

Introduction of Statistics & Probability

→ M. Nurul Islam

Ch-2 - movement to low staff

Ch-2

Statistics

- ↳ Sensors, Sample
 - ↳ Population (all possible sample)

→ Data

- 1. Quantitative (Height, weight)
 - 2. Qualitative
 - 2. Numeric, contiguous
 - 2. Qualitative (categorical)

↳ gender, → male, female

↳ status → yes, no

proposals → economic status or welfare issue: of oil and gas

MACS, Stage 3 early
~~late~~ gestation

Final

CONTINUATION

* Sample \oplus ? What do you understand by sample, describe it.

* Level of Measurement:

↳ Numerical

* Nominal Level of Measurement: Colors,

↳ **order** বড় যোগ এবং কৃতি মাঝ তাত্ত্বিক expressions করে,
যোগ, বিয়োগ, \times , \div , $=$

* Ordinal Level of Measurement: Health status, job rank

↳ Order/Rank ग्रन्थ विभाग

$\leq, \geq, =, \leq, \geq$

* Interval Level of Measurement: IQ test score.

↳ Interval \rightarrow 31, 50, 75 \rightarrow 2, 4, 6, 8 \rightarrow 0 means 0 IT

↳ Multiply/divide \rightarrow 31, 50, 75 \rightarrow \rightarrow [19=0] Ex

↳ temperature

* Ratio Level of Measurement: Height, weight

↳ Interval \rightarrow 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

↳ 0 means 0

\rightarrow 0 IT Meaning 0 IT

* table \rightarrow comparison 4 or Level \rightarrow 82-1

* Variable & Attribute:

↳ vari \rightarrow vary \rightarrow diff

↳ Quantitative \rightarrow Numerical
Qualitative

* Attribute: vari Qualitative variable \rightarrow distinct category

* Discrete

* Continuous

* Constant

* Summarizing and representing data:

1. Mean

size, freq, avg, \rightarrow 20, 30, 40

2. Median

\rightarrow Main influenced by extreme value

3. Mood
Mode

Middle point freq \rightarrow 20, 30, 40

\rightarrow can be different

2) \rightarrow Frequency distribution

5.1, 5.2, 5.4, 5.2, 5.3, 5.1, 5.2, 5.3, 5.2, 5.3, 5.1, 5.4, 6.00, 5.3,

5.5, 5.9, 5.8, 5.10, 5.7, 5.6, 5.4, 5.6, 5.11, 5.10, 5.6

\Rightarrow

Interval	Frequency	Border	Cumulative Frequency	Probability/Percent
5.00 - 5.50	15		15	15/25 =
5.5 - 5.10	8		23	8/25 =
5.10 - 5.60	2		25	2/25 =
	25			

$$P(X \approx 5.6) = \frac{8}{25}$$

Lower class Limit

Upper " "

* Class Mark = Mid point

* Instant steps

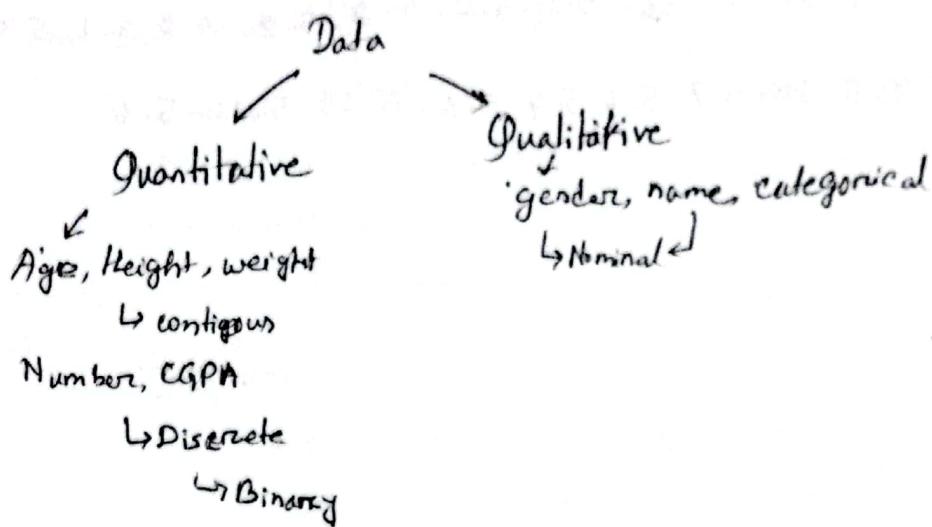
$$\text{Class interval} = \frac{L-S}{d} = 0.530$$

L = Large value

S = Small "

d = difference

Frequency
Distribution



Frequency Distribution:

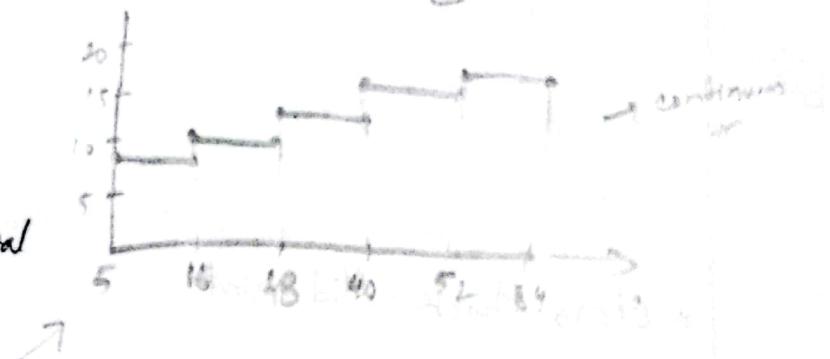
5, 8, 9, 9, 10, 11, 12, 16, 20, 25, 30, 31, 32, 40, 41, 42, 50, 51, 53, 60.

