name: - Shubham Kapoor

Roll No.: - 23020570048

Subject: - Probability for computing

Acknowledgement

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

I am deeply indebted to my mathema cs Professor Dr Aakash for his guidance and sugges ons for this project. The comple on 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.

INDEX

S.No	Topic
1.	Plotting and fitting of Binomial distribution and graphical representation of probabilities.
2.	Plotting and fitting of Multinomial distribution and graphical representation of probabilities.
3.	Plotting and fitting of Poisson distribution and graphical representation of probabilities
4.	Plotting and fitting of Geometric distribution and graphical representation of probabilities.
5	Plotting and fitting of Uniform distribution and graphical representation of probabilities.
6.	Plotting and fitting of Exponential distribution and graphical representation of probabilities.
7.	Plotting and fitting of Normal distribution and graphical representation of probabilities.
8	Calculation of cumulative distribution functions for Exponential and Normal distribution.
9.	Given data from two distributions, find the distance between the distributions.
10.	Application problems based on the Binomial distribution
11.	Application problems based on the Poisson distribution.
12.	Application problems based on the Normal distribution.
13.	Presentation of bivariate data through scatter-plot diagrams and calculations of covariance.
14	Calculation of Karl Pearson's correlation coefficients.
15.	To find the correlation coefficient for a bivariate frequency distribution

AS
7/5/24

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

Introduction:

- The Binomial distribution is used to model the probability of obtaining a certain number of successes in **n independent Bernoulli Trials**.
- It's PMF, variance & mean are given by:

$$
\begin{aligned}
\mathrm{E}\left[X^2\right] &= \sum_{x \in R_X} x^2 p_X(x) \\
&= 1^2 \cdot p_X(1) + 0^2 \cdot p_X(0) \\
&= 1 \cdot p + 0 \cdot (1-p) \\
&= p \\
\mathrm{E}[X]^2 &= p^2 \\
\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 = p - p^2 = p(1-p)
\end{aligned}
$$

- Graphical representation involves plotting the probability mass function (PMF) to show the probability of each possible outcome.
- A Binomial distribution can be approximated by a normal distribution when the number of trials are large, and by a Poisson distribution when the success probability is very low (typically <0.2).
- It is used in modelling anything having either a success or a failure, nothing in between, for example: a dice roll.
- In EXCEL, **BINOM.DIST(K,n,p,FALSE) .**

Application:

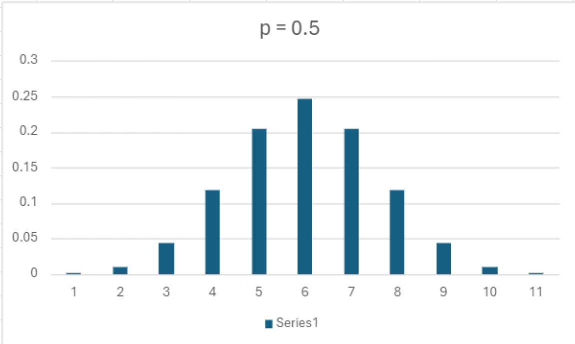
Problem statement:

- Suppose that an xyz electronic company produces both wired and wireless mice. The production output is 50% wired & 50% wireless mice. If we choose 10 mice at random & choosing a wireless mouse is a success, then plot the graph of this distribution with p = 0.5 as the success.

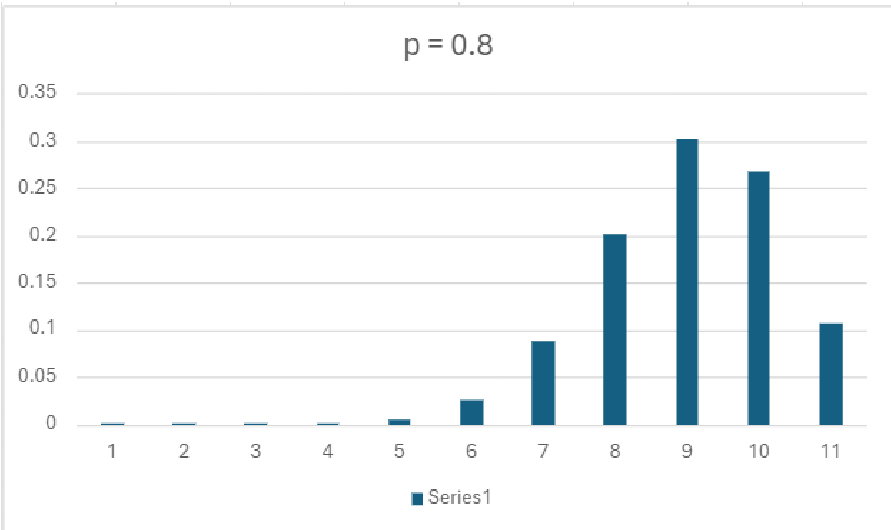
Answer:

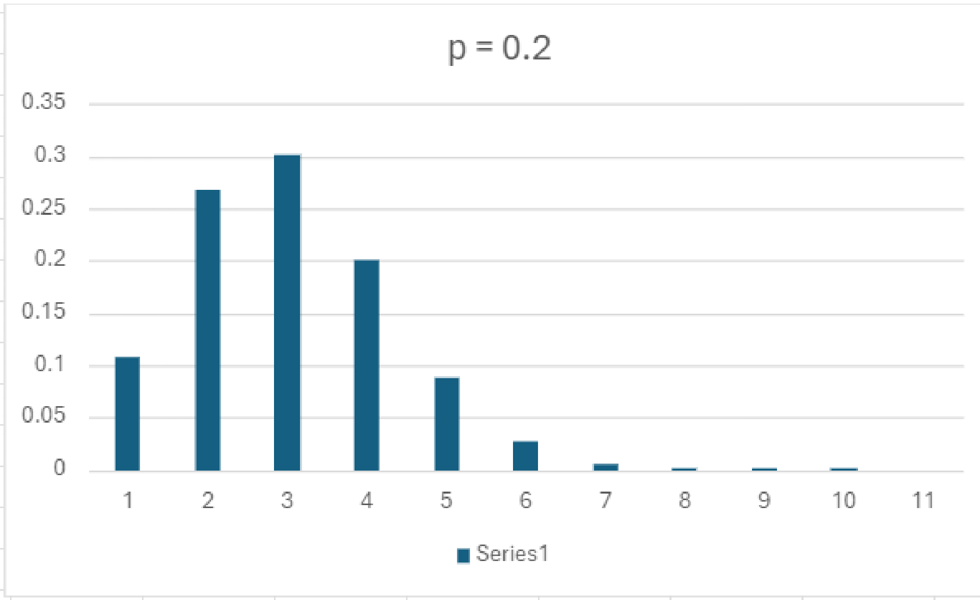
- Case I: When p = 0.5, the dist is symmetric about the expected value (n*p = 5), where the probability of success below the mean matches that above it.
- Case II: When p <0.5, the distribution is positively skewed, that is the probability of success below the mean is greater than those above it.
- Case III: When p > 0.5, the distribution is negatively skewed, that is the probability of success above the mean is greater than below above it.

p (success)	0.5
n	10
Successes:	Binomial Dist
0	0.000976563
1	0.009765625
2	0.043945313
3	0.1171875
4	0.205078125
5	0.24609375
6	0.205078125
7	0.1171875
8	0.043945313
9	0.009765625
10	0.000976563



-


p = 0.8
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
11
Series1

- 
p = 0.2
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
11
Series1

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

Introduction:

- The Multinomial distribution generalizes the Binomial distribution to more than 2 possible outcomes.
    - It can be graphically represented by plotting the PMFs of each possible outcome.
- Its PMF is given by, where p1, p2...pN are the probabilities of kth result & x1, x2..., xk the value of the Random variable which is desired.

