



PROBABILITY FOR COMPUTING

PRACTICAL FILE



RAMANUJAN COLLEGE



UNIVERSITY OF DELHI

SUBMITTED BY:

NAME : Mayank Nautiyal

ROLL NO. : 24570038

EXAMINATION ROLL NO :

24020570045

CLASS: BSC(H) COMPUTER
SCIENCE SEM 2

SUBMITTED TO:

DR. AAKASH

ASSISTANT PROFESSOR

(OPERATIONAL RESEARCH),

DEPARTMENT OF COMPUTER SCIENCE,

RAMANUJAN COLLEGE,

UNIVERSITY OF DELHI,

CR PARK MAIN ROAD, BLOCK H, KALKAJI

NEW DELHI 110019

Acknowledgement

I would like to take this opportunity to acknowledge everyone who has helped us in every stage of this project.

I am deeply indebted to my mathematics Professor, Dr Aakash for his guidance and suggestions in completing this project. The completion of this project was possible under his guidance and support.

I am also very thankful to my parents and my friends who have boosted me up morally with their continuous support.

At last but not least, I am very thankful to God almighty for showering his blessings upon me.

s.no	topic	Page	Sign.
1.	Plotting and fitting of Binomial distribution and graphical representation of probabilities.	5-7	
2	Plotting and fitting of Multinomial distribution and graphical representation of probabilities.	8-9	
3	Plotting and fitting of Poisson distribution and graphical representation of probabilities.	10-11	
4	Plotting and fitting of Geometric distribution and graphical representation of probabilities.	12-14	
5	Plotting and fitting of Uniform distribution and graphical representation of probabilities.	15-17	
6	Plotting and fitting of Exponential distribution and graphical representation of probabilities.	18-19	
7	Plotting and fitting of Normal distribution and graphical representation of probabilities.	20-22	
8	Calculation of cumulative distribution functions for Exponential and Normal distribution.	23	
9	Given data from two distributions, find the distance between the distributions.	24-25	
10	Application problems based on the Binomial distribution.	26	
11	Application problems based on the Poisson distribution.	27	
12	Application problems based on the Normal distribution.	28	
13	Presentation of bivariate data through scatter-plot diagrams and calculations of covariance.	29-33	
14	Calculation of Karl Pearson's correlation coefficients.	34	
15	To find the correlation coefficient for a bivariate frequency distribution.	35	
16	Generating Random numbers from discrete (Bernoulli, Binomial, Poisson) distributions.	36	
17	Generating Random numbers from continuous (Uniform, Normal) distributions.	37	
18	Find the entropy from the given data set.	38-39	

Binomial distribution

Binomial Distribution in Probability gives information about only two types of possible outcomes i.e. Success or Failure. Binomial Probability Distribution is a discrete probability distribution used for the events that give results in **‘Yes or No’ or ‘Success or Failure’**.

the probability of success (usually denoted as “p”) and the probability of failure (usually denoted as “q”) is constant for each trial.

Binomial Distribution Formula

The Binomial Distribution Formula which is used to calculate the probability, for a random variable $X = 0, 1, 2, 3, \dots, n$ is given as

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Probability Mass Function
for a Binomial

Probability that our
variable takes on the
value k

Mean $\mu = n \cdot p$

Variance $\sigma^2 = n \cdot p \cdot q$

Std. Dev. $\sigma = \sqrt{n \cdot p \cdot q}$

, $k = 0, 1, 2, 3, \dots$

Where ,

- p is probability of success
- q is probability of failure and $q = 1 - p$
- $p, q > 0$ such that $p + q = 1$
- n is the number of independent trials
- k is the number of success
- nCk is the number of ways to obtain k success in n trials.

Understanding Binomial Distribution Symmetry

- A binomial distribution is symmetric if its success probability, p , is equal to 0.5 (i.e., the probability of failure, q , is also 0.5). When $p = 0.5$, the distribution is said to be balanced, which means that the chances of success and failure are equal. In this case, the given binomial distribution has $p = 0.5$, which means that it is **symmetric**. The shape of the probability distribution will be **symmetric and not skewed**.

Determining the Skewness of the Distribution

- Since we know that the distribution is symmetric when $p = 0.5$, it follows that the **skewness is zero**. Skewness is a measure of the asymmetry of a distribution. A distribution is **positively skewed** (skewed right) when there are more smaller values, and **negatively skewed** (skewed left) when there are more larger values. A symmetric distribution has a skewness of zero because there is an equal balance of larger and smaller values in the distribution. This means that the shape of the probability distribution is not skewed to the left or the right. Therefore, the shape of the probability distribution in this case is symmetric due to $p = 0.5$.

how to implement in excel

= BINOM.DIST(number_s , trials, probability_s, cumulative)

= BINOM.DIST(k ,n ,p, FALSE)

The BINOM.DIST utilizes the accompanying contentions:

1. **Number_s** (required argument) – This is the number of successes in trials.
2. **Trials** (required argument) – This is the number of independent trials. It must be greater than or equal to 0.

3. **Probability_s** (required argument) – This is the probability of success in each trial.
4. **Cumulative** (required argument) – This is a logical value that determines the form of the function. It can either be:
 1. **TRUE** – Uses the cumulative distribution function.
 2. **FALSE** – Uses the probability mass function

1. Plotting and fitting of Binomial distribution and graphical representation of probabilities.

Binomial distribution

Binomial Distribution in Probability gives information about only two types of possible outcomes i.e. Success or Failure. Binomial Probability Distribution is a discrete probability distribution used for the events that give results in '**Yes or No**' or '**Success or Failure**'.

the probability of success (usually denoted as "p") and the probability of failure (usually denoted as "q") is constant for each trial.

Binomial Distribution Formula

The Binomial Distribution Formula which is used to calculate the probability, for a random variable $X = 0, 1, 2, 3, \dots, n$ is given as

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Probability Mass Function
for a Binomial

↑
Probability that our
variable takes on the
value k

$$\text{Mean } \mu = n \cdot p$$

$$\text{Variance } \sigma^2 = n \cdot p \cdot q$$

$$\text{Std. Dev. } \sigma = \sqrt{n \cdot p \cdot q}$$

, $k = 0, 1, 2, 3, \dots$

Where ,

- p is probability of success

- q is probability of failure and $q = 1 - p$
- $p, q > 0$ such that $p + q = 1$
- n is the number of independent trials
- k is the number of success
- nCk is the number of ways to obtain k success in n trials.

Understanding Binomial Distribution Symmetry

- A binomial distribution is symmetric if its success probability, p , is equal to 0.5 (i.e., the probability of failure, q , is also 0.5). When $p = 0.5$, the distribution is said to be *balanced*, which means that the chances of success and failure are equal. In this case, the given binomial distribution has $p = 0.5$, which means that it is **symmetric**. The shape of the probability distribution will be **symmetric and not skewed**.

Determining the Skewness of the Distribution

- Since we know that the distribution is symmetric when $p = 0.5$, it follows that the **skewness is zero**. Skewness is a measure of the asymmetry of a distribution. A distribution is **positively skewed** (skewed right) when there are more smaller values, and **negatively skewed** (skewed left) when there are more larger values. A symmetric distribution has a skewness of zero because there is an equal balance of larger and smaller values in the distribution. This means that the shape of the probability distribution is not skewed to the left or the right. Therefore, the shape of the probability distribution in this case is symmetric due to $p = 0.5$.

how to implement in excel

= BINOM.DIST(number_s, trials, probability_s, cumulative)

= BINOM.DIST(k, n, p, FALSE)

The BINOM.DIST utilizes the accompanying contentions:

5. **Number_s** (required argument) – This is the number of successes in trials.
6. **Trials** (required argument) – This is the number of independent trials. It must be greater than or equal to 0.
7. **Probability_s** (required argument) – This is the probability of success in each trial.

8. **Cumulative** (required argument) – This is a logical value that determines the form of the function. It can either be:

1. *TRUE* – Uses the cumulative distribution function.
2. *FALSE* – Uses the probability mass function

1. Plotting and fitting of Binomial distribution and graphical representation of probabilities.

Suppose there is a quiz competition in school and 50% chances are of getting answer right .If 10 students at random are participating and right answer is defined as a success. The probability distribution of the number of success during these 10 trials with probability $p=0.8$. then plot the graph of this probability in excel ?



