PRACTICAL FILE OF PROBABILITY FOR COMPUTING

RAMANUJAN COLLEGE

UNIVERSITY OF DELHI

DSC 06: PROBABILITY FOR COMPUTING

2024-25

SEMESTER 2

SUBMITTED BY	SUBMITTED TO
NAME: **GURPREET SINGH PAUL** ROLL NO. **2457022(24020570023)** COURSE: **B.Sc(H)Computer Science** SEMESTER:2	**Dr Aakash** Assistance Professor (Operational Research), Department of Computer Science, Ramanujan College, University of Delhi, CR Park Main Road, Block H, Kalkaji, New Delhi-110019

INDEX

Acknowledgement

I would like to express my heartfelt gratitude to everyone who contributed to the successful completion of this practical assignment.

First and foremost, I extend my sincere appreciation to my teacher, **Dr Aakash**, **Assistance Professor (Operational Research)**, **Ramanujan College, University of Delhi**, whose guidance, patience, and insightful feedback were invaluable throughout this process. Their dedication to fostering a deep understanding of mathematics has greatly enhanced my learning experience.

I am also grateful to my classmates for their encouragement and collaboration, which made the learning journey both engaging and enjoyable.

Name: Gurpreet Singh Paul

Roll no. 24020570023

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

Binomial Distribution:
The binomial distribution is a discrete probability distribution. It describes the outcome of binary scenarios, e.g. toss of a coin.

Binomial Distribution Formula:

$$P(x;n,p)={}^{n}c_{x}\, p^{x}(1-p)^{n-x}$$

Where:

- P(x; n, p) is the probability of x successes in n trials in an experiment which can result in exactly two outcomes (success or failure).
- p is the probability of success on an individual trial.
- n is the number of trials. x is the
- total number of successes.

Mean($\mu=np$)

Variance($\boldsymbol{\sigma}^2=npq$)

Implementation in Excel:

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

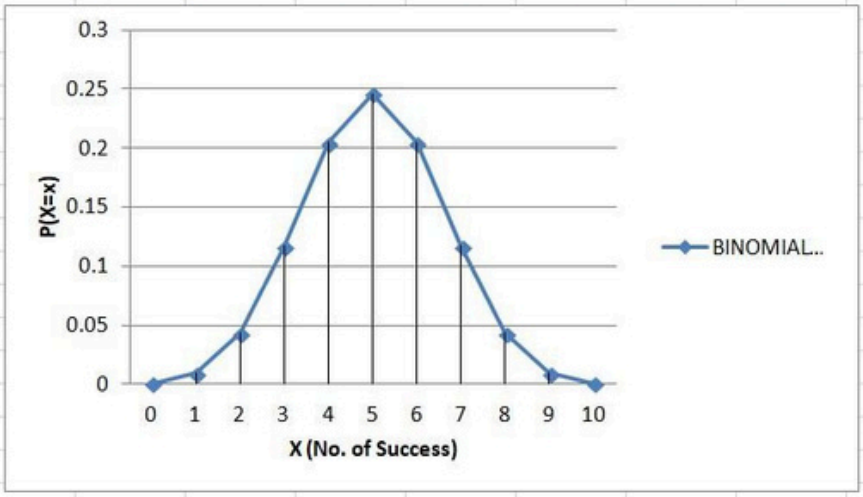
Where:

- number_s: number of successes. trials:
- total number of trials.
- probability_s: probability of success on each trial. cumulative: TRUE returns the cumulative
- probability; FALSE returns the exact probability

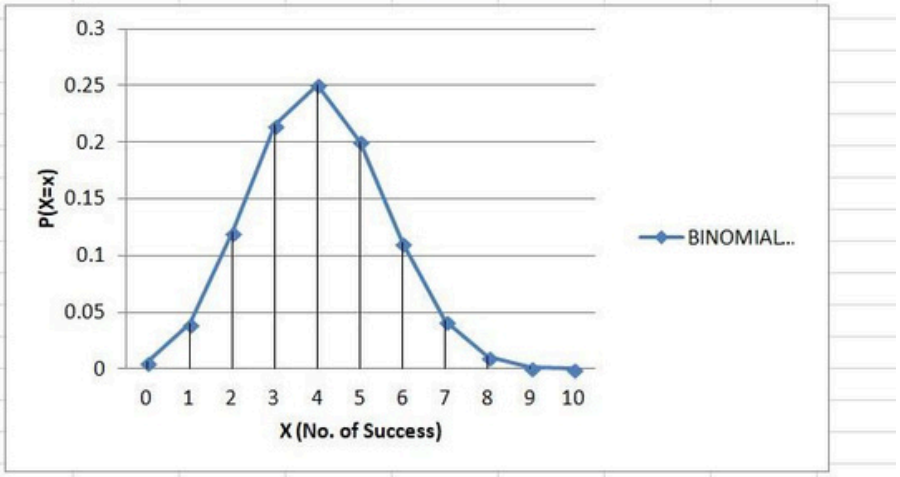
Skewness In Case Binomial Distribution Can Be Defined As Follows:

