}$$

$$E(x^2) = np + n^2 . p^2 - np^2$$

Let us substitute the values $E(x^2) = np + n^2 . p^2 - np^2$, and $[E(x)]^2 = (np)^2$, in the variance formula to find the variance of the binomial distribution.

Variance$\sigma^2 = E(x^2) - [E(x)]^2$

$$\sigma^2 = (np + n^2 . p^2 - n . p^2) - (np)^2$$

$$\sigma^2 = np + n^2 . p^2 - n . p^2 - n^2 . p^2$$

$$\sigma^2 = np - n . p^2$$

$$\sigma^2 = np(1-p)$$

$$\sigma^2 = npq$$

In R programming language, the binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of independent Bernoulli trials. Each trial has two possible outcomes: success or failure, and the probability of success remains constant.

The probability mass function (PMF) of the binomial distribution is given by:

$$P(X=x)=\binom{n}{x}P^{x}q^{n-x}$$

where:

- n is the number of trials,
- k is the number of successes,
- p is the probability of success in a single trial,
- C(n,x) is the binomial coefficient, which represents the number of ways to choose k successes out of n trials, and is calculated using the choose() function in R.

In R Programming Language, there are 4 built-in functions to for Binomial distribution. They are:

1. **dbinom(x, size, prob):probability mass function**

   Probability mass function $P(X=x)=\binom{n}{x}P^{x}q^{n-x}$

   This function calculates the probability mass function (PMF) of the binomial distribution.

   - x: The number of successes.
   - size: The total number of trials.
   - prob: The probability of success in a single trial.

   Example:

   dbinom(2, 5, 0.4)

2. **pbinom(q, size, prob): cumulative distribution function (cdf):**

   $$F(k)=\sum_{i=0}^{k}P(X=i)$$

   This function calculates the cumulative distribution function (CDF) of the binomial distribution, i.e., the probability of getting up to q successes.

   - q: The number of successes.
   - size: The total number of trials.
   - prob: The probability of success in a single trial.

Example:

```
pbinom(2, 5, 0.4)
```

3. **qbinom(p, size, prob): Quantile Function**

$$Q(p) = \begin{cases} 0 & \text{if } p < 1 - p \\ 1 & \text{if } p \geq 1 - p \end{cases}$$

This function calculates the quantile function (inverse CDF) of the binomial distribution, i.e., the number of successes that corresponds to a given probability.

- p: The probability.
- size: The total number of trials.
- prob: The probability of success in a single trial.

Example:

```
qbinom(0.8, 5, 0.4)
```

4. **rbinom(n, size, prob):**

This function generates random samples from a binomial distribution.

- n: The number of random samples to generate.
- size: The total number of trials.
- prob: The probability of success in a single trial.

Example:

```
rbinom(10, 5, 0.4)
```

3.Poisson Functions

The Poisson distribution represents the probability of a provided number of caseshappening in a set period of space or time if these cases happen with an identified constant mean rate.

In mathematical terms, for a discrete random variable and a realization X = x, the Poisson mass function f is given as follows, where λp is a parameter of the distribution.

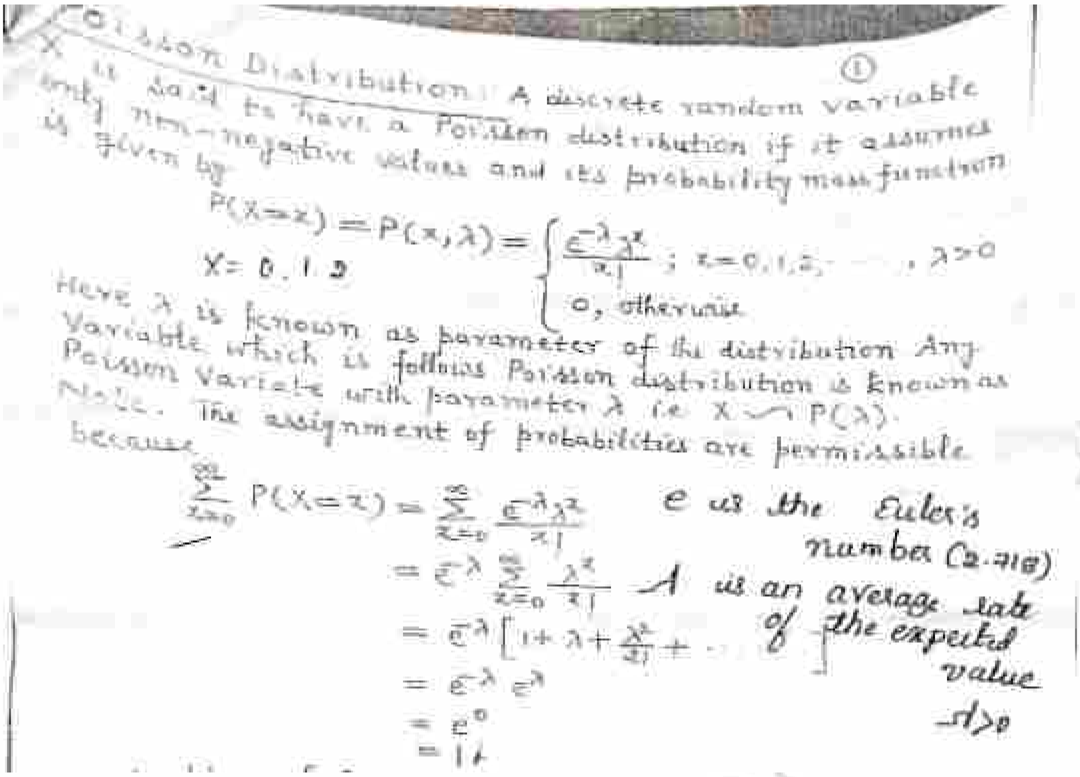
$$f(x) = \frac{\lambda_p^x \exp(-\lambda_p)}{x!}; \quad x = \{0, 1, \ldots\}$$

The notation

$$X \sim \text{POIS}(\lambda_p)$$

OR

The probability mass function is given by P(X = x) = $\frac{\lambda^{X} e^{z}}{x!}$



1

Poisson Distribution: A discrete random variable X is said to have a Poisson distribution if it assumes only non-negative values and its probability mass function is given by

$$P(X=x) = P(x,\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{x}}{x!} ; & x=0,1,2,\ldots, \lambda>0 \\ 0, & \text{otherwise} \end{cases}$$