CF

⇒

$$R = 60 - 5 = 55$$

$$\frac{55}{5} = 11 \rightarrow \text{Class interval}$$



Class Interval	Frequency	Frequency	CF	Percentage (%)
5-16	III	8	8	8/20 = 40%
17-28	II	2	10	10/20 = 50%
29-40	III	4	14	14/20 = 70%
41-52	III	9	18	18/20 = 90%
53-64	II	2	20	20/20 = 100%

Qualitative:

Cat, Cat, Cat, Cat, Cat, Hen, Hen.

Dog, Dog, Dog, Dog, Dog, Large, Large, Large, Large, Large, Small, Small,

Small

Relative frequency

Class	Frequency	Frequency	CF	Percentage (%)
Cat	III	5	5	5/28
Hen	III	10	15	10/28
Dog	III	5	20	5/28
Large	III	5	25	5/28
Small	III	3	28	3/28



Cumulative Frequency. 2 types: 1. <

$$\begin{array}{l} \leq 16 \rightarrow 8 \text{ Q.F} \\ \leq 28 \rightarrow 10 \text{ Q.F} \\ \leq 40 \rightarrow 14 \text{ Q.F} \end{array} \left. \begin{array}{l} \leq \\ \leq \\ \leq \end{array} \right\} \rightarrow \text{Less than C.Graph}$$

Upper Limit

$$\begin{array}{l} \geq 5 \rightarrow 20 \\ \geq 17 \rightarrow 12 \\ \geq 29 \rightarrow 10 \end{array} \left. \begin{array}{l} \geq \\ \geq \\ \geq \end{array} \right\} \rightarrow \begin{array}{l} \text{More} \\ \text{greater} \end{array} \text{ than C.Graph}$$

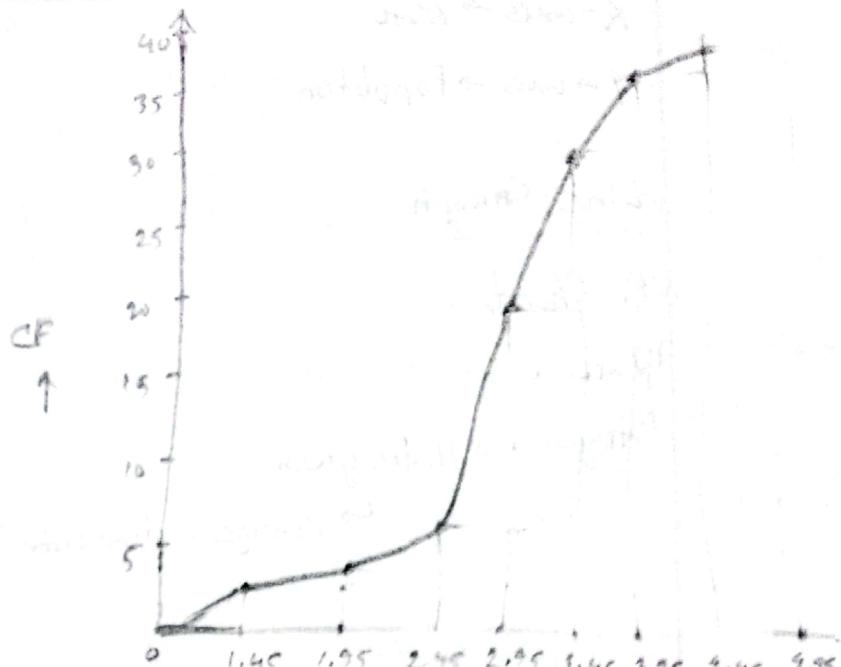
Lower Limit

Life length	Num of rats	CF	DF
1.45-1.95	2	2	40
1.95-2.45	1	3	37
2.45-2.95	4	7	32
2.95-3.45	15	22	22
3.45-3.95	10	32	7
3.95-4.45	5	37	3
4.45-4.95	3	40	2
	90		

Less than CF:

(a)

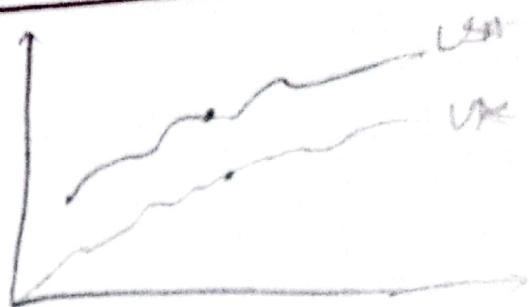
Class	CF
Less than 1.45	0
1.45 - 1.95	2
1.95 - 2.45	3
2.45 - 2.95	7
2.95 - 3.45	22
3.45 - 3.95	32
3.95 - 4.45	37
4.45 - 4.95	40



Less than Adv.

(b) More than CF:

X-axis \rightarrow Year
Y-axis \rightarrow Population

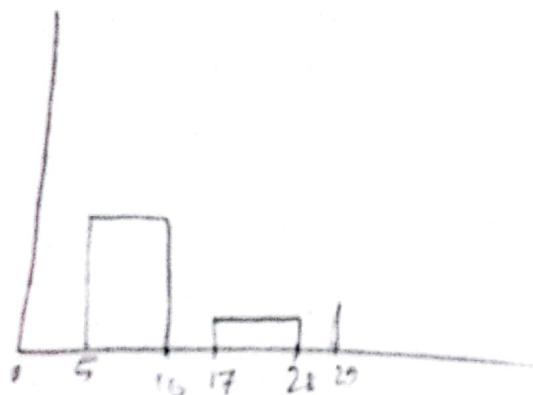


Line Graph:

* P. Chart,

* Bar \rightarrow Categorical

* Histogram * Histogram \rightarrow Discretized from CONTINUOUS
↳ Contiguous, discrete.