- If p=0.5, the binomial distribution will be symmetrical, regardless of the value of n.
- If p≠0.5, the distribution will be skewed.
- If p < 0.5, the distribution will be positively skewed or right-skewed. This means the bulk of the probability falls in the smaller numbers and the distribution tails off to the right.
- If p > 0.5, the distribution will be negatively skewed or left-skewed. This means the bulk of the probability falls in the larger numbers and the distribution tails off to the left.

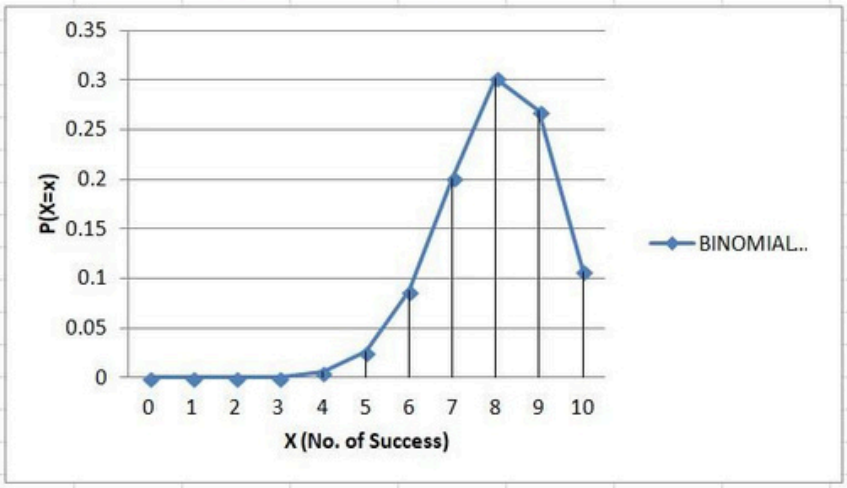
	A	B
1	Number of Trials	10
2	Probability of Success	0.5
3		
4	X (No. Of Success)	P(X=x)
5	0	0.000976563
6	1	0.009765625
7	2	0.043945313
8	3	0.1171875
9	4	0.205078125
10	5	0.24609375
11	6	0.205078125
12	7	0.1171875
13	8	0.043945313
14	9	0.009765625
15	10	0.000976563
16		



	A	B
1	Number of Trials	10
2	Probability of Success	0.4
3		
4	X (No. Of Success)	P(X=x)
5	0	0.006046618
6	1	0.040310784
7	2	0.120932352
8	3	0.214990848
9	4	0.250822656
10	5	0.200658125
11	6	0.111476736
12	7	0.042467328
13	8	0.010616832
14	9	0.001572864
15	10	0.000104858
16		



	A	B
1	Number of Trials	10
2	Probability of Success	0.8
3		
4	X (No. Of Success)	P(X=x)
5	0	1.024E-07
6	1	0.000004096
7	2	7.3728E-05
8	3	0.000786432
9	4	0.005505024
10	5	0.026424115
11	6	0.088080384
12	7	0.201326592
13	8	0.301989888
14	9	0.268435456
15	10	0.107374182
16		



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

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, p1, p2, ... , pk(so that pi ≥ 0 for i = 1 ,..., k and Σi=1kpi=1), and there are n independent trials. Then let the random variables Xi indicate the number of times outcome number i was observed over the n trials. Then X=(X1, X2 ,...,Xk) follows
a multinomial distribution with parameters n and p, where p=(p1, p2
,...,pk),

Multinomial Distribution Formula

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

$$\mathrm{Cov}(X_i, X_j) = -np_ip_j \quad (i \neq j)$$

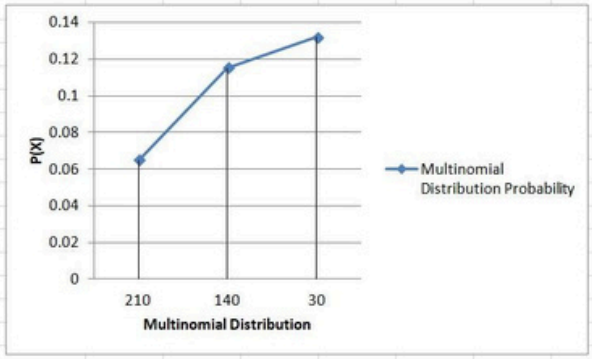
When X=(x_{1}, x_{2} ,...,x k) follows a multinomial distribution with the PMF given above, X_{i} follows a binomial distribution with n trials and success probability pi.