X= 0, 1, 2

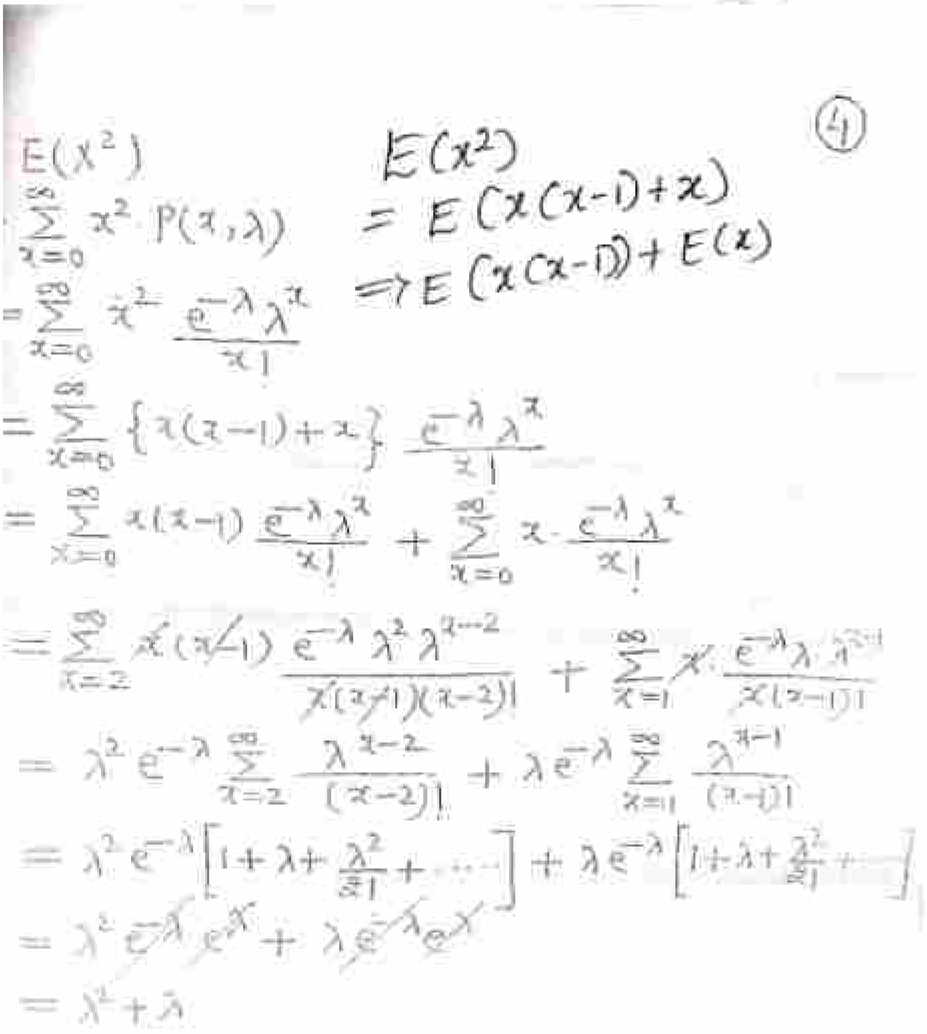
Here λ is known as parameter of the distribution Any Variable which is follows Poisson distribution is known as Poisson Variate with parameter λ i.e X ∼ P(λ). Note. The assignment of probabilities are permissible because

$$\sum_{x=0}^{\infty} P(X=x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!}$$
$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!}$$
$$= e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2!}+\cdots\right]$$
$$= e^{-\lambda} e^{\lambda}$$
$$= e^{0}$$
$$= 1$$

e is the Euler's number (2.718)

λ is an average rate of the expected value

λ>0



4

$E(x^2)$

$$\sum_{x=0}^{\infty} x^2 \cdot P(x,\lambda)$$
$$= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^{x}}{x!}$$
$$= \sum_{x=0}^{\infty} \{x(x-1)+x\} \frac{e^{-\lambda}\lambda^{x}}{x!}$$
$$= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^{x}}{x!} + \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x!}$$
$$= \sum_{x=2}^{\infty} \cancel{x}(\cancel{x-1}) \frac{e^{-\lambda}\lambda^{2}\lambda^{x-2}}{\cancel{x}(\cancel{x-1})(x-2)!} + \sum_{x=1}^{\infty} \cancel{x} \frac{e^{-\lambda}\lambda\,\lambda^{x-1}}{\cancel{x}(x-1)!}$$
$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$
$$= \lambda^2 e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2!}+\cdots\right] + \lambda e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2!}+\cdots\right]$$
$$= \lambda^2 \cancel{e^{-\lambda}}\,\cancel{e^{\lambda}} + \lambda \cancel{e^{-\lambda}}\cancel{e^{\lambda}}$$
$$= \lambda^2+\lambda$$

$E(x^2)$
$= E(x(x-1)+x)$
$\Rightarrow E(x(x-1)) + E(x)$

$$\mu_1 = 0$$

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

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

$$= \lambda$$

$$= \text{Variance}$$

$$\boxed{\text{Mean} = \text{Variance} = \lambda}$$

Built in function in Poissoin Distribution

1.dpois(x, lambda):

- This function calculates the probability mass function (PMF) of the Poisson distribution.
- x: The number of events.
- lambda: The average rate of occurrence.

Example:

dpois(2, lambda = 3)

2.ppois(q, lambda, lower.tail = TRUE):

This function calculates the cumulative distribution function (CDF) of the Poisson distribution, i.e., the probability of getting up to q events.

- q: The number of events.
- lambda: The average rate of occurrence.
- lower.tail: Logical, indicating whether to calculate the lower tail probability (default is TRUE).

Example:

ppois(2, lambda = 3)

3.qpois(p, lambda, lower.tail = TRUE):Quantile function

This function calculates the quantile function (inverse CDF) of the Poisson distribution, i.e., the number of events that corresponds to a given probability.

- p: The probability
- lambda: The average rate of occurrence.
- lower.tail: Logical, indicating

Example:

qpois(0.8, lambda = 3)

4.rpois(n, lambda):random sampling

This function generates random samples from a Poisson distribution.

- n: The number of random samples to generate.
- lambda: The average rate of occurrence.

Example:

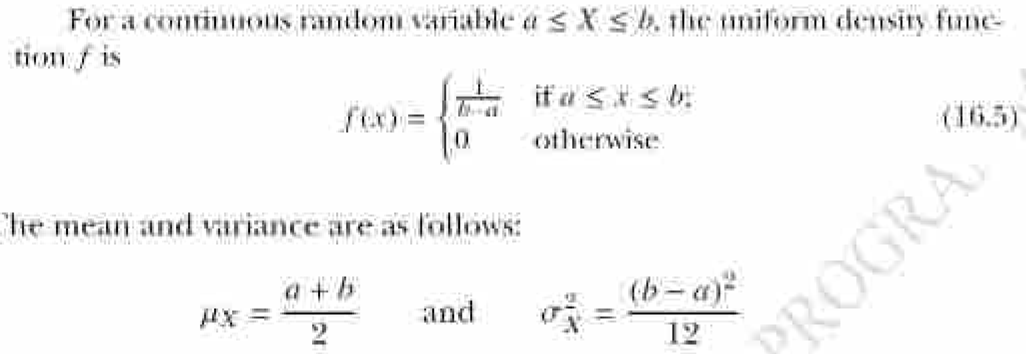
rpois(10, lambda = 3)

Common Probability Density Functions

Uniform Distribution

The continuous uniform distribution is also referred to as the probability distribution of any random number selection from the continuous interval defined between intervals a and b.

A uniform distribution holds the same probability for the entire interval. Thus, its plot is a rectangle, and therefore it is often referred to as rectangular distribution.

