Introduction to Statistical Methods

BITS Pilani

Pilani Campus

BITS Pilani
Pilani Campus

Course No: AIMLCZC418
Course Title: ISM
WEBINAR 1

Topics - Webinar

- Descriptive Statistics
  - Measures of Central Tendency
  - Measures of Variability
- Probability
  - Probability – Introduction and Basics
  - Conditional probability
  - Bayes' theorem

Measures of Central Tendency

A psychologist wrote a computer program to simulate the way a person responds to a standard IQ test. To test the program, he gave the computer 15 different forms of a popular IQ test and computed its IQ from each form

IQ Values: 134 136 137 138 138 143 144 144 145 146

146 146 147 148 153

Find the following Statistical measures:

i. Mean, median, and mode
ii. Rang, Variance and standard deviation
iii. Obtain the quartiles
iv. Determine the interquartile range.
v. Identify potential outliers, if any.

innovate achieve lead

Measures of Central Tendency

Arranging the data in ascending order:

134, 136, 137, 138, 138, 143, 144, 144, 145, 146, 146, 146, 147, 148, 153

N=15

134 1
136 1
137 1
138 2

S.No	Formula	Solution
Mean	$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$	μ = 143
Median	$p = \frac{n+1}{2}$	Median = 144
Mode	The mode is the value or values that occur most frequently in the data set. A data set can have more than one mode, and it can also have no mode	Mode = 146

innovate achieve lead

Measures of Central Tendency

$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{n}$

Range	$\text{Range} = x_n - x_1$	Minimum = 134 Maximum =153 Range R = 19
Variance	For a Population whole $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} = \frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2$ For a Sample part $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	Variance = 26 $x \quad x^2$ $\sum x \quad \sum x^2$
Standard deviation $= \sqrt{\text{Variance}}$	For a Population $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$ For a Sample $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$	Standard Deviation = 5.09901951

$$
\begin{aligned}
\sigma^2 &= \frac{\sum (X-\mu)^2}{N} \\
&= \frac{\sum (X^2 - 2\mu X + \mu^2)}{N} \\
&= \frac{\sum X^2}{N} - \frac{2\mu \sum X}{N} + \frac{N\mu^2}{N} \\
&= \frac{\sum X^2}{N} - 2\mu^2 + \mu^2
\end{aligned}
$$

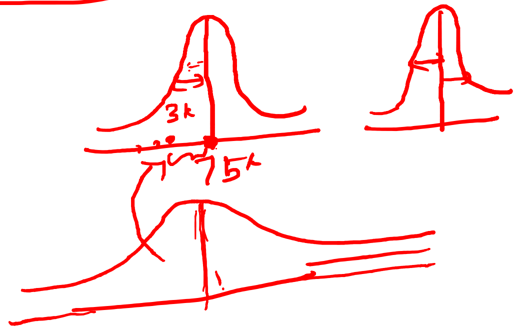
$$\boxed{\sigma^2 = \frac{\sum X^2}{N} - \mu^2}$$

$\sum_{i=1}^{N} \mu^2 = N\mu^2$

$\bar{x} = 75k$

$\sigma = 3k$

$\sigma = 25k$



Measures of Central Tendency

50 75

Quartiles	Quartiles separate a data set into four sections. The median is the second quartile Q_2. It divides the ordered data set into higher and lower halves. The first quartile, Q_1, is the median of the lower half not including Q_2. The third quartile, Q_3, is the median of the higher half not including Q_2.	Quartiles: Q_1 --> 138 Q_2 Median. --> 144 Q_3 --> 146
Interquartile range	IQR=Q3−Q1	Interquartile Range IQR =8
Potential outliers, if any.	$$\text{Upper Fence} = Q_3 + 1.5 \times IQR$$ $$\text{Lower Fence} = Q_1 - 1.5 \times IQR$$	none

Measures of Variability

For two observations 'a' and 'b', show that standard deviation is half of the distance between them.