how to implement in excel

Multinomial MULTINOMIAL(X1,X2,X3)

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

	A	B	C	D
2	X1	4	3	2
3	X2	0	1	1
4	X3	6	3	2
5				
6	P1	0.375	0.375	0.375
7	P2	0.125	0.125	0.125
8	P3	0.5	0.5	0.5
9				
10	Multinomial	210	140	30
11				
12	Multinomial Distribution Probability	0.064888	0.115356445	0.131835938
13				
14	P1^X1	0.019775391	0.052734375	0.140625
15	P1^X2	1	0.125	0.125
16	P1^X3	0.015625	0.125	0.25
17				
18				
19				
20				
21				



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

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

$$P(K) = e\_\boldsymbol{\lambda}(\boldsymbol{\lambda}^k / k!)$$

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

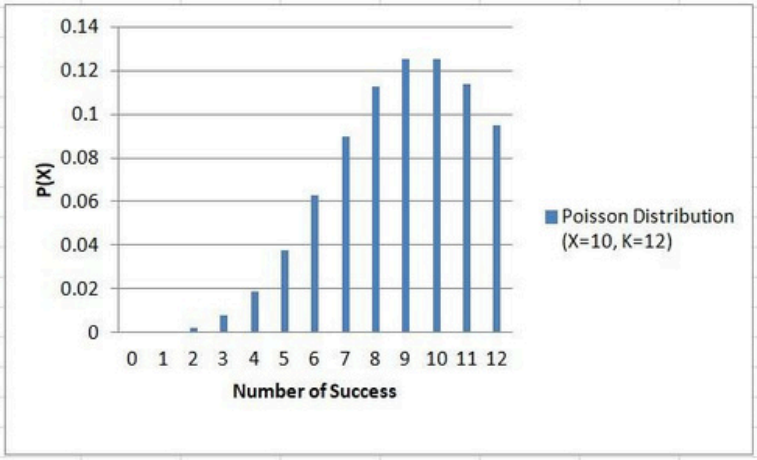
Where:

- P(X=k) is the probability of observing k events
- e is the base of the natural logarithm (approximately 2.71828)
- A mean number of success that occur during a specific interval,∋λnp k
- is the number of success how to implement in excel

POISSON.DIST(number_s,average,cumulative)

POISSON.DIST(k, **λ** ,FALSE)

	A	B
1	X (Mean number of Success)	10
2		
3	K (Number of Success)	Poisson Distribution
4	0	4.53999E-05
5	1	0.000453999
6	2	0.002269996
7	3	0.007566655
8	4	0.018916637
9	5	0.037833275
10	6	0.063055458
11	7	0.090079226
12	8	0.112599032
13	9	0.125110036
14	10	0.125110036
15	11	0.113736396
16	12	0.09478033
17		
18		



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

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

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

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

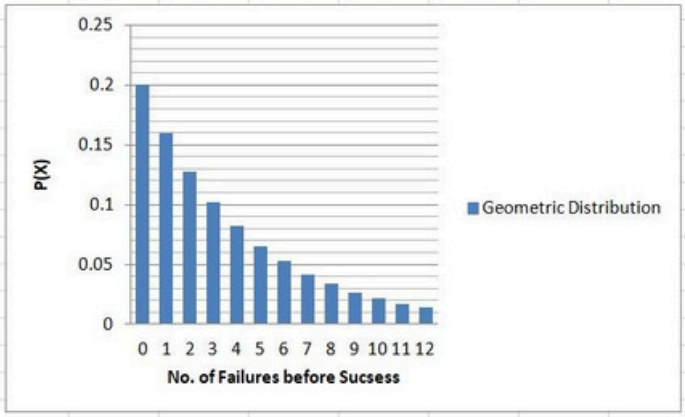
Geometric Distribution

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~G (p) represents a discrete random variable, X, with a geometric probability distribution.

.

	A	B
1	Probability of Success on Given Trial(P)	0.2
2		
3	No. Of Failures before first success	Geometric Distribution
4	0	0.2
5	1	0.16
6	2	0.128
7	3	0.1024
8	4	0.08192
9	5	0.065536
10	6	0.0524288
11	7	0.04194304
12	8	0.033554432
13	9	0.026843546
14	10	0.021474836
15	11	0.017179869
16	12	0.013743895



how to implement in excel
Probability = (1-p)^k*p

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

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 maximum. The distribution is written as U(a, b) 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].

