Lecture - 04

→ IQR (Python Code)

→ Probability.

→ Permutation and Combination.

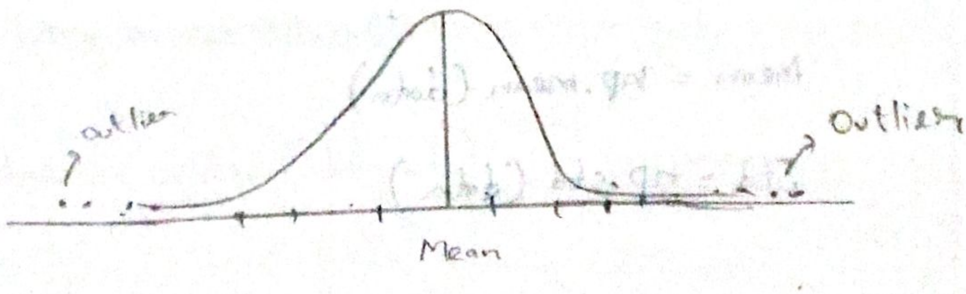
→ Confidence Interval

→ P Value

→ Hypothesis testing.

Outlier:

The outlier lies after the left and right of 3rd Standard deviation.



Python Code:

Example for outlier ~~find~~ ~~program~~.

```
import numpy as np

import matplotlib.pyplot as plt

% matplotlib inline

# Define data

dataset = [11, 10, 12, 14, 12, 15, 14, 13, 15, 102, 12, 14, 17, 19,
           107, 10, 13, 12, 14, 12, 108, 12, 11, 14, 13, 15, 10, 15,
           12, 10, 14, 13, 15, 10]

Outliers = [] # initialize empty list
```

~~def detect_outlier(data):~~

```
def detect_outlier(data):

    thresold = 3   # 3rd std away should be call outliers.

    Mean = np.mean(data)
    Std = np.std(data)

    for i in data:

        z_score = (i - mean)/std

        if np.abs(z_score) > thresold:

            Outliers.append(i)
```

return Outliers

detect_outlier (dataset)

Output:

[102, 107, 108]

IQR: [Finding IQR]

1. Sort the data.

2. Calculate Q1 and Q3.

3. IQR (Q3 - Q1)

4. Finding the lower fence (q1 - 1.5 (iqr))

5. Finding the upper fence (q3 + 1.5 (iqr))

Program:

```
# Using above dataset

dataset = Sorted ( dataset)


# finding Quarter Q1 and Q3

q1, q3 = np. Percentile (dataset, [25,75])

# finding IQR

IQR = (q3 - q1)
```

find lower and higher fence

```
lower_fence = Q1 - (1.5 * IQR)
Higher_fence = Q3 + (1.5 * IQR)

print (lower_fence, Higher_fence)

        7.5    ,    19.5
```

Condition for finding Outlier and removing

```
for i in dataset :

    if i < 7.5 and i > 19.5 :
        dataset.remove (i)
        ~~return i~~

    else :
        return "No outlier present"
```

PROBABILITY

Probability is a measure of the likelihood of an event.

Eg: Rolling a dice $\{1, 2, 3, 4, 5, 6\}$

To find probability of getting 6 in dice,

$$P(6) = \frac{\text{Number of way an event occur}}{\text{Number of possible Outcomes}}$$

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

Another Example,

Toss a Coin $\{H, T\}$

$P(H) = 1/2$

$P(T) = 1/2$

Addition Rule (probability, "or")

We need to Understand addition rule, first know about mutual exclusive and non-mutual exclusive events.

Mutual - Exclusive events:

Two events are Mutual exclusive if they Cannot Occur at the Same time.

Eg: Rolling a dice {1, 2, 3, 4, 5, 6}

Outcome - {1}, {2}, {3}, {4}, {5}, {6}

Cannot come - {1, 2}, {3, 4} [Two or more events cannot occur at same time]

Another Example: Tossing a coin - Either head (or) tail.

Non-Mutual Exclusive Events:

Multiple events can occur at the same time.

Eg: Deck of cards {Q, ♡} [Queen card with heart card]

↑

Two events occur.

Addition rule:

This rule is used to find probability of two events based on the type of event occur.

Problem:

Problem based on mutual exclusive events.

1. If I toss a Coin, what is the probability of the Coin landing on heads or tails ?

Soln:

The above events are Mutual Exclusive.

P(A) - probability of head event   P(B) - Probability of tail event.

P(A (or) B) ?

$P(A \text{ or } B) = P(A) + P(B)$ [Addition rule]

$= \frac{1}{2} + \frac{1}{2}$

$= 1$

$\boxed{P(A \text{ or } B) = 1}$

Another Example,

Rolling a dice.

P(1 or 3 or 6) ?

