Mode :

> The most frequent number occurring in the data set is known as the mode.

Eg :

60, 63, 64, 45, 63, 65, 70, 55, 63, 60, 65, 63.

2	—	60, ~~4~~, 60
4	—	63, 63, 63, 63
1	—	64
1	—	45
2	—	65, 65
1	—	70

55 — 1

63 → Mode.

Median :

1. Arrange the data in Ascending order.

2. The Data is odd means,

$$\frac{(n+1)}{2}$$

3. If even means,

$(n/2)\ (n/2+1)$

Eg:

60, 63, 45, 63, 65, 70, 55, 63, 60, 65, 63

1. Assending order:

45, 55, 60, 60, 63, 63, 63, 63, 65, 65, 70

$\Downarrow$ Median.

2. $n=11 \Rightarrow \frac{11}{2} = 5.5$ (6th element).

Eg:

26.3, 28.7, 27.4, 26.6, 27.4, 26.9

1. A.O:

26.3, 26.6, 26.9, 27.4, 27.4, 28.7.

2. $n=6 \Rightarrow n/2\ ,\ \frac{n}{2}+1$

$6/2\ ,\ 3+1 = 4 \Rightarrow \frac{3+4}{2}$

3

$$\frac{26.9 + 27.4}{2} = 27.15$$

Median = 27.15

Mean :

> It represents the average of the given collection of data.

> It is applicable for both continuous & discrete data.

> It is equal to the sum of all the values in the collection of data divided by the total no. of values.

Eg :

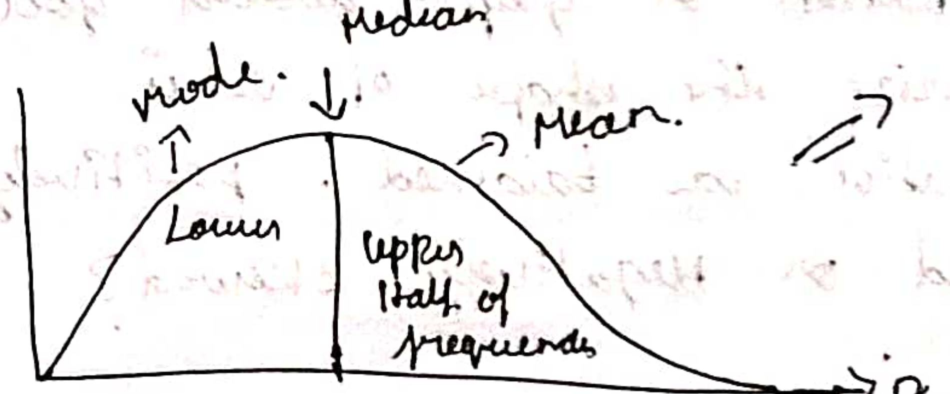
60, 63, 45, 63, 65, 70, 55, 63, 60, 65, 63.

$$\overline{x} = \frac{\Sigma x}{n}$$

$$= \frac{60+63+45+63+65+70+55+63+60+65+63}{11}$$

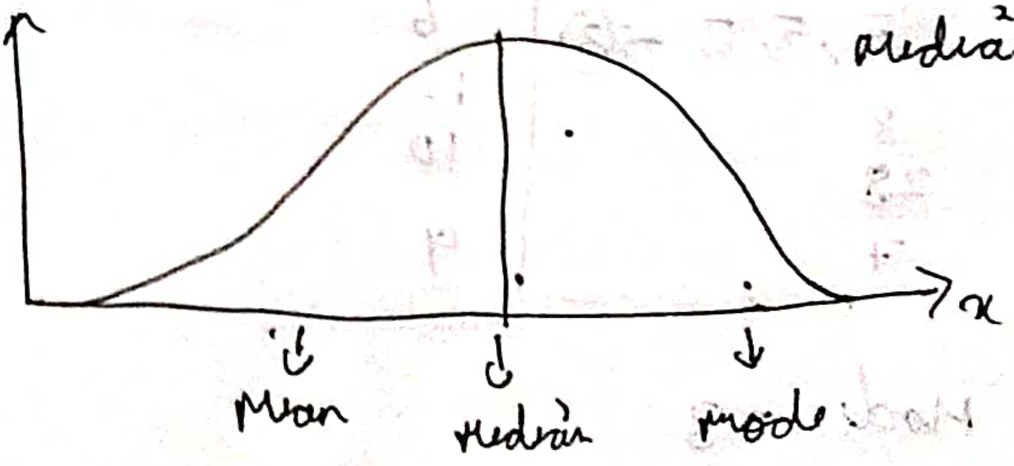
$$= \frac{797}{11} = 72.455.$$

properties :



$\Rightarrow$ mean > median

Positively skewed Distribution



median > mean.

Negatively skewed Distribution.

Q) During their first swim through a water maze. 15 Laboratory rats made the following no. of errors (blind alleyway entrances): 2, 17, 5, 3, 28, 7, 5, 8, 5, 6, 2, 12, 10, 4, 3.

a) Find the Mean, median, & mode for these data

b) Without constructing a frequency distribution or graph would you characterize the shape of this distribution as balanced, positively skewed or Negatively skewed?

a) Mode

2 12
17
5, 5, 5, → 3
3
28
7

8
6
12
10
4

Mode = 3

median,

1. Sort:

2, 2, 3, 3, 4, 5, 5, 5 6, 7, 8, 10, 12,
17, 28

⇓

Median = 5

$n \Rightarrow 15. \Rightarrow \frac{15}{2} = 7.5 = 8^{th}$ element.

Mean:

$$\frac{\Sigma x}{n} = 7.8$$

b) Mean Median

7.8 > 5

∴ since Mean is greater than the Median so it is positively skewed.

Describing Variability:

Range:

It is the difference between the maximum & the minimum observation of the distribution.

Range = Max − Min.

IQR:

> It defines the diff. b/w the 3^{rd} & the 1^{st} Quartiles.

> Q1 → 25^{th} percentile (Lower).

> Q2 → the Median (50%) "

> Q3 → 75^{th} percentile. (upper)

IQR = 75% − 25%.

Degree of Freedom : $n-1$

Standard Deviation : $= \sqrt{\text{variance}}$.

variance $= \frac{SS}{N}$

$V = \frac{SS}{N-1}$

Variance :

Population

sample.

Sum of Square =

$\Sigma(x-\mu)^2$.

(definition Formula).

$SS = \Sigma(x-\bar{x})^2$.

(Computation Formula)

$SS = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$

$SS = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$.