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

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

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

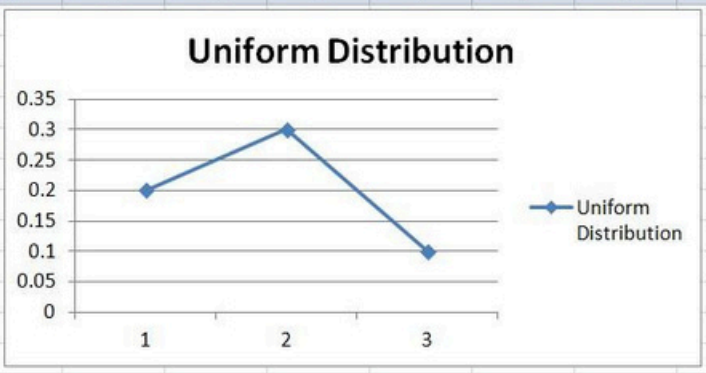
how to implement in excel

P = (x2 - x1) / (b - a)

For calculating probability, we need:

1. a: minimum value in the distribution
2. b: maximum value in the distribution
3. x1: the minimum value you're interested in
4. x2: the maximum value you're interested in

	A	B	C	D
1	a	10		
2	b	20		
3				
4		Case 1	Case 2	Case 3
5	X1	11	14	18
6	X2	13	17	19
7				
8	Uniform Distribution	0.2	0.3	0.1
9				
10				
11				
12				
13				



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

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.

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

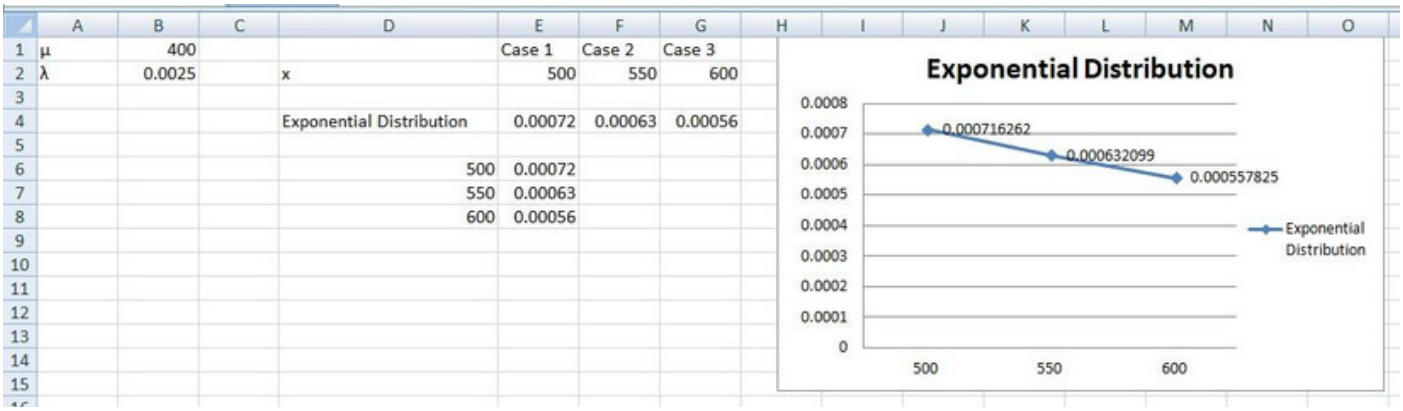
$$E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Here 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)



	A	B	C	D	E	F	G
1	μ	400			Case 1	Case 2	Case 3
2	λ	0.0025		x	500	550	600
3							
4				Exponential Distribution	0.00072	0.00063	0.00056
5							
6				500	0.00072		
7				550	0.00063		
8				600	0.00056		
9							
10							
11							
12							
13							
14							
15							

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

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.