Multinomial distribution

The multinomial distribution is a multivariate generalization of the [binomial distribution](#). Consider a trial that results in exactly one of some fixed [finite number](#) k of possible outcomes, with [probabilities](#) p_1, p_2, \dots, p_k (so that $p_i \geq 0$ for $i = 1, \dots, k$ and $\sum_{i=1}^k p_i = 1$), and there are n independent trials. Then let the random variables X_i indicate the number of times outcome number i was observed over the n trials. Then $X = (X_1, X_2, \dots, X_k)$ follows a multinomial distribution with parameters n and \mathbf{p} , where $\mathbf{p} = (p_1, p_2, \dots, p_k)$.

Multinomial Distribution Formula

$$p(x_1, x_2, \dots, x_k) = \left[\frac{n!}{x_1! \cdot x_2! \cdot \dots \cdot x_k!} \right] \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k}$$

Mean	$E\{X_i\} = np_i$
Variance	$\text{Var}(X_i) = np_i(1 - p_i)$ $\text{Cov}(X_i, X_j) = -np_i p_j \quad (i \neq j)$

When $\mathbf{X} = (x_1, x_2, \dots, x_k)$ follows a multinomial distribution with the PMF given above, X_i follows a binomial distribution with n trials and success probability p_i .

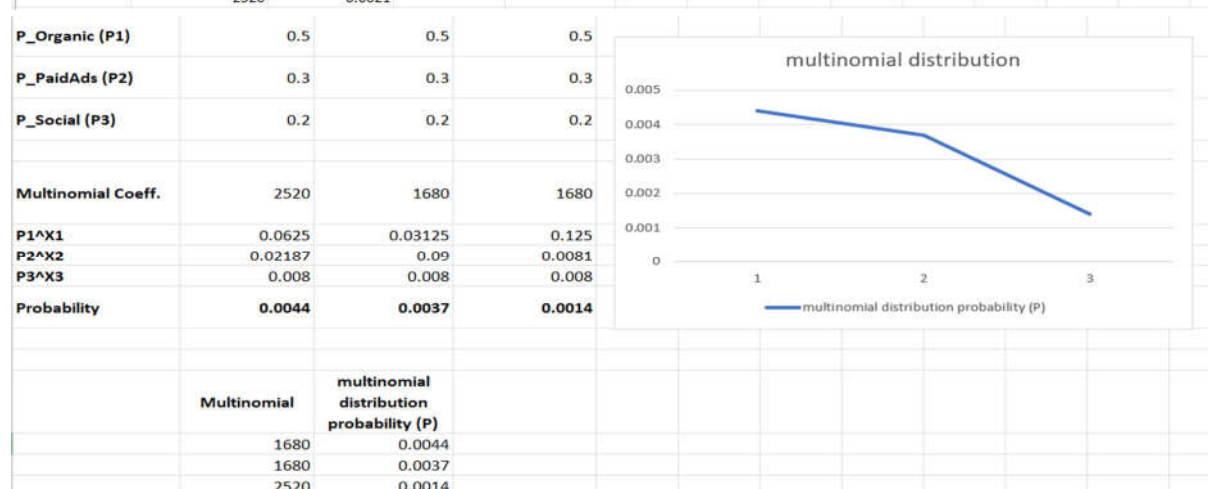
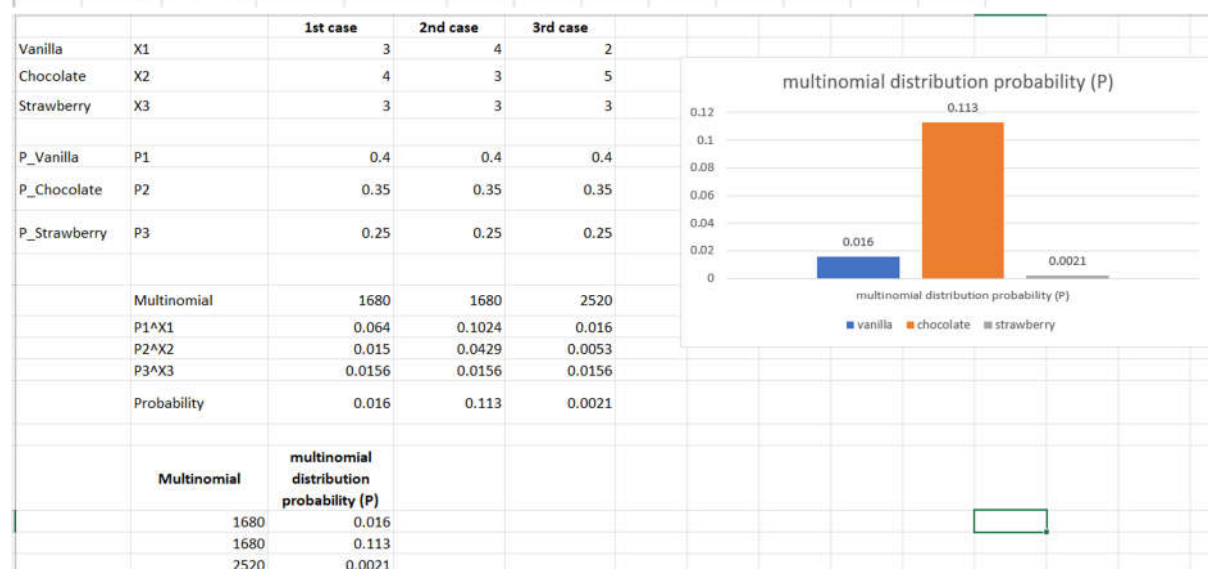
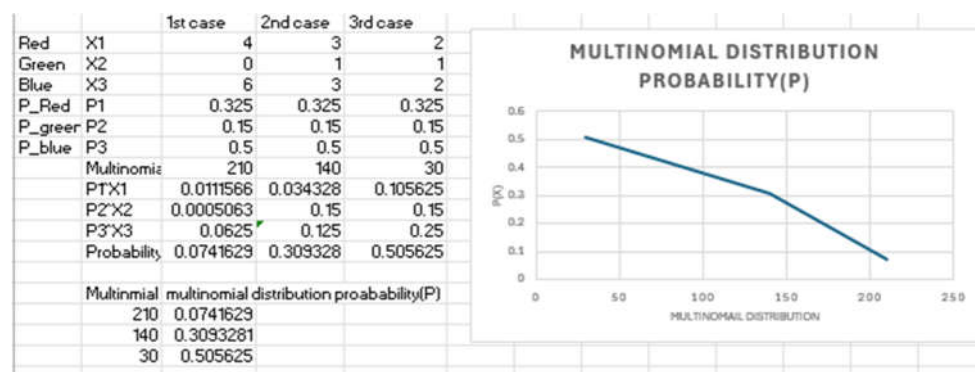
how to implement in excel

Multinomial = MULTINOMIAL(X1,X2,X3)

Probability = MULTINOMIAL*PRODUCT(p1^X1,p2^X2,p3^X3)

2. Plotting and fitting of Multinomial distribution and graphical representation of probabilities.

Suppose that a bag contain 8 balls 3 red, 1 green and 4 blue to reach in a bag to pull ball at random and then pull the ball back and pull out another ball .experiment is repeated at total of 10 time . What is the probability outcome will result in 4 red and 6 blues?



Poisson distribution

The Poisson distribution is a type of discrete probability distribution that determines the likelihood of an event occurring a specific number of times (k) within a designated time or space interval. This distribution is characterized by **a single parameter, λ (lambda)**, representing the average number of occurrences of the event.

