# Chapter 04: Correlation & Regression

## ☐ Correlation

The relation between one or more variables is called correlation .example: height & weight of a group of peoples.

## ☐ Correlation Coefficient

The numerical measurement of correlation is called correlation coefficient . denoted by

$r_{xy}$ =[(n Σ xy - ΣxΣy) / √{n Σ x^2 - (Σx)^2} * √{ nΣy^2 - (Σy)^2}]

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

## ☐ Properties of Correlation Coefficient

1. Correlation coefficient is independent of change of origin and scale of measurement.

2. Correlation coefficient lies between -1 to +1. i.e, $-1 < r_{xy} < 1$.

3. Correlation coefficient is symmetric. i.e, $r_{xy} = r_{yx}$

4. Correlation coefficient is the geometric mean of regression coefficients. $i.e,\ r_{xy} = \sqrt{b_{yx} \times b_{xy}}$

5. For two independent variable correlation coefficient is zero

6. It is always unit free.

☐ Regression

The probable movement of one variable in terms of the other variables is called regression.Example: The productions of paddy of amount y is dependent on rainfall of amount

☐ Regression Coefficient

The mathematical measures of regression are called the coefficient of regression

$$b_{yx} = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} x^2 - (\sum_{i=1}^{n} x)^2}$$

$$b_{xy} = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} y^2 - (\sum_{i=1}^{n} y)^2}$$

## ☐ Properties of Regression Coefficient

1. Regression coefficient is independent of change of origin but not of scale.

2. Regression coefficient lies between $-\infty$ to $+\infty$. i.e, $-<\infty \quad b_{yx} < \infty$.

3. Regression coefficient is not symmetric. i.e, $b_{xy} = b_{yx}$

4. The geometric mean of regression coefficients is equal to correlation coefficient

   i.e, $r_{xy} = \sqrt{b_{yx} \times b_{xy}}$

5. The arithmetic mean of two regression coefficient is greater than correlation

   Coefficient. i.e, $\left(\frac{b_{yx} + b_{xy}}{2}\right) \geq r_{xy}$

6. If one of regression coefficient is greater than unity the other must be less than

   unity. i.e, $b_{xy} \geq 1$ and $b_{yx} < 1$

7. Regression coefficient is not pure number.

☐ Difference between Correlation Coefficient and Regression Coefficient

Distinguish between correlation coefficient and regression coefficient.

| Correlation coefficient | Regression coefficient. |
|---|---|
| **1. The numerical value by which we measure the strength of linear relationship between two or more variables is called correlation coefficient.** | 1. The mathematical measures of regression are called the coefficient of regression. |
| **2. Correlation coefficient is independent of change of origin and scale of measurement.** | 2. Regression coefficient is independent of change of origin but not of scale. |
| **3. Correlation coefficient lies between -1 to +1.** | 3. Regression coefficient lies between -∞ to +∞ |
| **4. Correlation coefficient is symmetric. i.e, $r_{xy} = r_{yx}$** | 4. Regression coefficient is not symmetric. i.e, $b_{xy} = b_{yx}$ |
| **5. It is always unit free.** | 5. It is not unit free. |

☐ Interpretation

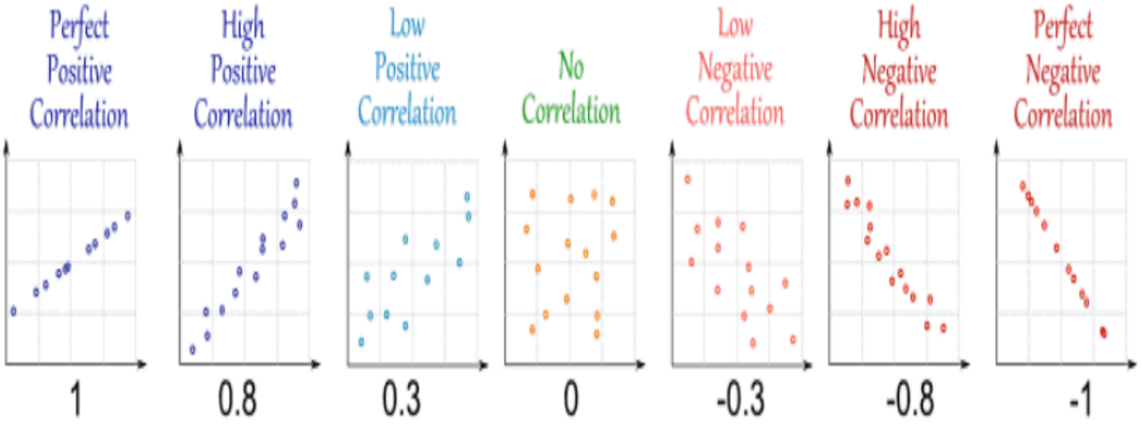
## Comment on Correlation Coefficient ($r_{xy}$)