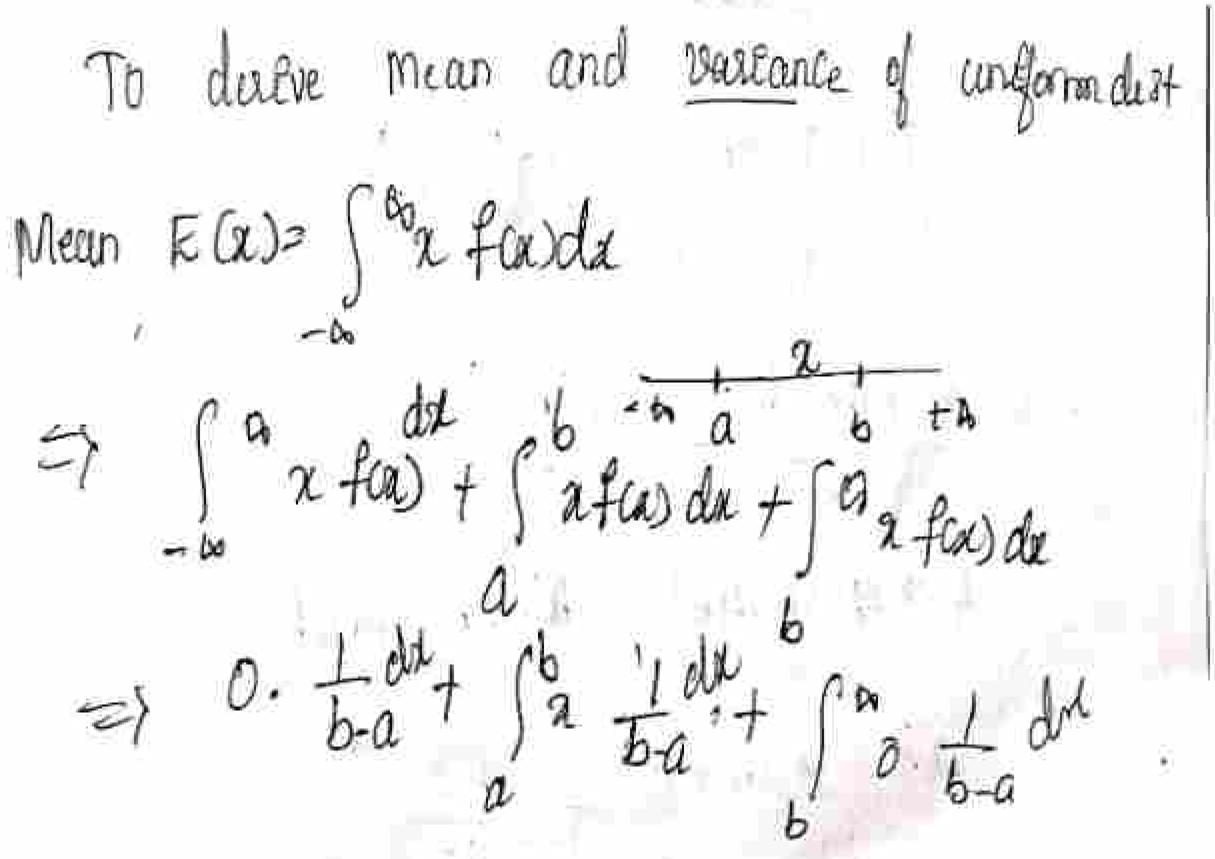


For a continuous random variable $a \leq X \leq b$, the uniform density function f is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b; \\ 0 & \text{otherwise} \end{cases} \qquad (16.5)$$

The mean and variance are as follows:

$$\mu_X = \frac{a+b}{2} \qquad \text{and} \qquad \sigma_X^2 = \frac{(b-a)^2}{12}$$

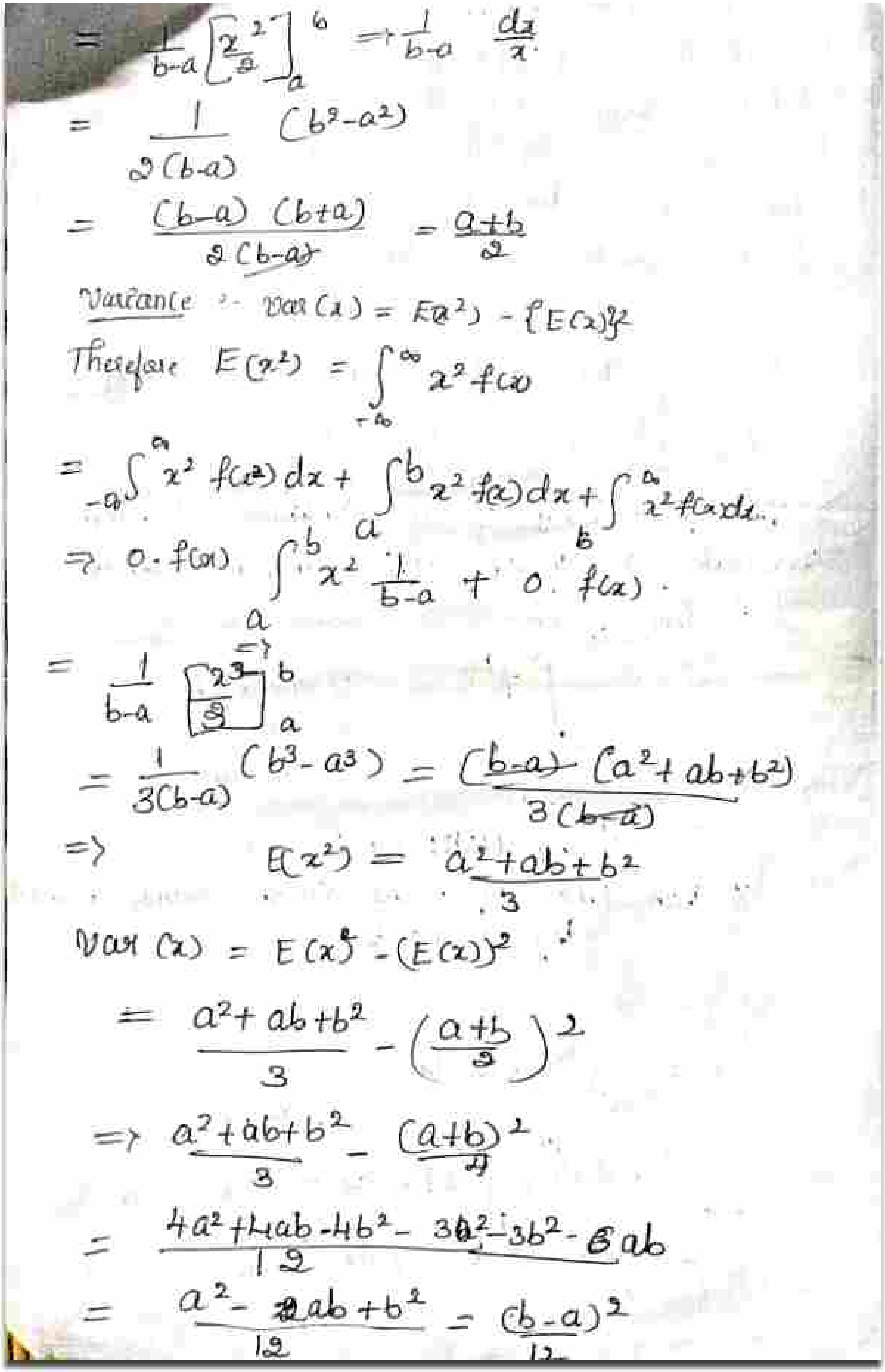


To derive mean and variance of uniform dist

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

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

$\Rightarrow 0 \cdot \frac{1}{b-a} dx + \int_{a}^{b} x \frac{1}{b-a} dx + \int_{b}^{\infty} 0 \cdot \frac{1}{b-a} dx$



$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \Rightarrow \frac{1}{b-a} \quad \frac{dx}{x}$

$= \frac{1}{2(b-a)} (b^2-a^2)$

$= \frac{(b-a)(b+a)}{2\cancel{(b-a)}} = \frac{a+b}{2}$

Variance :- $Var(x) = E(x^2) - \{E(x)\}^2$

Therefore $E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)$

$= \int_{-\infty}^{a} x^2 f(x) dx + \int_a^b x^2 f(x) dx + \int_b^{\infty} x^2 f(x) dx$..