\rightarrow Discrete

Mean
Weight 25

6.25

1

5.25

2.5

Mean

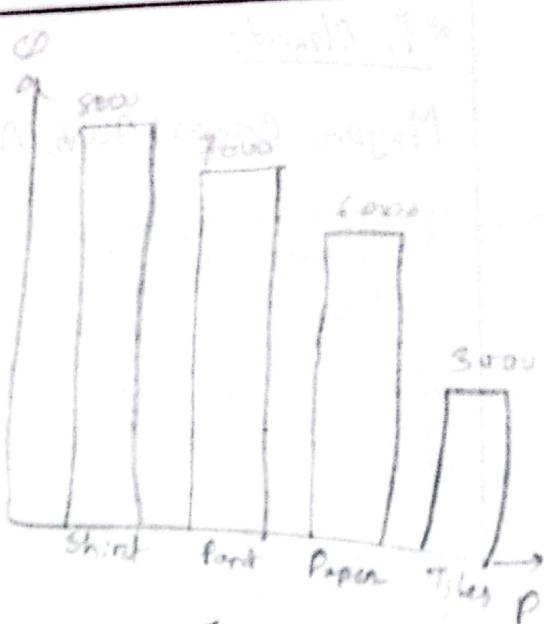
Weight

25

Contiguous or Categorical Bar Diagram

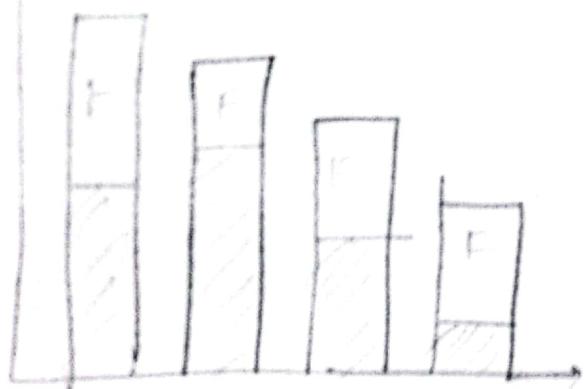
→ Categorical

Products	Quantity	Ratio (Shirt : Pant : Paper : Tiles)
Shirt	5000	4000 : 4000
Pant	7000	3000 : 3000
Paper	6000	3000 : 3000
Tiles	3200	1000 : 2000

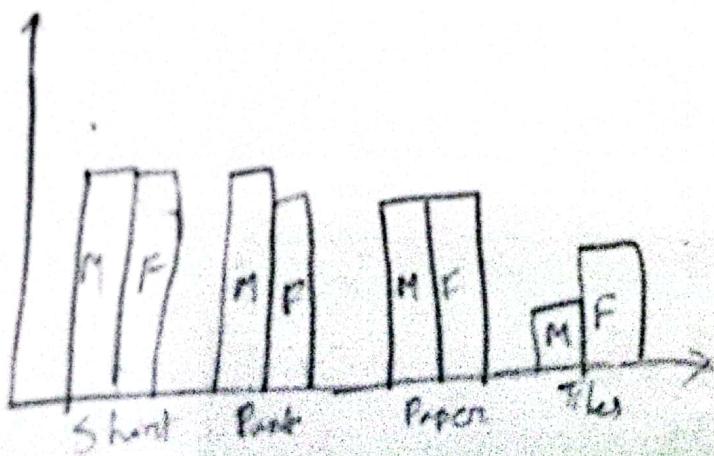


→ Bar Diagram

↳ Component
Multiple



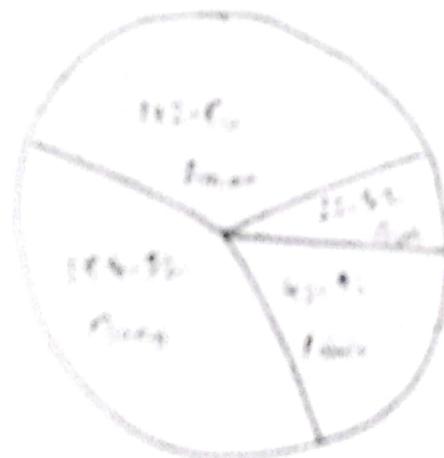
Multiple Bar Diagram:



4 P. Chart

Major Crops And Area, Hectare

Major Crops	Area (Hectare)	Rate (Rs/ha)	Value
Rice	18	115/100 = 0.115	2.070 236.279.33
Maize	5	115/100 = 0.115	0.575 236.279.33
Groundnut	1.5	115/100 = 0.115	0.173 236.279.33
Pulses	32	115/100 = 0.115	3.680 236.279.33
Total	200		



Ground Land 100%

Stem and Leaf Plot:

↳ Data nature (5.25)

84, 17, 38, 45, 47, 53, 76, 54, 75, 22, 66, 65, 55, 54, 51, 33, 39, 19, 54, 72

order order 200s 100s 20s 10s 2s 1s 0.5s 0.1s 0.05s 0.01s

Stem	Leaf
1	7 9
2	2
3	8 3 9
4	5 7
5	3 4 5 4 4
6	6 5
7	6 5 2
8	4



* Leaf 29 distribution (5.25) shape symmetric

Bivariate table:

↳ 2 categorical variables → Relation (5.25) → 2x2 categorical

		Economic Status	
		P	R
Gender	M	2	1
	F	1	2
Gender	M	2	1
	F	1	2
		P	R
Gender	M	2	1
	F	1	2
		P	R
Gender	M	2	1
	F	1	2
		P	R
Gender	M	2	1
	F	1	2

Gender	ES		
	R	M	P
M	2	1	2
F	2	2	1
M	4	3	3
F	2	2	1

$$P(F|R) = \frac{2}{4}$$

$$P(R|F) = \frac{2}{5}$$

* Possible Questions:

1. Data कि? Example, PICTURE
2. Scale, Variable कि? Example
3. Frequency, distribution \rightarrow table draw, histogram draw,
4. Class कि, Types, frequency, Class Limit / Boundary.
5. किसी Frequency table draw कैसे?

 - ↳ Cumulative,

6. Bar diagram:
 - ↳ Multiple
 - ↳ Stack
 - ↳ Component
7. Bar diagram & Histogram diff.

Page - 83
Exercise - 2
Date - 1/1/2022

Key:

Chi-square:

Association between 2 categorical variable

Central tendency \rightarrow Ch-3

Measure of central tendency

↳ Mean, Median, Mode

↳ Center as a Point of Measure

* Arithmetic mean क्या है?