| |
|---|
| Perfect Positive ($r_{xy}$ =1) |
| Perfect Negative ($r_{xy}$ = -1) |
| Strongly Positive ($0.7 \leq r < 1$) |
| Strongly Negative($-1 < r \leq -0.7$) |
| Moderate Positive ($0.3 \leq r < 0.7$) |
| Moderate Negative($-0.7 < r \leq -0.3$) |
| Weak Positive ($0 < r < 0.3$) |
| Weak Negative ($-0.3 < r < 0$) |
| Zero Correlation (0) |

☐ Scattered Diagram

**Scatter diagram**: The diagrammatic way of representing bivariate data is called scatter diagram.

Suppose, $(x_1,y_1)$, $(x_2,y_2)$..........$(x_n,y_n)$ are n pairs of observations. If the values of the variables x and y be plotted along the x-axis and y-axis respectively in the xy-plane, the diagram of dots so obtained is known as scatter diagram.




In positive correlation, both variables increase together.

In negative correlation, one variable increases while the other decreases.

In no correlation, there is no clear pattern in the data.

## ☐ *Importance of Correlation and Regression in Data Analysis*

In data analysis, understanding the relationship between variables is crucial,

- Correlation and regression are important statistical tools used to analyze relationships between variables.
- Correlation measures the strength and direction of the relationship between two variables.
- Regression helps to predict the value of one variable based on another variable.
- Both tools help in understanding how variables affect each other in real-world data.
- They are widely used in fields like computer science, business, economics, and healthcare for data analysis.
- Correlation is useful for selecting important variables in machine learning and removing irrelevant ones.
- Regression is used for forecasting and building predictive models.

- These methods help detect patterns, trends, and anomalies in data sets.
- They enable organizations and researchers to make informed and data-driven decisions.

In conclusion, correlation and regression are essential for analyzing data relationships and making accurate predictions, which are crucial for effective decision-making in many fields.

# Experiment, Outcome, Random Experiment, Sample Space (Definitions)

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

Tossing of a coin or throwing of a dice or the drawing of a cards etc, are the example of experiment.

Outcomes: The results of an experiment are called outcomes.
Random experiment: Experiments are called random experiments if the outcomes depend on chance and cannot be predicted with certainty.

Example: Tossing of a fair coin, throwing of dice etc are the examples of random experiments

Sample space: The collection or totality of all possible outcomes of a random experiment is called sample space. It is usually denoted by S or Ω.

**Sample point**: Each element of a sample space is called sample point.

**Event**: Any subset of a sample space is called event.

**Mutually exclusive events**: Two events are said to be mutually exclusive if they have no common points. If A and B are two mutually exclusive events,

Example, A ={1,3,5} and B ={2,4,6} then $A \cap B = \Theta$

**Non- mutually exclusive events**: Two events are said to be non-mutually exclusive if they have common points. If A and B are two mutually exclusive events,

Example, A ={1,2, 4} and B ={2,4,6} then $A \cap B = \{2,4\}$

**Conditional Probability**: If A and B are two events in S. Then the conditional probability of A for given value of B, denoted by P [A | B] is defined by

If 2 event mutually exclusive then p(a CAP b)=0

**Conditional Probability**: If A and B are two events in S. Then the conditional probability of A for given value of B, denoted by P [A | B] is defined by

$P[A \mid B] = P(AB)/P(B)$ ; $P[B]>0$

Similarly, $P[B \mid A] = P(AB)/P(A)$ ; $P[A]>0$

**Independent Event:** Two events A and B are said to be independent if and only if one of the following conditions holds:

(I) P[AB] = P[A]P[B]

(II) P [A | B ] = P[A]

(III) P [B | A ]= P[B]

**Dependent Event:** Two events A and B are said to be dependent if and only if one of the following conditions holds:

(I) P[AB] $\neq$ P[A]P[B]

(II) P [A | B ] $\neq$ P[A]

(III) P [B | A ] $\neq$ P[B]

**Definition of probability:**
There are four approaches of defining probability.

- Classical or mathematical or priori definition of probability.
- Empirical or statistical or posterior or frequency probability.
- Subjective probability.
- Axiomatic probability.