$\Rightarrow 0 \cdot f(x) \int_a^b x^2 \frac{1}{b-a} + 0 \cdot f(x)$.

$\Rightarrow$

$= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$

$= \frac{1}{3(b-a)}(b^3-a^3) = \frac{\cancel{(b-a)}(a^2+ab+b^2)}{3\cancel{(b-a)}}$

$\Rightarrow \quad E(x^2) = \frac{a^2+ab+b^2}{3}$

$Var(x) = E(x^2) - (E(x))^2$

$= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$

$\Rightarrow \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4}$

$= \frac{4a^2+4ab-4b^2-3a^2-3b^2-6ab}{12}$

$= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$

Built in function of uniform distribution

1. **Probability Density Function:**dunif() method in R programming language is used to generate density function. It calculates the uniform density function in R language in the specified interval (a, b).

   Syntax:

   dunif(x, min = 0, max = 1, log = FALSE)

   Parameter:

   x: input sequence

   min, max= range of values

   log: indicator, of whether to display the output values as probabilities.

   Example:

   dunif(0.5, min = 0, max = 1)

2. **Cumulative probability distribution**

   The punif() method in R is used to calculate the uniform cumulative distribution function, this is, the probability of a variable X taking a value lower than x (that is, x <=X)

   Syntax:**punif(q, min = 0, max = 1, lower.tail = TRUE)**

   - **q:** The value(s) at which to calculate the CDF.
   - **min:** The lower bound of the interval (default is 0).
   - **max:** The upper bound of the interval (default is 1).
   - **lower.tail:** Logical, indicating whether to calculate the lower tail probability (default is TRUE).

   Example:

   punif(0.7, min = 0, max = 1)

3. **Quantile function :**This function calculates the quantile function (inverse CDF) of the uniform distribution, i.e., the value that corresponds to a given probability.

   Synatx: qunif(p, min = 0, max = 1, lower.tail = TRUE)

   - **p:** The probability(s) at which to calculate the quantiles.
   - **min:** The lower bound of the interval (default is 0).
   - **max:** The upper bound of the interval (default is 1).
   - **lower.tail:** Logical, indicating whether to calculate the lower tail probability (default is TRUE).

     Example:

     qunif(0.8, min = 0, max = 1)

4.Random sampling: The **runif()** function in R programming language is used to generate a sequence of random following the uniform distribution.

Syntax:

runif(n, min = 0, max = 1)

Parameter:
n= number of random samples

min=minimum value(by default 0)

max=maximum value(by default

Example:

runif(10, min = 0, max = 1)

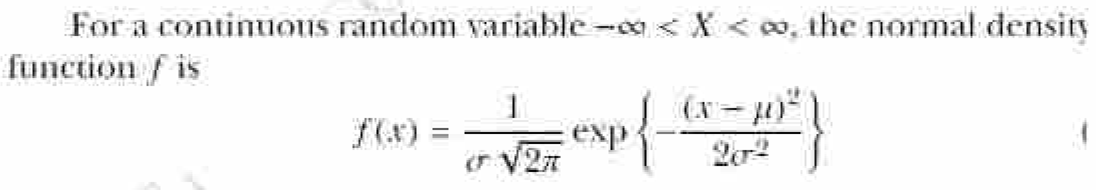
Normal Distribution in R

Normal Distribution is a probability function used in statistics that tells about how the data values are distributed.

For example, the height of the population, shoe size, IQ level, rolling a dice, and many more.

It is generally observed that data distribution is normal when there is a random collection of data from independent sources. The graph produced after plotting the value of the variable on x-axis and count of the value on y-axis is bell-shaped curve graph.

The graph signifies that the peak point is the mean of the data set and half of the values of data set lie on the left side of the mean and other half lies on the right part of the mean.