$$f(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} \quad p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

$$\text{for } x_i = 0,1,2,\cdots \text{ such that } \sum_{i=1}^{k} x_i = n$$

$$P[X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k]$$

Also,

$$E(X_i) = np_i$$

$$Var(X_i) = np_i(1-p_i)$$

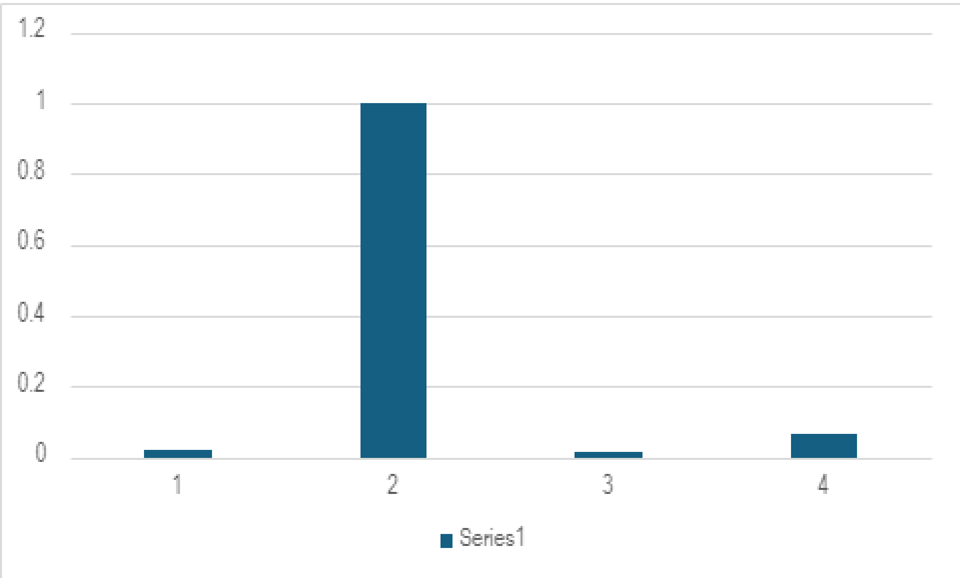
- The mean and variance are given by: E(Xi) = npi, & Var(Xi) = npi(1-pi).
    - In EXCEL, **MULTINOMDIST(R1, R2)**

Application:

- Suppose that a bag contains 8 balls, 3 red, 1 green & 4 blue. You reach at the bag and pull out a ball and put it back, and pull out another ball. This experiment is repeated a total of 10 times. What is the probability that the outcome will result in 4 red & 6 blue balls?
    - E1 : red ball is drawn, E2: a green ball, E3: blue ball
        - P1 : 3/8, P2: 1/8 , P3: 4/8

B15 =B11*PRODUCT(B12,B13,B14)

	A	B	C	D	E
1					
2	A	B	C		
3	Red x1	4			
4	P1	0.375			
5	Green x2	0			
6	P2	0.125			
7	Blue x3	6			
8	P3	0.5			
9	n (Total Trials)	10			
10					
11	multinomial	210			
12	p1^outcome1	0.019775391			
13	p2^outcome2	1			
14	p3^outcome3	0.015625			
15	prob	0.064888			

- 

-

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

Introduction:

- The Poisson distribution models the number of events occurring in a fixed interval of time or space, given a constant mean rate.
- Graphical representation involves plotting the probability mass function (PMF) to show the probabilities of different event counts.
- PMF given by:

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

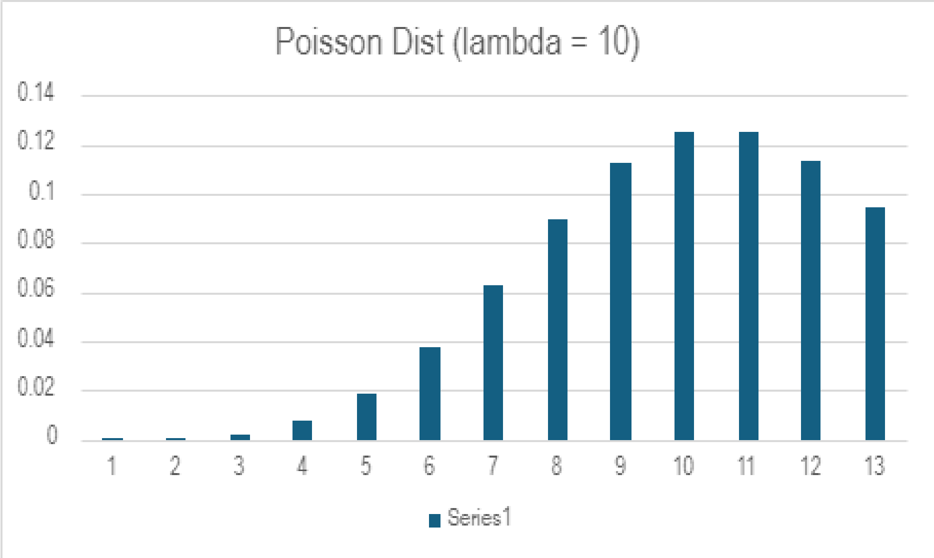
- λ- mean number of successes over a given interval
  $Var(X) = \lambda$
- Mean = λ, variance = λ
- Where lambda is the mean rate, and x is the number of events occurring in that interval.
- It is used in modelling radioactive decay, network failures, etc.
- EXCEL formula, **POISSON.DIST(K, λ, FALSE)**

Application:

Problem statement:

- An electronic store sells an average of 10 desktops in a week. Assuming that the purchases are as described above, then what is the probability that the store will have turned away potential customers if they stock 2 computers. How many computers should the store stock in order to make sure that it has a 99% probability of meeting a week's demands?

K (computers stocked)	Poisson dist		Lambda	10
0	4.53999E-05			
1	0.000453999			
2	0.002269996			
3	0.007566655			
4	0.018916637			
5	0.037833275			
6	0.063055458			
7	0.090079226			
8	0.112599032			
9	0.125110036			
10	0.125110036			
11	0.113736396			
12	0.09478033			

-


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

Introduction:

- The Geometric distribution models the number of trials needed to achieve the first success in a sequence of independent Bernoulli trials.
- Graphical representation includes plotting the probability mass function (PMF) to show the probability of each possible number of trials.
- PMF given by:

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

$$P(X > n) = (1-p)^n$$

$$\text{Mean } \mu = \frac{1}{p}$$

$$\text{Variance } \sigma^2 = \frac{1-p}{p^2}$$

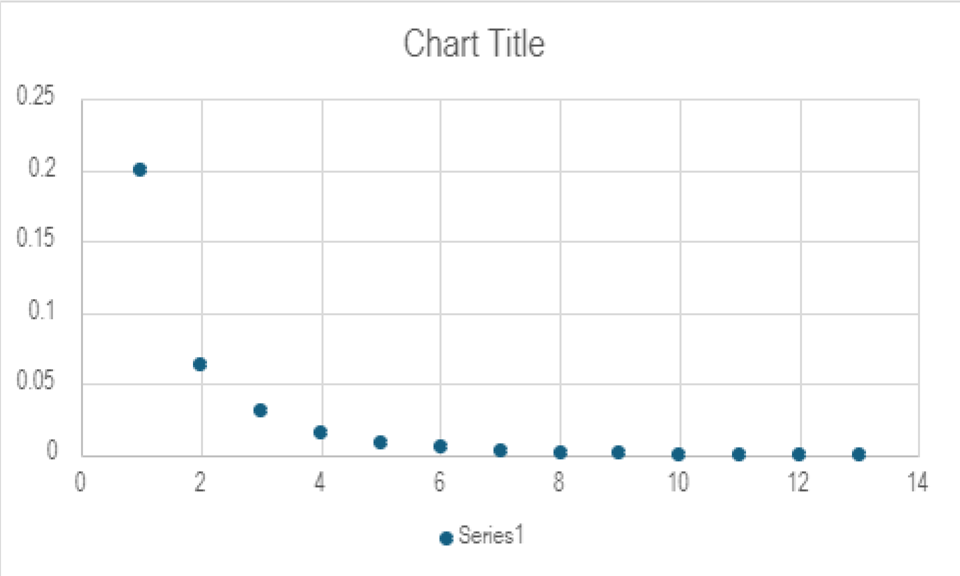
p – probability of success

- n- number of first successful trial
- Mean = 1/p, variance = (1-p)/p^2, p: prob of success
- It can be used to model the number of lottery tickets needed before the first win, or can be used in structural analysis, to calculate the number of earthquakes before a building/structure goes down.
- In EXCEL, **NEGBINOM.DIST(x, k, p, FALSE).**

