Syllabus

Mathematical Foundations.

Review of probability theory, Correlation, Dependence and Independence, Conditional Probability, Baye's Theorem, The Normal Distribution, The central Limit Theorem, Hypothesis and Inference: Statistical Hypothesis Testing, Confidence Intervals, P-hacking, Bayesian Inference.

Module-2
Mathematical Foundations

① Review of Probability theory:-

→ Probability theory is the study of uncertainty.

→ It describes how likely an event is to occur.

Probability is the measure of the likelihood that an event will occur in random experiment.

It ranges between 0 and 1.

Example:-

Tossing a coin.

The possible outcomes are head and tail.

Random Experiment:-

Random experiment is an experiment that is repeated under a identical condition for which we have knowledge for all positive outcomes.

Example:

Before throwing a coin or before rolling a die we don't know the result.

Sample Space (s)

A set of possible outcomes in any given random experiment is random sample space and it is denoted by 's'.

Example:

1) If two coins are tossed, the sample space

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

2) When two dies are thrown

$$S = \{(1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)$$
$$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$$
$$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$$
$$(4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6)$$
$$(5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (5,6)$$
$$(6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)\}$$

Generally a sample space is classified according to the number of elements contain.

1. discrete
2. continuous.

Discrete sample space :-

If a sample space has finitely or countable infinite number of elements they contain.

Continuous sample space :-

If the element of the sample space has a continuous then the sample space is said to be continuous.

Outcome :-

The result of an random experiment will be called an outcome.

Trial and Event :-

Any particular performance of an random experiment is called Trial.

Outcomes on combination of outcomes are called as Event.

Probability of an Event :-

Let 'S' be the sample space and 'A' be the event associated with random experiment then the probability of A is denoted as,

$$P(A) = \frac{\text{No. of favourable outcome of A}}{\text{Total no. of outcomes}}$$

$$P(A) = \frac{n(A)}{n(S)}$$

* If the probability of occurrence of an event is 0, it is called impossible event.

* If the probability of occurrence of an event is 1, it is called sure event

Simple Events

→ If the event E has only one sample point of a sample space it is called simple event.

→ It is an event that consists of exactly one outcome.

Eg: Throwing a die the possibility of two appearing on the die is an simple event

$$E = \{2\}.$$

Compound Events:-

If there is more than one sample point on a sample space, it is called compound event.

Eg: Throwing a die, possibility of an odd number is

$$E = \{1, 3, 5\}$$

There is more than one possibility. So it is called compound event.

Independent and Dependent Events

If the occurrence of any event is completely not affected by the occurrence of any other event, it is called independent event.

If the events which are affected by other events are known as dependent event.

Correlation:-

what is Correlation?

- statistical relationship between the two entities.
- It measures the extent to which two variables are linearly related.

For example:

The height and weight of a person are related and taller people tend to be heavier than shorter people.

There are three types of correlation

* Positive Correlation
* Negative Correlation
* No Correlation.

Positive Correlation:-

→ A positive correlation means that this linear relationship is positive.

→ Two variables increase or decrease in the same direction.

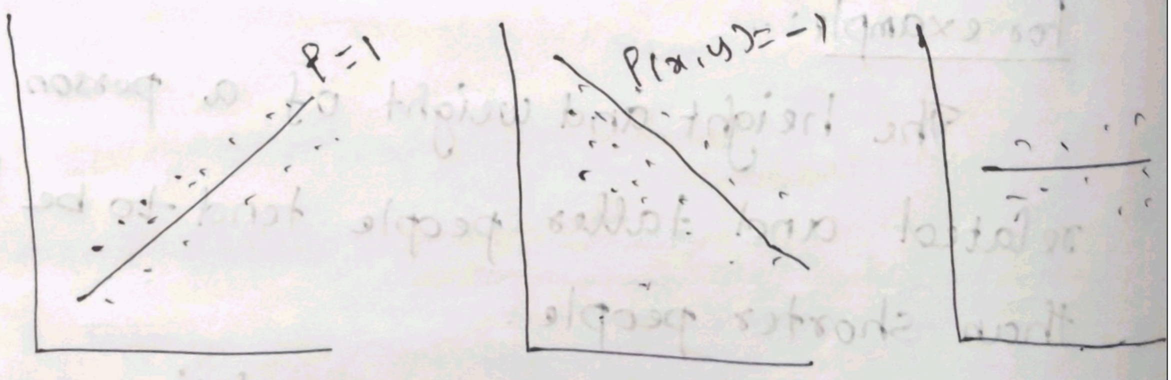
Negative Correlation.

The relationship line has a negative slope. The variables change in opposite direction.

(ie) one variable decreases while the other increases.

