Anmol Singh
M.Sc. 12
Roll no. 04

Biostatistics
Assignment - 2

3. P

1. $A = \{8.1, 5.9, 6.4, X\}$; $\bar{x} = 7.0$, $X = ?$

Arithmetic mean $= \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\bar{x} = \frac{8.1 + 5.9 + 6.4 + X}{4}$$

$$7 \times 4 = 20.4 + X$$

$$X = 28 - 20.4 = 7.6$$ Ans

2. $A = \{12, 14, 13, 15, 16, 24, 17, 11\}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{12 + 14 + 13 + 15 + 16 + 24 + 17 + 11}{8}$$

$$\bar{x} = 15.25$$

$$\log(GM) = \frac{1}{n} \sum_{i=1}^{n} \log(x_i) = \frac{9.371516}{8} = 1.1714$$

4.

taking anti log:

$$GM = 10^{\log(GM)} = 14.8401$$

Ratio of $\bar{x} : GM$

$$\frac{\bar{x}}{GM} = \frac{15.25}{14.8401} = 1.0276$$ Ans

3. $P_n = f(x) = 3x^3 + 7x$, interval $= [0,2]$, $b = 2$, $a = 0$

$$\mu = \frac{1}{b-a}\int_a^b f(x)\,dx$$

$$= \frac{1}{2-0}\int_0^2 3x^3 + 7x$$

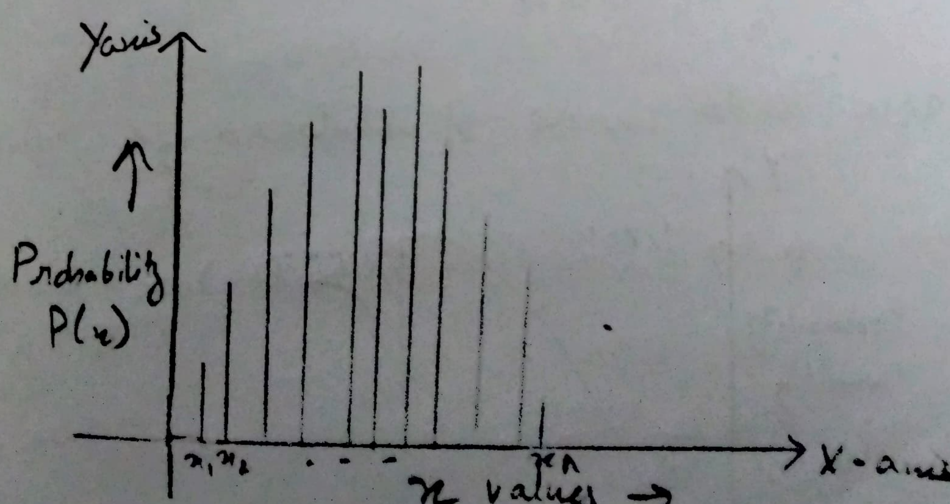
$$= \frac{1}{2}\int_0^2 \frac{3x^4}{4} + \frac{7x^2}{2}$$

$$= \frac{1}{2}\left[\frac{3x^4}{4} + \frac{7x^2}{2}\right]_0^2$$

$$= \frac{1}{2}\left[\frac{3\times 2^4}{4} + \frac{7\times 2^2}{2}\right]$$

$$\bar{x} = \frac{1}{2}\left[12 + 14\right] = \frac{26}{2} = \underline{\underline{13}} \text{ Ans}$$

4. In a probability distribution of discrete data, we represent X-axis as discrete values and the Y-axis as the probability of getting value x.



Also sum of these probabilities is 1

Comparing this to probability density in a continous probability distribution.

In a continous probability distribution, there can be infinite numbers in a given range, just like the concept of infinite nos. between 0 & 1 in number line.