Application:

- Suppose a programmer is waiting outside the PM's office for the PM's views on AI (whether he supports it or not). If the probability that the PM supports AI is 0.2, what is the probability that 4th PM is the first one to support AI.
-

X (success in ith trial)	Failures	Geom Dist	p(success)	0.2
1	0	0.2		
2	1	0.064		
3	2	0.03072		
4	3	0.016384		
5	4	0.00917504		
6	5	0.005284823		
7	6	0.00310043		
8	7	0.001842541		
9	8	0.001105525		
10	9	0.000668228		
11	10	0.000406283		
12	11	0.000248202		
13	12	0.00015223		

- 

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

Introduction

- The Uniform distribution represents outcomes that are equally likely over a specified range.
- Graphical representation involves plotting the probability density function (PDF) to show the uniform distribution over the specified range
- PDF, mean, SD & variance given by:

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

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

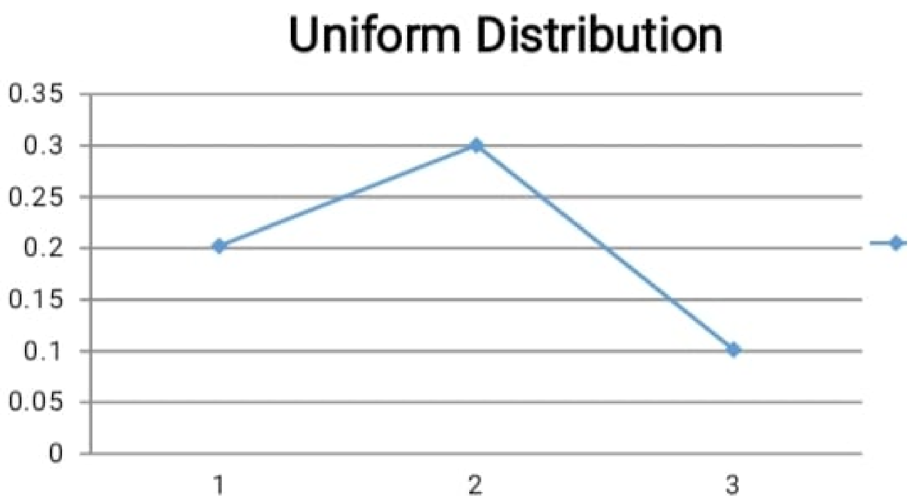
$$\text{Variance } \sigma^2 = \frac{(b-a)^2}{12}$$

- This is used in PRNGs to generate Random Numbers, and is also used in gaming, and simulations to generate terrains and maps, known as Procedural Random Generation. (PRGs).
- In EXCEL, **UNIFORM_DIST(x, α, β, cum)**

Application:

- A content search on an xyz search engine will show up the results every time in 20 secs. If you search something on xyz, what is the probability that it will show up in 8 seconds?

A	B	X	Y	Probability	G
0	20	0	8	0.4	0-8
15	25	17	19	0.2	17-19
120	170	150	170	0.4	150-170



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

Introduction:

- The Exponential distribution models the time between 2 events in a Poisson process, where events occur **continuously** and independently at a constant mean rate.
- It can be graphically represented by plotting the PDFs and the probability of the time intervals between the events.
- PDF given by:

$$
\begin{aligned}
\mathrm{E}[X^2] &= \int_0^\infty x^2 \lambda \exp(-\lambda x)dx \\
&= \left[-x^2 \exp(-\lambda x)\right]_0^\infty + \int_0^\infty 2x \exp(-\lambda x)dx \qquad \text{(integrating by parts)} \\
&= (0-0) + \left[-\frac{2}{\lambda} x \exp(-\lambda x)\right]_0^\infty + \frac{2}{\lambda}\int_0^\infty \exp(-\lambda x)dx \qquad \text{(integrating by parts agai} \\
&= (0-0) + \frac{2}{\lambda}\left[-\frac{1}{\lambda}\exp(-\lambda x)\right]_0^\infty \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

$$\mathrm{E}[X]^2 = \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

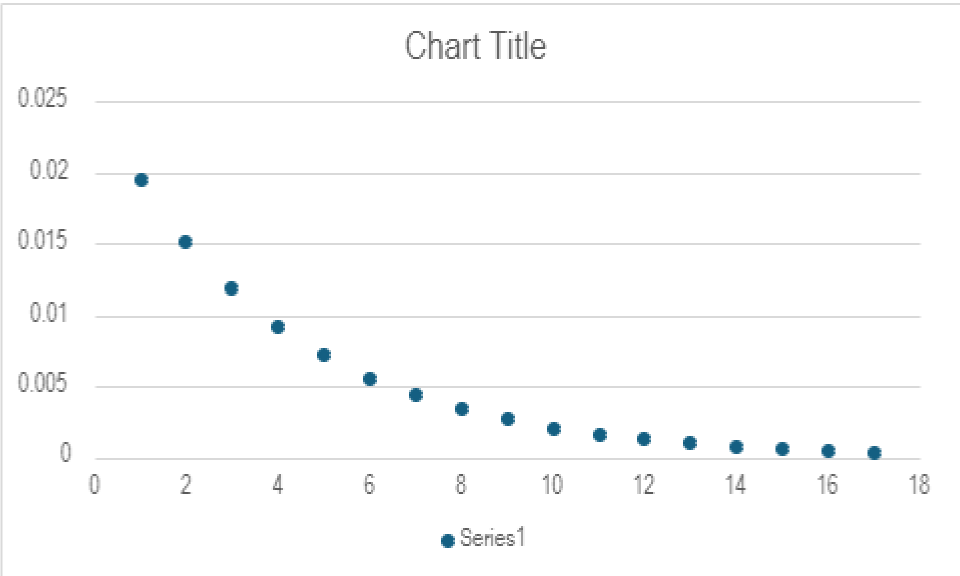
- Here, Lambda is the rate parameter, which is the frequency at which the events occur (basically the time b/w 2 events).
- It is memoryless, i.e.
- $$P(X > x + a | X > a) = P(X > x).$$
- Mean = 1/λ, Variance = 1/λ^2.
- In EXCEL, **EXPON.DIST(X, λ, cumulative)**

Application:

1. Suppose a computer goes on an automatic update every 40 weeks on average. After an update occurs, find the probability that it will take more than 50 weeks for the next update to arrive.