Eg: Find STD: $n = 7$

13, 10, 11, 7, 9, 11, 19. $\Rightarrow \frac{70}{7} = 10.$

X	$X - \mu$	$(X - \mu)^2$
13	3	9
10	0	0
11	1	1
7	−3	9
9	−1	1
11	+1	1
9	−1	1
$\Sigma 70$		$\Sigma 22.$

$$\text{Variance} = \frac{SS}{N.}$$

$$= \frac{22}{7}$$

$$= 3.14$$

$$\text{std. deviation} = \sqrt{\text{Var.}}$$

$$= \sqrt{3.14} = 1.77$$

Q) using the definition formula for the sum of squares, calculate the sample standard deviation for the following four scores: 1, 3, 4, 4.

~~(a) 1, 3, 7, 2, 0, 4, 7, 3~~

std. dev. $= \sqrt{var}$.

$var = \frac{SS}{N-1}$.

$SS = \Sigma (x - \bar{x})^2$.

$\bar{X} = \frac{\Sigma X}{N} = \frac{12}{4} = 3$

X	$x - \bar{x}$	$(x - \bar{x})^2$
1	-2	4
3	0	0
4	1	1
4	1	1
12	0	6

$SS = 6$.

$var = \frac{6}{3} = 2$.

$\sqrt{2} = 1.414$

S.D.

Q) As a first step toward modifying his study habits, Phil keeps daily records of his study time.

a) During the first two weeks, Phil's mean study time equals 20 hours per week. If he studied 22 hours during the first week, how many hours did he study during the second week?

$$\text{Mean} = \frac{1^{st} + 2^{nd}}{2} = 20.$$

$$\frac{22 + 2^{nd}}{2} = 20$$

$$2^{nd} = 18 \text{ hrs.}$$

b) During the 1^{st} 4 weeks, Phil's Mean study time equals 21 hours. If he studied 22, 18, & 21 hrs during the 1^{st}, 2^{nd} & 3^{rd} weeks, respectively, how many hours did he study during the 4th week?

$$\frac{x + y + w + z}{4} = 21.$$

$$\frac{22 + 18 + 21 + z}{4} = 21.$$

$$z = 23 \text{ hrs}$$

c) If the Information in (a) & (b) is to be used to estimate some unknown population characteristic, the notion of degree of freedom can be introduced. How many degrees of freedom are associated with (a) & (b)?

$df = n-1.$

a) $n=2 \Rightarrow df = 1$

b) $n=4 \Rightarrow df = 3.$

d) Describe the mathematical restriction that causes a loss of degrees of freedom in (a) & (b).

when all observation are expressed as deviations from their mean, the sum of all deviations. must equal to zero.

Q.) Determine the values of the range & the IQR for the following data set.

a) Retirement ages : 60, 63, 45, 63, 65, 70, 55, 63, 60, 65, 63.

Sort :

45, 55, 60, 60, 63, 63, 63, 63, 65, 65, 70.

Range : max - min :

70 - 45 = 25.

IQR : 75% $0.75 \times 11 = 8.25 = 9^{th}$

25% $0.25 \times 11 = 2.75 = 3^{rd}$

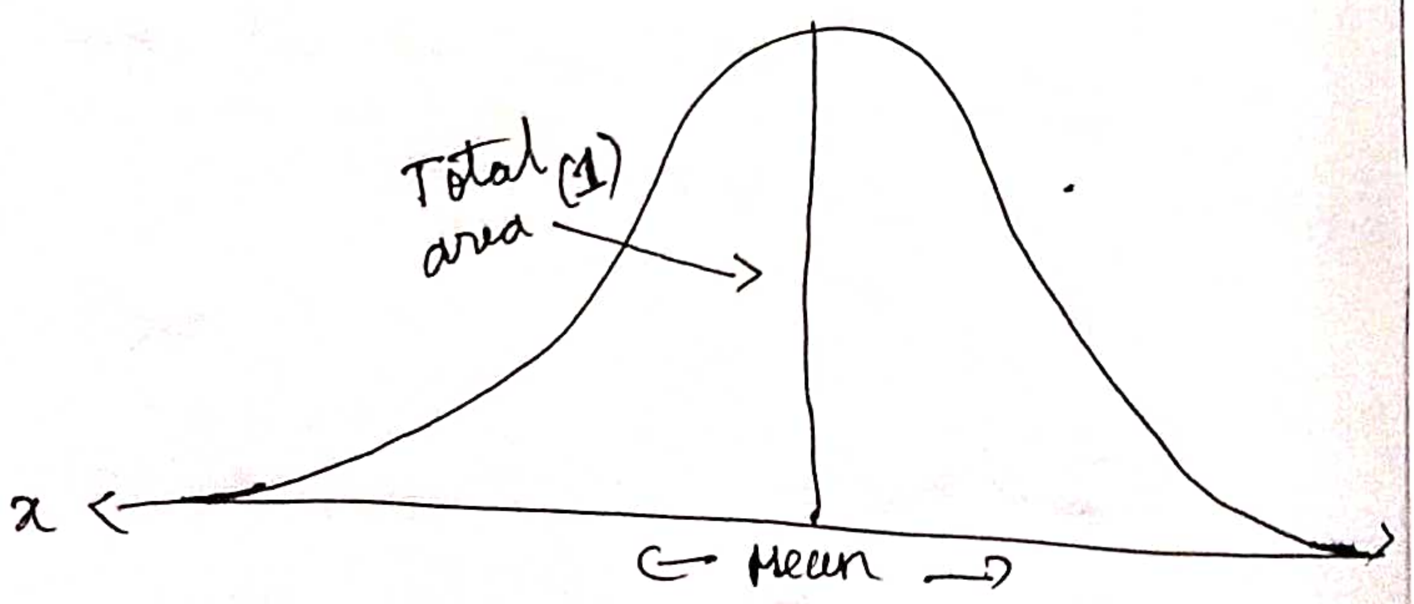
$3^{rd} = 60$ $9^{th} = 65$.