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

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

where, 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 right 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

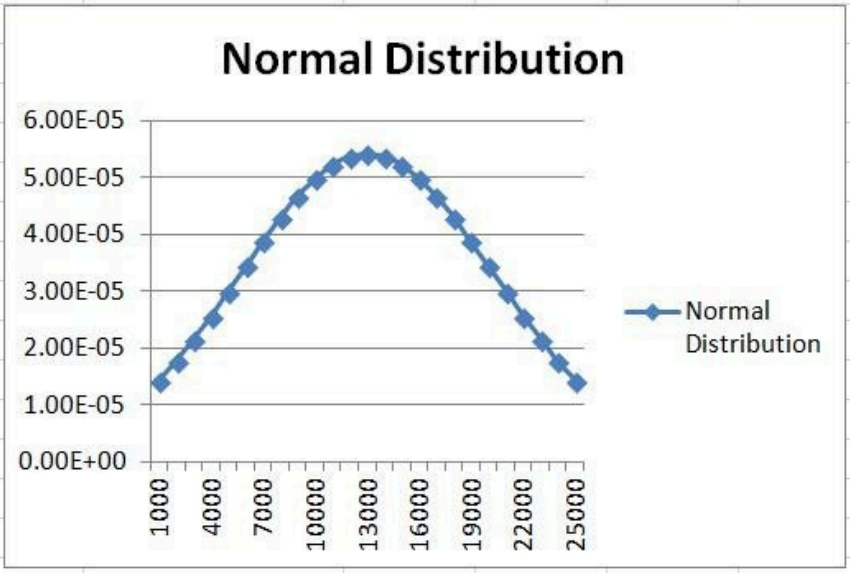
=STDEV(cell range)

4. Select a value for the distribution

5. Type the NORM.DIST function and fill

NORM.DIST(x,mean, standarddeviation,cumulative)
NORM.DIST(x,mean, standarddeviation, FALSE)

	B	C	D	E	F	G	H	I
1	Score	Normal Distribution		Mean	13000			
2	1000	1.43E-05		Standard Deviation	7359.801			
3	2000	1.77E-05						
4	3000	2.15E-05						
5	4000	2.57E-05						
6	5000	3.00E-05						
7	6000	3.45E-05						
8	7000	3.89E-05						
9	8000	4.30E-05						
10	9000	4.68E-05						
11	10000	4.99E-05						
12	11000	5.22E-05						
13	12000	5.37E-05						
14	13000	5.42E-05						
15	14000	5.37E-05						
16	15000	5.22E-05						
17	16000	4.99E-05						
18	17000	4.68E-05						
19	18000	4.30E-05						
20	19000	3.89E-05						
21	20000	3.45E-05						
22	21000	3.00E-05						
23	22000	2.57E-05						
24	23000	2.15E-05						
25	24000	1.77E-05						
26	25000	1.43E-05						



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

	A	B	C	D	E
1			NORMAL DISTRIBUTION		EXPONENTIAL DISTRIBUTION
2	Employee	Incentives	Normal Distribution	No. of days	Exponential Distribution
3	EMP1	1000	0.06859975	300	0.527633447
4	EMP2	2000	0.123838134	320	0.550671036
5	EMP3	3000	0.204480671	350	0.58313798
6	EMP4	4000	0.310147008	400	0.632120559
7	EMP5	5000	0.434415098	420	0.650062251
8	EMP6	6000	0.565584902	450	0.675347533
9	EMP7	7000	0.689852992	500	0.713495203
10	EMP8	8000	0.795519329	520	0.727468207
11	EMP9	9000	0.876161866	550	0.747160404
12	EMP10	10000	0.93140025	600	0.77686984
13					
14					
15	mean	5500		μ	400
16	standard d	3027.6504		λ	0.0025

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

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.

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

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

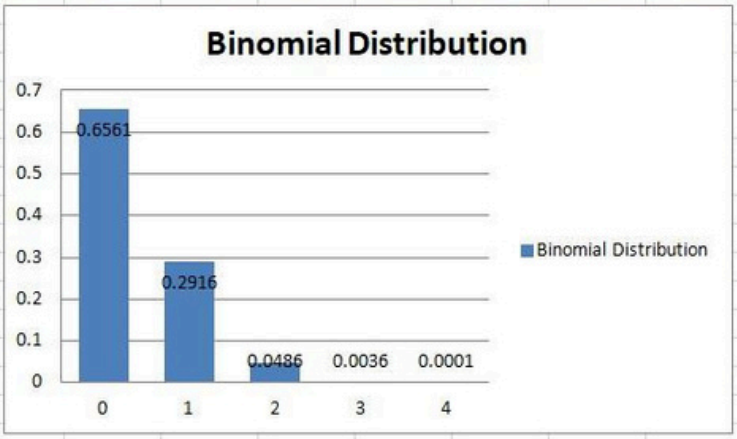
= SORT(SUM X MYZ(array_X, array_Y))