λ = 1/mu = 1/40 = 0.025

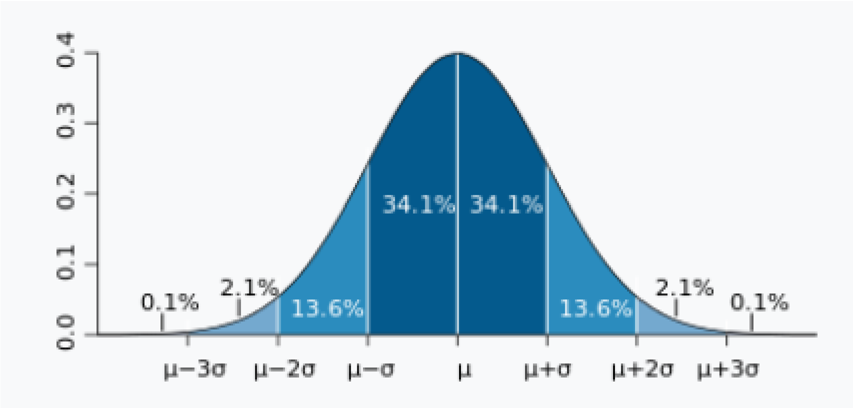
Weeks until the next update	Exponential Dist		
10	0.01947002	lambda	0.0250
20	0.015163266		
30	0.011809164		
40	0.009196986		
50	0.00716262		
60	0.005578254		
70	0.004344349		
80	0.003383382		
90	0.002634981		
100	0.002052125		
110	0.001598197		
120	0.001244677		
130	0.000969355		
140	0.000754935		
150	0.000587944		
160	0.000457891		
170	0.000356606		



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

Introduction:

- The Normal distribution is used to model an event having a large number of identical & independently distributed Random variables added together; this distribution takes the form of a Bell curve (e^(-x^2)).
- It is represented by N(mean,variance), the standard Normal dist. is given by: N(0,1), where the mean is 0 & variance is 1.



- It is the most important distribution, with each natural & social process following it.
- It's PDF is given by:

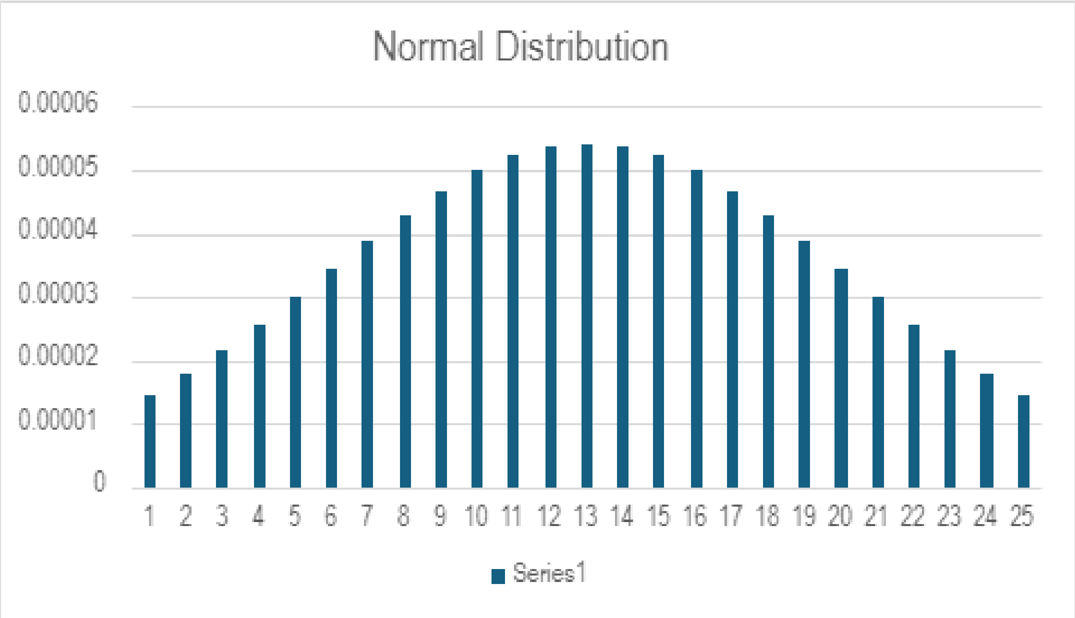
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Graphical representation involves plotting the probability density function (PDF) to show the bell-shaped curve of the distribution.
- Mean = μ , Variance = σ^2

- In Excel, **NORM.DIST(X, μ, σ, cumulative)**.
- Mean = **AVERAGE(C1)**
- STD. dev = **STDDEV(C2)**

Application:

		NORM_DIST		
EMP1	1000	1.43472E-05	MEAN	13000
EMP2	2000	1.77407E-05	STD DEV	7359.800722
EMP3	3000	2.15356E-05		
EMP4	4000	2.56641E-05		
EMP5	5000	3.00245E-05		
EMP6	6000	3.44833E-05		
EMP7	7000	3.88799E-05		
EMP8	8000	4.3035E-05		
EMP9	9000	4.6763E-05		
EMP10	10000	4.98843E-05		
EMP11	11000	5.22406E-05		
EMP12	12000	5.37075E-05		
EMP13	13000	5.42056E-05		
EMP14	14000	5.37075E-05		
EMP15	15000	5.22406E-05		
EMP16	16000	4.98843E-05		
EMP17	17000	4.6763E-05		
EMP18	18000	4.3035E-05		
EMP19	19000	3.88799E-05		
EMP20	20000	3.44833E-05		
EMP21	21000	3.00245E-05		
EMP22	22000	2.56641E-05		
EMP23	23000	2.15356E-05		
EMP24	24000	1.77407E-05		
EMP25	25000	1.43472E-05		



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

Introduction:

- The cumulative distribution function (CDF) gives the probability that a random variable takes on a value less than or equal to a given value.

$$F_X(x) = P(X \leq x)$$

$F_X(x)$ = function of X

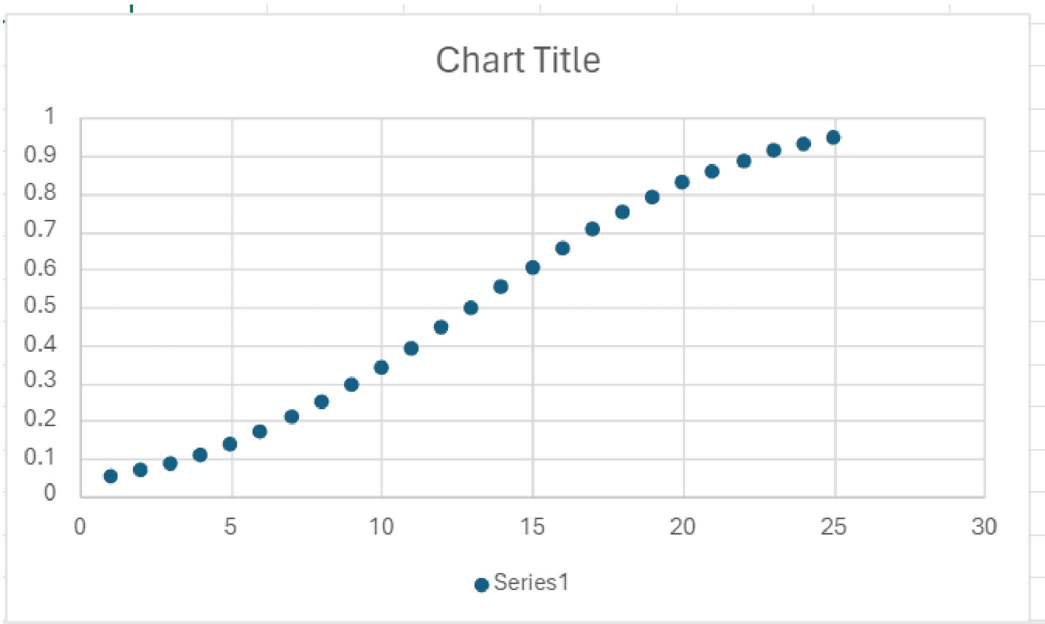
X = real value variable

P = probability that X will have a value less than or equal to x

- For exponential random var, it is given by: $P(X \leq x) = 1 - e^{(-\lambda x)}$
- For normal dist, it is given by:

$$F_X(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]$$

- In EXCEL, exponential dist CDF = **EXPON.DIST(X, λ, TRUE).**
- Normal, CDF = **NORM.DIST(X, μ, σ, TRUE).**
- **Normal:**