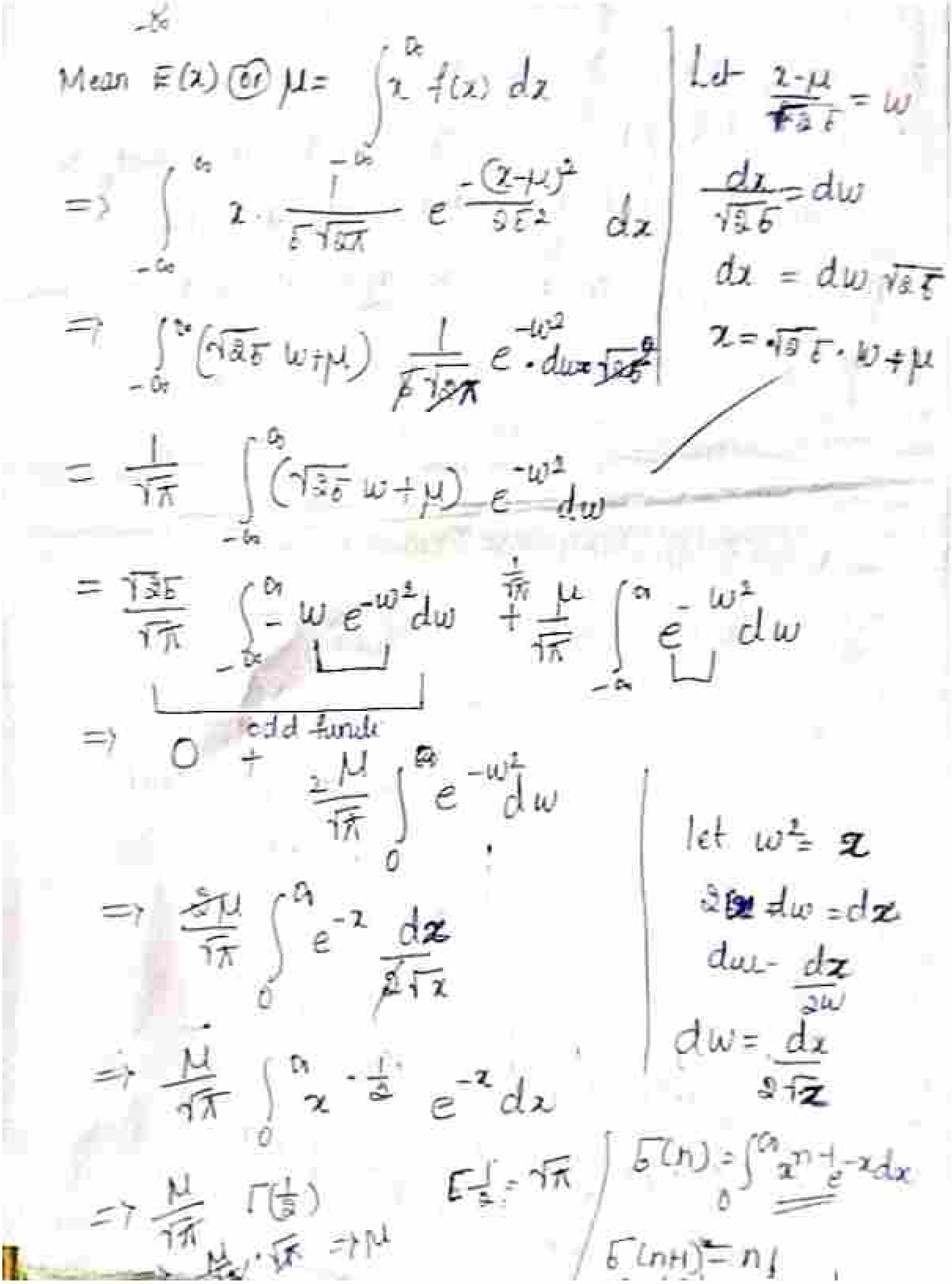


For a continuous random variable $-\infty < X < \infty$, the normal densit function f is

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

Mean=μ **Variance=σ^2**

To Derive mean and Variance of Normal Distribution



Mean $E(x)$ (or) $\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$

$\Rightarrow \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$

$\Rightarrow \int_{-\infty}^{\infty} (\sqrt{2}\sigma\, w+\mu)\, \frac{1}{\cancel{\sigma}\sqrt{\cancel{2}\pi}}\, e^{-w^2} \cdot dw\, \cancel{\sqrt{2}\sigma}$

$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma\, w+\mu)\, e^{-w^2}\, dw$

$= \underbrace{\frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} w\, e^{-w^2} dw}_{\text{odd funcn}} + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-w^2} dw$

$\Rightarrow 0 + \frac{2\mu}{\sqrt{\pi}} \int_{0}^{\infty} e^{-w^2} dw$

$\Rightarrow \frac{\cancel{2}\mu}{\sqrt{\pi}} \int_{0}^{\infty} e^{-x}\, \frac{dx}{\cancel{2}\sqrt{x}}$

$\Rightarrow \frac{\mu}{\sqrt{\pi}} \int_{0}^{\infty} x^{-\frac{1}{2}}\, e^{-x}\, dx$

$\Rightarrow \frac{\mu}{\sqrt{\pi}}\, \Gamma(\tfrac{1}{2})$ $\qquad \Gamma\tfrac{1}{2} = \sqrt{\pi}$