## Classical Probability (Definition)

the probability of the event A, denoted by P[A] is defined as

$$\text{Probability of the EVENT} = \frac{\text{number of favorable outcomes}}{\text{number of total outcomes}}$$

$$P(A) = \frac{N[A]}{N[S]} = \frac{m}{n} \quad ;0 \le P[A] \le 1$$

## Axiomatic Probability (Definition)

Suppose S is a sample space and A is an event of this sample space. Then the probability of the event A, denoted by P[A] must satisfy the following four axioms:

(i) $P(A) \geq 0$;

(ii) $P(S) = 1$;

(iii) If A and B are mutually exclusive events, then $P[A \cup B] = P[A] + P[B]$;

## Additive Law (Definition)

If A and B are two events, then

(i) $P[A \cup B] = P[A] + P[B] - P[AB]$ ; Not Mutually Exclusive

(ii) $P[A \cup B] = P[A] + P[B]$; Mutually Exclusive

If A, B, C are three non **mutually exclusive** events then,
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

If A, B, C are three **mutually exclusive** events then,
$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

value of the probability lies between 0 to 1

**If A and $\overline{A}$ are mutually exclusive then show that**

$$P(A) + P(\overline{A}) = 1$$

**Mathematical problem of Bayes' theorem:**

**In a factory, machines A, B and C manufacture respectively 25%, 35% and 40% of the product. Of the total of their output 5, 4 and 2 percent are defective bolts respectively. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machine (i) A and (ii) C.**

**Solution:** Let

$B_1$ : An item produced by machine A

$B_{2:}$ An item produced by machine B

$B_{3:}$ An item produced by machine C

D: Defective item produced by the machines.

We have, $P[B_1]= 25\%=0.25$, $P[B_2]= 35\%=0.35$, $P[B_3]= 40\%=0.40$,

And $P(D|B_1)= 5\%=0.05$, $P(D|B_2)= 4\%=0.04$, $P(D|B_3)= 2\%=0.02$,

We have to find, $P(B_1|D)$ and $P(B_3|D)$

$$P[B_i|D] = \frac{P[B_i]P[D/B_i]}{\sum_{i=1}^{n} P[B_i]P[D/B_i]};\ i = 1,2,\ldots\ldots\ldots.,n$$

$$P(B_1|D) = \frac{P(B_1)P(D/B_1)}{P(B_1)P(D/B_1)+P(B_2)P(D/B_2)+P(B_3)P(D/B_3)}$$

$$= \frac{(0.25)(0.05)}{(0.25)(0.05)+(0.35)(0.04)+(0.40)(0.02)}$$

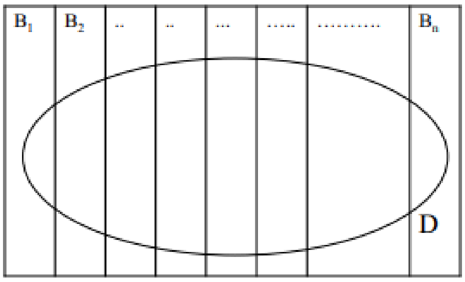
$$= 0.3623$$

Bayes' Theorem – Slides

## Statement of Bayes theorem

Let $B_1$, $B_2$,.................,$B_n$ be n mutually exclusive and exhaustive events in a random experiment and D be any event in S., then Bayes theorem state that

$$P[B_i|D] = \frac{P[B_i]P[D/B_i]}{\sum_{i=1}^{n} P[B_i]P[D/B_i]}; i = 1,2,..........,n$$




---

Chapter: 6 Random variables & probability functions

## Random Variable and Example

A random variable is a real valued function whose values are determined with the outcomes of a random experiment. It is usually denoted by X,Y,Z etc

examp: Let X be a random variable that represents the number of heads when two coins are tossed.