No Correlation:-

The variables behave very differently and thus have no linear relationship.



Example

X	Y
Height	Weight

X ↑ whether Y ↑ or not

X ↑ Y ↓

$$\begin{array}{llll} & & & Cov(x,y) \\ x\uparrow & y\uparrow & \rightarrow & +ve \\ x\uparrow & y\downarrow & \Rightarrow & -ve \end{array}$$

Covariance helps to find the <u>direction of relationship</u>.

Pearson Correlation Coefficient (CC)

① Covariance

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

② $CC = \rho_{(x,y)} = \dfrac{Cov(x,y)}{\sigma_x \sigma_y}$ → to know strength

$\sigma_x\sigma_y$ std deviation.

$\sigma_x \sigma_y \Rightarrow$ how much amout of positivity (or) negativity.

$\Rightarrow$ ~~what~~ How strong this is correlated.

Range of CC is $-1 \leq \rho \leq 1$

(or)

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

r - Coefficient of correlation.

$\bar{x}$ - Mean of x

$\bar{y}$ - Mean of y

$x_i\ y_i$ - Samples of variable x, y.

⇒ It defines the strength of the relationship between two variables and their association.

Spearman's Rank Correlation

Measures the strength and direction of association between two ranked variables.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

di – difference between the ranks of corresponding variables.

n – number of observations.

Covariance

– important topic in Data preprocessing.

Two random variables

Size	prize
1200 sqm	100k $
1800 sqm	200k $
1800 sqm	300 k$

Covariance

- $$\mathrm{Cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

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

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \underbrace{\mu_x}_{\text{Mean}}) * (x_i - \mu_x)$$

$$\mathrm{Cov}(x, x) = \mathrm{var}(x)$$

$x \uparrow \; y \uparrow = \boxed{}$ +ve

$x \uparrow \; y \downarrow = \boxed{-ve}$

On Programming Side,

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

Note :- just sum up x, y, x^2, y^2 and xy

x	1	2	3	4	5	6
y	2	4	7	9	12	14

$n = 6$ samples

x	y	xy	x^2	y^2
1	2	2	1	4
2	4	8	4	16
3	7	21	9	49
4	9	36	16	81
5	12	60	25	144
6	14	84	36	196
21	48	211	91	490
Σx	Σy	Σxy	Σx^2	Σy^2

$$P = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}}$$

$$= \frac{6(211) - 21 \times 48}{\sqrt{\left[6 \times 91 - (21)^2\right]\left[6 \times 490 - (48)^2\right]}}$$

$$r = \frac{1266 - 1008}{\sqrt{[546-441][2940-2304]}}$$

$$= \frac{258}{\sqrt{105 \times 636}} = \frac{258}{\sqrt{66780}}$$

$$= \underline{0.998}$$

Mean Median and Mode

<u>Mean</u> - The arithmetic average of all numbers.

<u>Median</u> - The value in the center when the numbers are arranged least to greatest

<u>Mode</u> - The most commonly appearing value.

Given: 9, 3, 1, 8, 3, 6.

1 3 3 6 8 9

$$\text{Mean} = \frac{1+3+3+6+8+9}{6} = \frac{30}{6} = \underline{\underline{5}}$$

$$\text{Median} = \frac{3+6}{2} = \frac{9}{2} = \underline{\underline{4.5}}$$

Mode = 3.

2) Given 5, 9, 2, 1, 7, 4, 7

1 2 4 5 7 7 9

$$Mean = \frac{1+2+4+5+7+7+9}{7}$$

$$= \frac{42}{7} = 6$$

Median = 5

Mode = 7

Correlation Coefficient

Birth Rate X	Death Rate Y
24	15
26	20
32	22
33	24
35	27
30	24

Solve

X	9	8	7	6	5	4	3	2	1
Y	15	16	14	13	11	12	10	8	9.

X	60	34	40	50	45	41	22	43
Y	75	32	34	40	45	33	12	30

Another Method.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

Correlation – measure of the strength of association among and between variables.

Correlation coefficient:-

– It expresses the strength of association between variables.

r vs r^2 (coefficient of determination)

↓

Correlation

It gives you a positive or negative number.