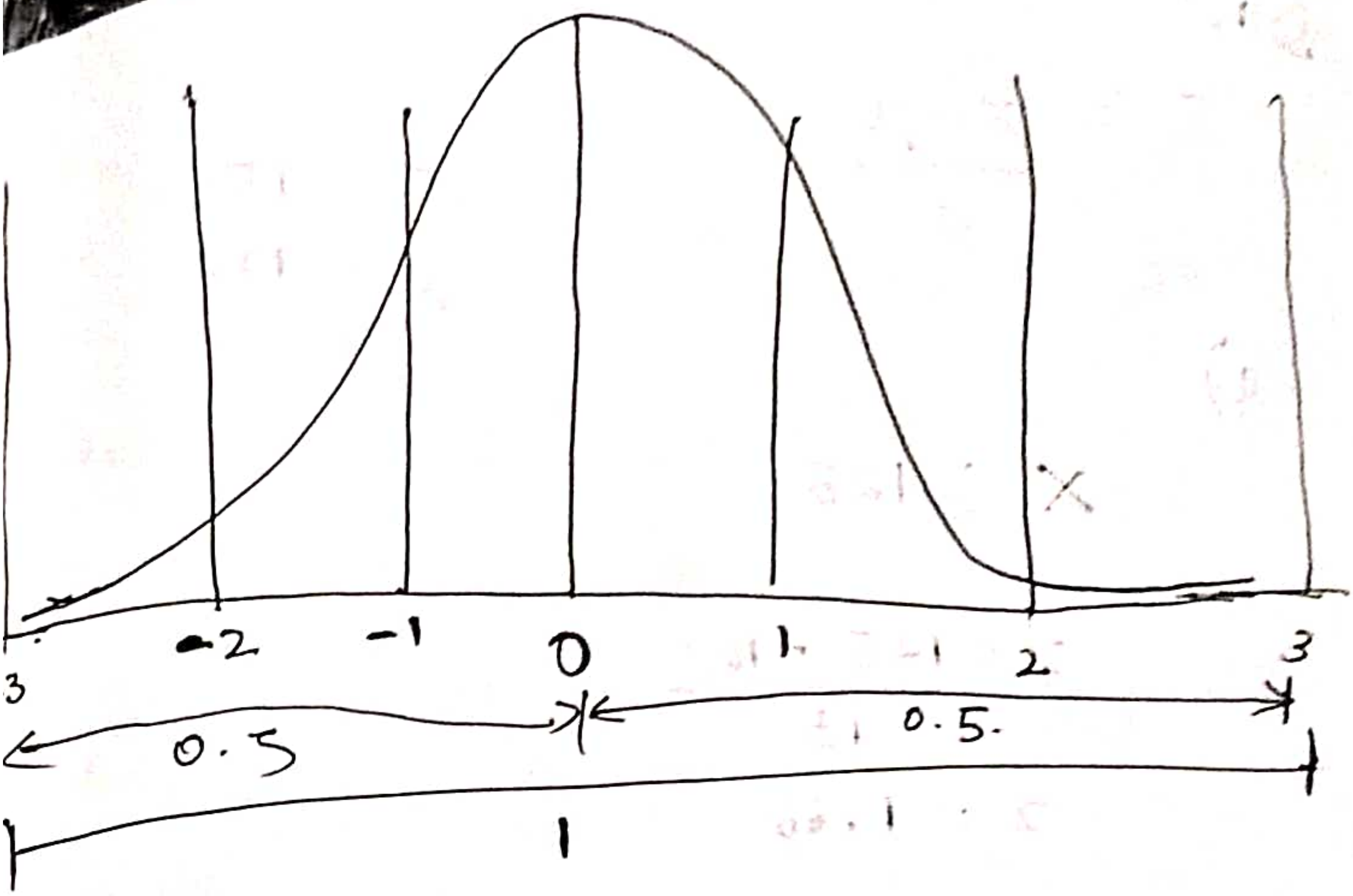
IQR $= 65 - 60$

$= 5$.

Normal distribution: (properties)

1. The Mean, Median, & Mode are equal
2. The Normal curve is bell-shaped and is symmetric about the Mean.
3. The Total area under the normal curve is equal to 1.
4. The Normal curve approaches, but never touches, the x-axis as it extends farther and farther away from the Mean.





Q) Scores on the Wechsler Adult Intelligence Scale (WAIS) approximate a normal curve with a mean of 100 & an standard deviation of 15. What proportion of IQ scores are

a) above 125?

b) below 82?

c) within 9 points of the mean?

d) more than 40 points from the mean?

Soln:

$$Z = \frac{x-\mu}{\sigma}$$

$\sigma = 15$

$\mu = 100$

a)

$x > 125$

$$Z = \frac{125 - 100}{15}$$

$Z = 1.66$

$P(Z > 1.66) = 0.5 - P(Z = 1.66)$

$= 0.5 - 0.4515$

$= 0.0485$

b) $x < 82$

$$Z = \frac{82 - 100}{15}$$

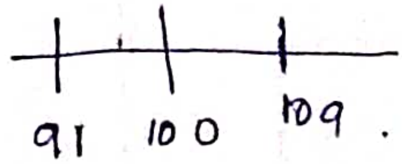
$= -1.2$

$P(Z < -1.2) = 0.5 - P(Z = -1.2)$

$= 0.5 - 0.3849$

$= 0.1151$

c) $\mu = 100$



$91 < x < 109$

$91 < x \Rightarrow Z_1 = \frac{x - \mu}{\sigma}$

$= \frac{91 - 100}{15} = -0.6$

$x < 109 \Rightarrow Z_2 = \frac{109 - 100}{15} = 0.6$

$p(-0.6 < z < 0.6) = 2p(z < 0.6)$.

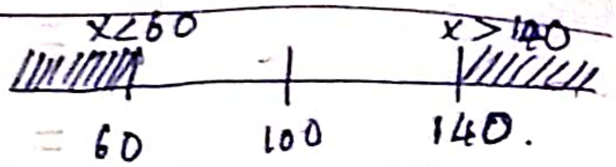
$= 2p(z = 0.6)$.

$= 2(0.2257)$

$= 0.4514$.

---

d) $\mu = 100$



$x > 140$, $x < 60$.