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Page 21

Ans

grouped \rightarrow class interval 0-100, 100-200, ...
ungrouped \rightarrow 1, 2, 3, ...

Age of marriage (x)	Frequency (f)	Eff.
11	17	187
12	28	336
13	37	481
14	52	728
20	8	160
	Σf_i	200

~~$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$~~ \rightarrow ungrouped

~~$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$~~ \rightarrow grouped

Class Interval	Frequency	Mid-point (m)	Eff.
17-20	17	$\frac{17+20}{2}$	
28-36	28	32	
52-60	52	56	

~~$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$~~ \rightarrow grouped

Q. 7. $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$ when values are not known
class interval by middle point of the

H.W.

* Pooled mean

\rightarrow 2 or group by Mean combine

Pooled Mean

↳ Merge M&G

↳ 2 Grp group for combined Mean

M	F
5.1	5.0
5.6	5.1
5.8	5.9
5.9	5.8
6.0	5.9
6.1	6.0
6.2	6.1
6.3	6.2
6.4	6.3
6.5	6.4
6.6	6.5
6.7	6.6
6.8	6.7
6.9	6.8
7.0	6.9
7.1	7.0
7.2	7.1
7.3	7.2
7.4	7.3
7.5	7.4
7.6	7.5
7.7	7.6
7.8	7.7
7.9	7.8
8.0	7.9
8.1	8.0
8.2	8.1
8.3	8.2
8.4	8.3
8.5	8.4
8.6	8.5
8.7	8.6
8.8	8.7
8.9	8.8
9.0	8.9
9.1	9.0
9.2	9.1
9.3	9.2
9.4	9.3
9.5	9.4
9.6	9.5
9.7	9.6
9.8	9.7
9.9	9.8
10.0	9.9
10.1	10.0
10.2	10.1
10.3	10.2
10.4	10.3
10.5	10.4
10.6	10.5
10.7	10.6
10.8	10.7
10.9	10.8
11.0	10.9
11.1	11.0
11.2	11.1
11.3	11.2
11.4	11.3
11.5	11.4
11.6	11.5
11.7	11.6
11.8	11.7
11.9	11.8
12.0	11.9
12.1	12.0
12.2	12.1
12.3	12.2
12.4	12.3
12.5	12.4
12.6	12.5
12.7	12.6
12.8	12.7
12.9	12.8
13.0	12.9
13.1	13.0
13.2	13.1
13.3	13.2
13.4	13.3
13.5	13.4
13.6	13.5
13.7	13.6
13.8	13.7
13.9	13.8
14.0	13.9
14.1	14.0
14.2	14.1
14.3	14.2
14.4	14.3
14.5	14.4
14.6	14.5
14.7	14.6
14.8	14.7
14.9	14.8
15.0	14.9
15.1	15.0
15.2	15.1
15.3	15.2
15.4	15.3
15.5	15.4
15.6	15.5
15.7	15.6
15.8	15.7
15.9	15.8
16.0	15.9
16.1	16.0
16.2	16.1
16.3	16.2
16.4	16.3
16.5	16.4
16.6	16.5
16.7	16.6
16.8	16.7
16.9	16.8
17.0	16.9
17.1	17.0
17.2	17.1
17.3	17.2
17.4	17.3
17.5	17.4
17.6	17.5
17.7	17.6
17.8	17.7
17.9	17.8
18.0	17.9
18.1	18.0
18.2	18.1
18.3	18.2
18.4	18.3
18.5	18.4
18.6	18.5
18.7	18.6
18.8	18.7
18.9	18.8
19.0	18.9
19.1	19.0
19.2	19.1
19.3	19.2
19.4	19.3
19.5	19.4
19.6	19.5
19.7	19.6
19.8	19.7
19.9	19.8
20.0	19.9
20.1	20.0
20.2	20.1
20.3	20.2
20.4	20.3
20.5	20.4
20.6	20.5
20.7	20.6
20.8	20.7
20.9	20.8
21.0	20.9
21.1	21.0
21.2	21.1
21.3	21.2
21.4	21.3
21.5	21.4
21.6	21.5
21.7	21.6
21.8	21.7
21.9	21.8
22.0	21.9
22.1	22.0
22.2	22.1
22.3	22.2
22.4	22.3
22.5	22.4
22.6	22.5
22.7	22.6
22.8	22.7
22.9	22.8
23.0	22.9
23.1	23.0
23.2	23.1
23.3	23.2
23.4	23.3
23.5	23.4
23.6	23.5
23.7	23.6
23.8	23.7
23.9	23.8
24.0	23.9
24.1	24.0
24.2	24.1
24.3	24.2
24.4	24.3
24.5	24.4
24.6	24.5
24.7	24.6
24.8	24.7
24.9	24.8
25.0	24.9
25.1	25.0
25.2	25.1
25.3	25.2
25.4	25.3
25.5	25.4
25.6	25.5
25.7	25.6
25.8	25.7
25.9	25.8
26.0	25.9
26.1	26.0
26.2	26.1
26.3	26.2
26.4	26.3
26.5	26.4
26.6	26.5
26.7	26.6
26.8	26.7
26.9	26.8
27.0	26.9
27.1	27.0
27.2	27.1
27.3	27.2
27.4	27.3
27.5	27.4
27.6	27.5
27.7	27.6
27.8	27.7
27.9	27.8
28.0	27.9
28.1	28.0
28.2	28.1
28.3	28.2
28.4	28.3
28.5	28.4
28.6	28.5
28.7	28.6
28.8	28.7
28.9	28.8
29.0	28.9
29.1	29.0
29.2	29.1
29.3	29.2
