ISM Assignment - 1

AFML Section - 6. Name - Sagar Powar

Bits ID - 2025AA05421

Q. 1

Sort the Data

26, 27, 28, 29, 29, (30), 31, 32, 33, 34, 35

1) Mean $= \frac{\text{Sum of all values}}{\text{Number of values}} = \frac{334}{11} = 30.36^{\circ}C$ (approx.)

2) Median - with 11 values, median will be 6th value.

i.e. **median = 30°C**

3) Range = Max value - min. value

$= 35 - 26$

Range = 9

4) Variance:

using population variance $= \sigma^2 = \frac{\sum (x-\mu)^2}{n}$

Total squared deviations = 84.55

Sr. No.	values	distance	squared distance
1	26	-4.36	19.04
2	27	-3.36	11.31
3	28	-2.36	5.59
4	29	-1.36	1.86
5	29	-1.36	1.86
6	30	-0.36	0.13
7	31	0.64	0.40
8	32	1.64	2.68
9	33	2.64	6.95
10	34	3.64	13.22
11	35	4.64	21.50

Population Variance:

$$\sigma^2 = \frac{\text{sum. of squared distance/deviations}}{n}$$

$$\sigma^2 = \frac{84.55}{11} = 7.69$$

(5) Q1

lower half : 26, 27, (28) 29, 29

$Q_1 = 28$

upper half : 31, 32, (33) 34, 35

$Q_3 = 33$

(II) Skewness

As mean (30.36) > median (30)

data has positive skew (right skew slightly)

(III) IQR outlier Detection

$IQR = Q_3 - Q_1 = 33 - 28 = 5$

1) lower fence $= Q_1 - 1.5(IQR) = 28 - 1.5(5) = 28 - 7.5$

$= 20.5$

2) upper fence $= Q_3 + 1.5(IQR) = 33 + 1.5(5) = 33 + 7.5$

$= 40.5$

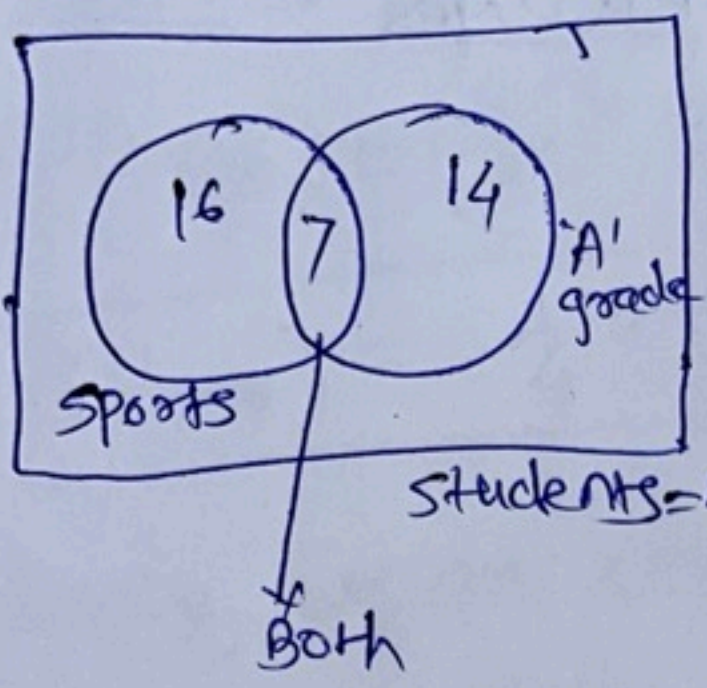
As there is no data outside 20.5 - 40.5,

so NO OUTLIERS by IQR method

Name - Sagar Powar
Bits ID - 2025AA05421

Section - 6

Q.2. Using ven diagram



given ↓

Sports = 7 (boys) + 9 (girls) = 16 (S)

'A' grade = 6 (boys) + 8 (girls) = 14 (A)

Both (S and A) = 3 (boys) + 4 (girls) = 7

Findout = P(S ∪ A)

$$P(S \cup A) = P(S) + P(A) - P(S \cup A)$$

$$= \frac{16}{40} + \frac{14}{40} - \frac{7}{40}$$

$$\boxed{P(S \cup A) = \frac{23}{40} = 0.575}$$

So poobability of a student is either involved in sports or scored an 'A' grade = 0.575 (57.5%)

Name - Sagar powar
BITS ID - 2025AA05424

Section - 6

Q3. Using given dataset, create below tables for each feature. with Depression (D) yes & D NO.