X : a b

$$\boxed{\sigma^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$\sigma^2 \geq 0$

$$= \frac{(a^2+b^2)}{2} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{a^2+b^2}{2} - \left(\frac{a^2+b^2+2ab}{4}\right)$$

$$= \frac{1}{4}\left[2a^2+2b^2-a^2-b^2-2ab\right]$$

$$\sigma^2 = \frac{1}{4}\left[a^2+b^2-2ab\right] = \frac{(a-b)^2}{4} \Rightarrow \sigma = \frac{|a-b|}{2}$$

Measures of Variability

For two observations ‘a’ and ‘b’, show that standard deviation is half of the distance between them.

Solution:

$$\sigma^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 = \frac{a^2+b^2}{2} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{2}(a^2+b^2) - \frac{1}{4}(a^2+b^2+2ab) = \frac{1}{4}[2(a^2+b^2) - a^2 - b^2 - 2ab]$$

$$= \frac{1}{4}(a^2+b^2-2ab) = \frac{1}{4}(a-b)^2 = [\frac{1}{2}(a-b)]^2$$

$$\Rightarrow s.d.(\sigma) = \sqrt{\left(\frac{a-b}{2}\right)^2} = \frac{1}{2}|(a-b)|; (\because \text{ s.d. is always } \geq 0)$$

$$\therefore s.d.(\sigma) = \frac{1}{2}[\text{ Distance between } a \text{ and } b]$$

Basic Probability

If $P(A) = 1/2$, $P(B) = 1/3$ and $P(A \cap B) = 1/5$ then find

a). $P(A \cup B)$ b). $P(A^c \cap B)$ c). $P(A \cap B^c)$

d). $P(A^c \cap B^c)$ e). $P(A^c \cup B^c)$ f). $P((A \cup B)^c)$

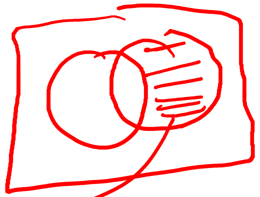
$$P(A \cap B) = \underbrace{P(A) + P(B) - P(A \cup B)}$$

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

b) $P(A^c \cap B) = P(B - (A \cap B))$

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

$=$



Basic Probability

If $P(A) = 1/2$, $P(B) = 1/3$ and $P(A \cap B) = 1/5$ then find

a). $P(A \cup B)$ b). $P(A^c \cap B)$ c). $P(A \cap B^c)$

d). $P(A^c \cap B^c)$ e). $P(A^c \cup B^c)$ f). $P((A \cup B)^c)$

d) $P(A^c \cap B^c) = P((A \cup B)^c)$

$= 1 - P(A \cup B)$

e) $P(A^c \cup B^c) = P((A \cap B)^c)$

$= 1 - P(A \cap B)$

innovate achieve lead

If $P(A) = 1/2$, $P(B) = 1/3$ and $P(A \cap B) = 1/5$ then find

a). $P(A \cup B)$ b). $P(A^c \cap B)$ c). $P(A \cap B^c)$

d). $P(A^c \cap B^c)$ e). $P(A^c \cup B^c)$ f). $P((A \cup B)^c)$

$P(A^c) = 1 - P(A)$

Solution

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/3 - 1/5 = 19/30 = 0.6333$

b) $P(A^c \cap B) = P(B - A) = P(B - (A \cap B)) = P(B) - P(A \cap B) = 1/3 - 1/5 = 2/15 = 0.1333$

c) $P(A \cap B^c) = P(A - B) = P(A - (A \cap B)) = P(A) - P(A \cap B) = 1/2 - 1/5 = 3/10 = 0.3$

d) $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 19/30 = 11/30 = 0.3667$