$P(1 \text{ or } 3 \text{ or } 6) = P(1) + P(3) + P(6)$

$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$

$P(1 \text{ or } 3 \text{ or } 6) = \frac{3}{6} \Rightarrow \frac{1}{2}$

Problem:

Problem based on non-mutual exclusive events.

1. You are picking a card randomly from a deck. What is the probability of choosing a card that is Queen or a heart?

The above events are non-mutual exclusive events.

$P(A)$ - probability of card that are Queen.

$P(B)$ - probability of card that are a heart.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$= \frac{16}{52}$$

$$\boxed{P(A \text{ or } B) = \frac{16}{52}}$$

$P(A) - \frac{4}{52}$ ↑ Queen

$P(B) - \frac{13}{52}$ ↑ Heart

$P(A \text{ and } B) = \frac{1}{52}$ ↑ Card with Queen and heart

Multiplication Rule:

We need to Understand multiplication rule, First Understand about independent and dependent events.

Independent Events:

It is the event, it does not depend upon the other event.

Eg: Rolling a dice $\{1, 2, 3, 4, 5, 6\}$

↑

Each and every are independent

Outcome like 1, 1, 2, 2, 6, 5 . . .

Dependent Event:

One event are depend upon the other events.

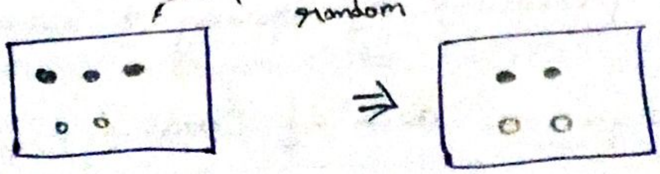
These are non-independent.

• → red ball

○ → Green ball

Eg:

→ pick one Green ball random



$P(\text{Green}) = \frac{3}{5}$ $\qquad$ $P(\text{red}) = \frac{2}{4}$

dependent

⇓ we apply Something in ML called as,

Naive Bayes (Conditional Probability)

Problem:

Problem based on Independent events.

1. What is the probability of rolling a "5" and then a "4" in a dice ?

The above event are independent event.

$P(A \text{ and } B) = P(A) * P(B)$ [multiplication rule]

$P(5 \text{ and } 4) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$

$$\boxed{P(5 \text{ and } 4) = \frac{1}{36}}$$

Problem:

Problem based on dependent event

1. What is the probability of drawing a Queen and then a Aces from a deck of Cards ?

The above events are dependent event.

Conditional Probability

$$P(A \text{ and } B) = P(A) * P(B/A)$$

Bayes theorem

$$P(Q \text{ and } A) = P(Q) * P(A/Q)$$

$$= \frac{4}{52} * \frac{4}{51}$$

$$\boxed{P(Q \text{ and } A) = \frac{16}{2652}}$$

Permutation and Combination:

Permutation:

→ Permutation is the arrangement of items in which order matters.

→ Number of ways of selection and arrangement of items in which order matters.

$$\boxed{{}^nP_r = \frac{n!}{(n-r)!}}$$

Combination:

→ Combination is the selection of items in which order does not matter.

→ Number of ways of selection of items in which order does not matters.

$$\boxed{{}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}}$$

Permutation Example:

I went to school trip to choclate factory. In Choclate factory six type of brands are available are dairy milk, 5 star, mars, kitkat, Milkybar and Snickers. Assign a task for Student, ~~whoever~~ whichever Chocalate you see first three chocalate need to be noted in diary.

Soln:

Student → $\underset{\text{1st noted}}{\underline{6 \text{ option}}}$ $\underset{\text{2nd noted}}{\underline{5 \text{ option}}}$ $\underset{\text{3rd noted}}{\underline{4 \text{ option}}}$

$= 6 \times 5 \times 4 = 120$ ways.

↓ Permutation

Diarymilk, milkybar, kitkat

kitkat, mikybar, dairymilk.

kitkat, Diarymik, milkybar.

___ ___ ___

___ ___ ___

upto 120 way permutation.

So,

$$^{n}P_{r} = \frac{n!}{(n-r)!} = \frac{6!}{(6-3)!}$$

$$= \frac{6 \times 5 \times 4 \times \cancel{3!}}{\cancel{3!}} = 120 \text{ ways}.$$

Permutation return all possible ways of items.