Then X can be 0, 1, or 2 depending on the outcome of the toss.

# Discrete & Continuous Random Variables (with Computer Science Examples)

Continuous random variable: A random variable is called continuous random variable if it can take any values between certain limits.

examp;Time taken by an algorithm to execute,

Response time of a server

Discrete random variable: A random variable is called discrete random variable if it can take only isolated values. examp:Number of times a function is called in a program.Number of errors in a program

## Probability Function – Definition

**Probability function:**

A function f(x) of a discrete random variable X is called a probability function if it satisfies the following two conditions:

$$(i)\, f(x) \geq 0$$

$$(ii) \sum f(x) = 1$$

# Probability Density Function – Definition

**Probability density function:**

A function f(x) of a discrete random variable X is called a probability function if it satisfies the following two conditions:

$$(i)\, f(x) \geq 0$$

$$(ii) \int_{-\infty}^{\infty} f(x)\,dx = 1$$

# Difference between PF and PDF

| No. | PF (Probability Function) | PDF (Probability Density Function) |
|---|---|---|
| 1 | Used for **discrete** random variables | Used for **continuous** random variables |
| 2 | Gives **probability of exact values** | Gives **probability density**, not exact value |
| 3 | **P(X = x)** is **non-zero** | **P(X = x) = 0** always |
| 4 | Total **sum** of all probabilities = 1 | Total **area under the curve** = 1 |

| | | |
|---|---|---|
| 5 | Probabilities are **added** | Probabilities are found using **integration** |
| 6 | Example: $P(X = 2) = 0.3$ | Example: $P(2 \le X \le 4) = \int f(x)\,dx$ (area) |

Mathematical Expectation – Definition & Properties

**Mathematical expectation of a random variable**

If X is a discrete or continuous random variable with probability function or probability density function f(x).then the mathematical expectation of X is usually denoted by E[X] or μ and defined by

$$(i)\mu = E(X) = \sum xf(x)$$

$$= \int_{-\infty}^{\infty} xf(x)dx$$

**Properties of mathematical expectation of a random variable:**

•If b is a constant then E[b] = b

•If X is a random variable with expectation E[X], then E[aX+b] = a E[X] + b, Where a and b constant.

•If X is a random variable with expectation E[X], then E[X-E(X)] = 0

•If X and Y are random variables then E[X+Y] = E[X] + E[Y]

•If X and Y are random variables then E[X-Y] = E[X] - E[Y]

# Variance of Random Variable – Definition & Properties

**Variance** of a random variable measures how much the values of the variable **spread out** or **deviate** from its **mean (expected value)**.

**Properties of varianceof a random variable:**

•If b is a constant then $V[b] = 0$

•If X is a random variable with expectation $E[X]$, then

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

•If X and Y are random variables then $V[X+Y] = V[X] + V[Y]$

•If X and Y are random variables then $V[X-Y] = V[X] + V[Y]$

•If X is a random variable, then $V[aX+b] = a^2 V[X]$ , Where a and b constant.

## {} boshbe na

## ,mathmatical expression (mean)

**Co-efficient variable[CV]=[SP(x) / E(x)]**

**sp(x)=** $\sqrt{V(x)}$

P(x=4)=? Range er moddhe na hole direct 0.

# Importance of Random Variable & Mathematical Expectation in CSE

**. Random Variable**

- Shows **random things** like how many users visit a website or when a computer might fail.
- Used in **random-based programs** to make them faster or better.
- Helps check how **reliable** a system is.
- Used to **simulate** or test systems before real use.

**2. Mathematical Expectation**

- Finds the **average result** of something, like how long a program usually takes to run.
- Helps guess the **usual work** a server or network will have.
- Used in **machine learning** to reduce mistakes.
- Helps plan **resources** by predicting what will be needed on average.

**Application problem:** A discrete random variable X has the following probability function:

| Values of X:x | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| f (x) | 0.12 | 0.18 | k | 0.30 | 0.16 |

Find the value of k, Compute (ii) P[X>3] ; (iii) P[1< X < 4] ; and (iv) P[X<1].