Poisson Distribution Formula

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Mean	$\mu = E(X) = \lambda$
Variance	$\sigma^2 = V(X) = \lambda$
Standard Deviation	$\sigma = \sqrt{\sigma^2} = \sqrt{\lambda}$

Where:

- $P(X=k)$ is the probability of observing k events
- e is the base of the natural logarithm (approximately 2.71828)
- λ mean number of success that occur during a specific interval, $\lambda = np$
- k is the number of success

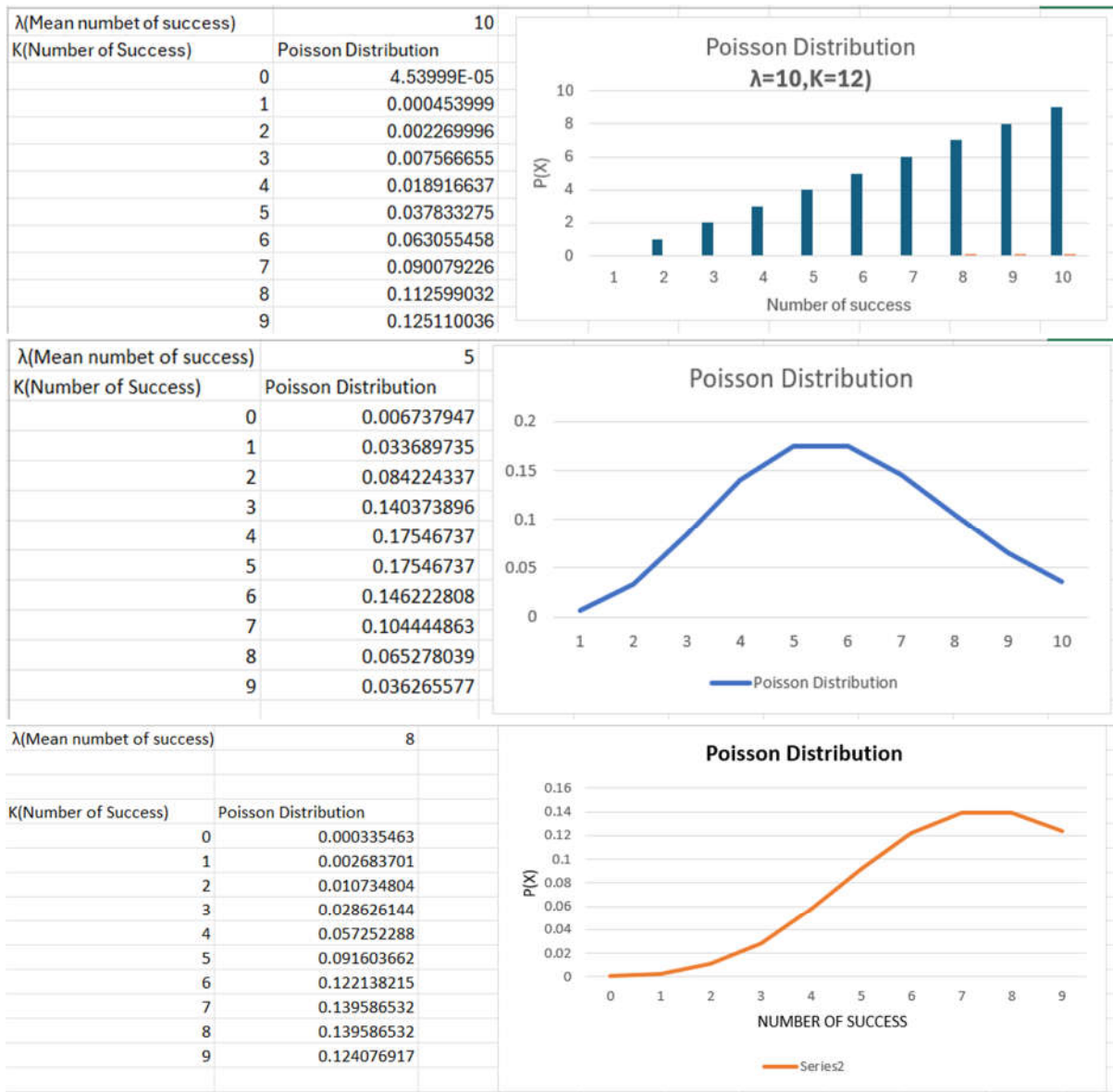
how to implement in excel

POISSON.DIST(number_s, average, cumulative)

POISSON.DIST(k , λ , FALSE)

3. Plotting and fitting of Poisson distribution and graphical representation of probabilities.

An electronic store sells on average 8 Desktop in a week. Assuming that purchases are as described above then what is the probability that the store will have turn away potential buyers before the end if the stock 9 computers ? how many computers should the store stock in order to make sure that it has 99% probability of being able to supply a week's demand ?



Geometric distribution

In a Bernoulli trial, the likelihood of the number of successive failures before success is obtained is represented by a geometric distribution, which is a sort of discrete probability distribution. A Bernoulli trial is a test that can only have one of two outcomes: success or failure. In other words, a Bernoulli trial is repeated until success is obtained and then stopped in geometric distribution.

A geometric distribution is a discrete probability distribution that indicates the likelihood of achieving one's first success after a series of failures. The number of attempts in a geometric distribution can go on indefinitely until the first success is achieved. Geometric distributions are probability distributions that are based on three key assumptions.

- *The trials that are being undertaken are self-contained.*
- *Each trial may only have one of two outcomes: success or failure.*
- *For each trial, the success probability, represented by p , is the same*

Geometric Distribution formula

$$P(X=k) = (1-p)^k p$$

Mean:	$\mu = E(X) = \frac{1}{p}$
Variance:	$\sigma^2 = V(X) = \frac{(1-p)}{p^2}$

where:

- **k**: number of failures before first success
- **p**: probability of success on each trial

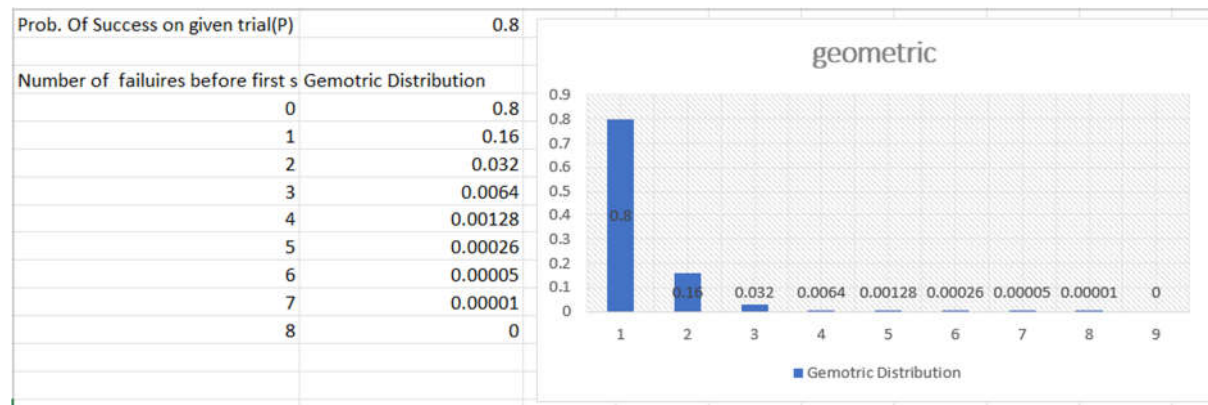
The chance of a trial's success is denoted by p , whereas the likelihood of failure is denoted by q , $q = 1 - p$ in this case. $X \sim G(p)$ represents a discrete random variable, X , with a geometric probability distribution.

how to implement in excel

Probability = $(1-p)^k * p$

4. Plotting and fitting of Geometric distribution and graphical representation of probabilities.

Suppose a programmer is waiting outside a PM office to take his views on artificial intelligence like they support AI or not .The probability that a PM supports the AI is $p = 0.8$. what is the probability that the fourth PM , the programmer talk to is the first PM to support AI ?



Uniform Distribution Function

A uniform distribution is a distribution that has constant probability due to equally likely occurring events. It is also known as rectangular distribution (continuous uniform distribution). It has two parameters a and b : a = minimum and b = maximum. The distribution is written as $U(a, b)$.

A uniform distribution is a type of probability distribution where every possible outcome has an equal probability of occurring. This means that all values within a given range are equally likely to be observed.

Uniform Distribution Formula

The [probability density function](#) (PDF) of a continuous uniform distribution defines the probability of a random variable falling within a particular interval. For a continuous uniform distribution over the interval $[a, b]$.