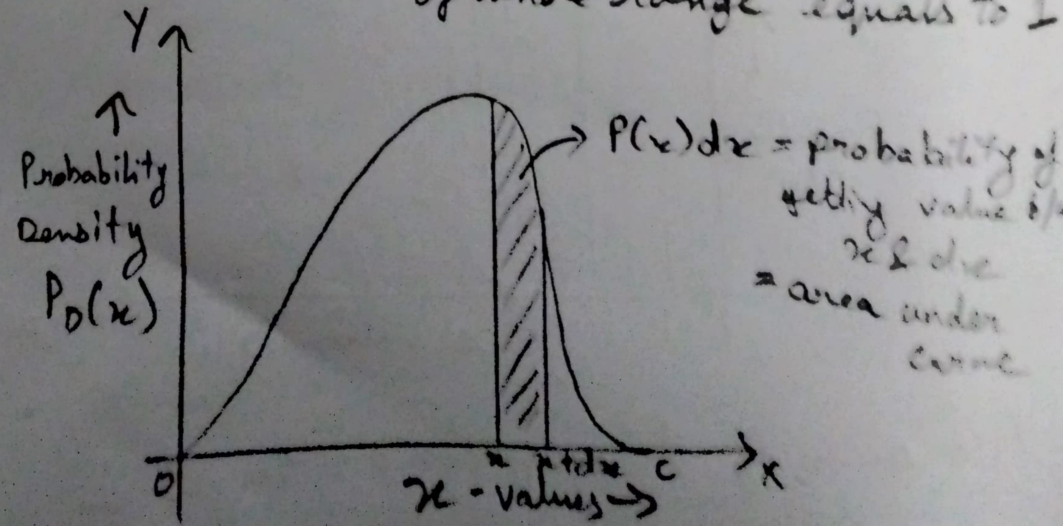
As such probability of each value is infinitesmally small, and should sum up to 1, we cannot represent such distribution same as discrete distribution.

Thus, insted of probability of observing a value x, we take probability of observing a value within a range / interval around x, this is called probability density $P_D(x)$.

$P_D(x)$ is probability per unit interval around x,
$P_D(x)\,dx$ gives probability of observing a value within dx around x.

Here probability of observing a value in an interval $[a, b]$ is $\int_a^b P_D(x)\,dx$.

Total area under curve of whole range equals to 1



	Discrete Distribution	Continous distribution
μ	$\sum_{x} x P(x)$	$\int_{x_{min}}^{x_{max}} x P_D(x) dx$
σ	$\sum_{x} (x-\mu)^2 P(x)$	$\int_{x_{min}}^{x_{max}} (x-\mu)^2 P_D(x) dx$

5: a) $T_{Fs} = \{44, 46, 49, 52, 55, 62, 67, 72, 77, 80, 83, 86, 88, 90, 92, 94, 99, 100, 101, 106\}$

$n=20$

$$\bar{x} = \sum_{i=1}^{n} \frac{(x_i)}{n} = \frac{1543}{20} = \underline{\underline{77.15}} \text{ mg/dL}$$