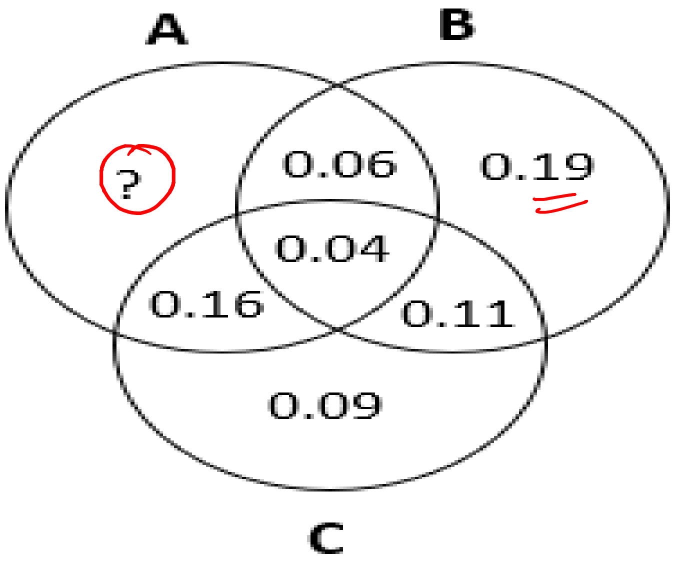
e) $P(A^c \cup B^c) = P((A \cap B)^c) = 1 - P(A \cap B) = 1 - 1/5 = 4/5 = 0.8$

f) $P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 19/30 = 11/30 = 0.3667$

Basic Probability

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

There are three events A, B and C. The probability of occurrence of at least one of them is 0.23. Using the probabilities given in the following Venn diagram, find the probability of the event A.



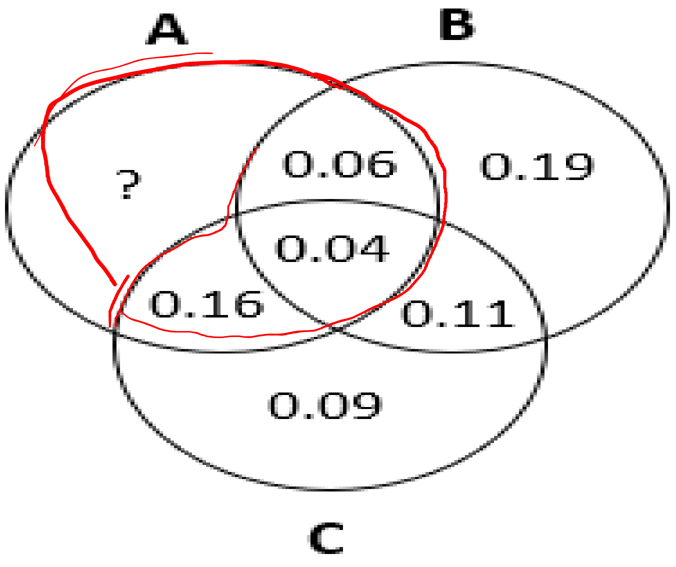
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC)$

$+ P(ABC)$

$P(A) =$

Basic Probability

There are three events A, B and C. The probability of occurrence of at least one of them is 0.23. Using the probabilities given in the following Venn diagram, find the probability of the event A.



Solution:

$P(A) = ?$

$P(B) = 0.19$

$P(C) = 0.09$

$P(A \cap B) = 0.06$

$P(A \cap C) = 0.16$

$P(B \cap C) = 0.11$

$P(A \cap B \cap C) = 0.04$

$P(A \cup B \cup C) = 0.23$

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

$$= 0.23 - 0.19 - 0.09 + 0.06 + 0.16 + 0.11 - 0.04 = \mathbf{0.24}$$

Basic Probability

A political leader has submitted his nomination to compete in two different electoral constituencies namely A1 and A2. The probability of wining in constituency A1 and A2 is 0.80 and 0.65 respectively. The probability of losing at least one of the constituencies is 0.35. What will be the probability that he will win in one of the constituencies?

$P(A) = 0.8$

$P(B) = 0.65$

$P(\bar{A} \cup \bar{B}) = 0.35$

$1 - P(A \cap B) = 0.35$

$P(A \cap B) = 0.65$

Required ~~Result~~ $= P(A) - P(A \cap B) = 0.8 - 0.65 = 0.15$

$= P(B) - P(A \cap B) = 0.65 - 0.65 = 0$

Basic Probability

A political leader has submitted his nomination to compete in two different electoral constituencies namely A1 and A2. The probability of wining in constituency A1 and A2 is 0.80 and 0.65 respectively. The probability of losing at least one of the constituencies is 0.35. What will be the probability that he will win in one of the constituencies?