29.4	29.3
29.5	29.4
29.6	29.5
29.7	29.6
29.8	29.7
29.9	29.8
30.0	29.9
30.1	30.0
30.2	30.1
30.3	30.2
30.4	30.3
30.5	30.4
30.6	30.5
30.7	30.6
30.8	30.7
30.9	30.8
31.0	30.9
31.1	31.0
31.2	31.1
31.3	31.2
31.4	31.3
31.5	31.4
31.6	31.5
31.7	31.6
31.8	31.7
31.9	31.8
32.0	31.9
32.1	32.0
32.2	32.1
32.3	32.2
32.4	32.3
32.5	32.4
32.6	32.5
32.7	32.6
32.8	32.7
32.9	32.8
33.0	32.9
33.1	33.0
33.2	33.1
33.3	33.2
33.4	33.3
33.5	33.4
33.6	33.5
33.7	33.6
33.8	33.7
33.9	33.8
34.0	33.9
34.1	34.0
34.2	34.1
34.3	34.2
34.4	34.3
34.5	34.4
34.6	34.5
34.7	34.6
34.8	34.7
34.9	34.8
35.0	34.9
35.1	35.0
35.2	35.1
35.3	35.2
35.4	35.3
35.5	35.4
35.6	35.5
35.7	35.6
35.8	35.7
35.9	35.8
36.0	35.9
36.1	36.0
36.2	36.1
36.3	36.2
36.4	36.3
36.5	36.4
36.6	36.5
36.7	36.6
36.8	36.7
36.9	36.8
37.0	36.9
37.1	37.0
37.2	37.1
37.3	37.2
37.4	37.3
37.5	37.4
37.6	37.5
37.7	37.6
37.8	37.7
37.9	37.8
38.0	37.9
38.1	38.0
38.2	38.1
38.3	38.2
38.4	38.3
38.5	38.4
38.6	38.5
38.7	38.6
38.8	38.7
38.9	38.8
39.0	38.9
39.1	39.0
39.2	39.1
39.3	39.2
39.4	39.3
39.5	39.4
39.6	39.5
39.7	39.6
39.8	39.7
39.9	39.8
40.0	39.9
40.1	40.0
40.2	40.1
40.3	40.2
40.4	40.3
40.5	40.4
40.6	40.5
40.7	40.6
40.8	40.7
40.9	40.8
41.0	40.9
41.1	41.0
41.2	41.1
41.3	41.2
41.4	41.3
41.5	41.4
41.6	41.5
41.7	41.6
41.8	41.7
41.9	41.8
42.0	41.9
42.1	42.0
42.2	42.1
42.3	42.2
42.4	42.3
42.5	42.4
42.6	42.5
42.7	42.6
42.8	42.7
42.9	42.8
43.0	42.9
43.1	43.0
43.2	43.1
43.3	43.2
43.4	43.3
43.5	43.4
43.6	43.5
43.7	43.6
43.8	43.7
43.9	43.8
44.0	43.9
44.1	44.0
44.2	44.1
44.3	44.2
44.4	44.3
44.5	44.4
44.6	44.5
44.7	44.6
44.8	44.7
44.9	44.8
45.0	44.9
45.1	45.0
45.2	45.1
45.3	45.2
45.4	45.3
45.5	45.4
45.6	45.5
45.7	45.6
45.8	45.7
45.9	45.8
46.0	45.9
46.1	46.0
46.2	46.1
46.3	46.2
46.4	46.3
46.5	46.4
46.6	46.5
46.7	46.6
46.8	46.7
46.9	46.8
47.0	46.9
47.1	47.0
47.2	47.1
47.3	47.2
47.4	47.3
47.5	47.4
47.6	47.5
47.7	47.6
47.8	47.7
47.9	47.8
48.0	47.9
48.1	48.0
48.2	48.1
48.3	48.2
48.4	48.3
48.5	48.4
48.6	48.5
48.7	48.6
48.8	48.7
48.9	48.8
49.0	48.9
49.1	49.0
49.2	49.1
49.3	49.2
49.4	49.3
49.5	49.4
49.6	49.5
49.7	49.6
49.8	49.7
49.9	49.8
50.0	49.9
50.1	50.0
50.2	50.1
50.3	50.2
50.4	50.3
50.5	50.4
50.6	50.5
50.7	50.6
50.8	50.7
50.9	50.8
51.0	50.9
51.1	51.0
51.2	51.1
51.3	51.2
51.4	51.3
51.5	51.4
51.6	51.5
51.7	51.6
51.8	51.7
51.9	51.8
52.0	51.9
52.1	52.0
52.2	52.1
52.3	52.2
52.4	52.3
52.5	52.4
52.6	52.5
52.7	52.6
52.8	52.7
52.9	52.8
53.0	52.9
53.1	53.0
53.2	53.1
53.3	53.2
53.4	53.3
53.5	53.4
53.6	53.5
53.7	53.6
53.8	53.7
53.9	53.8
54.0	53.9
54.1	54.0
54.2	54.1
54.3	54.2
54.4	54.3
54.5	54.4
54.6	54.5
54.7	54.6
54.8	54.7
54.9	54.8
55.0</td	

* 1, 2, 7, 12, 17, 19, 21, 34 \rightarrow Ungrouped Data

$$\frac{12+17}{2} = 14.5$$

Grouped Data:

$$\tilde{m} = l_m + \frac{h}{f_m} \left(\frac{n}{2} - F_{m-1} \right)$$

l_m = Median class $\&$ Lower Limit

h = difference

f_m = Median $\&$ Frequency

F_{m-1} = Median $\&$ Σf_{m-1} CF

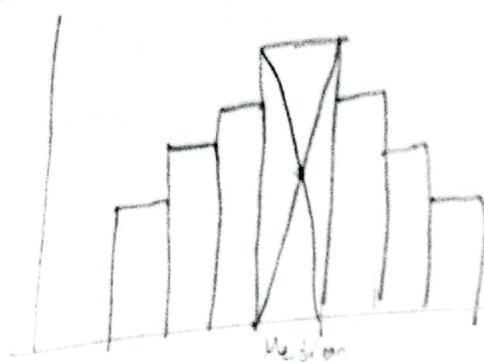
Age	f	CF
24.5-29.5	3	3
29.5-34.5	9	12
34.5-39.5	15	27
39.5-44.5	12	39
44.5-49.5	7	46
49.5-54.5	4	50
$n=50$		

$$\tilde{m} = 34.5 + \frac{5}{15} \left(\frac{50}{2} - 12 \right)$$

~~243.33~~

$$= 38.83$$

Histogram \rightarrow Normal Distribution:



Quartile:

$\frac{1}{4} \rightarrow n = 451.7$ Divisible by 25,

$$Q_1 = \frac{\frac{n}{4}th + (\frac{n}{4} + 1)th}{2} \rightarrow 25$$

$Q_2 = \text{Median} \rightarrow 50$

$$\frac{3}{4} \rightarrow Q_3 = \frac{3n}{4}th + (\frac{3n}{4} + 1)th$$

$n = 451.7$ Divisible by 25,

$\frac{15}{4} \rightarrow Q_1 = \text{Next higher integer}$

$\approx 4 \rightarrow Q_1$

$Q_2 = \text{Median}$

$Q_3 = \frac{3n}{4} \rightarrow \text{Next higher integer}$

$$= \frac{3 \times 15}{4} = 11.25 \approx 12 \leftarrow Q_3$$

19, 17, 19, 23, 27, 32, 40, 49, 54, 59, 71, 80,

$$\rightarrow n = 12$$

$$Q_1 = \frac{19+23}{2} = 21$$

~~$$Q_3 = \frac{54+59}{2} = 56.5$$~~

$$Q_1 = \frac{19+23}{2} = 21$$

$$Q_3 = \frac{54+59}{2} = 56.5$$

$$Q_2 = \frac{32+40}{2} = 36$$

$$Q_4 = 80$$

$$Q_{12} = l_n + \frac{h}{f_n} \left(-\frac{rn}{4} - F_{(n)-1} \right)$$

$$Q_1 = l_1 + \frac{h}{f_3} \left(-\frac{rn}{4} - F_{(3)-1} \right)$$

$$= 34.5 + \frac{5}{15} \left(-\frac{60}{4} - 12 \right)$$

$$= 34.67$$

$$\frac{50}{4} = 12.5$$

↳ 3rd class

$$\frac{3 \times 50}{4} = 37.5$$

↳ 4th class

$$Q_3 = 37.5 + \frac{5}{12} \left(-\frac{50}{4} - 27 \right)$$

$$= 43.875$$

$$\left(\frac{10}{(A+D)} \right) A + \frac{10}{(A+D)} D$$

Percentile:

$$P_p = l_p + \frac{h}{f_p} \left(\frac{p n}{100} - F_{(p)-1} \right)$$

$$P_1 = 29.5 + \frac{5}{9} \left(\frac{12 \times 50}{100} - 3 \right)$$

$$= 31.16$$

121. (Q5 ans)

P

$$\frac{p n}{100}$$

$$= \frac{12 \times 50}{100} = 6$$

$$\approx 31.16$$

75%

$$\frac{75 \times 50}{100} = 37.5$$

≈ 38

\hookrightarrow 4th class

Deciles:

$$D_1 = l_1 + \frac{h}{f_1} \left(\frac{D n}{10} - F_{(1)-1} \right)$$

7

Mode:

$$M_o = l_o + h \left(\frac{f_1 - f_2}{f_1 + f_2} \right)$$

l_o = Lower limit of the Modal Class

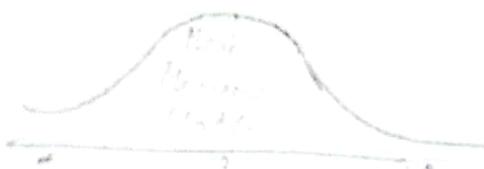
h = height of Modal class

$$M_o = 15$$

$$M_o = 34.5 + 5$$

$$M_o = 34.5 + 5 \left[\frac{(15-9)}{(15-9) + (15-12)} \right] = 37.83$$

Symmetric distribution \rightarrow Mean, Median, Mode \rightarrow same



Abnormal distribution \rightarrow Mean, Median, Mode 'same' \rightarrow

Right skewed:



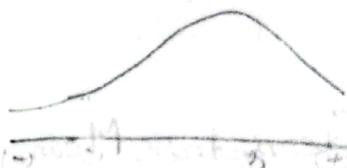
Mean $\rightarrow 318.625$ 27.5

Median $\rightarrow 277.5$

Mode \rightarrow Left side $\rightarrow 318.625$
 35 point

\rightarrow highest frequency

Left skewed



Mean $\rightarrow 218.625$ 157.5

Median $\rightarrow 217.5$

Mode \rightarrow Right side $\rightarrow 218.625$
 35 point

\rightarrow highest frequency

H.W.

- Arithmetic mean depends on both origin and scale
- Variance \rightarrow only scale but not origin

* C.G. 14/12

→ 5, 9, 8, 18, 10, 8, 12, 9

$$\rightarrow \frac{5}{7}, \frac{9}{7}, \frac{8}{7}, \frac{18}{7}, \frac{10}{7}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

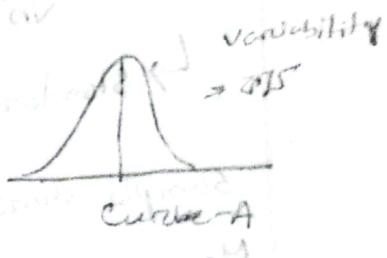
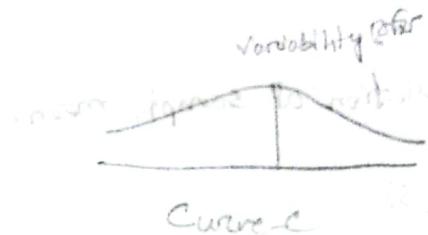
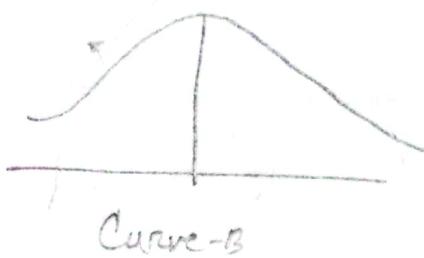


Geometric Mean, harmonic Mean

* Truncation, Truncation, Truncated Mean

Truncated

Disportion!