Solution: (i) Since,

$$\sum f(x) = 1$$

Or, (0.12+0.18+k+0.30+0.16) = 1

Or, 0.76+k = 1

Or, k =1-0.76

Or, k = 0.24

(ii) P[X>3] = P[X=4] = 0.16

(iii) P[1< X < 4]= P[X=2]+ P[X=3]

= k+0.30

= 0.24+0.30 = 0.54

(iv) P[X<1] = P[X=0]

= 0.12

**Assignment problem:** A discrete random variable X has the following probability function:

| Values of X:x | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| f (x) | 0.10 | 0.30 | 0.20 | 0.25 | 0.15 |

Find the value of (i) $P[X=1]$; (ii) $P[X>3]$ ; (iii) $P[1< X < 4]$ ; and (iv) $P[X<1]$.

Problem: A continuous random variable X has the following probability density function:

$$f(x) = kx^2 \; ; \; 0<X<1$$

i. Find the value of K

ii. probability that X lies between 0.2 and 0.50

iii. probability that X less than 0.30 and

iv. probability that X greater than 0.75

v. E[x]

vi. V[x]

vii. SD[x]

CHAPTER 7 & 8 DISTRIBUTION

# CHAPTER 8
# ESTIMATION AND TEST OF HYPOTHESIS

**Hypothesis**: A statement about the nature of a population.

Example: Students who eat breakfast will perform better on a stat exam than students who do not eat breakfast.

**Test of Hypothesis**: The statistical procedure which is used to verify any statement or assumption about population parameter on the basis of sample observations is known as test of significance.

**Null Hypothesis:** The hypothesis which we are going to test foe possible rejection under the assumption that it is true. Null hypothesis is denoted by

$H_0$ ; $H_0$: $\mu_1 = \mu_2$

**Alternative Hypothesis**: Each of all possible hypothesis other than null hypothesis is called alternative hypothesis and is usually denoted by $H_1$; $H_0$: $\mu_1 \neq \mu_2$

**Type-I error**: The error of rejecting $H_0$ (accepting $H_1$) which is true is called error of first kind or Type-I error. — $\alpha$

100

**Type –II error**: The error of accepting null hypothesis $H_1$when it is false is called error of $2^{nd}$ kind ~~or $H_1$~~ Type-II error is denoted by $\beta$

**Level of significance:** The probability of Type-I error denoted by $\alpha$ is called level of significance/ ~~Symbolically $\alpha$=(rejecting $H_0$|$H_1$ is~~ true) . ~~It is also known is size of a test~~

$\alpha = 5$

**P –value:**

The p value is the level at which the given value of the test statistic (such as t, F, ) would be on the border line between the $\alpha$ acceptance and rejection regions. The decision rules, which most researchers follow in stating their results, are as follows:

P <0.05 the results are regarded as statistically significant

P value is lest possibles values of alpha where we reject null hypothesis

# COMMONLY USED TEST STATISTIC

## 1. The normal test (z-test):

If t test value >30 then will do z test

## 2. The t test:

a. To test the significance of population mean μ

b. To test the significance in the difference between two population means

c. To test the equality of two correlated means

d. To test the significance of population correlation coefficient

e. To test the significance of the regression coefficient

## 3.Chi-square ($\chi^2$)test

a. To test the significance of a specified population variance
b. To test the goodness of fit of a distribution
c. To test the independence of attributes

## 4. F-test:

a. To test the equality of several population means
b. To test the equality of two population variance
c. To test the significance of regression

Chap :7

**Theoretical distributions are**

1. Binomial distribution
2. Poisson distribution

} Discrete distribution

3. Normal distribution $\longrightarrow$ Continuous distribution

## Binomial Distribution

### Binomial Distribution

A discrete random variable X is said to have binomial distribution if its probability function is as follows:

$$f(x) = {}^{n}C_{x}\ p^{x}\ q^{n-x}\ ;\ x = 0,1,2,\ldots\ldots\ldots\ldots\ldots,n$$

Where n and p are the parameters of the distribution and p + q = 1

N er man 30 er choto or soman hote hbe beshi hole poisson distribution hbe
X is no. of success, n is no. of trial,p probability of success and q probability of faillure

## Assumption of Binomial Distribution

**Assumption of binomial distribution:**

**1. Fixed Number of Trials (n)**
The experiment consists of a fixed number of trials, say n.
**2. Two Possible Outcomes**
Each trial results in one of **two outcomes**: "success" or "failure" (binary outcome).
**3. Constant Probability of Success (p)**
The **probability of success P**, remains **constant** for each trial.
**4. Independence of Trials**
Each trial is **independent**, meaning the outcome of one trial does **not affect** the outcome of another.
**5. Discrete Random Variable**
The random variable X, which denotes the **number of successes**, takes **discrete** values from 0 to n.