Assume that A, B be the events defined as follows:

A : "Winning in constituency A1"

B : "Winning in constituency A2"

Given:

P(A) = 0.80, P(B) = 0.65

and $P(\bar{A} \cup \bar{B}) = 0.35$

Now, $\because P(\bar{A} \cup \bar{B}) = 0.35$

$\therefore P(\overline{A \cap B}) = 0.35$

$\Rightarrow 1 - P(A \cap B) = 0.35$

$\Rightarrow P(A \cap B) = 0.65$

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

= 0.80 + 0.65 – 0.65

$\therefore P(A \cup B) = 0.80$

Then, $P(\textit{He will win in one of the constituencies}) = P(A \cup B) - P(A \cap B)$

$= 0.80 - 0.65$

$\therefore P(\textit{He will win in one of the constituencies}) = 0.15$

$P(\textit{He will win in constituency A1 ONLY}) = P(A) - P(A \cap B)$

= 0.80 – 0.65

= 0.15

$P(\textit{He will win in constituency A2 ONLY}) = P(B) - P(A \cap B)$

= 0.65 – 0.65

= 0

Independent events

Are every mutually exclusive independent events?

A & B

Mut exlu : $A \cap B = \phi$, $\boxed{P(A \cap B) = 0}$

Independ. : $P(A/B) = P(A)$

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

Test independence.

Let 2 events are mutually exclusive

$P(A \cap B) = 0$

$P(A)P(B) = 0$

$\boxed{P(A) = 0 \quad \text{or} \quad P(B) = 0}$

$P(A \cap B) = 0 \not\Rightarrow$ Mutually e.

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0$

$P(\text{pass}) = \%$

$P(\text{pass} / \ldots\ldots) = \%$ } events dependent.

Independent events

Let A and B be the two possible outcomes of an experiment and suppose P(A) = 0.4, P(B) = p and P(AUB) = 0.7

(i) For what choice of 'p' are A and B mutually exclusive?

(ii) For what choice of 'p' are A and B independent?

i) $P(A \cup B) = P(A) + P(B) - \cancel{P(A \cap B)}^{0}$

$P(B) = p$

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

$P(A)P(B)$

'p'

Independent events

Let A and B be the two possible outcomes of an experiment and suppose P(A) = 0.4, P(B) = p and P(AUB) = 0.7

(i) For what choice of 'p' are A and B mutually exclusive?

(ii) For what choice of 'p' are A and B independent?

(i) If A and B are mutually exclusive then $P(A \cap B) = 0$

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

$0.7 = 0.4 + P(B) - 0$

$\therefore P(B) = 0.7 - 0.4 = 0.3$

(ii) If A and B are independent then $P(A \cap B) = P(A).P(B)$

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

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

$$\begin{aligned} 0.7 &= 0.4 + P - 0.4 \cdot P \\ &= 0.4 + P(1 - 0.4) \\ &= 0.4 + 0.6P \\ 0.6P &= 0.7 - 0.4 = 0.3 \\ P &= \frac{0.3}{0.6} = \frac{1}{2} = 0.5 \end{aligned}$$

$\therefore P(B) = 0.5$

Answer. (i) p=0.3, (ii) p=0.5

Conditional Probability

F

A

In an online shopping survey, 30% of persons made shopping in Flipkart, 45% of persons made shopping in Amazon and 5% made purchases in both. If a person is selected at random, find

i) the probability that he makes shopping in at least one of two companies $P(F \cup A)$

ii).the probability that he makes shopping in Amazon given that he already made shopping in Flipkart. $P(A/F) = \frac{P(A F)}{P(F)}$

iii).the probability that the person will not make shopping in Flipkart given that he already made purchase in Amazon.

$P[F^c/A] =$

Conditional Probability

In an online shopping survey, 30% of persons made shopping in Flipkart, 45% of persons made shopping in Amazon and 5% made purchases in both. If a person is selected at random, find

i) the probability that he makes shopping in at least one of two companies

ii).the probability that he makes shopping in Amazon given that he already made shopping in Flipkart.

iii).the probability that the person will not make shopping in Flipkart given that he already made purchase in Amazon.