	A	B	C	D	E	F	G
1	Trials	Binomial Distribution	Poisson Distribution		Distance		
2	1	0.08957952			0.178548	n	0.4
3	2	0.20901888			0.155393	k	8
4	3	0.27869184			0.271542	λ	0.4
5	4	0.2322432			0.231528		
6	5	0.12386304			0.123806		
7	6	0.04128768			0.041284		
8	7	0.00786432			0.007864		
9	8	0.00065536			0.000655		
10							

10. Application problems based on the Binomial distribution.

Ques: Antibiotics occasionally cause nausea as a side effect. A major drug company has developed a new antibiotic called Phe-Mycin. The company claims that, at most, 10 percent of all patients treated with Phe-Mycin would experience nausea as a side effect of taking the drug. Suppose that we randomly select n = 4 patients and treat them with Phe-Mycin. Each patient will either experience nausea (which we arbitrarily call a success) or will not experience nausea (a failure). We will assume that p, the true probability that a patient will experience nausea as a side effect, is.10, the maximum value of p claimed by the drug company. Furthermore, it is reasonable to assume that patients' reactions to the drug would be independent of each other. Let x denote the number of patients among the four who will experience nausea as a side effect. It follows that x is a binomial random variable, which can take on any of the potential values 0, 1, 2, 3, or 4. That is, anywhere between none of the patients and all four of the patients could potentially experience nausea as a side effectSuppose that we wish to investigate whether p, the probability that a patient will experience nausea as a side effect of taking Phe-Mycin, is greater than. 10, the maximum value of p claimed by the drug company. This assessment will be made by assuming, for the sake of argument, that p equals .10, and by using sample information to weigh the evidence against this assumption and in favor of the conclusion that p is greater than.10. Suppose that when a sample of n=4 randomly selected patients is treated with PheMycin, three of the four patients experience nausea. Because the fraction of patients in the sample that experience nausea is 3/4.75, which is far greater than .10, we have some evidence contradicting the assumption that p equals.10.

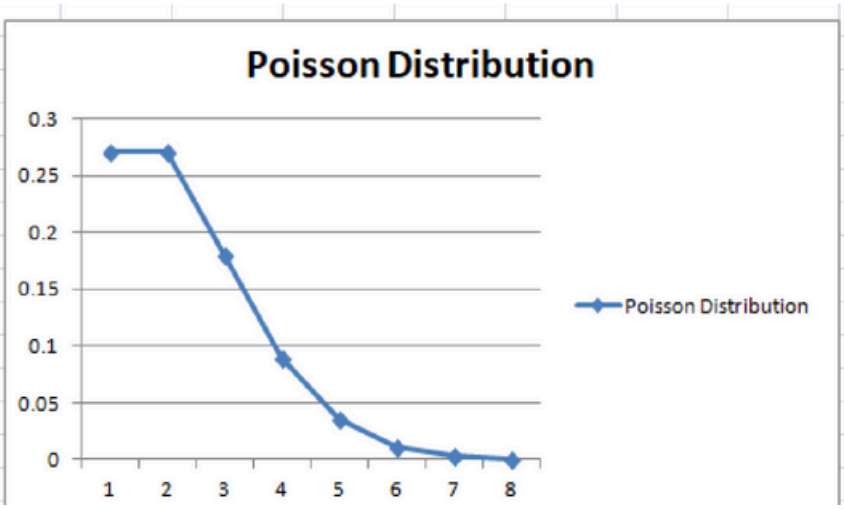
	A	B
1	n(Number of Trials)	4
2	p(Probablity of Success on given tria	0.1
3		
4		
5	k(Number of Success)	Binomial Distribution
6	0	0.6561
7	1	0.2916
8	2	0.0486
9	3	0.0036
10	4	0.0001



11. Application problems based on the Poisson distribution.

Ques: In a cafe, the customer arrives at a mean rate of 2 per min. Find the probability of arrival of 5 customers in 1 minute using the Poisson distribution formula.