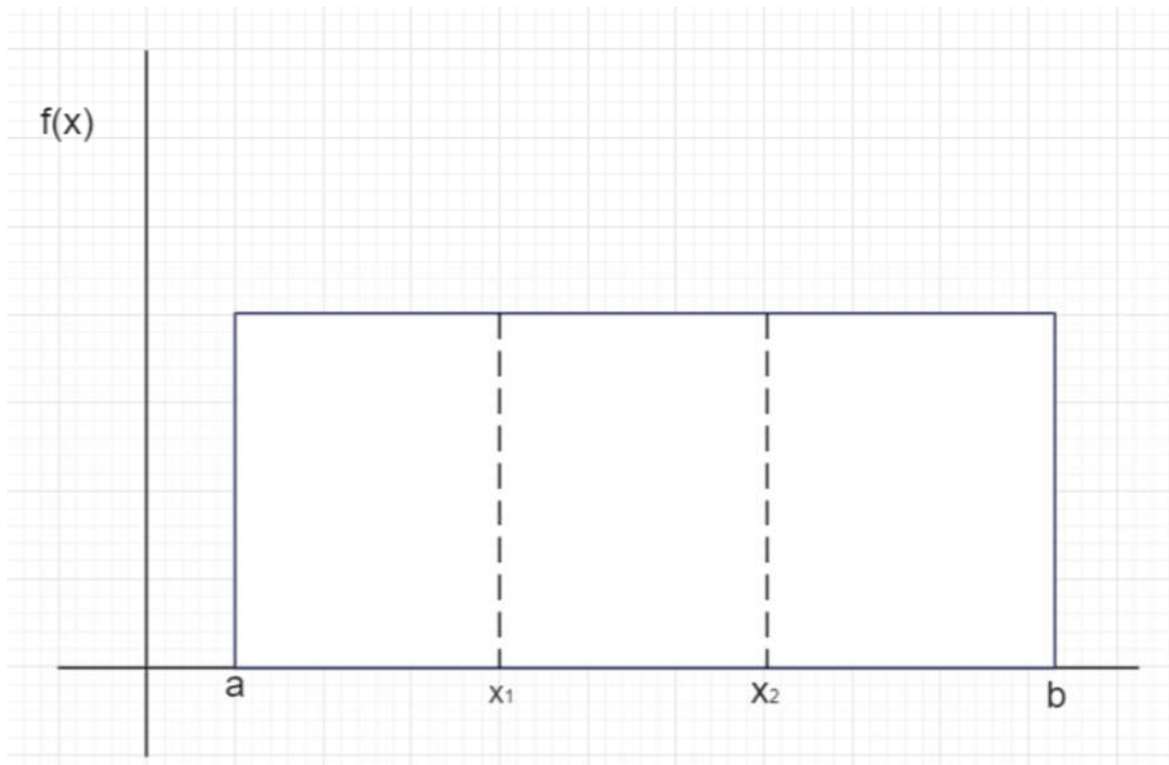
$$f(x) = \frac{1}{b - a} \text{ for } a \leq x \leq b$$

$$\text{Mean } \mu = \frac{a+b}{2}$$

$$\text{Variance } \sigma^2 = \frac{(b-a)^2}{12}$$

how to implement in excel

$$P = (x_2 - x_1) / (b - a)$$

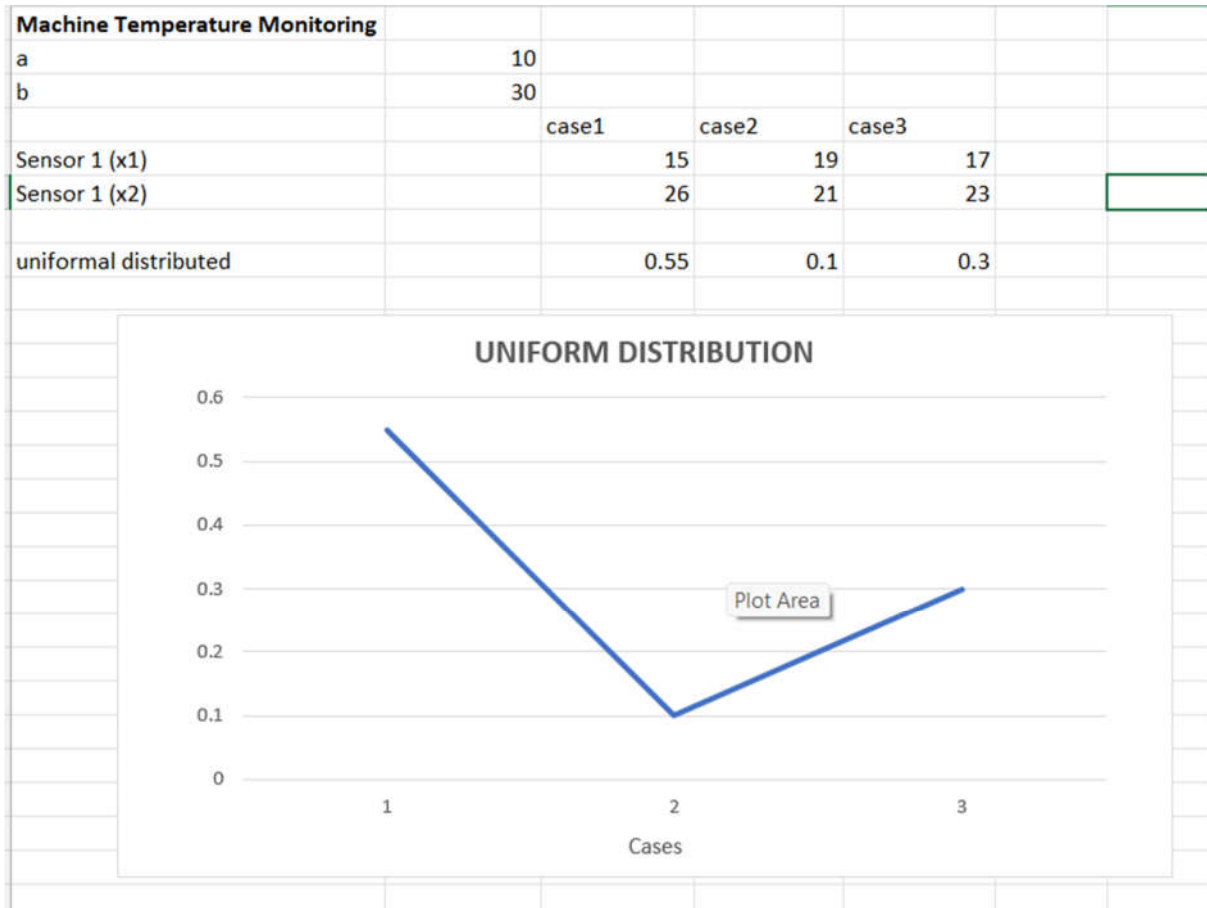


For calculating probability, we need:

- 1. a : minimum value in the distribution*
- 2. b : maximum value in the distribution*
- 3. x_1 : the minimum value you're interested in*
- 4. x_2 : the maximum value you're interested in*

5. Plotting and fitting of Uniform distribution and graphical representation of probabilities.

A sound sensor on x1x2 sensor takes 10 seconds and it will show up the result every time in 30 seconds. If you will use x1x2 sensor then what is the probability that a sensor will show up in 15-26 seconds ?



Exponential random variable

The support (set of values the Random Variable can take) of an Exponential Random Variable is the set of all positive real numbers. Suppose we are posed with the question- How much time do we need to wait before a given event occurs? The answer to this question can be given in probabilistic terms if we model the given problem using the Exponential Distribution. Since the time we need to wait is unknown, we can think of it as a Random Variable. If the probability of the event happening in a given interval is proportional to the length of the interval, then the Random Variable has an exponential distribution. The support (set of values the Random Variable can take) of an Exponential Random Variable is the set of all positive real numbers.