Solution: Given $P(F)$ = 30% = 0.30 $P(A$

$)$ = 45% = 0.45

$P(F \cap A)$ = 5% = 0.05

i) $P(F \cup A) = P(F) + P(A) - P(F \cap A)$

= 0.30 + 0.45 − 0.05 = 0.7

ii) $P(A \mid F) = \frac{P(A \cap F)}{P(F)}$

$= \frac{0.05}{0.30} = 0.167$

iii) $P(F' \mid A) = \frac{P(F' \cap A)}{P(A)}$

$P(F' \cap A) = P(A) - P(A \cap F)$

$= 0.45 - 0.05$

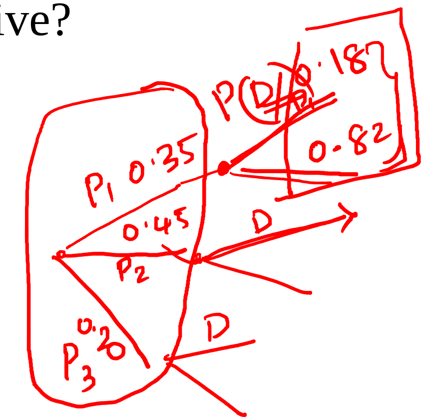
$= 0.40$

$P(F' \mid A) = \frac{0.40}{0.45} = 0.88$

Total Probability

$P(X \cup Y \cup Z) = \cdots P(X \cap Y)$ joint

Of 1000 car parts produced, it is known that 350 are produced in one plant, 450 parts in a second plant, and 200 parts in a third plant. Also, it is known that the probabilities are 0.18, 0.21, and 0.11 that the parts will be defective if they are produced in the first, second and third plants respectively. What is the probability that a randomly picked part from this batch is defective?