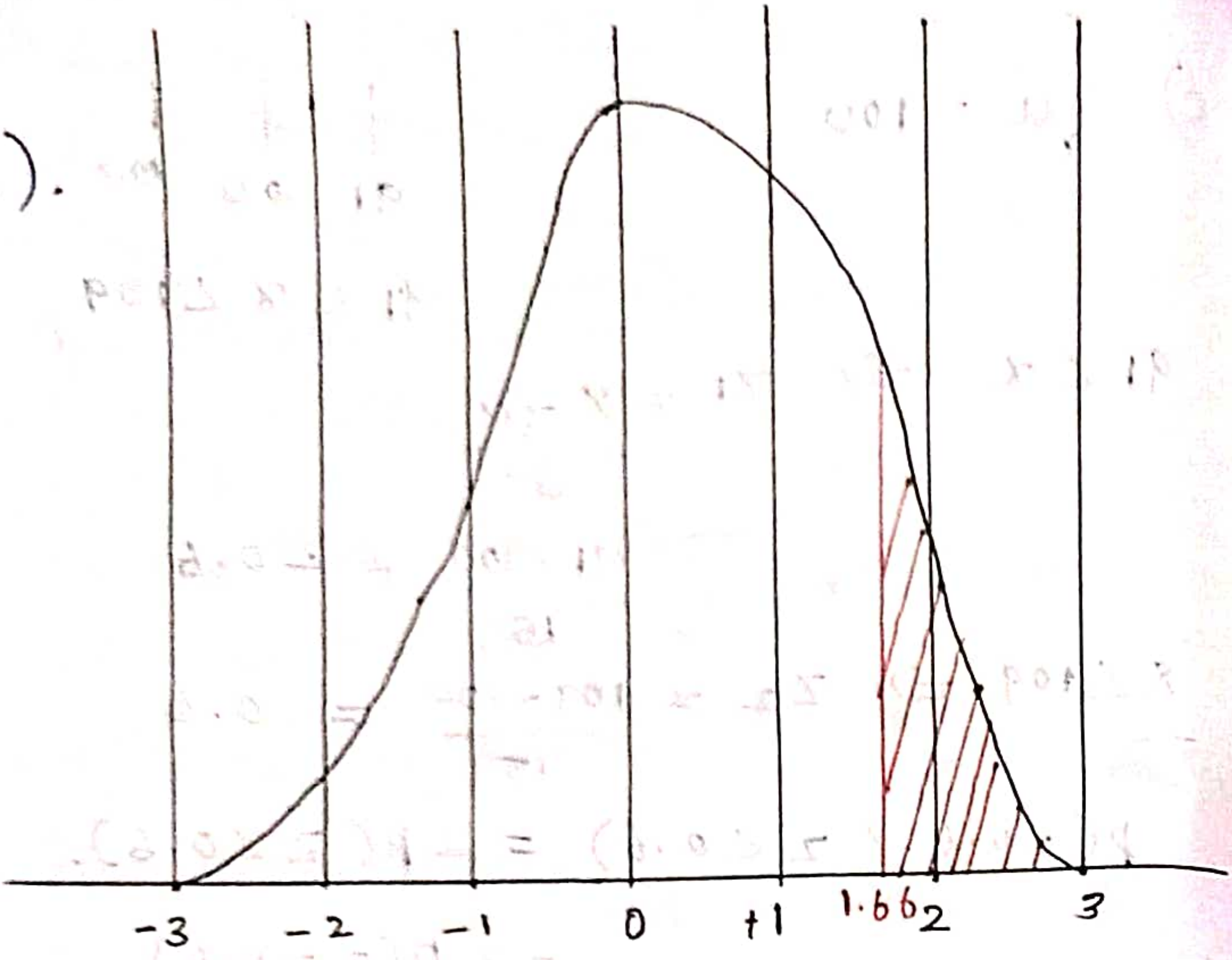
$Z_1 = \frac{140 - 100}{15} = 2.66$

$Z_2 = \frac{60 - 100}{15} = -2.66$

$P(Z_1 > 2.66) = 0.5 - P(z = 2.66)$.

$= 0.5 - 0.4961 = 0.0039$.

a).