$\Rightarrow \frac{\mu}{\cancel{\sqrt{\pi}}} \cdot \cancel{\sqrt{\pi}} \to \mu$

Let $\frac{x-\mu}{\sqrt{2}\sigma} = w$

$\frac{dx}{\sqrt{2}\sigma} = dw$

$dx = dw\,\sqrt{2}\sigma$

$x = \sqrt{2}\sigma \cdot w + \mu$

let $w^2 = x$

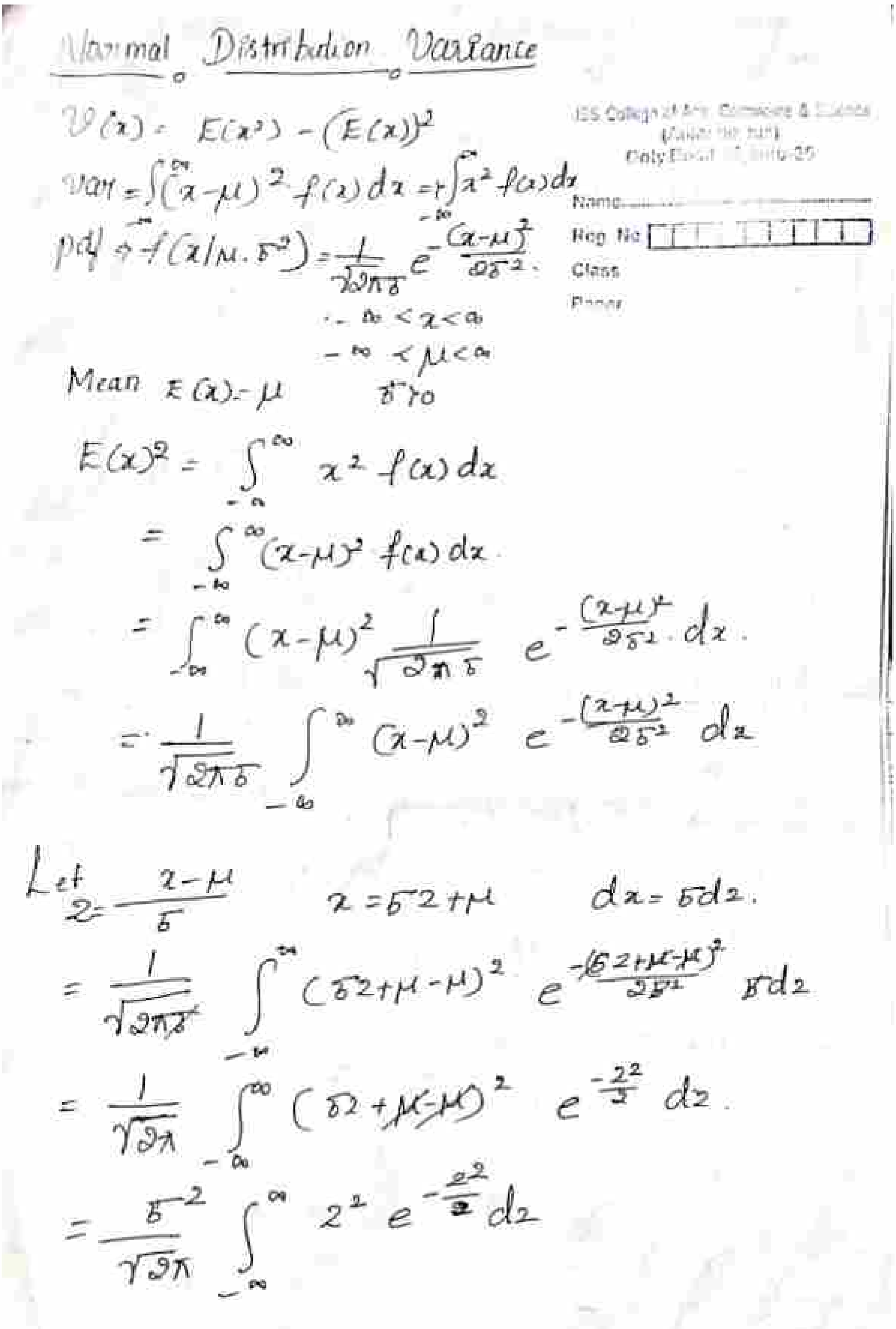
$2w\, dw = dx$

$dw = \frac{dx}{2w}$

$dw = \frac{dx}{2\sqrt{x}}$

$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$

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



Normal Distribution Variance

$V(x) = E(x^2) - (E(x))^2$

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

pdf → $f(x/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

$-\infty < x < \infty$

$-\infty < \mu < \infty$

Mean $E(x) = \mu$ $\quad \sigma > 0$

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

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

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

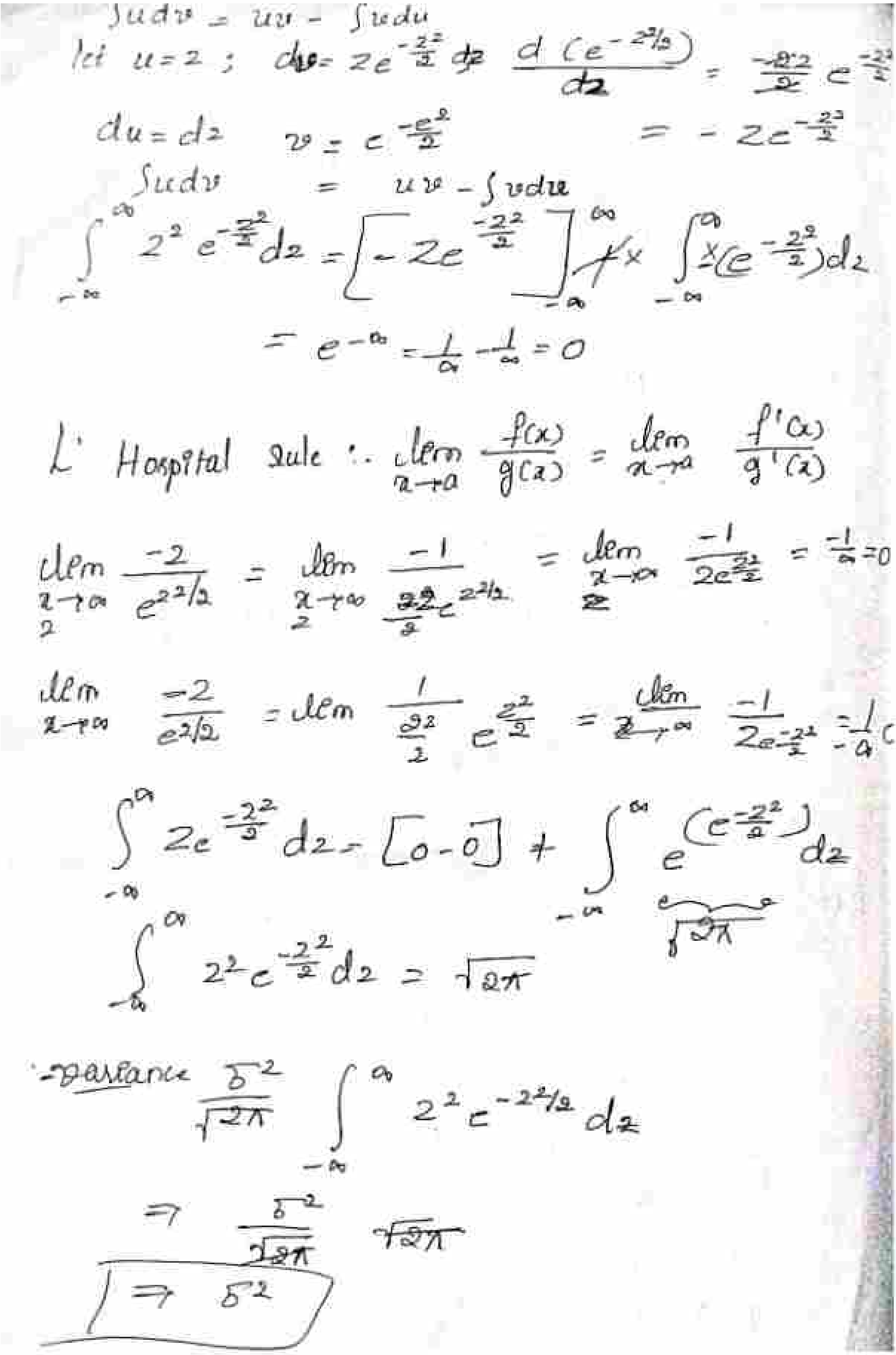
$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dz$$

Let $z = \frac{x-\mu}{\sigma}$ $\quad x = \sigma z + \mu \quad dx = \sigma dz.$

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

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

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



$\int u\,dv = uv - \int v\,du$

let $u = z$; $dv = z e^{-\frac{z^2}{2}} dz$ $\quad \frac{d\left(e^{-z^2/2}\right)}{dz} = \frac{-2z}{2} e^{-\frac{z^2}{2}}$

$du = dz \quad v = e^{-\frac{z^2}{2}} \quad = -z e^{-\frac{z^2}{2}}$

$\int u\,dv = uv - \int v\,du$

$$\int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz = \left[-z e^{-\frac{z^2}{2}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \left(e^{-\frac{z^2}{2}}\right) dz$$

$$= e^{-\infty} = \frac{1}{\infty} - \frac{1}{\infty} = 0$$

L' Hospital rule :- $\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$

$$\lim_{\substack{x\to\infty\\ z}} \frac{-z}{e^{z^2/2}} = \lim_{\substack{x\to\infty\\ z}} \frac{-1}{\frac{2z}{2} e^{z^2/2}} = \lim_{\substack{x\to\infty\\ z}} \frac{-1}{z e^{\frac{z^2}{2}}} = \frac{-1}{\infty} = 0$$

$$\lim_{x\to\infty} \frac{-z}{e^{z^2/2}} = \lim \frac{1}{\frac{z^2}{2}} e^{\frac{z^2}{2}} = \lim_{z\to\infty} \frac{-1}{z e^{-\frac{z^2}{2}}} = \frac{1}{-\infty}$$

$$\int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = [0 - 0] + \int_{-\infty}^{\infty} \underbrace{e^{\left(e^{-\frac{z^2}{2}}\right)}}_{\sqrt{2\pi}} dz$$

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

variance $\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$

$$\Rightarrow \frac{\sigma^2}{\cancel{\sqrt{2\pi}}} \cancel{\sqrt{2\pi}}$$

$$\boxed{\Rightarrow \sigma^2}$$

Built in function of Normal Distribution:

1.dnorm(): This function calculates the probability density function (PDF) of the normal distribution.

Synatx: dnorm(x, mean = 0, sd = 1)

x: The value(s) at which to evaluate the PDF.

mean: The mean of the distribution (default is 0).

sd: The standard deviation of the distribution (default is 1).

Example: dnorm(1, mean = 2, sd = 1)

2.pnorm():function is the cumulative distribution function which measures the probability that a random number X takes a value less than or equal to x.

Syntax: pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)

- **q:** The value(s) at which to calculate the CDF.
- **mean:** The mean of the distribution (default is 0).
- **sd:** The standard deviation of the distribution (default is 1).
- **lower.tail:** Logical, indicating whether to calculate the lower tail probability (default is TRUE).

Example:pnorm(1, mean = 2, sd = 1)

3.qnorm(): This function calculates the quantile function (inverse CDF) of the normal distribution, i.e., the value that corresponds to a given probability.

