TASK - 09: BWT - Data - Science.

Statistics:

① **Mean:** The mean average of sets of numbers.

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Example: 2, 4, 6, 8, 10

$$\bar{x} = \frac{2+4+6+8+10}{5} = 6$$

② **Median:** Middle value of sets of numbers

if $\boxed{n \text{ is odd,}}$
Median $\Rightarrow$ arrance in accending order and middle is mean.

example:

S = {7, 2, 4, 1, 5}
S = 1, 2, 4, 5, 7

$\boxed{\text{Median} = 4}$

if $\boxed{n \text{ is even: } \frac{\text{Sum of middle numbers}}{2}}$

example: $S = \{4, 3, 2, 1\}$

$$S = \frac{3+2}{2} = 2.5$$

③ Mode:

The most frequently number occurs in dataset

example: 5, 3, 4, 3, 2, 3, 7, 3

$\boxed{\text{Mode}(3) = 4}$

④ Normal Distribution:

A probability distribution that is symmetric about means, showing the data near the mean are more frequent in occurance.

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

example:

Mean (μ) = 50.

Standard daviation (σ) = 5

Prob b/w 45 & 55 = ?

so

convert 45 and 55 to z-score.

$$z_1 = \frac{45-50}{5} = -1$$

$$z_2 = \frac{55-50}{5} = 1.$$

$$P(45 < X < 55) = P(-1 < Z < 1)$$

$P(-1 < Z < 1)$ or

$$f(45) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{(45-50)^2}{2 \cdot 5^2}} = \frac{1}{5\sqrt{2\pi}} e^{-0.5}$$

$$f(55) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{(55-50)^2}{2 \cdot 5^2}} = \frac{1}{5\sqrt{2\pi}} e^{-0.5}$$

$$P(45 < X < 55) = \int_{45}^{55} \frac{1}{5\sqrt{2\pi}} e^{-\frac{(45-50)^2}{2 \cdot 5^2}} dx.$$

$$\boxed{P = 0.6827}$$

example:

z-score = ?

$\mu = 50$, $\sigma = 10$.

$$z = \frac{X-\mu}{\sigma} = \frac{60-50}{10} = 1$$

(5) Binomial Distribution:

Describes the number of success in fixed number of independent Bernoulli trial..

$$\boxed{P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

example: $n = 10$, $k = 6$, $p = 0.5$

n → no. of trials, k → number of success, p → probability of success.

$$P(X=6) = \binom{10}{6} (0.5)^6 (0.5)^{10-6}$$

$$\boxed{P(X=6) = 0.2051}$$

example:

$n = 8$, $k = 5$, $p = 0.8$

$$P(0.6) = \binom{8}{5} (0.6)^5 (1-0.5)^{8-5}$$

$$\boxed{P(0.6) = 0.2787}$$

⑥ Poisson Distribution:

Describes number of events occuring within given time interval.

$$\boxed{P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}}$$

example:

$\lambda = 5$ (Avg no. of calls per hr)
$k = 3$ (number of calls.)

$$P(X=3) = \frac{e^{-5} \cdot 5^3}{3!}$$

$$\boxed{P(3) = 0.1404}$$

example: $\lambda = 2$ (Avg no. of misprint)

$k = 0$

$$P(0) = \frac{e^{-2} \cdot 2^0}{0!}$$

$$\boxed{P(0) = 0.1353}$$

(7) Uniform Distribution:

All outcomes are equally likely within certain interval.

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

example:

$a = 1$ (lower bound)
$b = 10$ (upper bound)

$$f(x) = \frac{1}{10-1} = \frac{1}{9}$$

Suppose for $X < 4$.

$$P(X<4) = f(x) = \frac{1}{9}$$

$$\boxed{P(X<4) = 0.111}$$

example:

$a = 15$, $b = 25$.

$$f(x) = \frac{1}{b-a} = \frac{1}{25-15}$$

$$\boxed{P(X=20) = 0.1}$$

Probability:

① Basic Probability:

Likelihood of Event.

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

example:

Probability of rolling a 3 on a fair six-sided die.

$$P(\text{rolling of } 3) = \frac{1}{6}$$

example:

Probability of drawing a Ace from shuffled deck of 52 cards.

Total Ace cards in deck = 4.