## Poisson Distribution

### Poisson distribution $n \le 30$ $n \to \infty$

A discrete random variable X is said to have Poisson distribution if its probability function is as follows:

$$f(x) = \frac{e^{-m}\, m^{x}}{x!}\ ;\ x = 0,1,2,\ldots\ldots\ldots\ldots\ldots\ldots,\infty$$

Where m is the parameter of the distribution and e = 2.718

m = average

$e^{-m}\ m^{n}$

N er man barle p o barbe tkn binomial possession hbe

# Practical Application of Poisson Distribution (in CSE)

**Poisson distribution** is used to count how many times something happens in a fixed time, when the events happen **randomly**.

---

📌 **Uses in Computer Science (CSE):**

1. ✅ **Network Traffic**
   - To count how many **data packets** arrive in 1 second.
   - Helps in avoiding **network overload**.
2. ✅ **Website Requests**
   - To know how many **users visit a website** in a minute.
   - Helps in **server management**.
3. ✅ **System Errors**
   - To find how many **errors** happen during data transfer.
   - Useful in **error checking and fixing**.
4. ✅ **CPU or Task Scheduling**
   - To model how often tasks arrive in a **CPU queue**.
   - Helps in **task planning**.

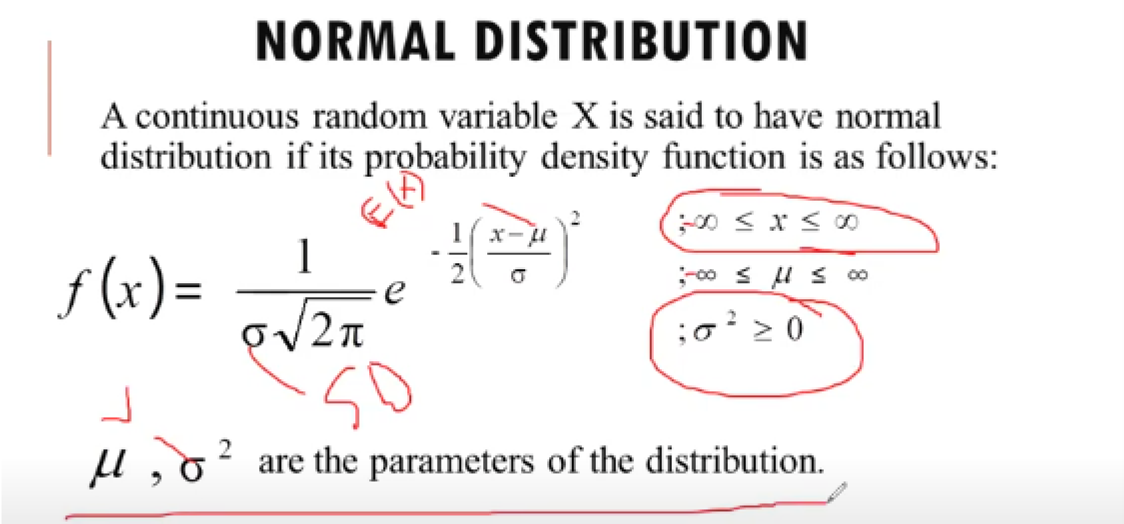
5. ✅ **Cloud Server Load**
    - To count how many **requests** are sent to cloud services.
    - Used in **load balancing**.
6. ✅ **Hardware or System Failures**
    - To predict how many **times a system may fail** in a month.
    - Helps in **maintenance planning**.

---

### ✅ In Simple Words:

Poisson distribution helps in **counting random events** like requests, errors, or failures.
It is very useful in **making computer systems faster, safer, and more efficient.**

Normal Distribution



Importance of Normal Distribution

I. **Importance of normal distribution:**

II. In practice under certain condition most of the probability and sampling distributions can be approximated by normal distribution.

III. According to central limit theorem, if mean and variance of a distribution exist, then the distribution converted to normal distribution.