$P_1 D$

$P_1 D'$

$\vdots$

1

$P(P_1) = \frac{350}{1000} = 0.35$

$P(D/P_1) = 0.18$

$P(D/P_2) = 0.21$

$P(D/P_3) = 0.11$

$$P(D) = P(DP_1 \cup DP_2 \cup DP_3)$$

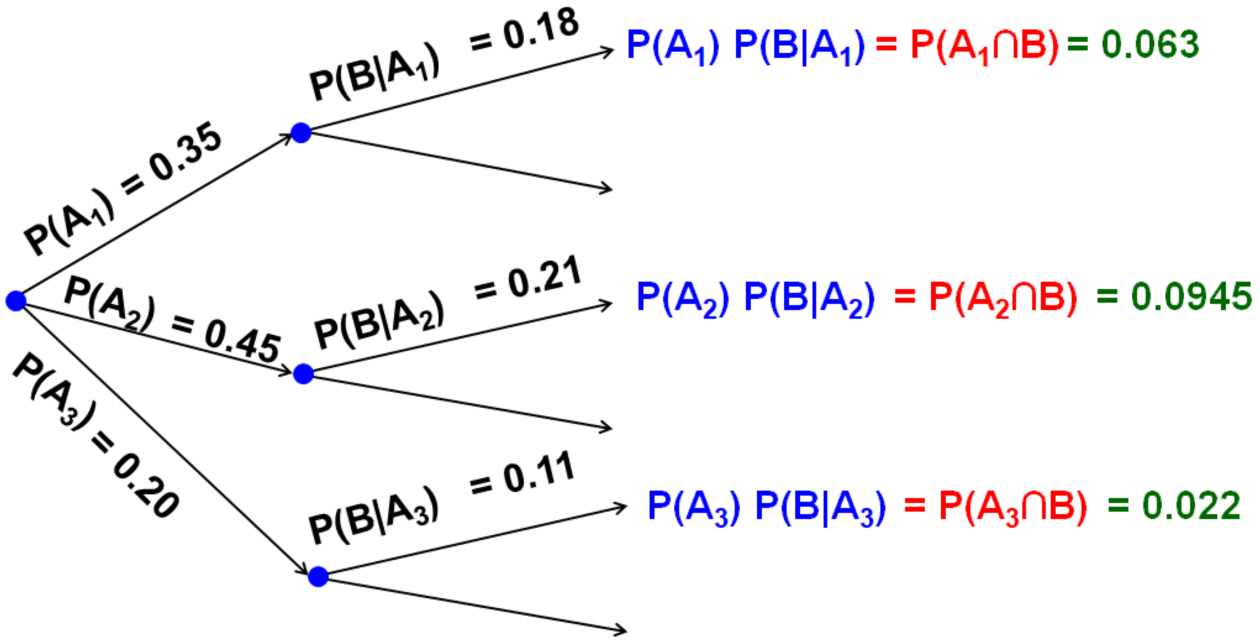
$$= P(DP_1) + P(DP_2) + P(DP_3)$$

$$= 0.35 \times 0.18 + 0.45 \times 0.21 + 0.2 \times 0.11$$

$$=$$

Total Probability

Of 1000 car parts produced, it is known that 350 are produced in one plant, 450 parts in a second plant, and 200 parts in a third plant. Also, it is known that the probabilities are 0.18, 0.21, and 0.11 that the parts will be defective if they are produced in the first, second and third plants respectively. What is the probability that a randomly picked part from this batch is defective?



$$P(B) = P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + P(A_3)\,P(B|A_3) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_2 \cap B) = 0.1795$$

Bayes Theorem

A manufacturing company produces certain types of output by 4 machines i.e A, B, C and D. Machine A produces 15%, Machine B produces 30 % and Machine C produces 30% of daily production. Based on experience it is observed that 1% of the output by Machine A is defective. Similarly, the defectives by other machines are 2% ,3% and 4% respectively. An item is drawn at random and found to be defective. Is it possible to find the defective item is produced by which Machine? If so, find it

d

Let A be the event that item produced by Machine A

Let B be the event that item produced by Machine B

Let C be the event that item produced by Machine C

Let D be the event that item produced by Machine D

Let S be the event that item is defective

Given P(A)=0.15, P(B) =0.30, P(C)=0.3, P(D)=0.25

$\rightarrow$ P(S/A)= 0.01; P(S/B)= 0.02; P(S/C)= 0.03; P(S/D)= 0.04

A B C D $\rightarrow 1-(\ )$

Total prob

$$\rightarrow P(S) = P(S/A)P(A) + P(S/B)P(B) + P(S/C)P(C) + P(S/D)P(D)$$

$$= 0.15x0.01 + 0.3x0.02 + 0.3x0.03 + 0.25x0.04$$

$$= 0.0265$$

$$P(A/S) = \frac{P(A\cap S)}{P(S)}$$

$$P(B/S)$$

Bayes Theorem

A manufacturing company produces certain types of output by 4 machines i.e A, B, C and D. Machine A produces 15%, Machine B produces 30 % and Machine C produces 30% of daily production. Based on experience it is observed that 1% of the output by Machine A is defective. Similarly, the defectives by other machines are 2% ,3% and 4% respectively. An item is drawn at random and found to be defective. Is it possible to find the defective item is produced by which Machine? If so, find it

The defective item is produced bv Machine A : P(A/S) = (0.15x0.01)/0.0265 = 0.0566

The defective item is produced bv Machine B : P(B/S) = (0.3x0.02)/0.0265= 0.2264

The defective item is produced bv Machine C : P(C/S) = (0.3x0.03)/0.0265= 0.3396

The defective item is produced bv Machine D : P(D/S) = (0.25x0.04)/0.0265= 0.3774

Mach "D"

Bayes Theorem

An insurance company insured 1,000 taxi drivers, 2,000 car drivers and 3,000 truck drivers. The probability of their accident is 0.05, 0.15 and 0.1 respectively. One of the insured persons meets with an accident. What is the probability that he is a truck driver?

Solution: Given

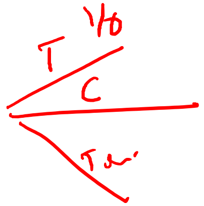
	Insured	P(accident)
Taxi drivers	1000	0.05
Car drivers	2000	0.15
Truck drivers	3000	0.1

Let

$$P(A) = P(\text{ taxi driven }) = \frac{1000}{6000} = \frac{1}{6}$$

$$P(B) = P(\text{ car driven }) = \frac{2000}{6000} = \frac{2}{6} = \frac{1}{3}$$

$$P(c) = P(\text{ Guck driven }) = \frac{3000}{6000} = \frac{1}{2}$$



$$P\left(\frac{TD}{A}\right) = \frac{P(\ldots)}{P(A)}$$

$\downarrow$ Tot.