This distribution can be used to solve following type of real life problems-

- How long does a shop owner need to wait until a customer enter a shop.
- How long will a battery continue to work before it dies.
- How long will a computer continue to work before it breakdown.

$$f(x) = \begin{cases} \lambda e^{-\lambda * x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Here λ is the rate parameter and its effects on the density function .

e is a constant roughly equal to 2.718

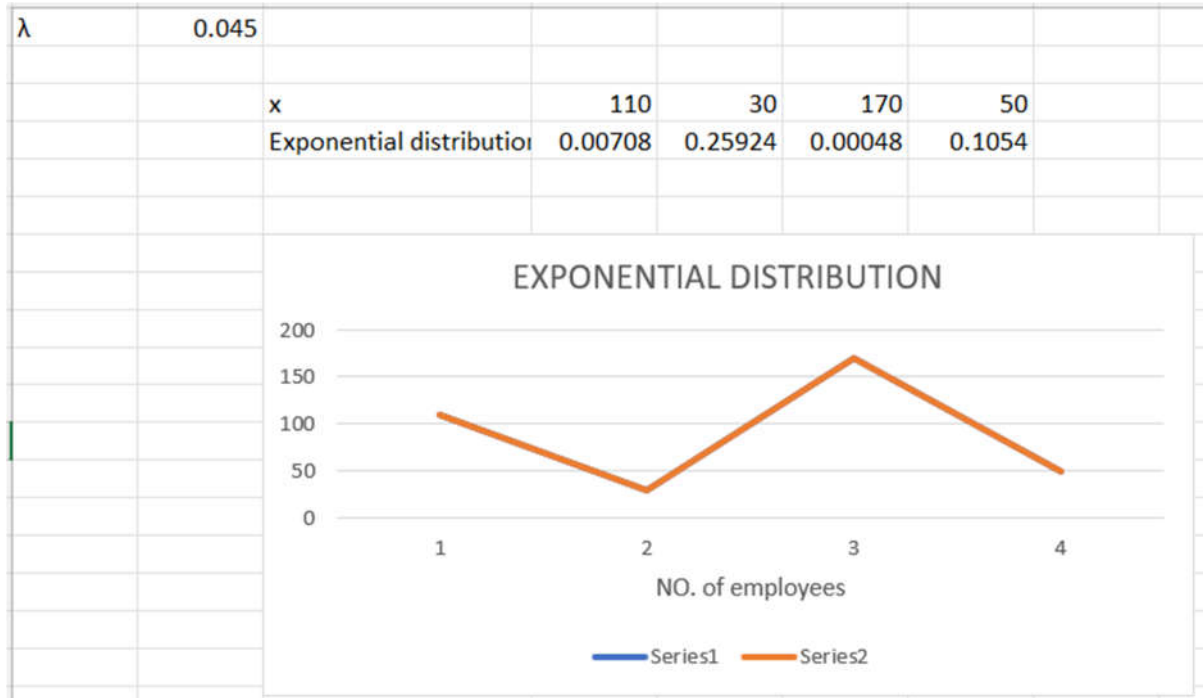
How to Implement in excel

EXPON.DIST(X, lambda, cumulative)

EXPON.DIST(X, lambda, FALSE)

6. Plotting and fitting of Exponential distribution and graphical representation of probabilities.

Suppose a big building constructing takes 100 days on average after an update occur find the probability that it will construct in more than 110 days and so on as in data ?



Normal Distribution

We define Normal Distribution as the probability density function of any continuous random variable for any given system. Now for defining Normal Distribution suppose we take $f(x)$ as the probability density function for any random variable X .

$$f(x) \geq 0 \quad \forall x \in (-\infty, +\infty),$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{where,} \quad \bullet \quad x$$

is Random Variable

- μ is Mean
- σ is Standard Deviation

Properties of Normal Distribution

- For normal distribution of data, mean, median, and mode are equal, (i.e., Mean = Median = Mode).
- Total area under the normal distribution curve is equal to 1.
- Normally distributed curve is symmetric at the center along the mean.
- In a normally distributed curve, there is exactly half value to the right of the central and exactly half value to the left side of the central value.
- Normal distribution is defined using the values of the mean and standard deviation.
- Normal distribution curve is a Unimodal Curve, i.e. a curve with only one peak

how to implement in excel

1. Input your data set into an Excel spreadsheet

2. Find the mean of your data set

=AVERAGE(cell range)

- "cell range" is a required component and the range of cells where your data exists, such as cells A1 through A64. You can write this in the function as A1:A64.

3. Find the standard deviation of your data set

`=STDEV(cell range)`

4. Select a value for the distribution

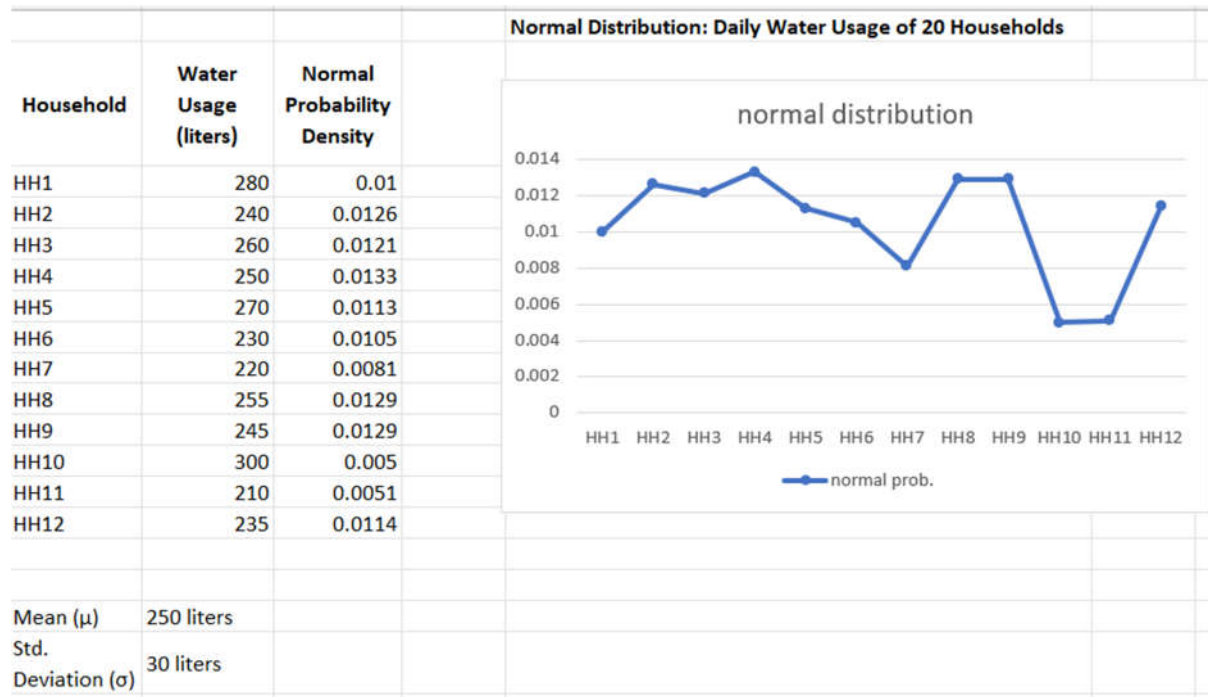
5. Type the *NORM.DIST* function and fill

`NORM.DIST(x, mean, standard deviation, cumulative)`

`NORM.DIST(x, mean, standard deviation, FALSE)`

7. Plotting and fitting of Normal distribution and graphical representation of probabilities.

Plot the normal graph according to the data of households and their water usage as shown below .



8. Calculation of cumulative distribution functions for Exponential and Normal distribution.

Calculate normal graph and exponential cumulative distribution according to the data of as students and their exam scores shown below .

Delivery Time (min)	Normal Distribution (Cumulative Probability)				Time Between Calls (sec)	Exponential Distribution (Cumulative Probability)
20	0.0062				60	0.2592
30	0.0668				90	0.3624
35	0.1587				120	0.4512
40	0.3085				150	0.5276
45	0.5				180	0.5934
50	0.6915				210	0.6502
55	0.8413				240	0.699
60	0.9332				270	0.7411
65	0.9772				300	0.7774
Mean (μ): 45 minutes					$\lambda = 0.00556$	
Standard Deviation (σ): 10 minutes						

Euclidean distance

Euclidean distance is the distance between two real distinct value .It is calculate by the square root of the sum of the squared difference elements in two vectors.

$$\text{Euclidean Distance} = |X - Y| = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2}$$

X: Array or vector X

Y: Array or vector Y

x_i: Values of horizontal axis in the coordinate plane

y_i: Values of vertical axis in the coordinate plane

n: Number of observations

how to implement in excel

= SORT(SUM X MYZ(array_X , array_Y))

9. Given data from two distributions, find the distance between the distributions.

In our case , we take data from the binomial and poisson distribution to find the euclidean distance between them.

trials	binomial distribution	poisson distribution	distance			
1	0.301989888	0.259867336	0.099021679			
2	0.301989888	0.049374794	0.09724737			
3	0.176160768	0.006254141	0.033432984	prob.	0.2	
4	0.066060288	0.000594143	0.004564722	k	9	
5	0.016515072	4.51549E-05	0.000278906	λ	0.38	
6	0.002752512	2.85981E-06	7.64781E-06			
7	0.000294912	1.55247E-07	8.7221E-08			
8	0.000018432	7.37422E-09	3.39467E-10			