Syntax: qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)

- **p:** The probability(s) at which to calculate the quantiles.
- **mean:** The mean of the distribution (default is 0).
- **sd:** The standard deviation of the distribution (default is 1).
- **lower.tail:** Logical, indicating whether to calculate the lower tail probability (default is TRUE).

Example:

qnorm(0.8, mean = 2, sd = 1)

4.rnorm(): This function generates random samples from a normal distribution.

Syntax: rnorm(n, mean = 0, sd = 1)

- **n:** The number of random samples to generate.
- **mean:** The mean of the distribution (default is 0).
- **sd:** The standard deviation of the distribution (default is 1).

Example:

rnorm(10, mean = 2, sd = 1)

Student's t-distribution

The Student's t-distribution is a continuous probability distribution generally used when dealing with statistics estimated from a sample of data.

Any particular t-distribution looks a lot like the standard normal distribution— it's bell-shaped, symmetric and it's centered on zero. The difference is that while a normal distribution is typically used to deal with a population, the t-distribution deals with sample from a population.

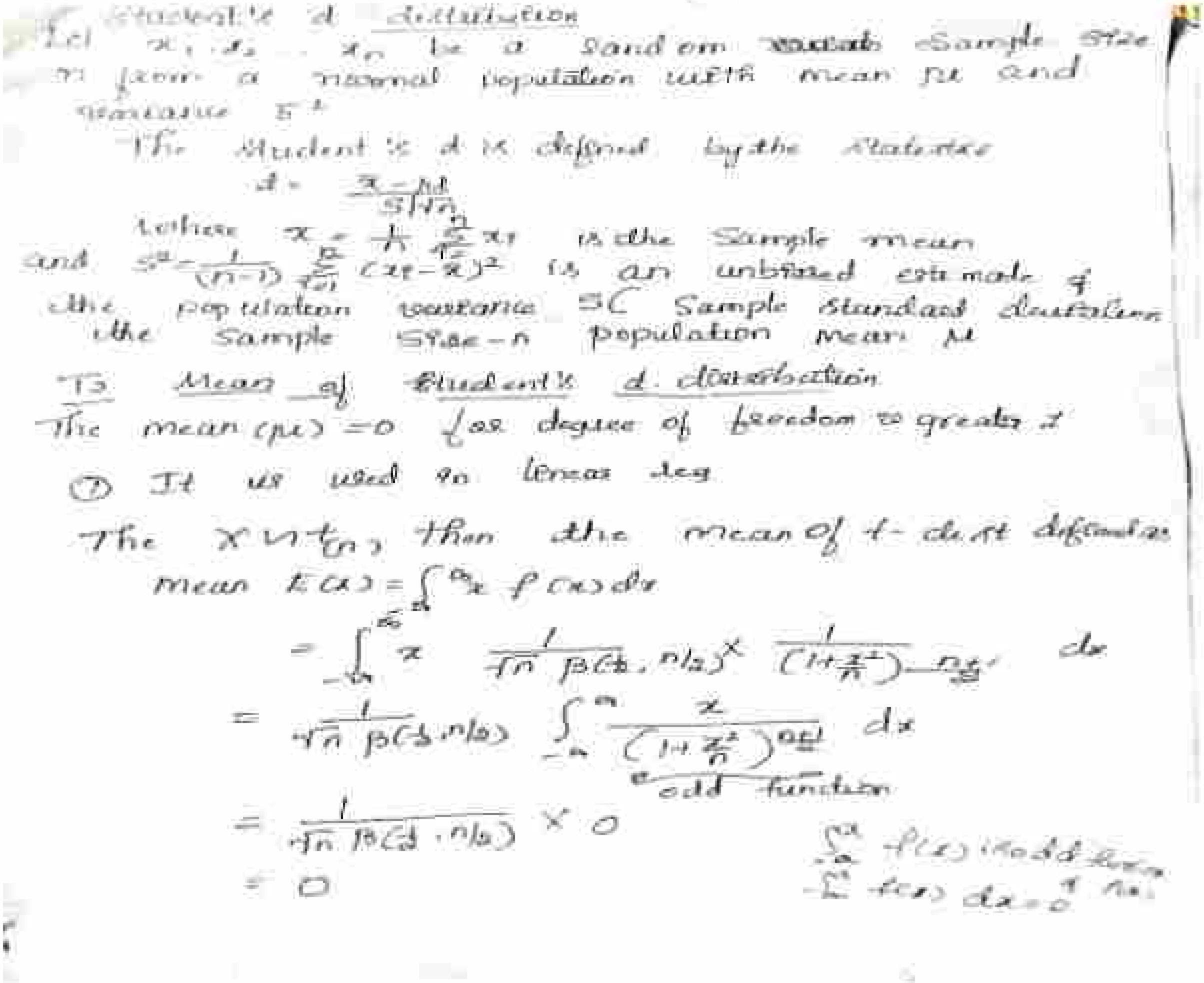
The probability density function (PDF) of the t-distribution with df degrees of freedom is given by:

$$f(x) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\,\Gamma\left(\frac{df}{2}\right)}\left(1+\frac{x^2}{df}\right)^{-\frac{df+1}{2}}$$

where

- df is the degrees of freedom and Γ
- Γ is the gamma function.

Mean =0 variance= $\frac{n}{n-2}$



Student's t distribution

Let $x_1, x_2 \ldots x_n$ be a random ~~variab~~ sample size n from a normal population with mean μ and variance σ^2

The Student's t is defined by the statistic

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean

and $S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is an unbiased estimate of the population variance S(Sample standard deviation the sample size - n population mean μ

1) Mean of Student's t distribution

The mean (μ) = 0 for degree of freedom is greater 1

① It is used in linear reg

The $X \sim t_{(n)}$, then the mean of t-dist defined as

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

$$= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{n}\,\beta(\frac{1}{2}, n/2)} \times \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}} dx$$

$$= \frac{1}{\sqrt{n}\,\beta(\frac{1}{2}, n/2)} \underbrace{\int_{-\infty}^{\infty} \frac{x}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}}}_{\text{odd function}} dx$$

$$= \frac{1}{\sqrt{n}\,\beta(\frac{1}{2}, n/2)} \times 0$$

$$= 0$$

$\int_{-a}^{a} f(x)$ is odd func

$\int_{-a}^{a} f(x)\, dx = 0$ if f is

Derivation of t distribution

The student t dist with $\sqrt{n} = n-1$ degrees of freedom with probability density function

$$f(x) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad -\infty < x < \infty$$

Derivation of Student's t distr

t is defined by the statistics

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{\sqrt{n}(x-\mu)}{S}$$

(or) $$t^2 = \frac{n(\bar{x}-\mu)^2}{S^2}$$

$S^2 = ns^2\,(n-1)$

$$\Rightarrow t^2 = \frac{n(\bar{x}-\mu)^2}{ns^2/(n-1)}$$ [Since $ns^2 = (n-1)S^2$]

$$\Rightarrow t^2 = \frac{n(n-1)(\bar{x}-\mu)^2}{ns^2}$$

$$\Rightarrow \frac{t^2}{(n-1)} = \frac{(\bar{x}-\mu)^2/(\sigma^2/n)}{ns^2/\sigma^2}$$

Since $x_1, x_2 \ldots x_n$ is random sample from normal dist with mean μ and variance σ^2

then the Sample mean $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Then $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ is Normal variate