Ans

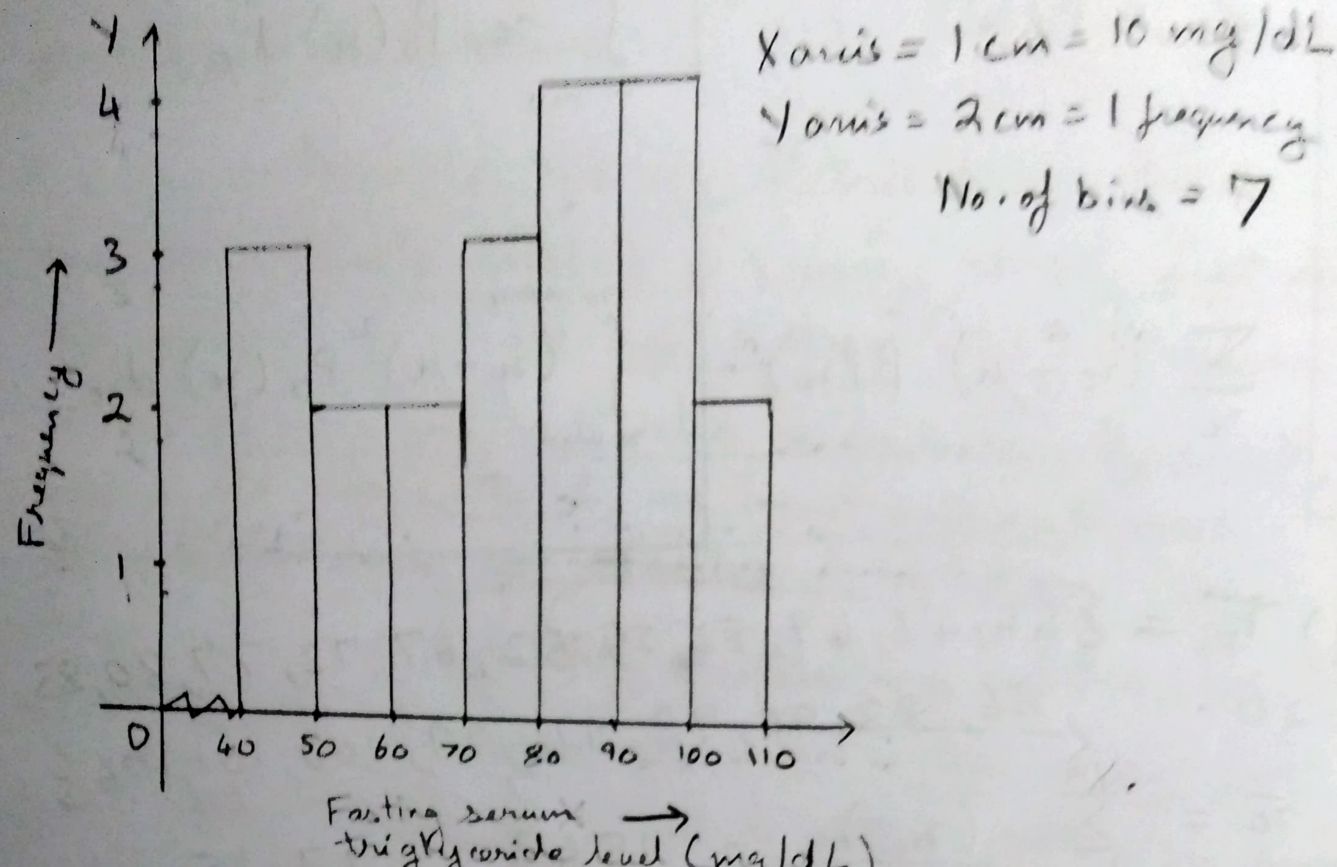
$$S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

$$= \frac{1}{19}\left((44-77.15)^2 + (46-77.15)^2 + \cdots + (106-77.15)^2 \right)$$

$$S = \sqrt{402.76579} = \underline{\underline{20.069}} \text{ mg/dL}$$

Ans

5. Frequency Histogram, bins width = 10



6. As this case has 2 mutually exclusive outcome.
It is Bernoulli Trials with 5 events.
Prob. of success = 0.9 (∴ 90% given)
no. of failure to observe = 1, no. of success = 4

$P_B(x, n, p) = p^x (1-p)^{n-x}$

$P_B(4, 5, 0.9) = 0.9^4 (1-0.9)^{5-4}$

$\underline{P_B(4, 5, 0.9) = 0.06561}$

7. Given $\mu = 2$ SNPs / 1000 kb
For 3000 kb

$\mu = \frac{2}{1000} \times 3000 = 6$ SNPs / 3000 kb

Random event of observing 12 SNPs in 3000kb has probability:

$$P_p(x, \mu) = \frac{\mu^x e^{-\mu}}{x!}$$

$$P_p(12, 6) = \frac{6^{12} \times e^{-6}}{12!}$$

$$= \frac{2.17678 \times 10^9 \times 2.478752 \times 10^{-3}}{12!}$$

$$P_p(\overset{x}{12}, \overset{\mu}{6}) = \underline{\underline{0.0112644}} \text{ Ans}$$