- **Exponential:**

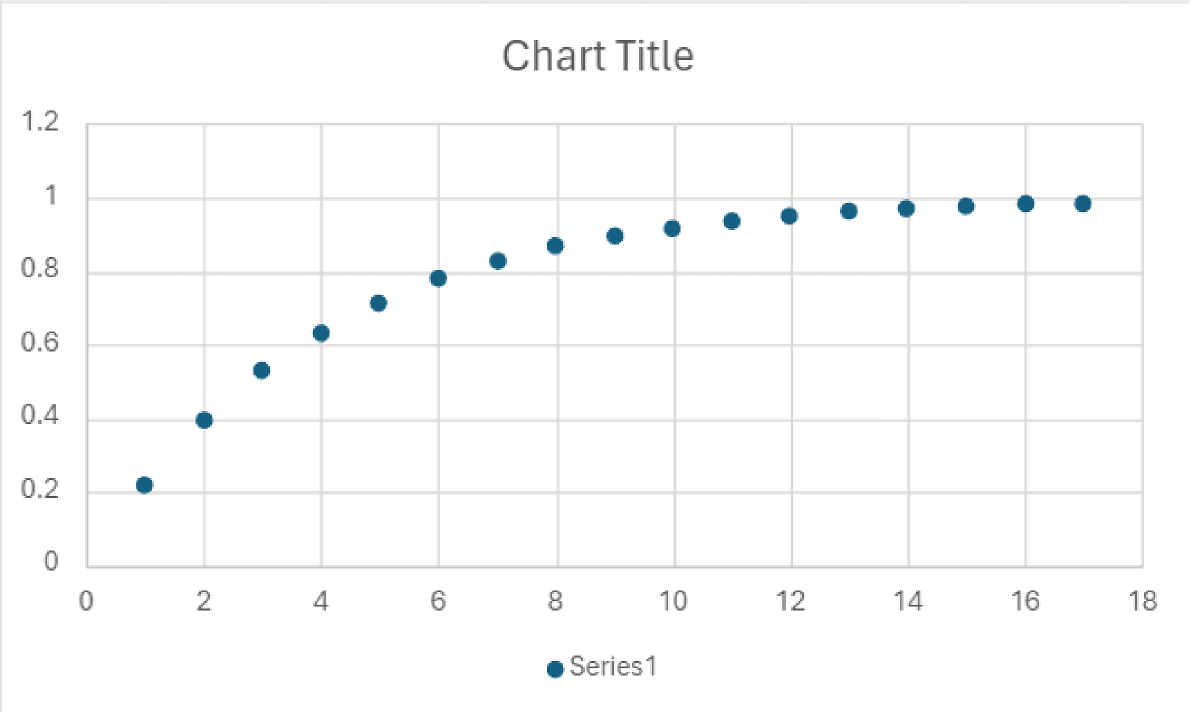
- 

Chart Title
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
14
16
18
Series1

9. Given data from two distributions. Calculate the distance between the 2 distributions.

Introduction:

- We'll be using **Euclidean Distance** for the distance between 2 real valued vectors. It is calculated as the square root of the sum of the squares of the differences of the vectors' components.

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

p, q	= two points in Euclidean n-space
q_i, p_i	= Euclidean vectors, starting from the origin of the space (initial point)
n	= n-space

- In Excel, function: **SQRT(SUMXMY2(array_x,array_y))**

Binomial	**Poisson**
X	**Y**
0.1	0.3
0.3	0.5
0.4	0.8
0.7	0.11
0.8	0.14
0.1	0.19
0.15	0.25
0.28	0.27
0.2	0.30
0.21	0.35

Binomial	Poisson
X	Y
0.1	0.3
0.3	0.5
0.4	0.8
0.7	0.11
0.8	0.14
0.1	0.19
0.15	0.25
0.28	0.27
0.2	0.3
0.21	0.35

10. Applications problem based on Binomial distribution.

Case Study: Euclidean dist: **1.035133**

The phe-mycin Case: Drug side effects antibiotics occasionally cause nausea as a side effect and a major drug company has developed a new antibiotic called phemycin. The company claims that at most, 10% 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 phemycin. Each patient will either experience nausea(success) or will not experience nausea (Failure). We will assume that p, the true probability that a patient will experience nausea as a side effect is 0.10, the maximum value of p claimed by the drug company. Let x denote the number of patients among the four who will experience nausea as a side effect. It follows that c is a binomial random variable which can take on any of the potential values 0,1,2,3,4.

Suppose that when a sample *n = 4* randomly selected patients is treated with phe-mycin, three of the four will experience nausea *3/4 =0.75* which is far greater than 0.10, we have some evidence contradicting the assumption that p equals 0.10 x = 3 or x=4 are mutually exclusive:

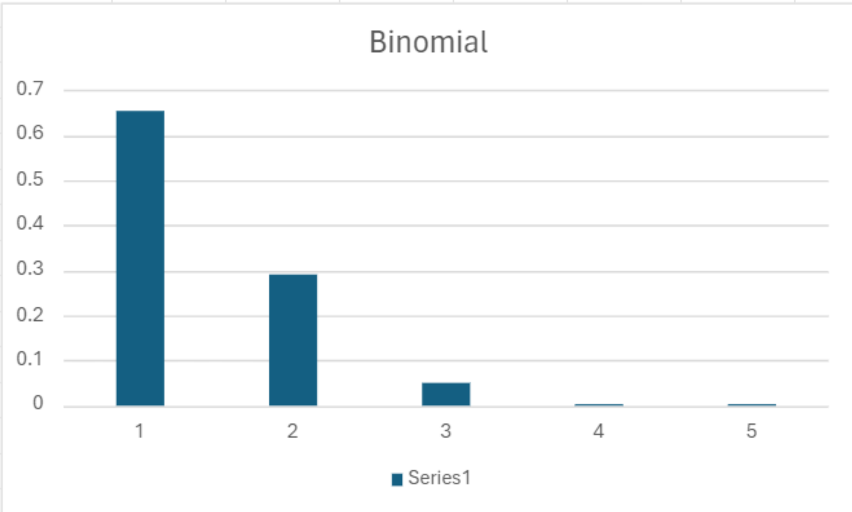
P(x >= 3) = P(x = 3) + P(x = 4) =0.0036 +0.0001 =0.0037

For instance, the probability that none of the four randomly selected patients experience nausea is:

P(0) = P(x = 0) =4!/0!(4-0)!*(0.10)(0.9)^4-0 =(0.9)^4=0.6561

P(X=3)=0.0036 says that 3 out of 4 would experience nausea.