* Mean 1270 after observation 20 20

* Measure of Dispersion:

1. Variance
 2. Range \rightarrow upper, 1
 3. Inter quartile
 4. Mean Deviation:

↳ Observation (25) Mean \approx 14.75

$$M_p(a) = \frac{\sum |x_i - a|}{n}$$

$$M_D(\tilde{m}) = \frac{\sum |x_i - \tilde{m}|}{n} \quad \tilde{m} = \text{Median}$$

$$M_D \text{ (mode)} = \frac{\sum |x_i - M_0|}{n}$$

group data \rightarrow

$$\frac{\sum |f(x_i) - a|}{n}$$

→ table - 4. 2

$s^2 \rightarrow \sqrt{s^2} = s \rightarrow$ standard deviation

standard error: $\frac{\text{standard deviation}}{\sqrt{n}}$

$$SD = \sqrt{\text{Var}}$$

$$SE = \frac{SD}{\sqrt{n}}$$

↳ Standard deviation of sample mean.

Sample variance (S^2)

Mean \Rightarrow Mean or Standard deviation

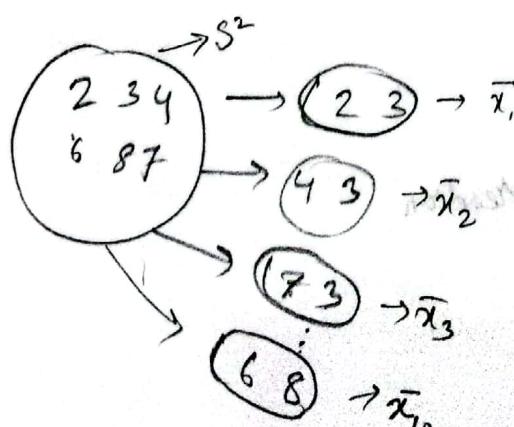
Variance of Mean

Sample variance deviation (SD) = $\sqrt{S^2}$

$\rightarrow M_1, S_1$ } Mean \Rightarrow Mean

$\rightarrow M_2, S_2$ } Mean (2nd mean) \Rightarrow Mean (1st mean diff.)

$$\frac{\sum (x_i - \bar{x})^2}{n}$$



$$\frac{(2 - 5)^2 + (3 - 5)^2}{2} = (2.5) \text{ aM}$$

$$\frac{(4 - 5)^2 + (3 - 5)^2}{2} = (2.5) \text{ aM}$$

$$\frac{(7 - 5)^2 + (3 - 5)^2}{2} = (8.0) \text{ aM}$$

* Variance \Rightarrow Property

* Co-efficient of variation: $CV = \frac{SD}{\bar{x}}$

↳ Adv. \rightarrow 2 Br variable w.r.t Mean

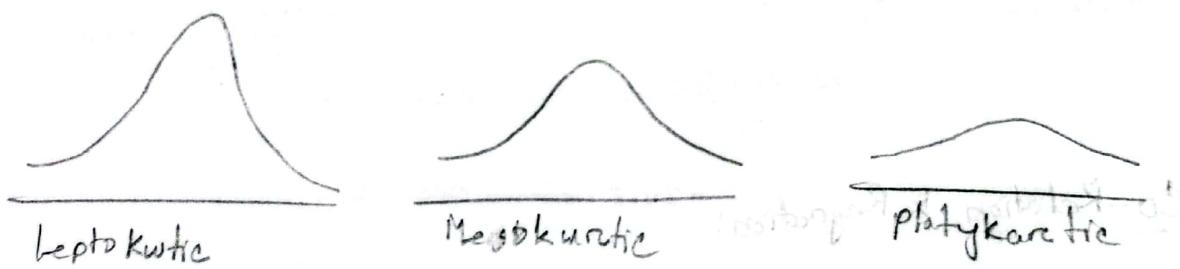
↳ 232 Page

Skewness, Quartiles

↳ 237

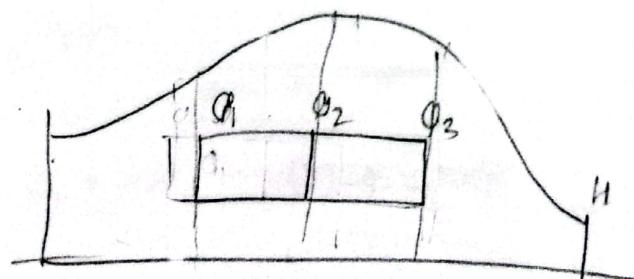


Quartiles: picked 3rd & 4th



Box plot: (Math)

Smallest measurement, Q_1, Q_2, Q_3 , highest measurement



Inner Fench:

$$g_i \pm 1.5 (10^6)$$

Inner Fench ≈ 2720 out layer

Outer Fench:

$$g_i \pm 3 (10^6)$$

Outer Fench ≈ 2720 extreme value 225.