8. We know: S → Studied P → Passed

$P(P|S) = 0.8$, $P(P|NS) = 0.4$

$P(S) = 0.6$, $P(P) = ?$, $P(NS) = 1 - P(S) = 0.4$

$$P(\text{Pass}) = P(S) \times P(P|S) + P(NS) * P(P|NS)$$

$$= 0.6 \times 0.8 + 0.4 \times 0.4$$

$$P(\text{Pass}) = 0.48 + 0.16 = 0.64$$

By Bayes' Theorem ⇒ Prob. that she studied given she passed:

$$P(S|P) = \frac{P(P|S) \cdot P(S)}{P(P)} = \frac{0.48}{0.64} = 0.75$$

Ans $\underline{\underline{75\%}}$

OKb

9. Given $P(\text{Glaucoma}) = 0.007$

$P(\text{Diabetes}) = 0.02$

$P(G \cap D) = 0.0008$

a) $P(G \cup D) = P(G) + P(D) - P(G \cap D)$

$= 0.007 + 0.02 - 0.0008$

$P(G \cup D) = \underline{\underline{0.0262}}$ Ans

b) $P(G|D) = \dfrac{P(G \cap D)}{P(D)}$

$P(G|D) = \dfrac{0.0008}{0.02} = \underline{\underline{0.04}}$ Ans

c) $P(D|G) = \dfrac{P(G \cap D)}{P(G)} = \dfrac{0.0008}{0.007}$

$P(D|G) = \underline{\underline{0.11428}}$ Ans

10. $\mu = 3.2\,kg$, $\sigma^2 = 1.21\,kg$ $\sigma = 1.1\,kg$

We can use the Z transform of Gaussian Distribu

(1) $P(x > 6.5\,kg)$

$Z = \dfrac{x - \mu}{\sigma} = \dfrac{6.5 - 3.2}{1.1} = \dfrac{3.3}{1.1} = 3$

$$P_n(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$z = \frac{x-\mu}{\sigma} = 3$$

$$= \int_{6.5}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$p(>z) = \int_{3}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$p(z>3) = 0.00135$$

(2) $P(x < 2.2kg)$

$$z = \frac{2.2 - 3.2}{1.1} = -0.90\overline{90} \approx -0.91$$

Similar to above :

$$P(<z) = \int_{-\infty}^{-0.91} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$P(<z) = 0.18141$$ Ans

tion

(3) $P(2.8 < x < 3.9)$

$$z_{2.8} = \frac{2.8 - 3.2}{1.1} = -0.3\overline{6}$$

$$z_{3.9} = \frac{3.9 - 3.2}{1.1} = 0.63\overline{63} \approx \underline{\underline{0.64}}$$

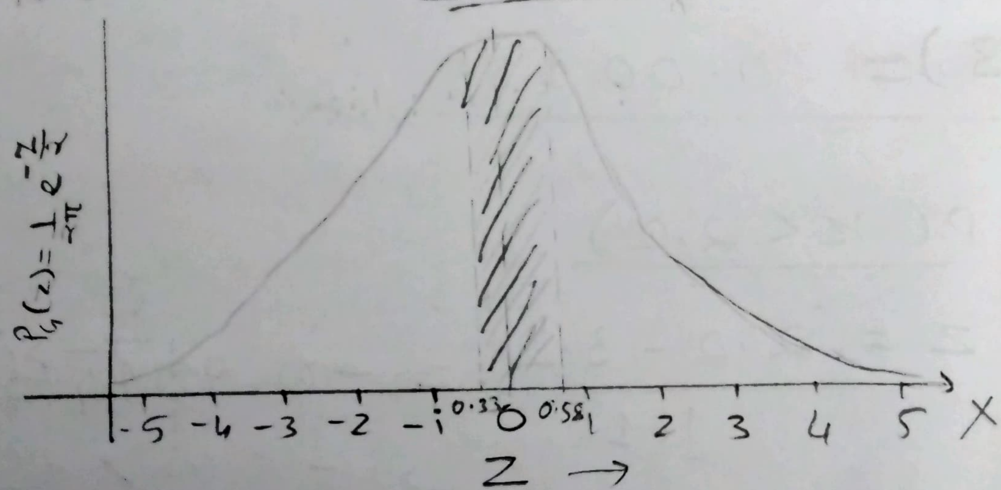
$P(z)$:

$$P(< Z_{2.8}) = 0.35942$$

$$P(< Z_{3.9}) = 0.73891$$

$$P(Z_{3.9} > Z > Z_{2.8}) = P(< Z_{3.9}) - P(< Z_{2.8})$$

$$= \underline{0.37949} \text{ Ans}$$



11. $\mu = 350$, $\sigma = 110$, $P_4(230 > x > 170) = ?$

Z transform of Gaussian Distribution:

$$Z_{170} = \frac{x - \mu}{\sigma} = \frac{170 - 350}{110} = -1.6\overline{363} \approx -1.64$$

$$Z_{350} = \frac{x - \mu}{\sigma} = \frac{230 - 350}{110} = -1.0\overline{90}$$

$P(x < Z_{170}) = 0.0505$

$P(x < Z_{350}) = 0.13786$

$P(-1.64 < x < -1.09) = \underline{0.087354}$ Ans

12. $\mu = 3.2, \ \sigma^2 = 1.21, \ \sigma = 1.1$

$P(x > 6.5\text{ kg}) = ?$

Applying Z transform of Gaussian Distribution:

$$Z_{6.5} = \frac{x - \mu}{\sigma} = \frac{6.5 - 3.2}{1.1} = 3$$

From Z table:

$P(x > Z_{6.5}) = 0.00135$

13.

Box 1	Box 2	Box 3
8 T1	11 T1	6 T1
6 T2	10 T2	8 T2
5 T3	7 T3	4 T3
19	28	18

Given that: $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$

There are 2 tissues picked, if both are T2 prob. that both are from

① B1 ② B2 ③ B3

i) For 1st tissue:

$$P_1(B1|T2) = \frac{P(T2|B1)\,P(B1)}{P(T2)}$$

$$= \frac{P(T2|B1)\,P(B1)}{P(T2|B1)\,P(B1) + P(T2|B2)\,P(B2) + P(T2|B3)\,P(B3)}$$

$$= \frac{6/19 \times \cancel{1/3}}{(6/19 \times \cancel{1/3}) + (10/28 \times \cancel{1/3}) + (8/18 \times \cancel{1/3})}$$

$$= 0.2826$$

For 2nd tissue: $B1 \rightarrow 18$, $T2$ in $B1 \rightarrow 5$

$$P_2(B1|T2) = \frac{P(T2|B1)\,P(B1)}{P(T2|B1)P(B1) + P(T2|B2)P(B2) + P(T2|B3)P(B3)}$$

$$= \frac{5/18 \times \cancel{1/3}}{(5/18 \times \cancel{1/3}) + (10/28 \times \cancel{1/3}) + (8/18 \times \cancel{1/3})}$$

$$= 0.25735$$

Thus for selecting 2, T2 tissues from B1 is

$$P(B1|T2) = P_1(B1|T2) \times P_2(B1|T2)$$

$$= \underline{\underline{0.07272}}$$

Ans

First pick:

ii) $P_1(B2|T2) = \dfrac{P(T2|B2)\,P(B2)}{P(T2)}$

$$= \frac{P(T2|B2)\,P(B2)}{P(T2|B1)\,P(B1) + P(T2|B2)\,P(B2) + P(T2|B3)\,P(B3)}$$

$$= \frac{10/28 \times \cancel{1/3}}{(6/19 \times \cancel{1/3}) + (10/28 \times \cancel{1/3}) + (8/18 \times \cancel{1/3})}$$

$P_1(B2|T2) = 0.31962$

Second pick: T2 in B2 → 9, total in B2 → 27

$P_2(B2|T2) = \dfrac{P(T2|B2)\,P(B2)}{P(T2)}$

$$= \frac{9/27 \times \cancel{1/3}}{(6/19 \times \cancel{1/3}) + (9/27 \times \cancel{1/3}) + (8/18 \times \cancel{1/3})}$$

$P_2 = 0.30481$

For selecting T2 tissue twice from B2 is

$P(B2|T2) = P_1(B2|T2) \times P_2(B2|T2)$

$= \underline{\underline{0.09742}}$ Ans

iii) For first pick:

$$P_1(B3|T2) = \frac{P(T2|B3)\,P(B3)}{P(T2)}$$

$$= \frac{P(T2|B3)\,P(B3)}{P(T2|B1)\,P(B1) + P(T2|B2)\,P(B2) + P(T2|B3)\,P(B3)}$$

$$= \frac{8/18 \times \cancel{1/3}}{(6/19 \times \cancel{1/3}) + (10/28 \times \cancel{1/3}) + (8/18 \times \cancel{1/3})}$$