n	4
p	0.1
X	
0	0.6561
1	0.2916
2	0.0486
3	0.0036
4	0.0001



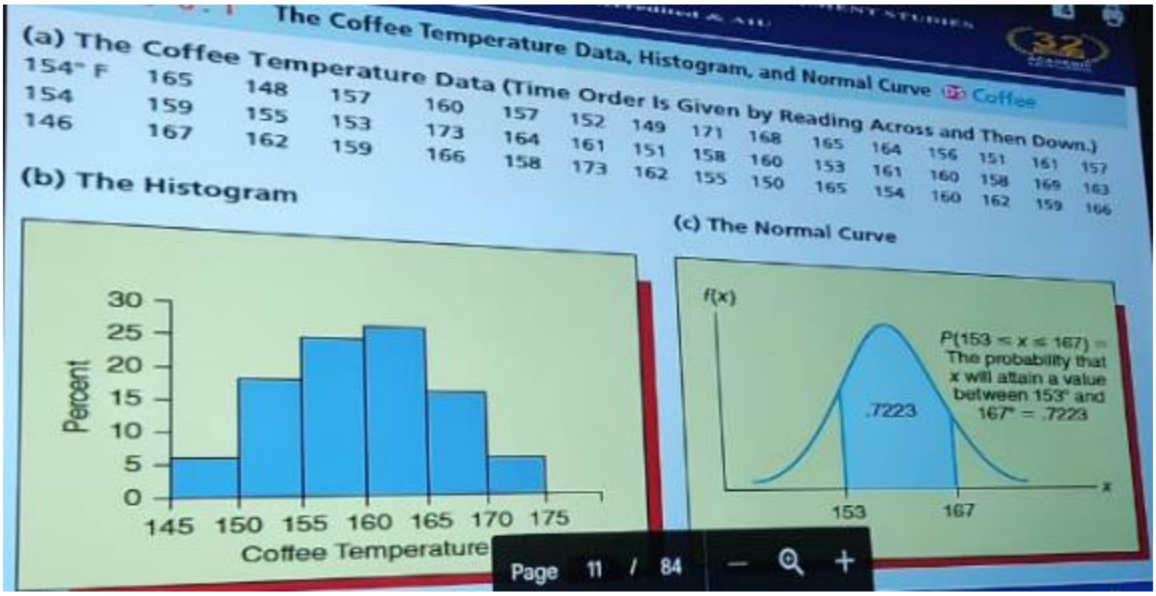
11. Applications problems based on Normal Distribution.

Case Study:

According to the website of the American Association for Justice, Stella Liebeck of Albuquerque, New Mexico, was severely burned by McDonald's coffee in February 1992. Liebeck, who received third-degree burns 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°F before the trial to about 158° 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 .

Make a time series plot of the coffee temperatures, and assuming process consistency, estimate limits between which most of the coffee temperatures at the restaurant would fall.



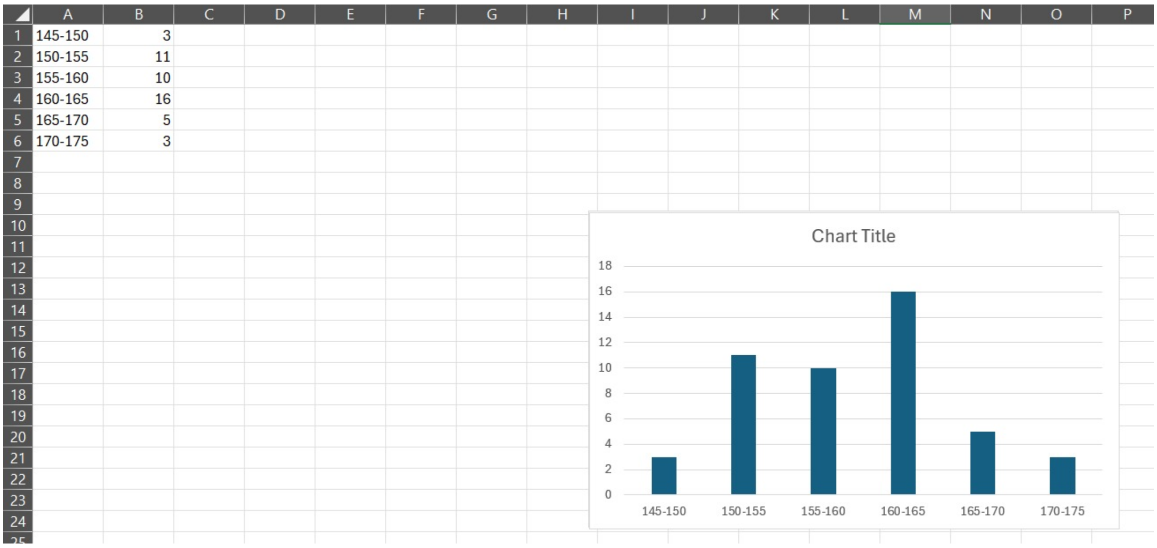
Answer:
One minus the probability would represent the proportion of coffee served by the restaurant that had the temperature outside the range **153 -167**

It follows that the probability that x will be in between 153 and 167 is the area under the temperature normal curve between 153 and 167

The area is <u>0.7223</u>
P(153<=x<=167)=5.7223

In conclusion, we estimate that 72.23% of the coffee served at the restaurant is within the range of temperature that is best and 27.77% of the coffee served is not in this range.

If management wishes a very high percentage of the coffee
Served to taste best,it must improve the coffee making process by better controlling temperature.



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

Bivariate Analysis:

The term Bivariate Analysis refers to the analysis of 2 variables. The objective of this is to understand the relationship between 2 variables. There are 3 common ways to perform this:

- Scatter Plots
- Correlation Coefficients
- Simple Linear Regression

Covariance in excel: **COVARIANCE.P(array_1,array_2)**, where arrays are the range of integer values.

Few things to remember about these arguments:

- If the array contains text or logical values, then they are ignored by the COVARIANCE function.
- The data set should have the same number of sample points (same size)
- The dataset should not be empty, nor should the std. dev of its values be equal to 0.

$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

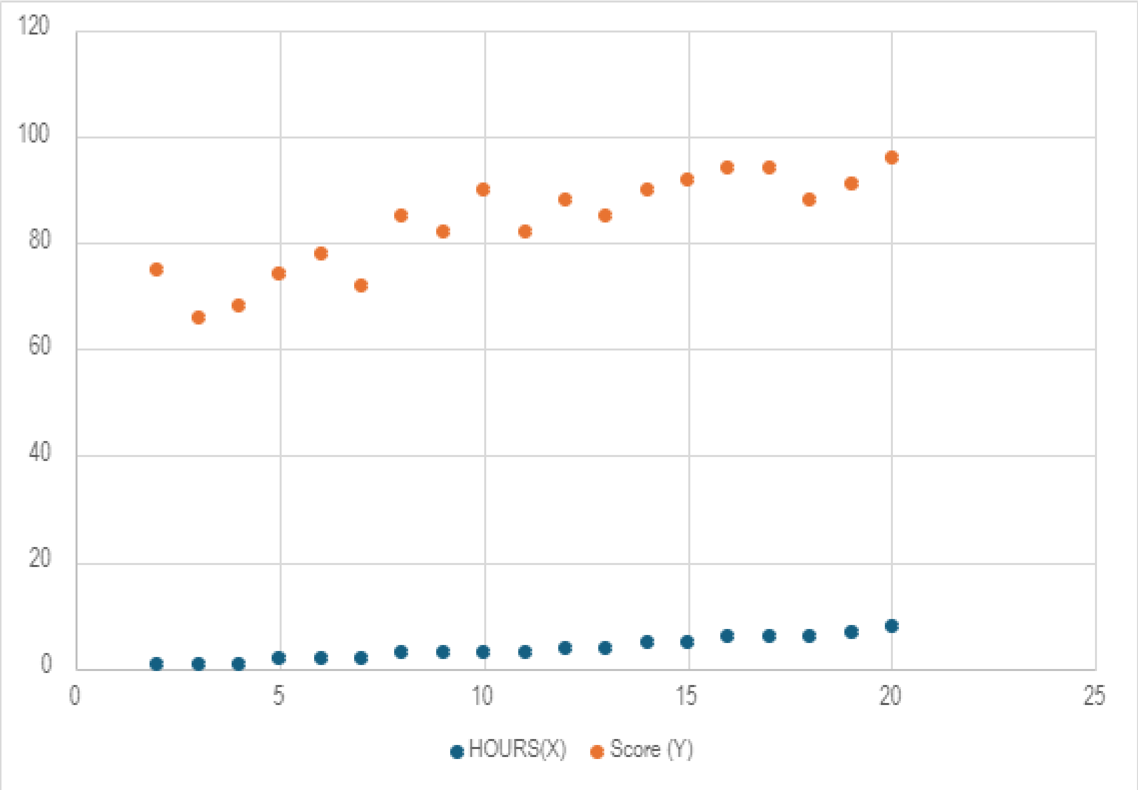
- n : sample size.
- Covariance is a measure to indicate the extent to which 2 random variables change together.
- The sign of covariance shows the linear relation b/w the variables.
- It is a measure of correlation.
- It is affected by change in scale.