IV. Nomal distribution is the basis of all the sampling distribution.

V. Assumption of normaility is the basis of all the test of significance in applied statistics.

VI. Normal distributions find its application in industrial statistics such as quality control.

# Importance of Poisson Distribution in CSE

## ✅ Importance of Poisson Distribution in Computer Science (CSE)

Poisson distribution is used to count how many times an event happens in a fixed time or space. It is useful when the events are random and not related to each other.

🔹 **Uses in CSE:**

1. It is used to count how many **data packets arrive at a router** in one second.

2. It helps predict how many **users visit a website** in a given time.
3. It is used to check how many **errors happen during data transmission**.
4. It helps estimate how often a **system or device might fail**.
5. It is used to model the **number of tasks entering a CPU or server queue**.

---

### ✅ Importance of Normal Distribution in Computer Science (CSE)

Normal distribution is a bell-shaped curve where most values are close to the average. Many real-world things follow this pattern.

🔹 **Uses in CSE:**

1. It is used to study **CPU and memory usage patterns**.

2. It helps measure and analyze **system performance**, like average response time.

3. It is used in **machine learning** to understand how data is spread.

4. It helps find **anomalies**, which are values far from the average.

5. It is useful in **software testing and quality control**.

---

Chap :8

in the following contingency table.

| Results | Job satisfaction | |
|---|---|---|
| | Yes | No |
| Excellent | 20 | 70 |
| Good | 45 | 65 |

Compute the value of Chi-square for the above data and comment on 5% level significance.

sol:

Ho: There is no relationship between results and job satisfaction
H1: There is a relationship between results and job satisfaction

We know, $\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Observe

Expecte

Computation table

| Results | Job satisfaction | | Total |
|---|---|---|---|
| | Yes | No | |
| Excellent | 20 ($O_{11}$) | 70($O_{12}$) | 90 – $R_1$ |
| Good | 45($O_{21}$) | 65($O_{22}$) | 110 – $R_2$ |
| Total | 65 – $C_1$ | 135 – $C_2$ | 200 – N |

Here, $E_{11} = \frac{90 \times 65}{200} = 29.25,$

$E_{12} = \frac{90 \times 135}{200} = 60.75,$

$\frac{R_1 \times C_1}{N}$

Here, $E_{11} = \frac{90 \times 65}{200} = 29.25,$

$E_{12} = \frac{90 \times 135}{200} = 60.75,$

$E_{21} = \frac{110 \times 65}{200} = 35.75,$

$E_{22} = \frac{110 \times 135}{200} = 74.25$

$$\text{Therefore, } \chi^2 = \frac{(20-29.25)^2}{29.25} + \frac{(70-60.75)^2}{60.75} + \frac{(45-35.75)^2}{35.75} + \frac{(65-74.25)^2}{74.25}$$

$$= 2.93 + 1.41 + 3.75 + 1.15 = 9.24$$

$$\text{Therefore, } \chi^2 = \frac{(20-29.25)^2}{29.25} + \frac{(70-60.75)^2}{60.75} + \frac{(45-35.75)^2}{35.75} + \frac{(65-74.25)^2}{74.25}$$

$\chi^2_{cal} = 2.93 + 1.41 + 3.75 + 1.15 = 9.24$

Desion: $\chi^2_{cal} = 9.24$, $\chi^2_{tab} = 3.84$

$\chi^2_{cal} > \chi^2_{tab}$ then the null hypothe[illegible] is re[illegible]

null hypothesis rejected sice Xtable < X0 its decision

| | |
|---|---|
| 8. | A software engineer records an average of 5 login failures per hour on university portal. Find the probability that<br>i) Exactly 4 login failures occur<br>ii) At most 2 login failures occur<br>iii) At least 3 login failure occur. |

=3.84)

$f(x) = \frac{e^{-m} m^x}{x!}$ ; $x = 0, 1, 2 \ldots$

$f(x) = \frac{e^{-5} 5^x}{x!}$ ; $x = 0, 1, \ldots$ | $m = 5$

i : $f(x = 4) = \frac{e^{-5} 5^4}{4!}$

ii $f(x \le 2) = f(x=0) + f(x=1) + f(x=2)$

Activate Windows
Go to Settings to activate Windows.