Combination Example:

Combination return only unique possible ways of items.

$$^{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{6!}{3!(6-3)!} = \frac{6 \times 5 \times 4 \times \cancel{3!}}{3! \times \cancel{3!}}$$

$$= \frac{\cancel{6}^{2}\times 5\times \cancel{4}^{2}}{1\times\cancel{2}\times\cancel{3}_{1}}$$

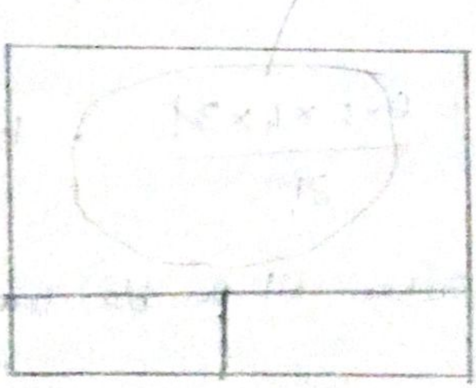
$$= 2\times 5\times 2 = 20 \text{ ways [unique ways]}$$

__Concept of P-value:__

P-value refers to the probability value of data fall in distribution.

__Example,__

This region Considered as more touch.



MOUSE PAD IN LAPTOP

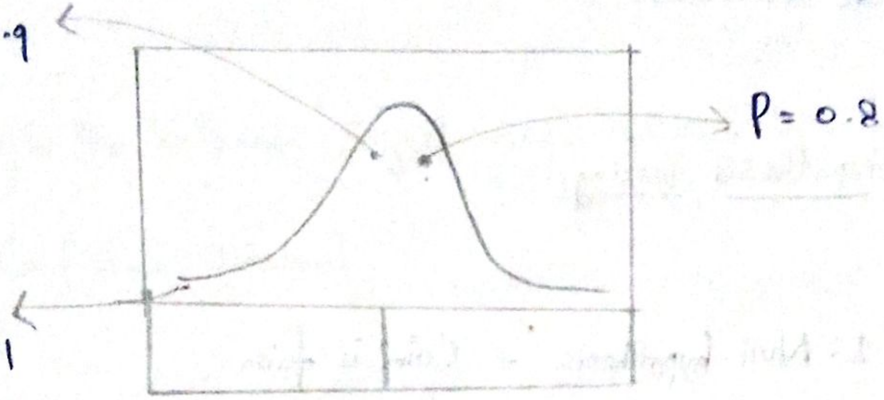
Every 100 time I touch the mouse pad 80 times I touch this specific region.

Consider Same as,

P=0.9

P = 0.8

P=0.01



The above one represents the how much data fall in the region in the terms of Probability.

Combine topic Hypothesis testing, Confidence Interval, Significance Value.

Coin → Test wheather this coin is a fair Coin or not by performing 100 toses.

Coin is fair, $P(H) = 0.5$, $P(T) = 0.5$

50 times head (or) tail we consider the Coin is fair.

Consider 50 times head, we Consider the Coin is fair.

To test the Coin, we Use the hypothesis testing on statistics.

<u>Hypothesis Testing:</u>

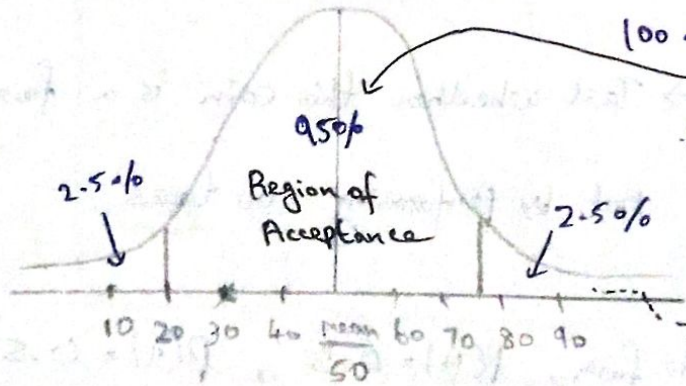
1. Null hypothesis - Coin is fair.
2. Alternate hypothesis - Coin is not fair.
3. Experiment.
4. Reject or accept the null hypothesis.

$\boxed{CI = 95\%}$

Significance Value (α) = 0.05

$1 - 0.05 = 0.95$

$100 - 5 = 95\%$



$5\% = 2.5\% + 2.5\%$

[Two tail test]

→ Region of rejection

20, 75 → Critical Value

50 times → head [we definitely say it is fair coin]

We Cannot perform 100 experiment because it is difficult.

So, In domain expert define Something called as Significance Value for normal distribution.

With the help of Significance Value, we define the Confidence Interval.

With the help of Confidence Interval, we Conclude the Coin is fair Coin (or) not.

Then, I toss a coin 30 times. Then Say the Coin is fair Coin or not.

The Confidence interval with 20-75. So, obiviously. The Coin is fair.

Without 100 experiment we say that the coin is fair or not based on the Confidence Interval.

The data fall outside the Confidence Interval. we Say the Coin is not fair.

Significance value $\alpha = 0.05$

15 heads - Coin is not fair

80 heads - Coin is not fair

40 heads - Coin is fair