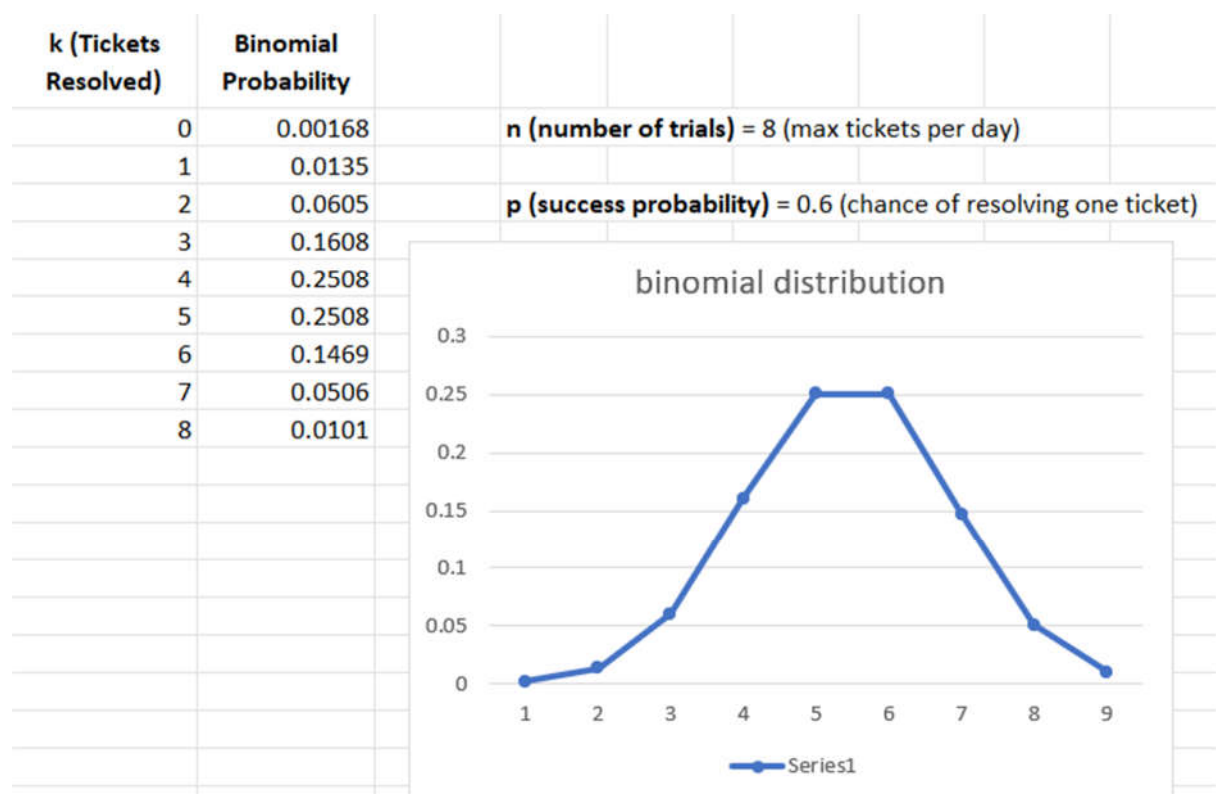
10. Application problems based on the Binomial distribution.

A customer support agent can handle a maximum of 8 support tickets per day. For each ticket, the probability of successfully resolving it is 0.6. The number of tickets resolved in a day follows a binomial distribution with:

- $n = 8$ trials (maximum tickets per day)
- $p = 0.6$ (probability of resolving a single ticket)

Based on this information:

- What is the probability that the agent resolves exactly 5 tickets in a day?
- What is the probability that the agent resolves at least 6 tickets in a day?
- Is it likely for the agent to resolve 2 or fewer tickets in a day? Justify your answer using the binomial probabilities.

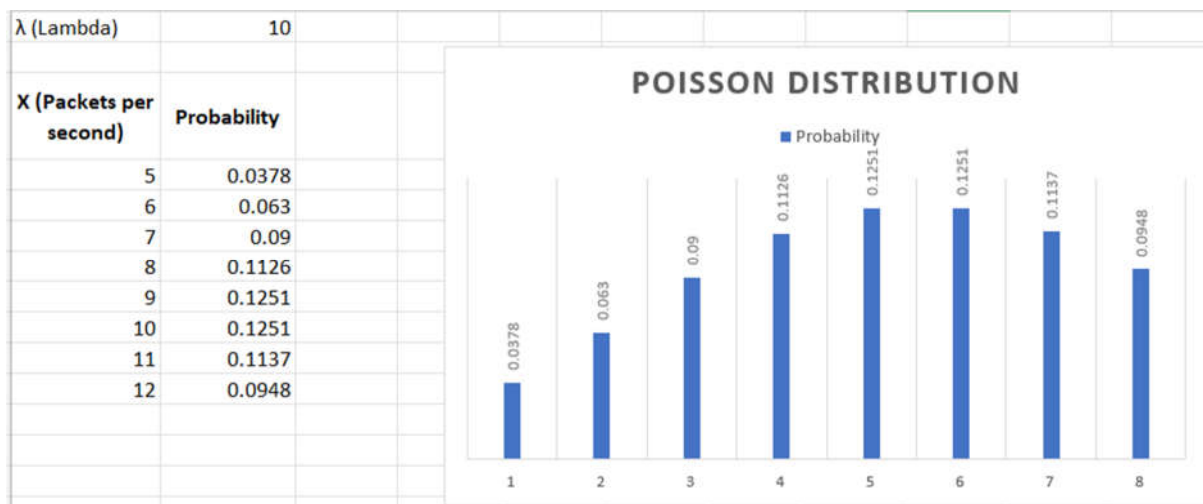


11. Application problems based on the Poisson distribution.

Ques:. : A network server receives data packets at an average rate of **10 packets per second**. The arrival of packets follows a **Poisson distribution** with a mean (λ) of **10**.

Using this model:

- What is the probability that exactly **9 packets** are received in a given second?
- What is the probability that the server receives **between 7 and 11 packets inclusive** in a second?
- Based on the distribution, would receiving **only 5 packets** in a second be considered unusual? Explain using the probability values.



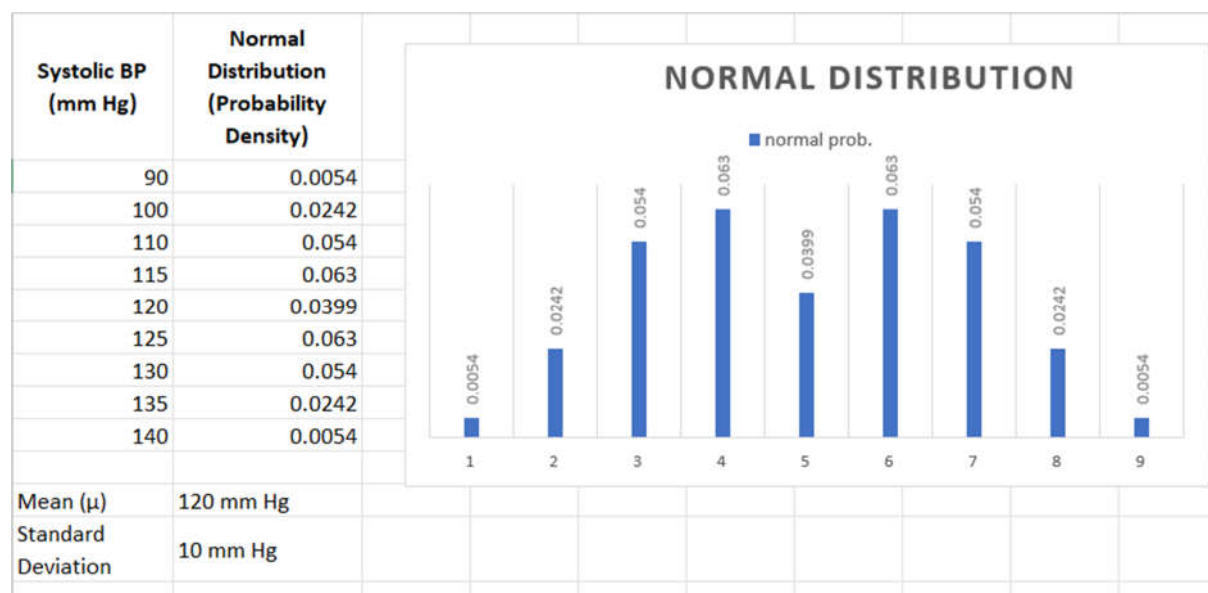
12. Application problems based on the Normal distribution.