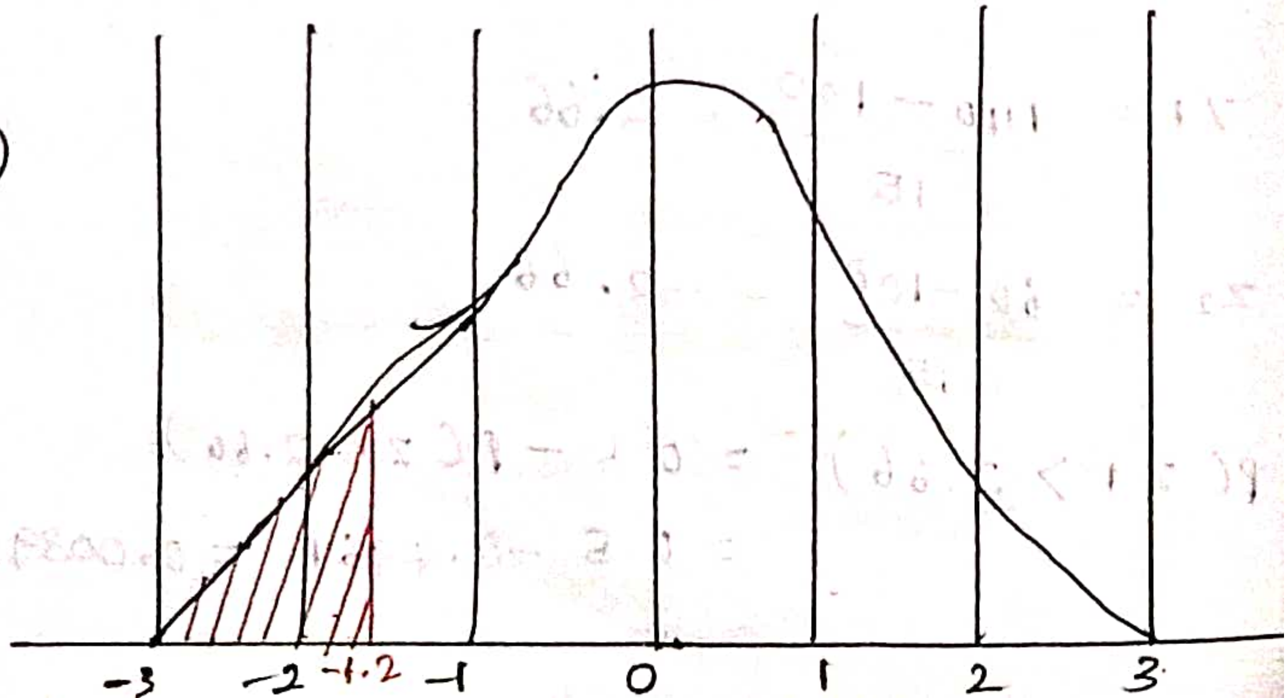


$p(z_2 < -2.66) = 0.5 - p(z = 2.66)$

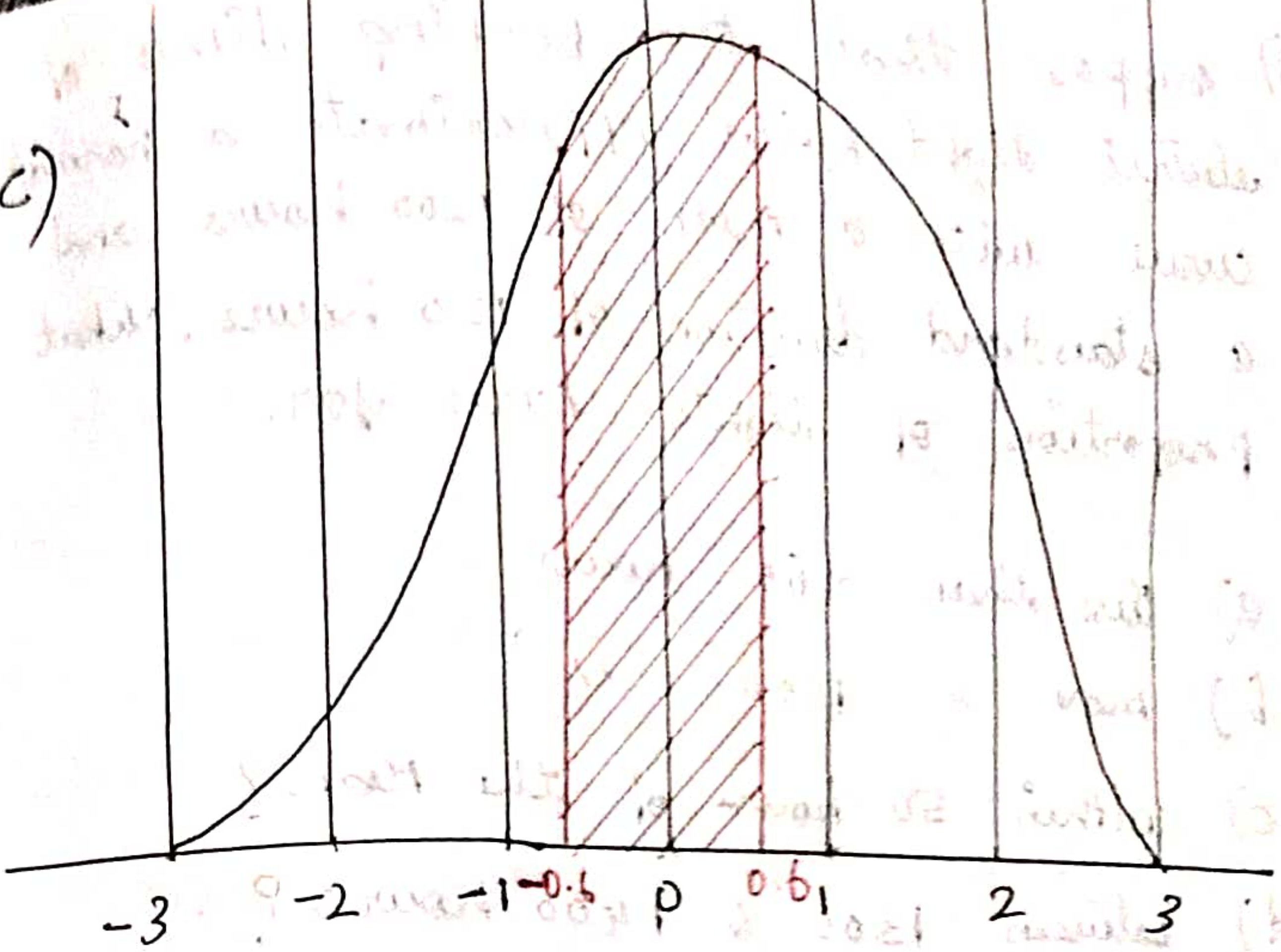
$= 0.5 - 0.4961$

$= 0.0039$

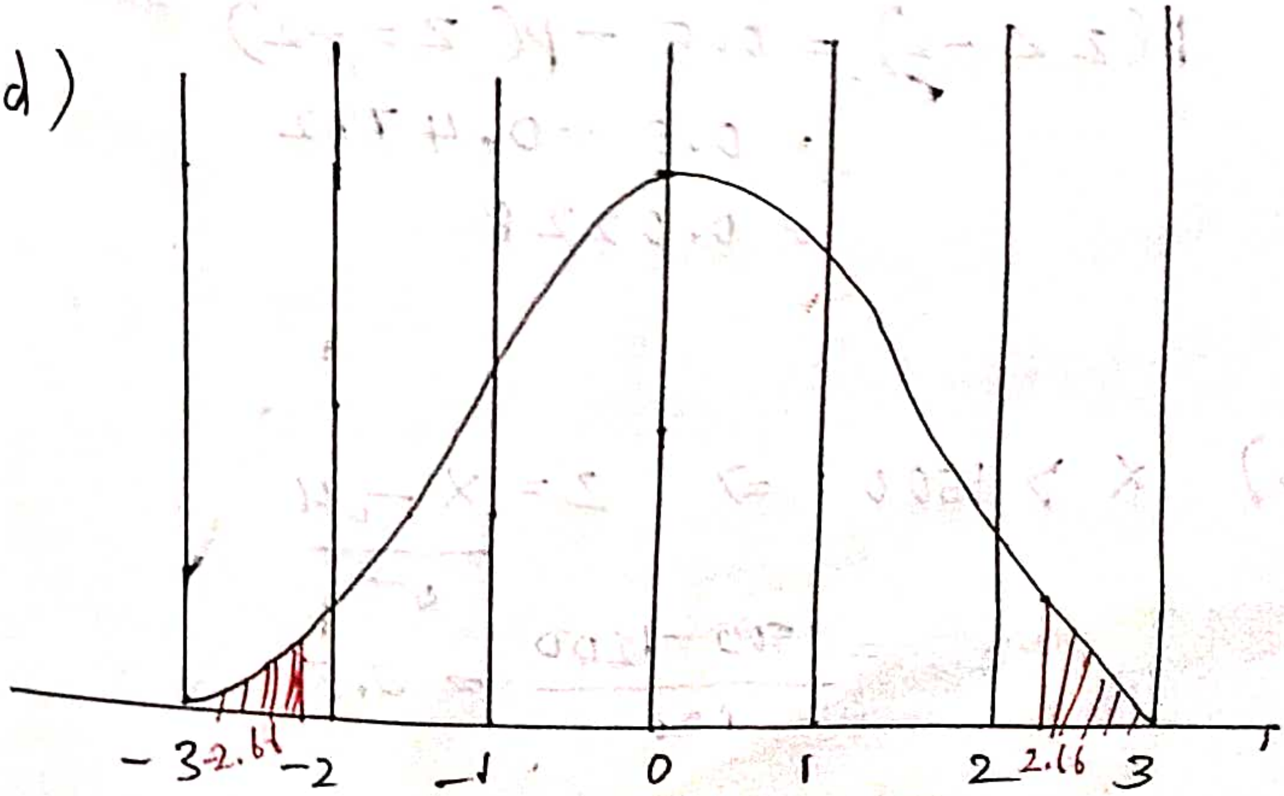
b)



c)



d)



Q) suppose that the burning times of electric light bulbs approximate a normal curve with a mean of 1200 hours and a standard deviation of 120 hours. what proportion of lights burn for.

a) less than 960 hours

b) more " 1500 "

c) within 50 hours of the Mean?

d) between 1300 & 1400 hours?

soln:

a) $x < 960 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{960-1200}{120}$