Dataset:

Perform bivariate analysis in EXCEL using the following dataset, of 2 variables: X : hours spent studying. Y: exam score of 20 students.

In EXCEL, **CORREL(array1,array2) = 0.890 (strong +ve correlation)**

HOURS(X)	Score (Y)			
			CORRELATION	0.890469676
1	75			
1	66			
1	68			
2	74			
2	78			
2	72			
3	85			
3	82			
3	90			
3	82			
4	88			
4	85			
5	90			
5	92			
6	94			
6	94			
6	88			
7	91			
8	96			



13. Calculation of Karl Pearson's correlation coefficients.

Introduction:

- Correlation is a measure used to represent how strongly two random variables are related to each other.
- Correlation refers to the scaled form of covariance.
- Correlation can range from -1 to 1.
- Correlation measures both the strength and the direction of linear relationship between the 2 variables.
- It is not affected by changes in the scale.
- **Pearson's Correlation Coeff.** is given by:

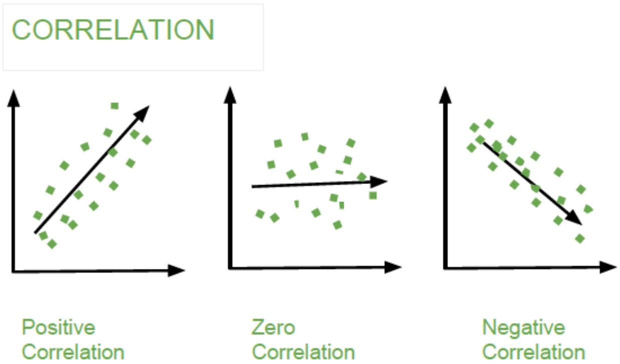
x_i are the **individual values** of one **variable** e.g. age

y_i are the **individual values** of the other **variable** e.g. salary

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

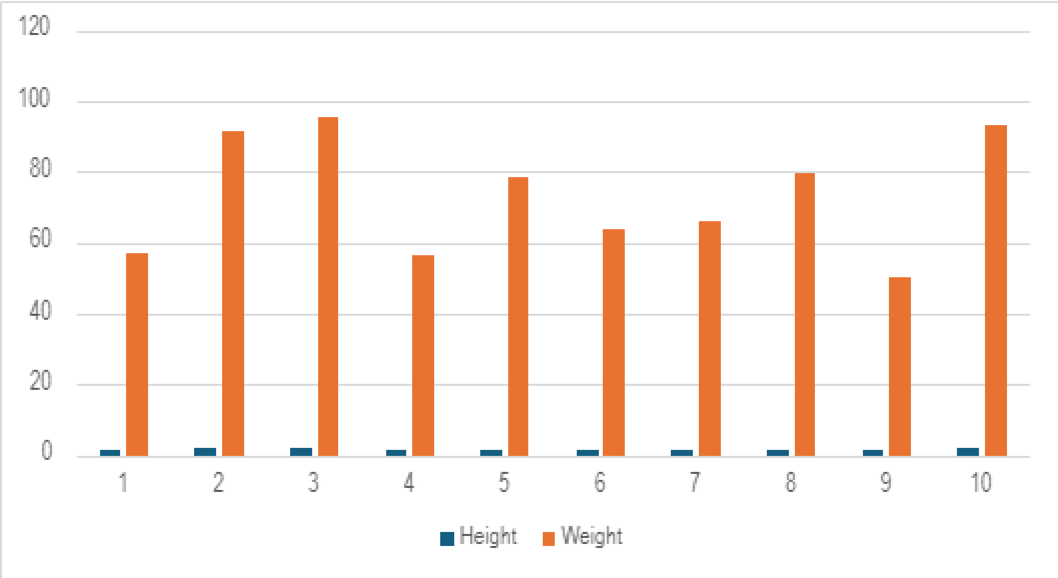
- Where r is the Pearson correlation coefficient,

$\bar{x}$ and $\bar{y}$ are respectively the **mean values** of the two variables.



- In EXCEL, **PEARSON(array1, array2)**

Gender	Height	Weight		
F	1.62	57.14	Pearson Correlation Coeff.	0.975635069
M	1.83	91.69		
M	1.89	95.27		
F	1.55	56.16		
F	1.74	78.52		
M	1.6	63.75		
F	1.6	66.09		
M	1.72	79.52		
F	1.54	50.22		
M	1.82	93.39		

- 

120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
Height
Weight

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

Introduction:

- Bivariate Frequency Distribution: A series of data showing the frequency of 2 variables simultaneously is called bivariate frequency dist. In other words, the frequency distribution of 2 variables is called Bivariate, examples: sales and advertisements, weight & height of an individual, etc.
- Significant in business, reason:
    - Decision Making
    - Market segmentation
    - Risk assessment
    - Resource Allocation

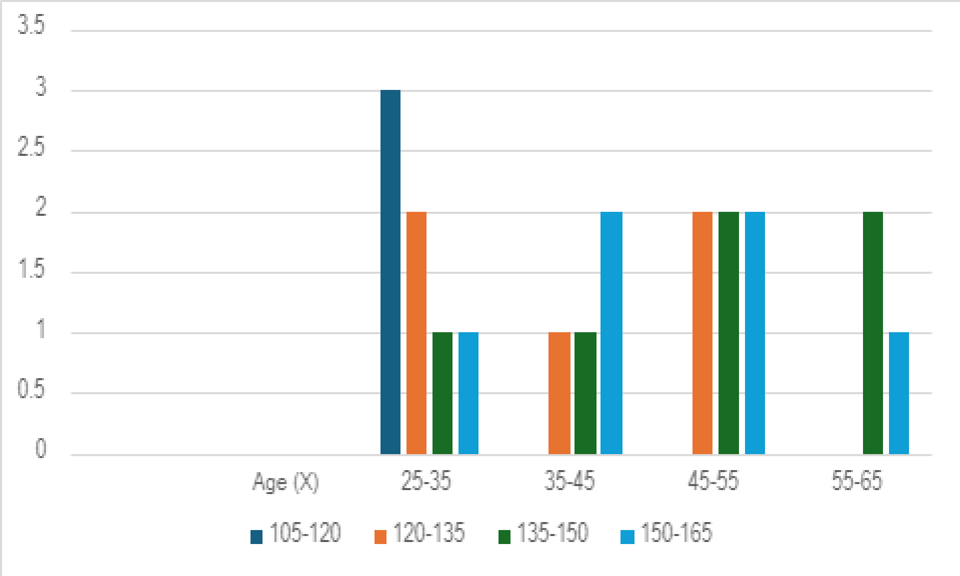
Question:

Prepare frequency distribution for the following data.