$P_1 = 0.39775$

For selecting T2 second time: T2 in B3 → 7, Total = 17.

$$P_2(B3|T2) = \frac{P(T2|B3)\,P(B3)}{P(T2)}$$

$$P_2(B3|T2) = \frac{P(T2|B3)\,P(B3)}{P(T2|B1)\,P(B1) + P(T2|B2)\,P(B2) + P(T2|B3)\,P(B3)}$$

$$= \frac{7/17 \times \cancel{1/3}}{(6/19 \times \cancel{1/3}) + (10/28 \times \cancel{1/3}) + (7/17 \times \cancel{1/3})}$$

$= 0.3796$

Selecting T2 from B3 both times ⇒

$P(B3|T2) = P_1 \times P_2$

$= \underline{\underline{0.15099}}$ Ans

14. $B = \{82, 86, 89, 80, 78, 83, 77, 94, 90, 88, 82, 82\}$

To find median data should be sorted in ascending order:

$n = 12$

$B = \{77, 78, 80, 82, 82, 82, 83, 86, 88, 89, 90, 94\}$

$$\text{Mean}(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{1011}{12} = \underline{\underline{84.25}} \text{ mm Hg}$$

Ans

$$\text{Stand. Deviation}(s) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$s = \sqrt{\frac{1}{11} \times \left((77-84.25)^2 + \cdots + (94-84.25)^2\right)}$$

$$s = \sqrt{\frac{294.25}{11}}$$

$$\underline{\underline{s}} = \sqrt{26.75} = \underline{\underline{5.1720}} \text{ mm Hg}$$

$$\text{Median} \rightarrow \text{For even} = \frac{\left(\frac{n}{2}\right)^{th} \text{term} + \left(\frac{n}{2}+1\right)^{th} \text{term}}{2}$$

$$= \frac{6^{th} \text{term} + 7^{th} \text{term}}{2}$$

$$\text{Median} = \frac{82+83}{2} = \underline{\underline{82.5}} \text{ mm Hg}$$

Ans

15. GFR of 12 patients :

Sorted $G = \{ \overset{1}{18}, \overset{2}{32}, \overset{3}{36}, \overset{4}{37}, \overset{5}{42}, \overset{6}{43}, \overset{7}{48}, \overset{8}{58}, \overset{9}{60}, \overset{10}{62}, 67 ...$

a) $\text{Mean } (\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{591}{12} = \underline{\underline{49.25}}$ ml/min Ans

$$\text{Std. dev. } (s) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$s = \sqrt{\frac{1}{11} \times ((18-49.25)^2 + \ldots + (88-49.25)^2)}$$

$$s = \sqrt{\frac{3864.25}{11}}$$

$$s = \sqrt{351.2954} = \underline{\underline{18.74287}} \text{ ml/min}$$

Ans

For even no. of values :

$$\text{Median} = \frac{\frac{n}{2}^{th} \text{ term} + (\frac{n}{2}+1)^{th} \text{ term}}{2}$$

$$= \frac{6^{th} \text{ term} + 7^{th} \text{ term}}{2} = \frac{43+48}{2} = \frac{91}{2} = \underline{\underline{45.5}}$$ Ans

1st Quartile $= (\frac{1}{4} \times n+1)^{th}$ term $= 3.25$, thus we will take mean of 3^{rd} & 4^{th} term

$$1\text{st Quartile} = \frac{3^{rd} \text{ term} + 4^{th} \text{ term}}{2} = \frac{36+37}{2} = \underline{\underline{36.5}}$$ Ans

3^{rd}

3^{rd}

b)

3rd Quartile : $\left(\frac{3}{4}(n+1)\right)^{th}$ term $= \underline{\underline{9.75}}$

So, we will take avg. of 9th and 10th term.

3rd Quartile $= \frac{60+62}{2} = \underline{\underline{61}}$ Ans

Quantile

Name : Anmol Singh

Date :

Q 15 - b)

Box and Whisker Plot