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

$df = N-2$

Student	$(n=7)$ X Hours studied	Y $(n=7)$ Exam Score
Jenn	2	58
Tan	4	32
Lew	5	63
Chelsea	7	87
Miyuki	3	67
Hamid	1	45
Juan	6	68

$N = 7+7 = \underline{14}$

$$\text{Mean } \bar{x} = \frac{2+4+5+7+3+1+6}{7}$$

$$= \frac{28}{7} = 4$$

$$\boxed{\bar{x} = 4}$$

$$\bar{y} = \frac{58+32+63+87+67+45+68}{7}$$

$$= \frac{420}{7}$$

$$\boxed{\bar{y} = 60}$$

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
2	58	2-4=-2	-2	4
4	32	0	-28	0
5	63	1	3	3
7	87	3	27	81
3	67	-1	7	-7
1	45	-3	-15	45
6	68	2	8	16
		0	0	142

$\sum(x-\bar{x})$

$\sum(x-\bar{x})(y-\bar{y})$

$(x-\bar{x})^2$	$(y-\bar{y})^2$
4	4
0	784
1	9
9	729
1	49
9	225
4	64
28	1864

$\sum(x-\bar{x})^2$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{142}{\sqrt{28}\sqrt{1864}}$$

$$= \frac{142}{5.29 \times 43.17}$$

$$= \frac{142}{228.46} = 0.62$$

positive

$\boxed{r = 0.62}$

(OR).

x	y	xy	x^2	y^2
2	58	116	4	3364
4	32	128	16	1024
5	63	315	25	3969
7	87	609	49	7569
3	67	201	9	4489
1	45	45	1	2025
6	68	408	36	4624

$\Sigma x = 28$ $\Sigma y = 420$ $\Sigma xy = 1822$ $\Sigma x^2 = 140$ $\Sigma y^2 = 27064$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{7 \times 1822 - 28 \times 420}{\sqrt{[7 \times 140 - (28)^2][7 \times 27064 - (420)^2]}}$$

$$= \frac{12{,}754 - 11760}{\sqrt{(980 - 784)}\sqrt{(189448 - 176400)}}$$

$$= \frac{994}{\sqrt{196}\sqrt{13048}}$$

$$= \frac{994}{14 \times 114.23}$$

$$= \frac{994}{1599.2}$$

$$= 0.62215$$

$$\boxed{P = 0.6}$$ ~~wrong~~ ~~answer~~

Ans = 0.6

Probability

Dependent and Independent Events

Dependent

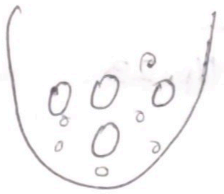
Two events A, B are dependent occuring of event A will affect the probability of Event B.

Example

Bag contains 5 black and 6 green balls. Randomly pick a ball from bag two times

A -> 1^{st} time should be black ball

B -> 2^{nd} time should be black



1^{st} time when we select a ball

$$\boxed{p(\text{black ball}) = \frac{5C_1}{11C_1}}$$

2^{nd} time -> selecting black ball second depend on 1^{st} time selected ball.

Independent Events

Two events A, B, ~~oo~~ occurrence of one event not affect the probability of other event.

$$P(A \cap B) = P(A) \cdot P(B)$$

Ex: Toss a coin two times

A -> 1^{st} time should be head

B -> 2^{nd} time should be head.

$$P(A \cap B) = P(A)(P(B)$$

$$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Types of Events

Impossible Event:-

Impossible event is an event that cannot happen.

Here E is empty set ie) $E = \varphi$

Example:

single throw of a die 

⇒ Getting a number multiple of 7 ⇒ Impossible

⇒ Number turns up is odd or even -
↳ event will happen

Event $E = \{1, 2, 3, 4, 5, 6\}$

Sure Event

A sure event is that ~~en~~ event which always happen whenever activity performed. That event contains whole sample space.

Simple Event

If an event E has only one sample point of a sample space it is called a simple (or elementary) event.

Example

Rolling a die

$S = \{1, 2, 3, 4, 5, 6\}$

Event → getting an odd number greater than 4.

$E = \{5\}$ → one sample point

So it is called simple event.

pblm

In a single throw of a die which of the following is a simple event?

a) Getting even number

b) Getting number multiple of 3

✓c) Getting multiple of 6

d) Getting odd number

a) 2, 4, 6   b) 3, 6   c) 6

d) 1, 3, 5

Compound Event:-

If an event has more than one sample point, it is called a Compound event.

Eg:- Tossing a coin twice

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

Event → atleast one head appear

$E = \{HH, HT, TH\}$

In a single throw of a die which of the following is a compound event?

a) Getting even number less than 3

b) Getting number multiple of 3 ✓

c) Getting multiple of 6

d) Getting odd number less than 3.

a) $\{2\}$ b) $\{3, 6\}$ $\{6\}$

$\{1\}$

Mutually Exclusive Events

Two events are said to be mutually exclusive events when both cannot occur at the same time.

* They don't have common outcome

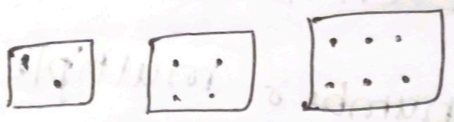
1) Tossing a coin.