$= -2.$

$p(z < -2) = 0.5 - p(z = -2)$

$= 0.5 - 0.4772$

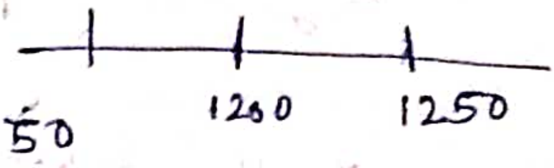
$= 0.0228$

b) $x > 1500 \Rightarrow z = \dfrac{x-\mu}{\sigma}$

$= \dfrac{1500-1200}{120} = 2.5$

$$p(z > 2.5) = 0.5 - p(z = 2.5)$$
$$= 0.5 - 0.4938$$
$$= 0.0062.$$

c)



$$1150 < x < 1250.$$

$$z_1 = \frac{x-\mu}{\sigma} = \frac{1150-1200}{120}$$
$$= -0.41$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{1250-1200}{120}$$
$$= 0.41$$

$$p(-0.4 < z < 0.41) = 2p(z = 0.41)$$
$$= 2(0.1591)$$
$$= 0.3182.$$

d)

$1300 < X < 1400$

$$z_1 = \frac{x-\mu}{\sigma} = \frac{1300-1200}{120} = 0.83$$

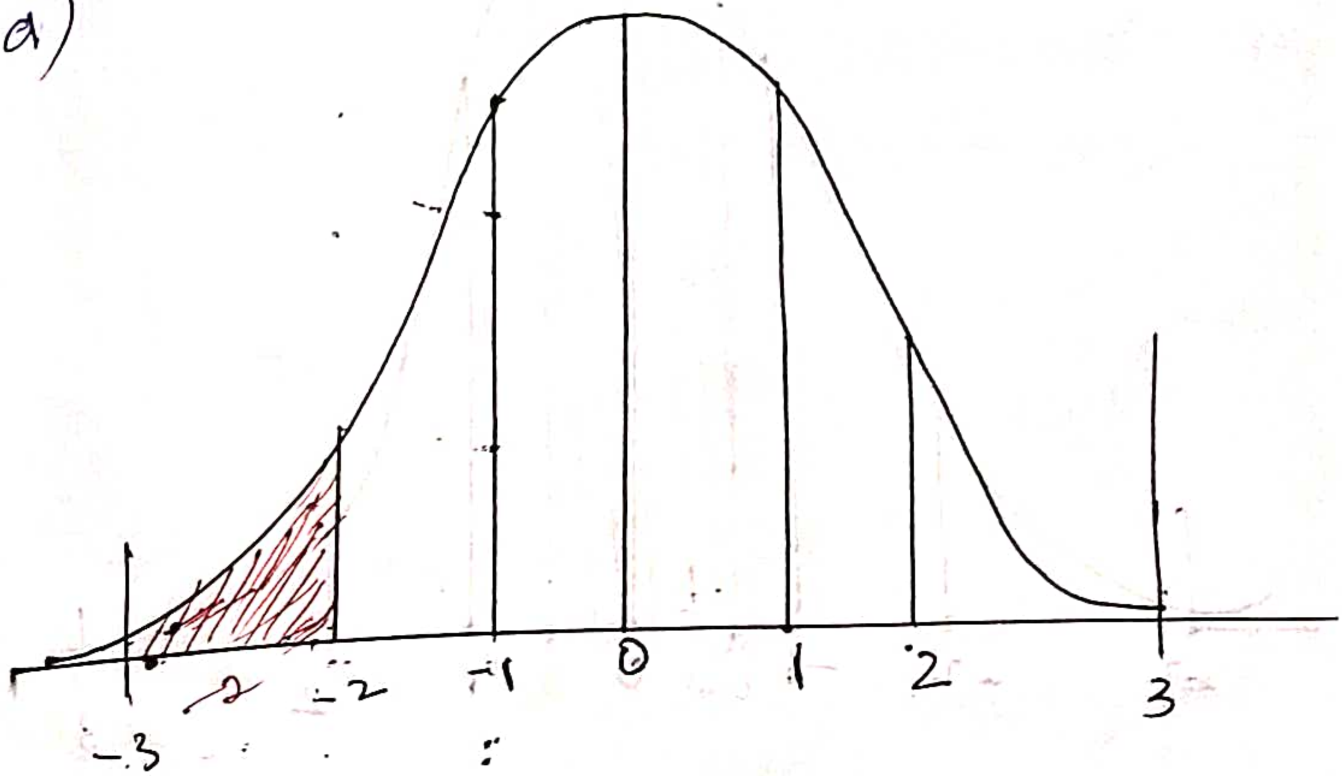
$$Z_2 = \frac{x-\mu}{\sigma} = \frac{1400-1200}{120} = 1.66$$