Ques:: The systolic blood pressure of adults in a certain population is normally distributed with:

- Mean (μ) = 120 mm Hg
- Standard deviation (σ) = 10 mm Hg

Using this distribution:

- What is the probability density of a person having a systolic BP of 110 mm Hg?
- Compare the likelihood (based on probability density) of systolic BP being 115 mm Hg versus 130 mm Hg.
- Is it more common for individuals to have a systolic BP between 110 and 130 mm Hg, or outside this range? Justify your answer using the data and graph.



Bivariate Data/ Bivariate Analysis

Bivariate analysis is one of the statistical analysis where two variables are observed. One variable here is dependent while the other is independent. These variables are usually denoted by X and Y. So, here we analyse the changes occurred between the two variables and to what extent.

The term bivariate analysis refers to as the analysis of two variables . the objective of bivariate analysis to understand the relationship between two variables. There are three common way to analysis the bivariate analysis –

1. Scatter plots

2. Correlation Coefficient

3. Simple linear Regression(SLR)

Bivariate frequency distribution

A series of statistical data showing the frequency of two variables simultaneously is called Bivariate frequency distribution. In other words, the frequency distribution of two variable is called Bivariate frequency distribution. For example: sales and advertisement expenditure , weight and height of an individual.

Why bivariate frequency distribution is significant in business research ?

- 1. Decision Making*
- 2. Market-segmentation*
- 3. Risk-assessment*
- 4. Resource allocation*

how to implement in excel

= COVARIANCE.P(array1,array2)

The COVARIANCE.P function used the following arguments array1, this is range or array of integer value. array2 is also the second range or values.

Few things to remember about argument

1. If the given array contain text or logical value then are ignore by the Covariance function in excel.

2. The data should contain numbers, names, array or references that are numeric .IF the some cell do not contain numeric data they are ignored.

3. The data set should be same size with the same number of data points.

4. The data set should not be empty nor should the standard Deviation of the value equal .

$$\text{Cov}(X,Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

\bar{X} and \bar{Y} are the sample mean of the two set of values and n is the sample size.

5. Covariance is measure to indicate the extent to which two random variable in tandem.

6. Correlation is the measure used to represent how strongly two random variable are strongly related to each other.

7. Covariance is nothing but a measure of correlation.

8. Correlation referred to the scaled form of covariance.

9. Covariance can ary between $-\infty$ to $+\infty$ and correlation range between -1 to +1 .

10. Covariance indicate the direction of the linear relationship between variables .

11. Correlation on the other hand measure both the strength and direction of the linear relationship between two variables.

12. Covariance is affected by change in scale.

13. Correlation is not affected by the change in scale.

Pearson Correlation Coefficient formula

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

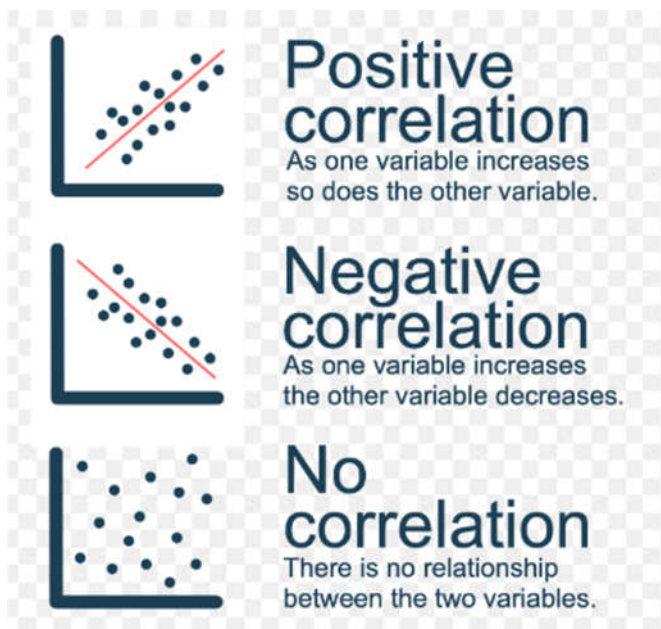
r = correlation coefficient

x_i = values of the x-variable in a sample

\bar{x} = mean of the values of the x-variable

y_i = values of the y-variable in a sample

\bar{y} = mean of the values of the y-variable



= PEARSON(array,array2)

Scatter plots

Scatter plots are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a **Cartesian system**. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis. These plots are often called **scatter graphs** or **scatter diagrams**.

Scatter plots instantly report a large volume of data. It is beneficial in the following situations

–

- For a large set of data points given
- Each set comprises a pair of values

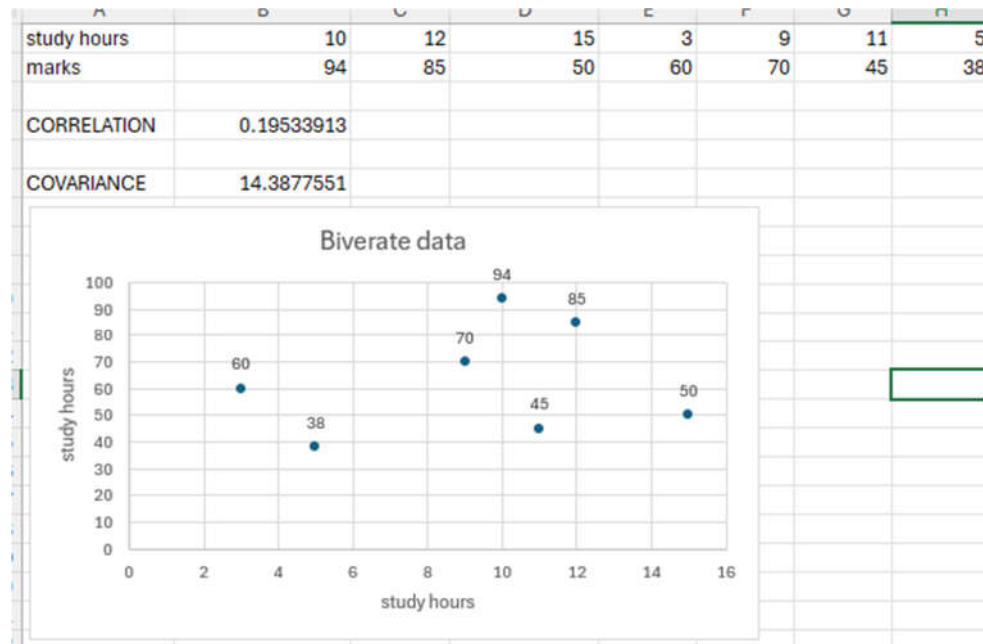
- *The given data is in numeric form*

13. Presentation of bivariate data through scatter-plot diagrams and calculations of covariance.

How to perform variant analysis and experiment using the following dataset having two variables .

a) hours spent studying and

b) exam score receive by students



Click insert tab along the top ribbon then click scatter chart within chart group.