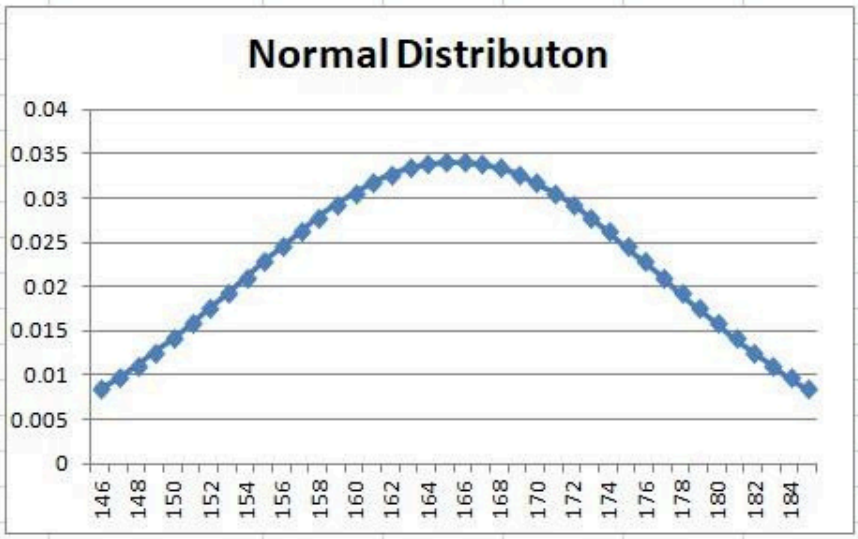
1	λ	2
2	x	5
3		
4		
5	x	Poisson Distribution
6	1	0.270670566
7	2	0.270670566
8	3	0.180447044
9	4	0.090223522
10	5	0.036089409
11	6	0.012029803
12	7	0.003437087
13	8	0.000859272
14		
15		



12.Application problems based on the Normal distribution.

Ques:: According to the website of the American Association for Justice, ^11 Stella Liebeck of Albuquerque, New Mexico, was severely burned by McDonald's coffee in February 1992. Liebeck, who received third-degree bums over 6 percent of her body, was awarded $160,000 in compensatory damages and $480,000 in punitive damages. A post-verdict investigation revealed that the coffee temperature at the local Albuquerque McDonald's had dropped from about 185 degree F before the trial to about 158 degree after the trial. This case concerns coffee temperatures at a fast-food restaurant. Because of the possibility of future litigation and to possibly improve the coffee's taste, the restaurant wishes to study the temperature of the coffee it serves. To do this, the restaurant personnel measure the temperature of the coffee being dispensed (in degrees Fahrenheit at a randomly selected time during each of the 24 half-hour periods from 8 a.m. to 7:30 p.m on a given day. This is then repeated on a second day, giving the 48 coffee temperatures in excel.

	Temp.	Normal Distribution		mean	165.5
2	146	0.008489973		Standard Deviatio	11.69045
3	147	0.00975627			
4	148	0.011129703			
5	149	0.012603918			
6	150	0.014169346			
7	151	0.015813073			
8	152	0.017518825			
9	153	0.019267081			
10	154	0.021035318			
11	155	0.022798407			
12	156	0.02452913			
13	157	0.026198838			
14	158	0.027778202			
15	159	0.029238056			
16	160	0.030550272			
17	161	0.031688662			
18	162	0.032629841			
19	163	0.033354025			
20	164	0.033845722			
21	165	0.034094282			
22	166	0.034094282			
23	167	0.033845722			
24	168	0.033354025			
25	169	0.032629841			
26	170	0.031688662			
27	171	0.030550272			
28	172	0.029238056			
29	173	0.027778202			
30	174	0.026198838			
31	175	0.02452913			
32	176	0.022798407			
33	177	0.021035318			
34	178	0.019267081			
35	179	0.017518825			
36	180	0.015813073			
37	181	0.014169346			
38	182	0.012603918			
39	183	0.011129703			



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

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-assessment4. Resource allocation how to implement in excel

= COVARIANCE.P(array1,array2)

The COVARIANCE.P function used the following arguments array 1, 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 donot 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.

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

X and 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 relatedto 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 -00 to +00 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 relationshipbetween 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

Positive correlation
As one variable increases so does the other variable.

Negative correlation
As one variable increases the other variable decreases.

No correlation
There is no relationship between the two variables.