Total cards = 52.

$$P(\text{drawing of Ace}) = \frac{4}{52} = \frac{1}{13}$$

② Conditional Probability:

The probability of event occuring given that another event has already occurred.

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

example: Probability that a student studied given that they passed..

$P(Pass) = 0.70$, $P(Studied \mid Pass) = ?$

$$P(Studied \mid Pass) = \frac{P(Studiend \cap Pass)}{P(Pass)}$$

$$P(Studied \cap Pass) = 0.5 * 0.70 = 0.35$$

$$P(Studied \mid Pass) = \frac{0.35}{0.7} = 0.5$$

example:

$$P(Red \mid Blue) = \frac{P(Red \cap Blue)}{P(Blue)}$$

$$P(Blue) = \frac{7}{12}$$

$$P(Red \cap Blue) = P(Red) \cdot P(Blue \mid Red)$$

$$P(Blue \mid Red) = \frac{7}{11}$$

$$P(Red \cap Blue) = \frac{5}{12} * \frac{7}{11} = \frac{5}{24}$$

$$P(Red \mid Blue) = \frac{5/24}{7/12} = \frac{5}{14}$$

(3) Independent Events:

Two events Independent if occurance of one does not affect the occurance of other.

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

Example:

$P(\text{Heads}) = 0.5$

$P(\text{4 on die}) = 1/6$

$P(\text{Heads} \cap \text{4 on die}) = P(\text{Heads}) \times P(\text{4 on die})$

$= 0.5 \times \frac{1}{6}$

$$\boxed{P(\text{Heads} \cap \text{4 on die}) = 0.0833}$$

example:

$P(\text{Rain}) = 0.3$

$P(\text{Rain of Sat} \cap \text{Rain on Sun}) = ?$

$P(\,''\,) = P(\text{Rain}) \cdot P(\text{Rain}) = 0.3 \times 0.3$

$$\boxed{P(\,''\,) = 0.09}$$

(4) **Bayes Theorem:** (Also called Niave Bayes in ML)

Used to find probability of an even given prior knowledge of conditions related to the event.

$$\boxed{P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}}$$

Example:

$P(+\text{ test} \mid \text{has disease}) = ?$

$P(-\text{ test} \mid \text{no disease}) = ?$

0.5% population has disease of

$P(\text{Has disease}) = 0.005$.

$P(\text{has disease} \mid + \text{test}) = ?$

$$P(1) = \frac{P(+\text{test} \mid \text{Has disease}) \cdot P(\text{has disease})}{P(+\text{test})}$$

$P(+\text{test} \mid \text{no disease}) = 1 - P(-\text{test} \mid \text{no disease})$

$= 1 - 0.01 = 0.99$

$P(+\text{ test}) = (0.99 \times 0.005) + (0.01 \times 0.99)$

$= 0.0149$

$$P(\text{has disease} \mid + \text{test}) = \frac{0.99 \cdot 0.005}{0.0149} = 0.333$$

Example:

$P(\text{cond}) = 1 - P(\text{No Cond})$

$P(\text{No cond}) \neq 1 - P(\text{cond})$

E_1 = event of spam email.

E_2 = event of non-spam email.

A = event of dedecting spam email

$P(E_1) = 0.5$, $P(E_2) = 0.5$

$P(A, E_1) = 0.99$ & $P(A, E_2) = 0.05$

$$P(E_2, A) = \frac{P(A|E_2)\,P(E_2)}{P(A|E_1)\,P(E_1) + P(A|E_2)\,P(E_2)}$$

$$P(E_2|A) = \frac{0.05 * 0.5}{0.99 * 0.5 + 0.05 * 0.5}$$

$P(E_2|A) = 0.048 \Rightarrow 4.8\%$ emails are not spam.

(5) Law of Total Probability.

Provides a way to break down complex probabilities into simpler, conditional prob-

$$\boxed{P(B) = \sum_i P(B|A_i)\,P(A_i)}$$

example: $P(B) = 0.45)$

$P(\text{no rain}) = P(B') = 1 - P(B) = 1 - 0.45 = 0.55$

$P(A|B) = 0.42$

$P(A|B') = 0.90$

$P(A) = P(B)\,(P(A|B) + P(B')\,P(A|B')$

$= 0.45 * 0.42 + 0.55 * 0.9$

$\boxed{P(A) = 0.684}$