$$p(0.83 < z < 1.66) = p(z=1.66) - p(z=0.83)$$

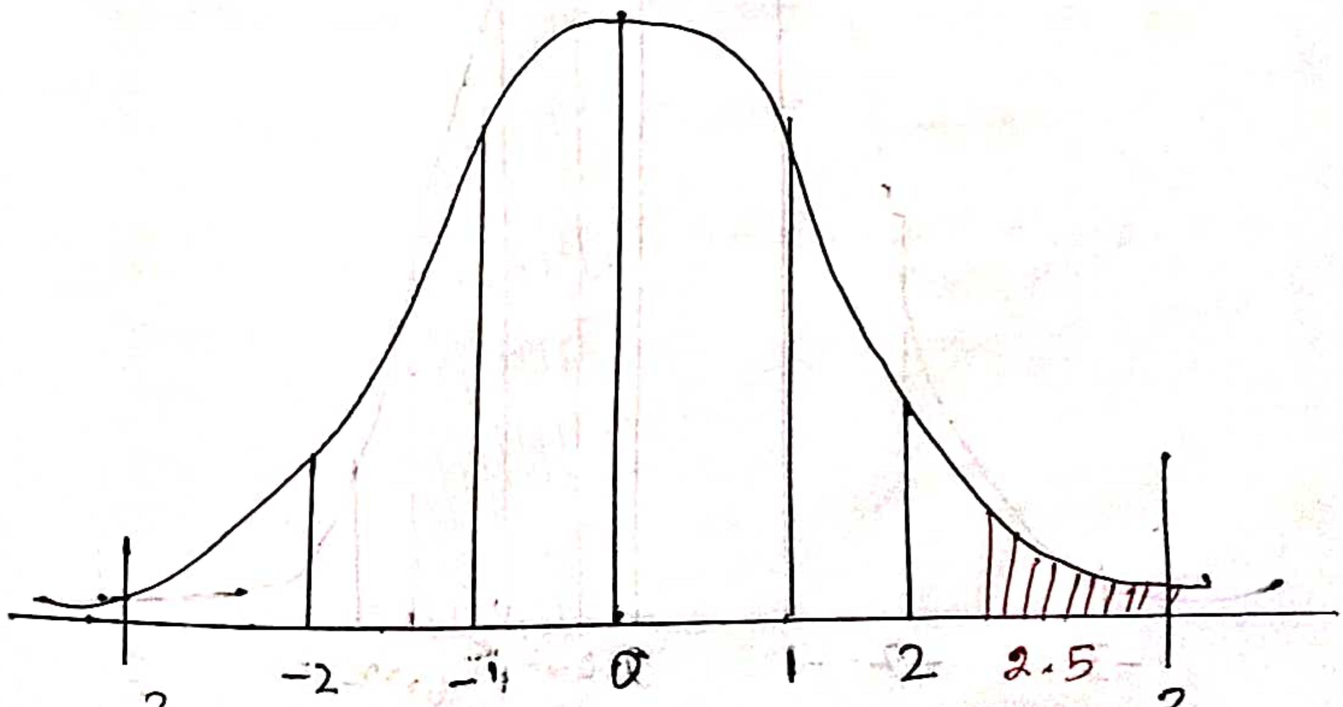
$$= 0.4515 - 0.2967$$

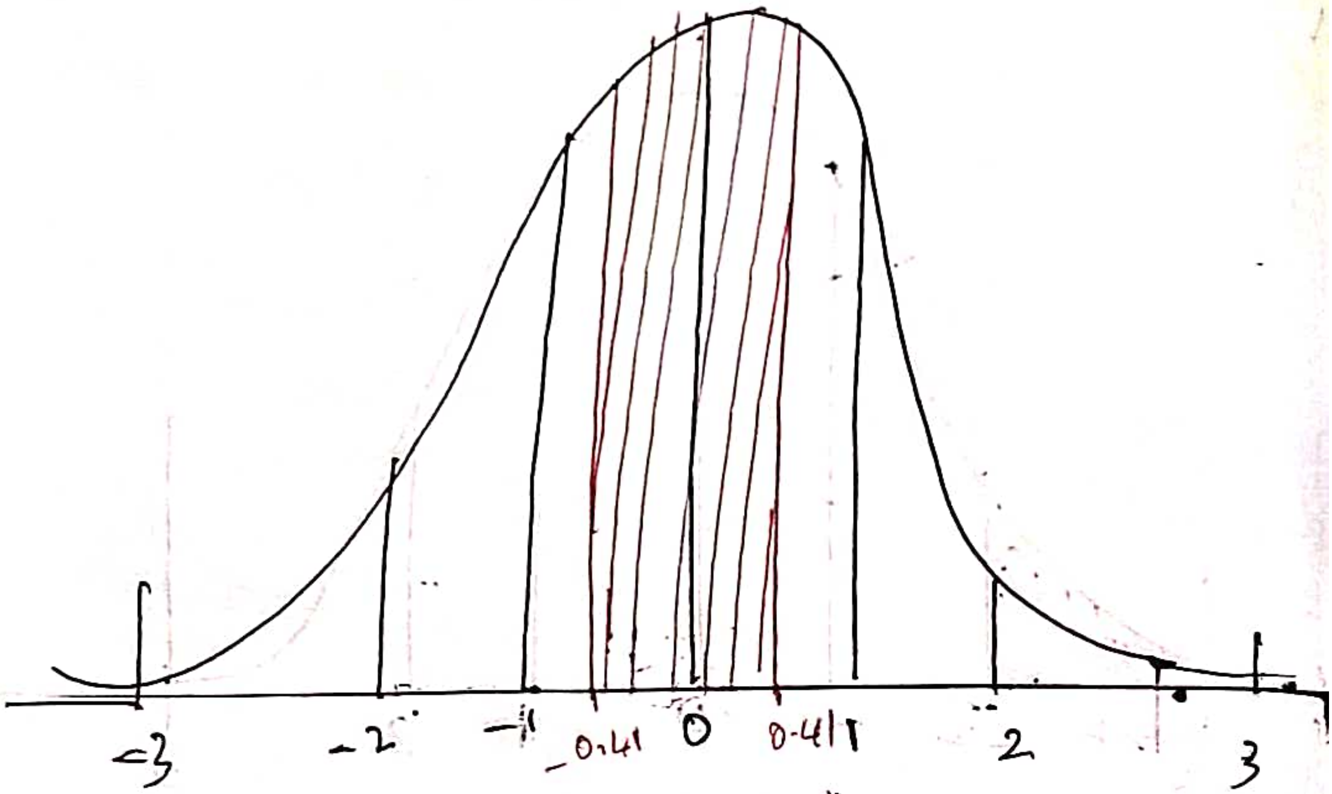
$$= 0.1548.$$

a)