CORRELATION in excel-

= **CORREL(hours, score)**

COVARIANCE in excel –

COVARIANCE.P(hours, score)

14. Calculation of Karl Pearson's correlation coefficients.

X	Y1	Y2	Y3	
	2	38	63	10
	3	56	59	35
	8	74	68	15
	11	33	79	24
	10	40	85	12
Karl pearson's correlation	XY1	-0.14493936		
	XY2	0.889803569		
	XY3	-0.140953713		

15. To find the correlation coefficient for a bivariate frequency distribution.

Marks	16_18	18_20	20_22	22_24	total	
10_20	5	1	1	0	7	
20_30	2	0	4	2	8	
30_40	1	3	5	3	12	
40_50	2	2	2	1	7	
50_60	3	0	3	1	7	
60_70	0	1	0	5	6	
total	13	7	15	12		

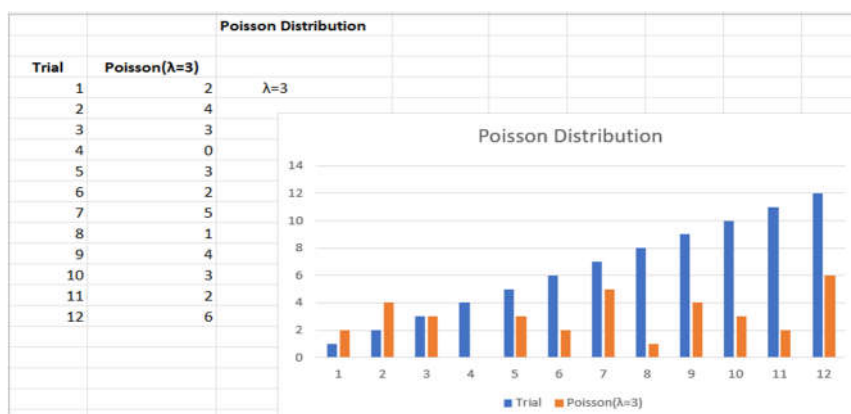
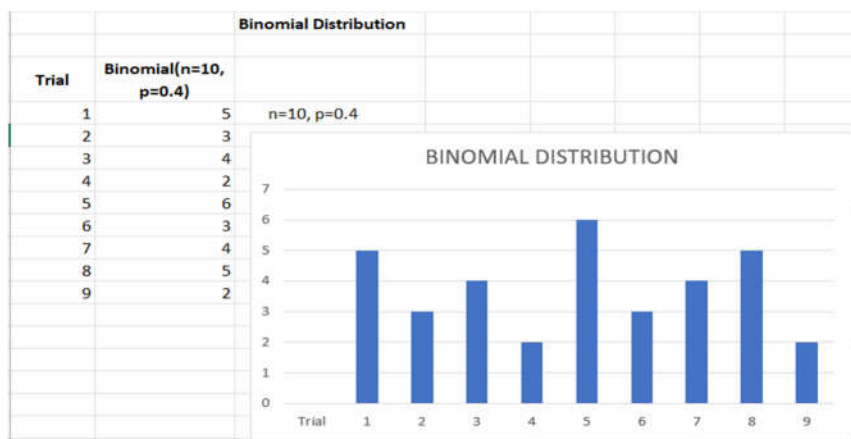
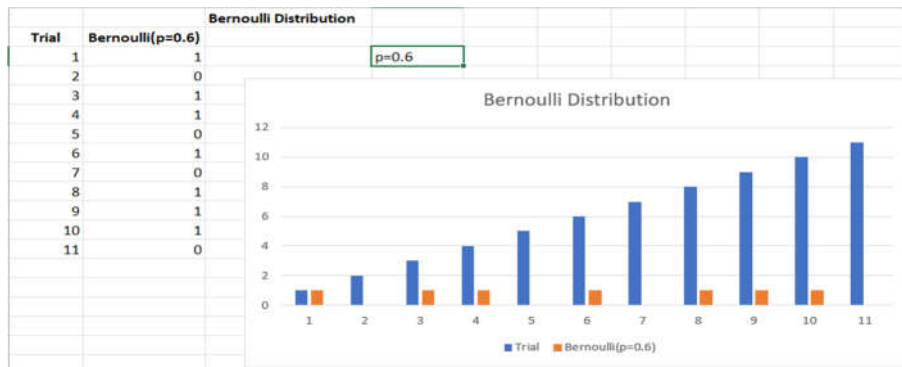
marginal frequency distribution of X:		
marks	total	
0	15	7
1	25	8
2	34	12
3	33	7
4	53	7
5	42	6

marginal frequency distribution of Y:		
age in years	total	
0	18	13
1	12	7
2	11	15
3	14	12

16. Generating Random numbers from discrete (Bernoulli, Binomial, Poisson) distributions.

How to implement in excel-

= **BINOM.INV (1,P, RAND())** will generate 1 or 0 with chance of 1 being P random number

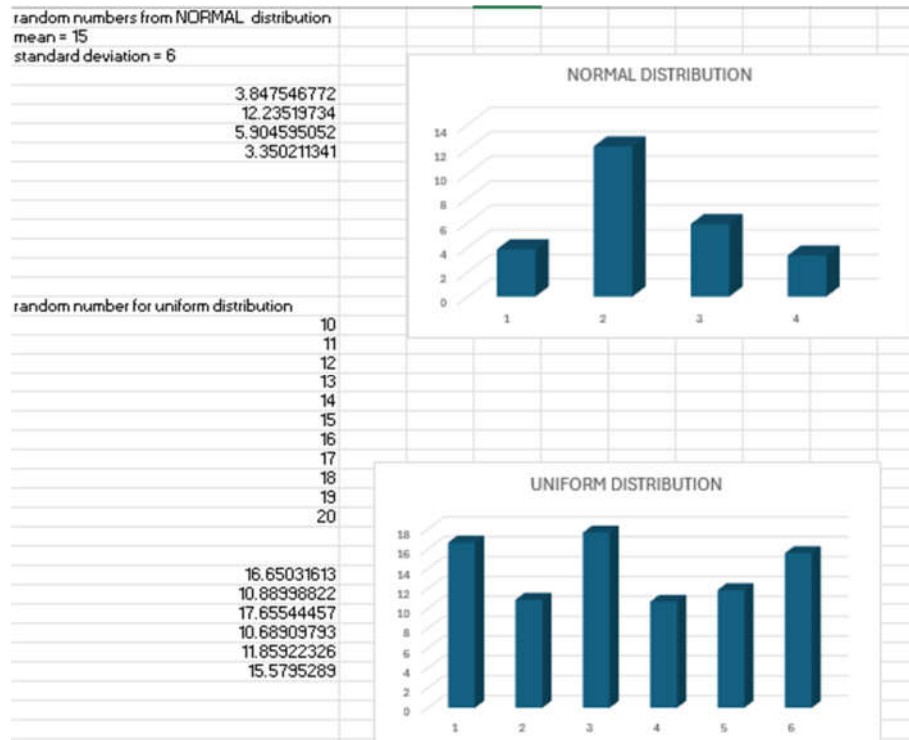


17. Generating Random numbers from continuous (Uniform, Normal) distributions.

How to implement in excel-

= NORMINV(RAND(),B2,C2)

Where this RAND() function create your probability . B2 provides you mean , C2 refers your standand deviation.



Entropy

The entropy of a random variable is the average level of information, surprise, or uncertainty inherent to the variable's possible outcomes. Given a discrete random variable X which takes value in the alphabet x and distributed according to the $P: x[0,1]$. The entropy is $H[X]$

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

The choice of base for log varies for different applications.

Base 2 gives the unit of bits while base e gives **natural units**.

Base e gives the units of $H(X)$.

An equivalent definition of entropy is the expected value of the self information of a variable.

Two bits of entropy

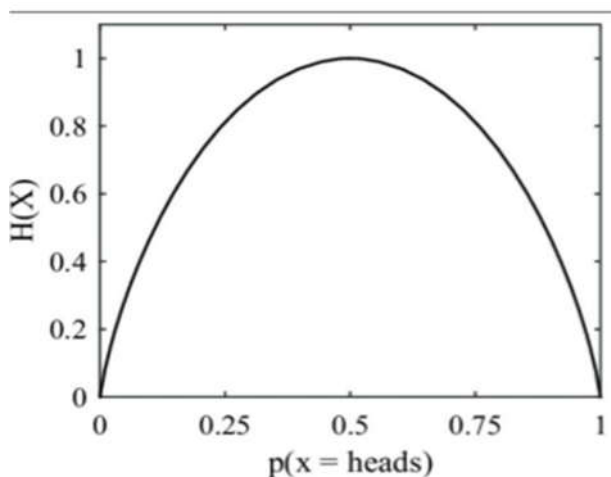
In case of a fair coin tosses, the information entropy in bits base 2 logarithm the number of possible outcomes with two coin, there are 4 possible outcomes and two bits of entropy

$\{HH, HT, TH, TT\}$

Generally information entropy is the average amount of information conveyed by an event, when considering all possible outcomes.

Example –

Entropy $H(X)$ of a coin flip measured in bits, graph versus the bias of the coin where $X=1$ represent entropy is the result of heads.



Here the entropy is at most 1 bit and to communicate the outcome of a coin flip (2 possible values – H or T) which requires an average of 1 bit (exactly 1 bit for a fair coin)

The result of a fair dice (6 possible values) have entropy $\log 6$ bits.

18. Find the entropy from the given data set.