A → Head

B → Tail.

2) A single throw of dice.

A →  odd.

B →  even.

In a single throw of a die

3) A → Getting number multiple of 3 {3, 6}

B → Getting a even number. {2, 4, 6}

A & B are not mutually exclusive

Qn: Which of the following events are mutually exclusive events when a coin is tossed twice?

a) A = exactly two heads B = atleast one head

b) A = exactly ~~one~~ two heads B = exactly two tails.

c) A = atleast one head B = atleast one tail.

If a coin tossed twice.

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

a) $A = \{\underline{HH}\}$ $B = \{\underline{HH}, HT, TH\}$

b) $A = \{HH\}$ $B = \{TT\}$

c) $A = \{HH, \underline{HT}, \underline{TH}\}$ $B = \{\underline{HT}, \underline{TH}, TT\}$

<u>Note:</u> If they have common outcome $\Rightarrow$ Not mutually exclusive

<u>Equally Likely Events</u>

Events which have the same chances of occurrence

(or)

when the likelihood of happening of two events are same they are known as equally likely events.

Equally likely

1) Tossing a coin

A = head $\quad n(A) = 1$

B = tail $\quad n(B) = 1$

2) A single throw of dice

A) odd $\quad n(A) = 3$

B) even $\quad n(B) = 3$

3) A single throw of a die

A → Getting a number less than 4 {1, 2, 3}

B → Getting a even number {2, 4, 6}

Independ Events

Independent events are those events whose occurrence is not dependent on any other event. (not related).

Eg: when a die is thrown twice, the result of the first throw does not affect the result of the second throw.

When two events E and F are independent

$$P(E,F) = P(E)P(F)$$

$P(A) + P(A') = 1$ $\quad (0 \leq P(A) \leq 1)$

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

$P(A \cup B) = P(A) + P(B)$ (Mutually exclusive)

$P(A \cap B) = P(A)\,P(B)$ (Independent Events)

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Baye's Formula

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

pblm

Consider an example of ~~o~~ rolling a die

If A is the event "the number appearing is odd" and B the event "the number appearing is a multiple of 3". Then,

$S = \{1, 2, 3, 4, 5, 6\}$, $n(S) = 6$.

$P(A)$ = No. appearing is odd $= \{1, 3, 5\}$

$$= \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

[illegible]

Conditional Probability

An ordinary die rolled once.
What is the sample space?

$$S = \{1, 2, 3, 4, 5, 6\}$$

What is the probability that a 2 is rolled?

$$P(2) = \frac{1}{6}$$

Suppose the die come up with an even number. What is the reduced sample space?

$$S_R = \{2, 4, 6\}$$

$p(B) =$ No. appearing is a multiple of 3

$(B) = \{3, 6\}$

$n(B) = 2$

$p(B) = \frac{n(B)}{n(S)} = \frac{2}{6} = \frac{1}{3}$

$p(A \cap B) = \frac{\{3\}}{n(S)} = \frac{1}{6}$

Conditional Probability

An ordinary die rolled once

what is the sample space?

$S = \{1, 2, 3, 4, 5, 6\}$

what is the probability that a 2 is rolled?

$p(2) = \frac{1}{6}$

Suppose the die come up with an even number. what is the reduced sample space?

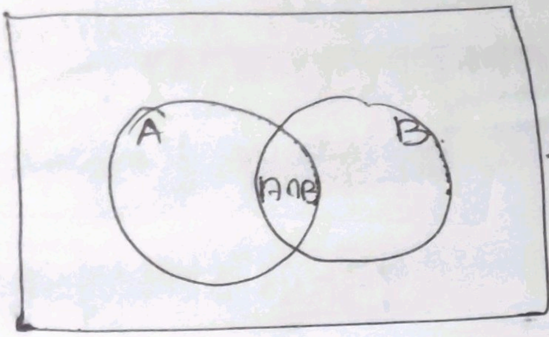
$S_B = \{2, 4, 6\}$

Given an even number is rolled, what is the probability that it is a 2?

$$P(2|\text{Even}) = \frac{1}{3}$$

Conditional Probability Definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{for } P(B) > 0$$



→ entire sample space

$A = \{1, 2, \underline{3, 4, 5}\}$

$S = \{1, 2, 3, 4, 5, 6\}$

$B = \{\underline{3, 4, 5}, 6\}$

What is the conditional probability of A given B?

~~What is the co~~

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{3/6}{4/6}$$

$P(A \cap B) = 3/6$

$$= \frac{3}{4}$$

$$P(A) = \frac{5}{6}.$$