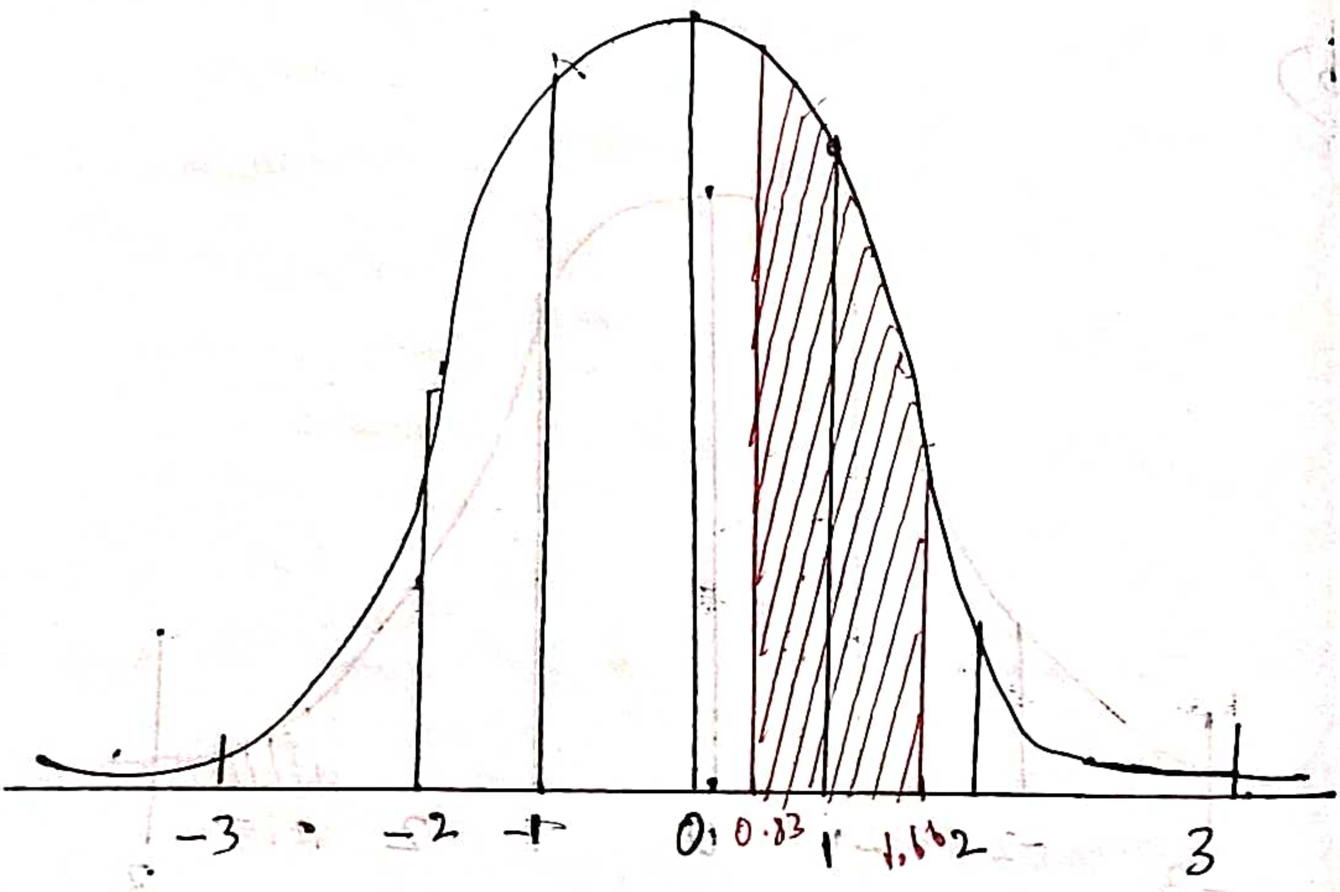


b)





-3
-2
-1
-0.41
0
0.41
1
2
3



-3
-2
-1
0
0.83
1
1.66
2
3

Frequency distribution.

Q.) construct the frequency distribution for : 17, 19, 22, 26, 29, 39, 17, 20, 23, 31, 12, 20, 27, 34, 18, 28, 15, 23, 18, 32, 29, 10, 22, 20.

1. sort :
10, 12, 15, 17, 17, 18, 18, 19, 20, 20, 20, 22, 22, 23, 23, 26, 27, 28, 29, 29, 31, 32, 34, 39.

2. Range = max − min = 39 − 10 = 29

3. No. of total groups = 6 (Assume)

4. width $= \frac{29}{6} = 4.83 \approx 5$

Class	Frequency
10 − 14	2
15 − 19	6
20 − 24	7
25 − 29	5
30 − 34	3
35 − 39	1
sum	24

Outliers: (Extreme value (or) un related data)

Q) Identify any outliers in each of the following sets of data collected from nine college student.

Summer Income	AGE	Family size	GPA
\$6,450	20	2	2.30
\$4,820	19	4	4.00
\$5,650	61	3	3.~~86~~ 56
\$1,720	32	6	2.89
\$600	19	18	2.15
\$0	22	2	3.01
\$3,482	23	6	3.09
\$25,700	27	3	3.50
\$8,548	21	4	3.20

outliers are a summer income of \$25,700, an age of 61; and a family size of 18. No outliers for GPA.

Relative frequency distribution:

Q) GRE scores for a group of graduate school applicants are distributed as follows:

$\therefore 0.005 = 0.01$

$\therefore 0.004 \neq 0.01$

GRE	f	R.f.
725 - 749	1	$1/200 = 0.005$ (0.01
700 - 724	3	0.02 (× 100 only
675 - 699	14	0.07 If they
650 - 674	30	0.15 ask in percentage)
625 - 649	34	0.17
600 - 624	42	0.21
575 - 599	30	0.15
550 - 574	27	0.14
525 - 549	13	0.07
475 - 499	4	0.02
500 - 524	2	0.01
	200	1.02

Formula : $r.f = f / \Sigma f_i$ f → Frequency

Cumulative Frequency distribution :-

Find cumulative frequency and %.

Q) GRE scores for a group of graduate school applicants are distributed as follows :

(*) → start with smallest group.

GRE	f	CF	CF%
725 - 749	1	200	100%
700 - 724	3	199	99.5%
675 - 699	14	196	98%
650 - 674	30	182	91%
625 - 649	34	152	76%
600 - 624	42	118	59%
575 - 599	30	76	38%
550 - 574	27	46.	23%
525 - 549	13	19	10%
500 - 524	4	6	3%
475 - 499	2	2	1%
Total	200		

$$C.F\% = \frac{C.F}{\Sigma f} \times 100$$

Q) Movie ratings reflect ordinal measurement because they can be ordered from most to least restrictive: NC-17, R, PG-13, PG and G. The ratings of some films shown recently in San Francisco are as Follow:

PG	PG	PG	PG-13	G
G	PG-13	R	PG	PG
R	PG	R	PG	R
NC-17	NC-17	PG	G	PG-13

a) construct a frequency distribution.

b) convert to relative frequencies, expressed as percentages.

c) construct a cumulative frequency distribution.

d) Find the approximate percentile rank for those films with a PG rating.

Rating	frequency	Relative Freq (%)	C.F	CF(%)
NC17	2	$\frac{2}{20} \times 100 = 10\%$	18+2=20	$\frac{20}{20} \times 100 = 100\%$
R	4	$\frac{4}{20} \times 100 = 20\%$	14+4=18	$\frac{18}{20} \times 100 = 90\%$
PG-13	3	$\frac{3}{20} \times 100 = 15\%$	11+3=14	$\frac{14}{20} \times 100 = 70\%$
PG	8	$\frac{8}{20} \times 100 = 40\%$	3+8=11	$\frac{11}{20} \times 100 = 55\%$
G	3	$\frac{3}{20} \times 100 = 15\%$	3	$\frac{3}{20} \times 100 = 15\%$
	20			
		$\frac{f}{\Sigma f} \times 100$		$\frac{cf}{\Sigma f} \times 100$

d) => $\frac{11}{20} \times 100$

$= 55\%$

Coefficient of correlation :

Formula:　　　　2. Direct Method.

$$r = \frac{n(\Sigma x_i y_i) - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n(\Sigma x_i^2) - (\Sigma x_i)^2 \cdot n(\Sigma y_i^2) - (\Sigma y_i)^2}}$$

$$\text{Rank } (\rho) = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}.$$

population correlation coefficient formula :

~~1. A~~

Methods of Calulating Karl Pearson's coefficient of correlation.

1. Actual Mean Method.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$