Bayes Theorem

An insurance company insured 1,000 taxi drivers, 2,000 car drivers and 3,000 truck drivers. The probability of their accident is 0.05, 0.15 and 0.1 respectively. One of the insured persons meets with an accident. What is the probability that he is a truck driver?

Let E be the event that person meets the accident

$$P\left(\frac{E}{A}\right) = \frac{5}{100} \quad , P\left(\frac{E}{B}\right) = \frac{15}{100} \quad , P\left(\frac{E}{C}\right) = \frac{1}{10}$$

Thus the required probability is

$$P(C/E) = \frac{P(C)\cdot P(E/C)}{P(A)\cdot P(E/A)+P(B)\cdot P(E/B)+P(C)\cdot P(E/C)}$$

$$= \frac{1/20}{1/120+1/20+1/20}$$

$$P\left(\frac{C}{E}\right) = \frac{6}{13} = 0.4615$$

Bayes Theorem

A manufacturer has three machine operators A, B and C. The first operator A produce 1% defective items, whereas the other two operators B and C produce 5% and 7% defective items respectively. A is on the job for 50% of the time, B is on the job for 30% of the time. A defective item is produced, what is the probability that it was produced by A, B, C? Based on this write your observations.

Let the event that items were produced by operator A be A, by operator B be B, by operator C be C, and item produced defective be D.

P(A) = Probability of item is produced by operator A. $=50\%=\frac{50}{100}=0.5$

P(B) = Probability of item is produced by operator B. $=30\%=\frac{30}{100}=0.3$

P(C) = Probability of item is produced by operator C. $=20\%=\frac{20}{100}=0.2$

P(D|A) = Probability of a defective item produced by operator A. $=1\%=\frac{1}{100}=0.01$

P(D|B) = Probability of a defective item produced by operator B. $=5\%=\frac{5}{100}=0.05$

P(D|C) = Probability of a defective item produced by operator C. $=7\%=\frac{7}{100}=0.07$

$$P(A|D) = \frac{P(A).P(D|A)}{P(A).P(D|A)+P(B).P(D|B)+P(C).P(D|C)}$$

On putting the values in the above formula, we get

$$P(A|D) = \frac{0.5\times0.01}{(0.5)\times(0.01)+(0.3)\times(0.05)+(0.2)\times(0.07)}$$
$$= \frac{0.005}{0.005+0.015+0.014}$$
$$= \frac{0.005}{0.034} = \frac{5}{34} = 0.14705$$

Thus, the probability that it was produced by A = $\frac{5}{34}$ = 0.14705

Probability that the defective item is produced by B is

$$P(B|D) = \frac{P(B).P(D|B)}{P(A).P(D|A)+P(B).P(D|B)+P(C).P(D|C)}$$

On putting the values in the above formula, we get

$$P(B|D) = \frac{0.3\times0.05}{(0.5)\times(0.01)+(0.3)\times(0.05)+(0.2)\times(0.07)}$$
$$= \frac{0.015}{0.005+0.015+0.014}$$
$$= \frac{0.015}{0.034} = \frac{15}{34} = 0.44117$$

Thus, the probability that it was produced by B= $\frac{15}{34}$ = 0.44117

Probability that the defective item is produced by C is

$$P(C|D) = \frac{P(C).P(D|C)}{P(A).P(D|A)+P(B).P(D|B)+P(C).P(D|C)}$$

On putting the values in the above formula, we get

$$P(C|D) = \frac{0.2\times0.07}{(0.5)\times(0.01)+(0.3)\times(0.05)+(0.2)\times(0.07)}$$
$$= \frac{0.014}{0.005+0.015+0.014}$$
$$= \frac{0.014}{0.034} = \frac{14}{34} = 0.41176$$

Thus, the probability that it was produced by C= $\frac{14}{34}$ = 0.41176

THANK YOU!