χ^2 is variate with 1 degree freedom

ns^2/σ^2 is χ^2_{n-1} variate with (n-1) of freedom

Bernoulli – $\mu = p$ $v = pq$

Binomial – np Vario $= npq$

Poisson – var. is mean

Variance: If $X \sim t_n$ then variance of t-distr is defined as

Variance $= V(x) = E(x^2) - [E(x)]^2$

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

$$\Rightarrow \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{n}\, \beta(\frac{1}{2},\frac{n}{2})} \times \frac{1}{(1 \times \frac{x^2}{n})^{\frac{n+1}{2}}}\, dx$$

$$\Rightarrow \frac{1}{\sqrt{n}\, \beta(\frac{1}{2},\frac{n}{2})} \int_{-\infty}^{\infty} \underbrace{\frac{x^2}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}}}_{\text{Even function}}\, dx$$

$$\Rightarrow \frac{1}{\sqrt{n}\, \beta(\frac{1}{2},\frac{n}{2})} \times 2\int_{0}^{\infty} \frac{x^2}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}}\, dx.$$

$\Rightarrow$ $y = \frac{x^2}{n}$ $\quad x^2 = ny \quad x = \sqrt{ny} \quad dx = \frac{\sqrt{n}}{2\sqrt{y}}\, dy$

When $x = -\infty$ $\quad y = \frac{x^2}{n} \Rightarrow y = \infty$

$x = \infty \quad y = \infty$

$$\Rightarrow \frac{1}{\sqrt{n}\, \beta(\frac{1}{2},\frac{n}{2})} \times \cancel{2}\int_{0}^{\infty} \frac{ny}{(1+y)^{\frac{n+1}{2}}} \times \frac{\cancel{\sqrt{n}}}{\cancel{2}\sqrt{y}}\, dy$$

$$= \frac{n}{\beta(\frac{1}{2},\frac{n}{2})} \int_{0}^{\infty} \frac{y^{1-\frac{1}{2}}}{(1+y)^{\frac{n+1}{2}}}\, dy$$

$$= \frac{n}{\beta(\frac{1}{2},\frac{n}{2})} \int_{0}^{\infty} \frac{y^{1/2}}{(1+y)^{\frac{n+1}{2}}}\, dy \quad \text{--- } (1)$$

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

$\therefore n = \frac{1}{2} + 1 = \frac{3}{2}$

$$\int_{0}^{\infty} \frac{x^{m-1}}{(1+x)^{m+n}}\, dx = \beta(m,n) \quad \text{---} \ (2)$$

$$V(x) = \frac{n}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^\infty \frac{y^{1/2}}{(1+y)^{\frac{n+1}{2}+1-1}}\, dy$$

$$= \frac{n}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^\infty \frac{y}{(1+y)^{3/2+(\frac{n}{2}-1)}}\, dy \quad \text{---} \, \textcircled{3}$$

$$= \frac{n}{\beta(\frac{1}{2}, \frac{n}{2})} \times \beta(3/2, n/2-1) \qquad \Bigg/ \; \beta(m,n) = \frac{\Gamma m \, \Gamma n}{\Gamma m+n}$$

Beta Dist

$$= \frac{n}{\frac{\Gamma\frac{1}{2}\,\Gamma\frac{n}{2}}{\Gamma\frac{n+1}{2}}} \times \frac{\Gamma 3/2 \, \Gamma \frac{n}{2}-1}{\Gamma \frac{n}{2}-1+3/2}$$

$$\rightarrow \frac{\frac{n}{\cancel{2}}}{\frac{(n-2)}{\cancel{2}}} = \frac{n}{n-2}$$

$$V(x) = \frac{n}{\frac{\Gamma\frac{1}{2}\,\Gamma\frac{n}{2}}{\cancel{\Gamma \frac{n}{2}+\frac{1}{2}}}} \times \frac{\Gamma 3/2 \, \Gamma\frac{n}{2}-1}{\cancel{\Gamma \frac{n}{2}+\frac{1}{2}}}$$

$$\Rightarrow \frac{n \, \Gamma 3/2 \, \Gamma \frac{n}{2}-1}{\Gamma\frac{1}{2}\, \Gamma\frac{n}{2}} \qquad \Gamma n = (n-1)\,\Gamma n-1$$

$$= \frac{n \cdot \frac{1}{2} \, \cancel{\Gamma 1/2} \, \cancel{\Gamma n/2-1}}{\cancel{\Gamma 1/2} \, (\frac{n}{2}-1) \, \cancel{\Gamma n/2-1}}$$

$$= \frac{\frac{n}{\cancel{2}}}{(n-2)/\cancel{2}}$$

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

Built in function in student's t distribution

1.dt(): This function calculates the probability density function (PDF) of the t-distribution.

Syntax: dt(x, df)

- **x:** The value(s) at which to evaluate the PDF.
- **df:** The degrees of freedom.

Example:

```
dt(2, df = 10)
```

2.Pt(): This function calculates the cumulative distribution function (CDF) of the t-distribution, i.e., the probability of getting a value up to q.

Syntax: pt(q, df, lower.tail = TRUE)

q: The value(s) at which to calculate the CDF.

df: The degrees of freedom.

lower.tail: Logical, indicating whether to calculate the lower tail probability (default is TRUE).

Example:

```
pt(2, df = 10)
```

3.qt(): This function calculates the quantile function (inverse CDF) of the t-distribution, i.e., the value that corresponds to a given probability.

Syntax: qt(p, df, lower.tail = TRUE)

- **p:** The probability(s) at which to calculate the quantiles.
- **df: The** degrees of freedom.
- **lower.tail:** Logical, indicating whether to calculate the lower tailprobability (default is TRUE).

Example:

```
qt(0.8, df = 10)
```

4.rt(): This function generates random samples from a t-distribution.

Syntax:rt(n, df)

n: The number of random samples to generate.

df: The degrees of freedom.

Example:

```
rt(10, df = 10)
```

MODULE 3 QUESTIONS

2 marks questions:

1.What is statistics.

2.What is probability.

3.What is Probability density function.

4.what is statistical computing.

5.what is probability mass function.

3 marks questions:

1.What is statistical computing. Explain Types of statistics.

2.What is random variable. Explain Types of random variable.

3.Explain student's t distribution.

4.define Bernoulli distribution. Derive Bernoulli distribution mean.

5.Explain Probability density function with three distributions.

5 marks questions:

1.What is Bernoulli Distribution. Write Bernoulli distribution mass function. Derive Bernoulli distribution mean (μ_x)=p and variance (σ^2)=pq

2.define data visulation.Explain pie charts, Bar charts, Histogram with example.

3.Derive binomial distribution mean.

4.Explain probability density function and three distribution of density function in statistis.

5.Derive uniform distribution of mean and variance.

6.Explain Binomial distribution built-in functions.

10 marks questions:

1.Write a R program To find mean,meadian,variance,range using function.

2.To derive mean (μ_x) and variance $(\sigma^2)$of the poission distribution.

3.To implent binomial distribution variance(σ^2)=npq using probability mass function.

4.Derive Normal distribution. Mean and variance

5.write a R program to implent using bernoulli,binomial,poisson distribution mean and variance.