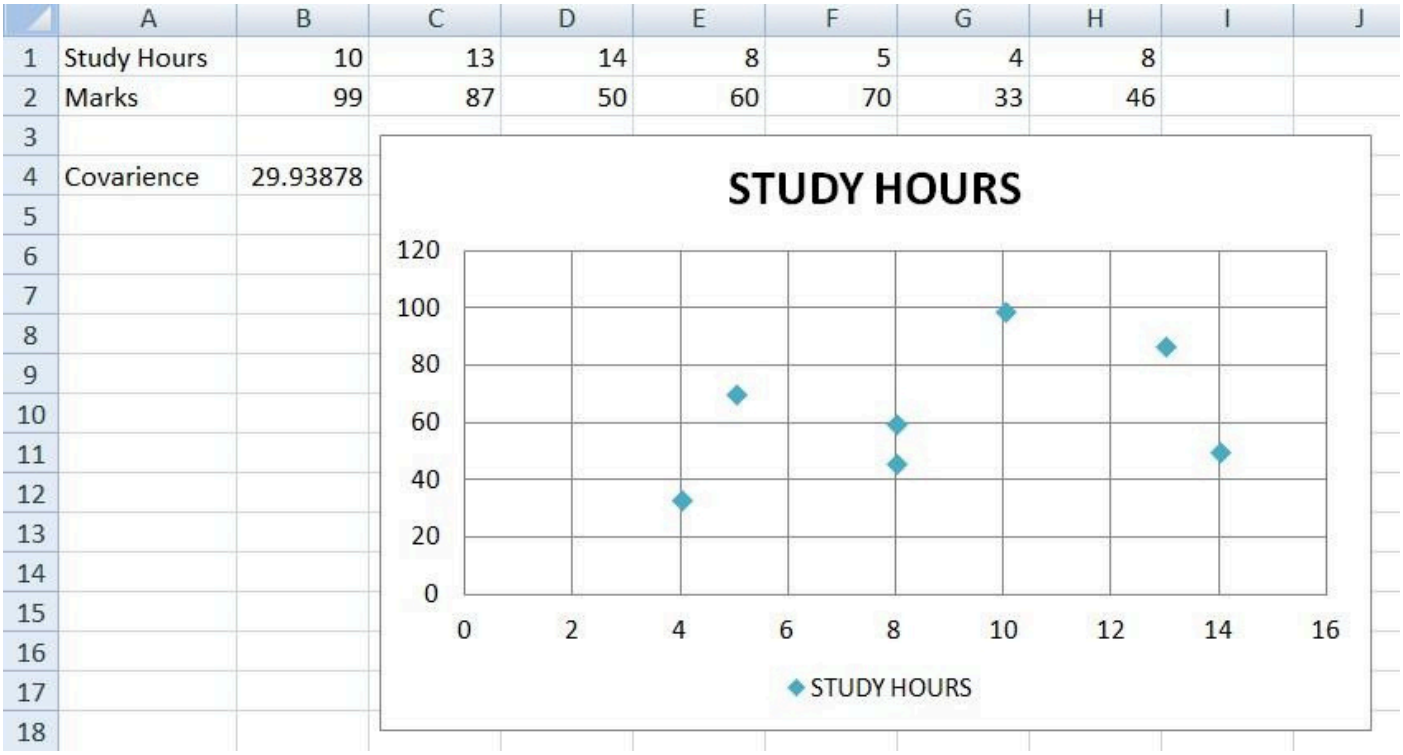
= PEARSON(arry:arry2)

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



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

CORRELATION in excel and COVARIANCE in excel -

= CORREL(hours,score)

= COVARIANCE.P(hours, score)

14. Calculation of Karl Pearson's correlation coefficients.

	A	B	C	D	E
1	X	Y1	Y2	Y3	
2	2	80	65	10	
3	5	95	69	30	
4	6	76	60	15	
5	8	58	95	25	
6	10	67	80	10	
7					
8	Karl Pearson's Correlation	XY1	-0.6273178		
9		XY2	0.63322478		
10		XY3	0.01814885		
11					

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

	A	B	C	D	E	F	G	H	I
1			Age in Years					Marginal Frequency Distribution of X:	
2	Marks	16_18	18_20	20_22	22_24	Total		Marks	Total
3	10_20	2	1	1	0	4		15	4
4	20_30	3	2	3	1	9		25	9
5	30_40	3	3	5	6	17		35	17
6	40_50	2	2	3	4	11		45	11
7	50_60	0	1	2	2	5		55	5
8	60_70	0	1	2	1	4		65	4
9									
10								Age in Years	Total
11	correlation coefficient along x-axis							17	10
12	-0.18773							19	10
13								21	16
14	correlation coefficient along Y-axis							23	14
15	0.774597								

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

1	Random numbersfrom Binomial distributions
2	N=10
3	P=0.4
4	5
5	4
6	4
7	5
8	4

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

= NORMINV(RAND(),B2,C2)

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

	A	B	C
1	Random numbers from Normal distributions		
2	mean	20	
3	standard Deviation	5	
4			
5		25.03399	
6		6.594571	
7		22.82169	
8		10.33076	

18 Find the entropy from the given data set.

The entropy of a random variable is the average level of information, surprise, or uncertainly inherent to the variable's possible outcomes. Given a discrete random 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 o gives the units of H(X)

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

	A	B	C	D
1		P(X)	log2P(X)	H(X)
2	HH	1	0	0
3	HT	0.5	-1	0.5
4	TH	0.5	-1	0.5
5	TT	0	0	0