The required data is: (45, 141); (26, 130); (62, 150); (28, 114); (55, 138); (36, 120); (48, 142); (40, 139); (28, 105); (32, 135); (31, 153); (37, 151); (59, 149); (50, 151); (48, 121); (47, 126); (33, 131); (42, 154); (49, 151); (34, 118).

Where, X: age in years (interval of 4) and y denotes the corresponding blood pressure.

Blood Pressure (Y)	105-120	120-135	135-150	150-165	T
Age (X)					
25-35	3	2	1	1	7
35-45		1	1	2	4
45-55		2	2	2	6
55-65			2	1	3
T	3	5	6	6	20
Marginal Freq Dist of X					
Req. Interval (X)	25-35	35-45	45-55	55-65	
Freq	7	4	6	3	
Marginal Freq Dist of Y					
Req. Interval (Y)	105-120	120-135	135-150	150-165	
Freq	3	5	6	6	



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

Introduction:

- Generating from Binomial:

- In excel, formula = **BINOM.INV(1, p, rand())**, will generate 1 or 0 with a chance of 1 being p.

p	Random Number
0.1	0
0.2	0
0.3	0
0.4	0
0.5	1
0.6	0
0.7	0
0.8	1
0.9	1

-

Random Number
0
0
1
0
0
0
1
1
0

Random Number
0
1
1
1
0
0
0
1
0

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

Introduction:

- Generating from Normal:
- In excel, random numbers can be generated from a given Normal distribution, using the formula, **NORMINV(RAND(),B2,C2)**, where **B2** is the mean, and **C2** is the std. dev.
- Example: If the average weekly sales of an electronic company for laptops is $2000, which fluctuates about $500 up or down, use the random function to establish probability.
  - Mean = 2000
  - STD. dev = 500
  - **NORMINV(RAND(),2000,500)**

STD DEV	500
MEAN	2000
Random Num	2388.15

STD DEV	500
MEAN	2000
Random Num	926.4369

STD DEV	500
MEAN	2000
Random Num	2051.593

STD DEV	500
MEAN	2000
Random Num	1289.387

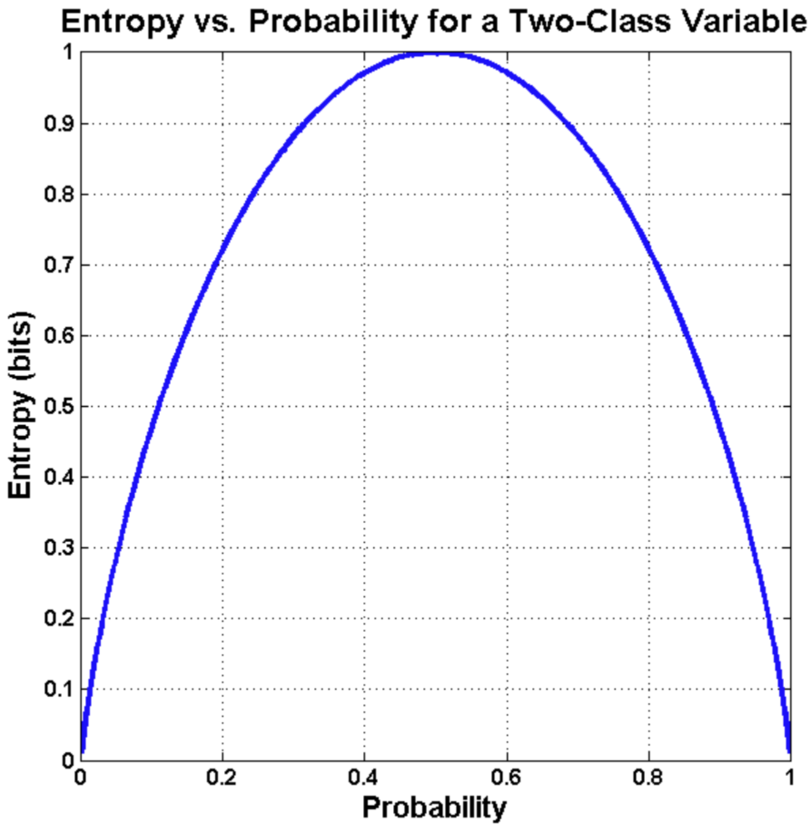
17. Find the entropy from the given dataset.

Introduction:

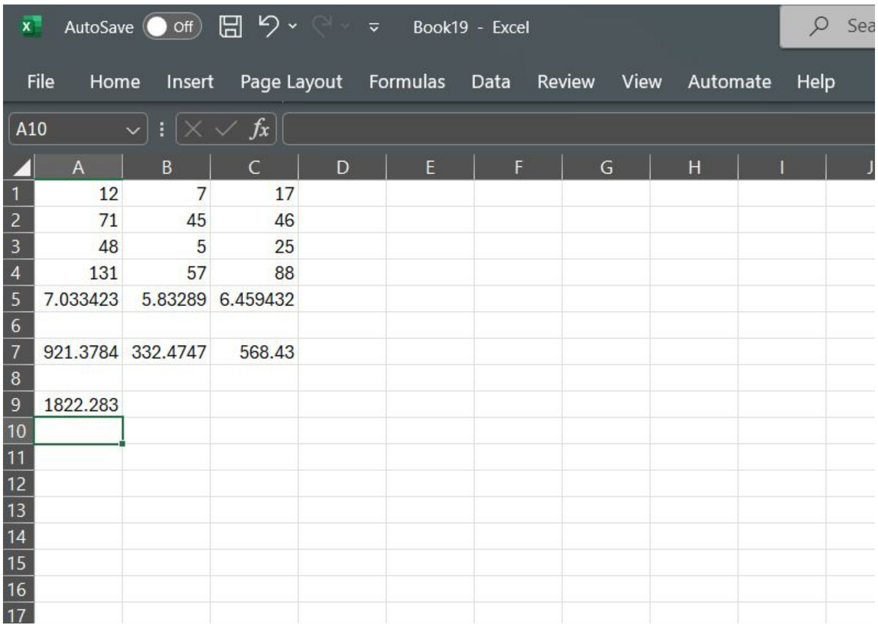
- The entropy of a random variable is the average level of information, uncertainty, surprise, randomness to a variable's possible outcomes.
- Given a random variable X, which takes values in the alphabet X & distributed as per P : X -> [0,1], the entropy is defined as:

$$\mathrm{H}(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

- Where the logarithm's base can be 2 (shannons/bits), e (nat), or 10 (hartleys).
- An equivalent definition of entropy is the expected value of the self information of a variable.
- In case of 2 fair coin tosses, the information entropy exhibits **Base-2** logarithm of the no. of possible outcomes with 2 coins, these are the possible outcomes with 2 bits of entropy:
  - **HH, HT, TH, TT.**
- Generally, entropy is the average information covered by an event when considering all possible outcomes.
- Entropy, H(X) of a coin flip measured in bits.



- For a fair coin, the entropy is 1 due to 2 possible outcomes.
- The result of a fair die, having 6 possible outcomes, has entropy calculated as log_base2(6 bits).



	A	B	C
1	12	7	17
2	71	45	46
3	48	5	25
4	131	57	88
5	7.033423	5.83289	6.459432
6			
7	921.3784	332.4747	568.43
8			
9	1822.283		

- Here, the first 3 rows are inputs. 4th row is the sum of (al:a3)..., 5th row denotes log_base2(sum): calculated by **LOG(number, [base]).**
- **Finally,** the 7th row lists the individual entropies, and the last row is the sum of all the entropies.

SUBMITTED TO :-

DR. AAKASH