Trouble sleeping	Dyes	DNO
Yes	2/2	0/2
NO	0/2	2/2

low energy	Dyes	DNO
yes	1/2	1/2
NO	1/2	1/2

Anxiety	Dyes	DNO
yes	2/2	0/2
NO	0/2	2/2

$P(D_{yes}) = \frac{2}{4} = 0.5$ & $P(D_{NO}) = \frac{2}{4} = 0.5$

A New person has,

$X_1 \Rightarrow$ Trouble Sleeping = yes

$X_2 \Rightarrow$ Lower energy = NO

$X_3 \Rightarrow$ Anxiety = yes

As per Naive bayes classifier, compute below scores,

$$P(D_{yes}/X_1X_2X_3) = \frac{P(X_1/D_{yes}) \cdot P(X_2/D_{yes}) \cdot P(X_3/D_{yes}) \cdot P(D_{yes})}{P(X_1X_2X_3)}$$

Name - Sagar Powar Section - 6

BITS ID - 2025AA05421

Now, $P(X_1/Dyes) = \frac{2}{2}$, $P(X_2/Dyes) = 1/2$ & $P(X_3/Dyes) = \frac{2}{2}$

$$P(Dyes) = \left(\frac{2}{2} \times \frac{1}{2} \times \frac{2}{2}\right) \times \frac{2}{4}$$

$$\boxed{P(Dyes) = \frac{1}{4} = 0.25}$$

similarly, we can compute $P(D_{NO})$ scores as below.

$$P(D_{NO}/X_1X_2X_3) = \frac{P(X_1/D_{NO}) \cdot P(X_2/D_{NO}) \cdot P(X_3/D_{NO}) \cdot P(D_{NO})}{P(X_1, X_2X_3)}$$

here, $P(X_1/D_{NO}) = \frac{0}{2}$, $P(X_2/D_{NO}) = \frac{1}{2}$ & $P(X_3/D_{NO}) = \frac{0}{2}$

so, $\boxed{P(D_{NO}) = \left(\frac{0}{2}\right) \times \frac{1}{2} \times \frac{0}{2}) \times \frac{2}{4} = 0}$

As, $P(Dyes) > P(D_{NO})$, we can say that a new person Has Depression

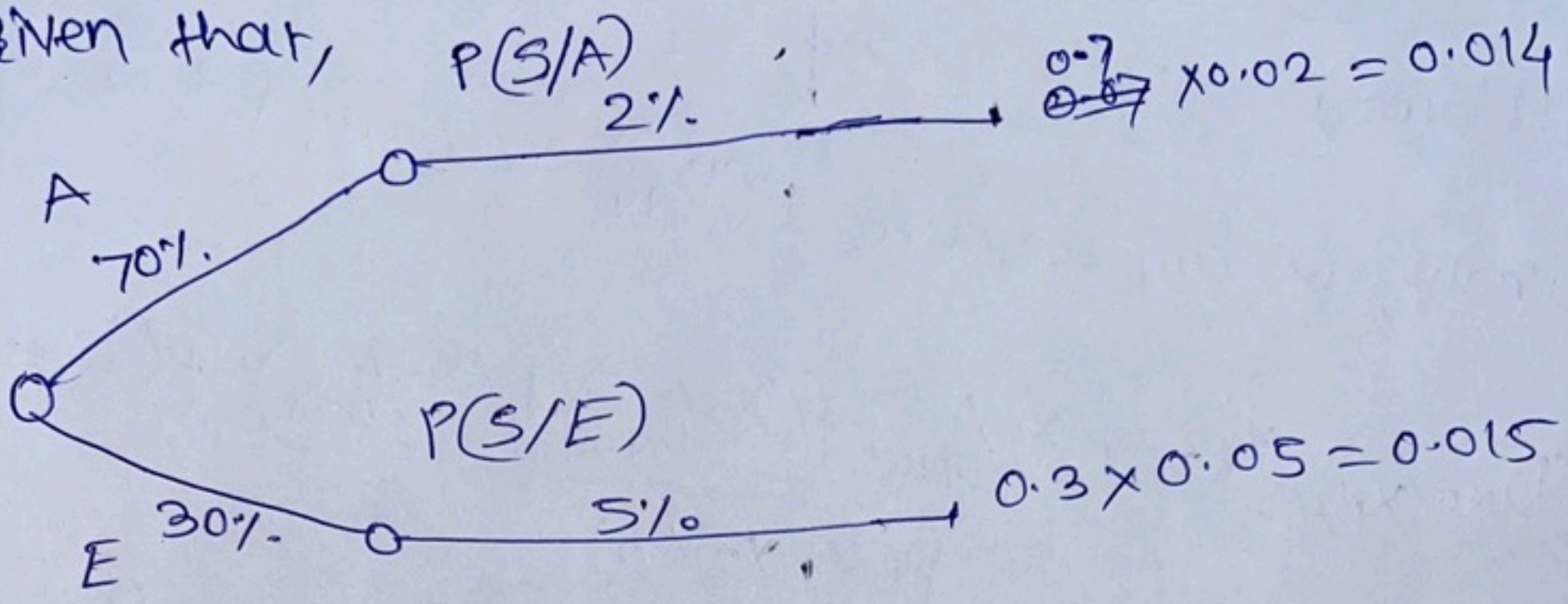
Prediction ⇒ Person has Depression

Name - Sagar Powar
Bits ID - 2025AA05421

Section - 6

Q.4. Let,
A = students in Arts
E = students in Engineering
S = students scored above 95%.

Given that,



we have to find, $P(A/S)$

Now, $P(S) = P(S/A) \cdot P(A) + P(S/E) + P(E)$ → Total probability

$= 0.014 + 0.015$

$P(S) = 0.029$

As per Baye's theorem,

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

$$\boxed{P(A/S) = \frac{0.014}{0.029} = 0.4828}$$

So, there is ~~48%~~ 48.3% chance that student is from Arts.