20-3-24

Lea-

ch-5

X/2017 Data & centre free measure \rightarrow Central tendency

* Data variability measure \rightarrow Dispersion

Co-Relation & Regression \rightarrow 2 variable relationship

Co-Relation:

$$\text{Corr}(x, y)$$

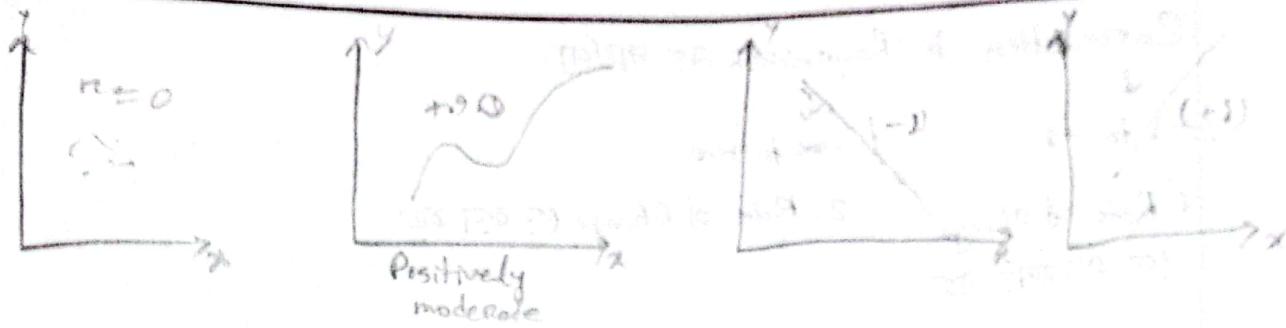
$$\frac{x}{y}$$

$$-1 \leq \text{Corr}(x, y) \leq 1$$

$$\left. \begin{array}{l} x(\text{Steady hours}) \\ y(\text{cgpa}) \end{array} \right\}$$

Range (-1, 1)

Middle $\rightarrow 0$



$$\text{Correlation} = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)} \sqrt{\text{Var}(y)}}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

$$= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

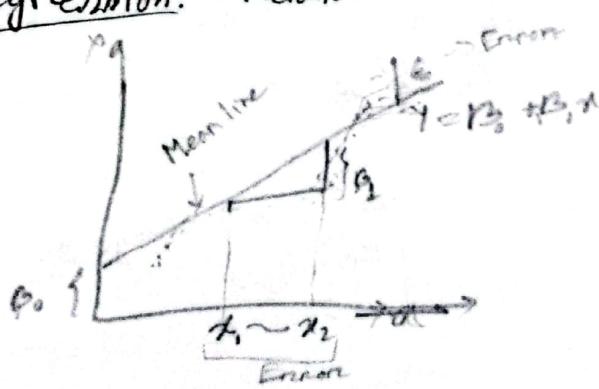
285 page 223

287

5.7

Sales \rightarrow Math

Regression: \rightarrow Relation is linear $\rightarrow -\infty$ to $+\infty$; B_1 = rate of change



$$Y = m x + c$$

$$m = \text{rate}$$

c = constant

$$B_1 = \frac{\Delta Y}{\Delta X} = \text{rate of change}$$

$B_1 \rightarrow +ve \rightarrow$ increase in x \rightarrow increase in y
 $\rightarrow -ve \rightarrow$ increase in x \rightarrow decrease in y

$x=0 \rightarrow$ intercept

BM₂
 $y = c - B_1 x$
 Help
 Exam
 70% CGPA

Correlation & Regression \Rightarrow 9.12 (a)1. $-1 \rightarrow +1$ 2. $-\infty \rightarrow +\infty$ 1. Rate of Change
(\approx constant) \Rightarrow 2. Rate of Change (\approx 0.07 2025)

Dependence

$$E(Y/x) = \beta_0 + \beta_1 x$$

$$Y = \beta_0 + \beta_1 x + \epsilon \rightarrow \text{Regression Line}$$

↳ Error

$$E(Y/x) = \hat{\beta}_0 + \hat{\beta}_1 x \rightarrow ; E(\epsilon/x) = 0 - \hat{\beta}_1 x$$

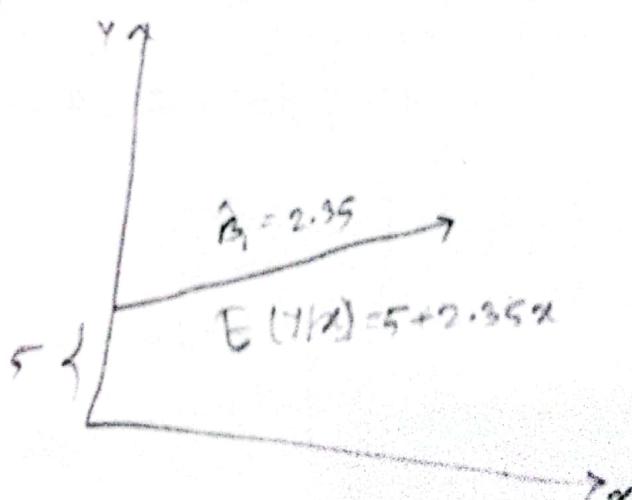
↳ Mean \approx 0 (0.07) for 0 error↳ ϵ - Error term $0 \rightarrow$ ~~big~~ Error term

$$\frac{0 - \text{big}}{\text{Error}} = +\infty$$

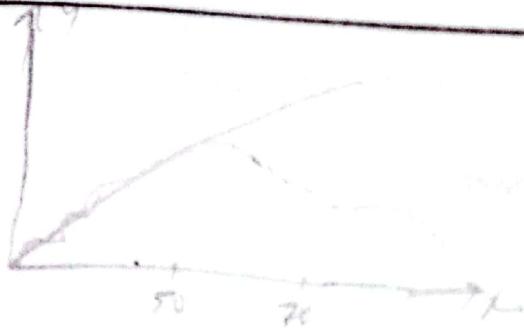
Error = 0

x (Height) cm	y (Age) Year
100	10
110	12
120	14
130	16
140	18
150	20

$$\text{Mean age} \\ \beta_0 = 5 \rightarrow x=0; \\ \hat{\beta}_1 = 2.35$$



With 1 unit change \Rightarrow
Mean Height
2.35 increase with 1 year age
TBS



Mean line or conditional distribution

Conditional distribution of Y given $X=x$ is $N(\mu_x, \sigma_x^2)$

Conditional distribution of X given $Y=y$ is $N(\mu_y, \sigma_y^2)$



* Regression Properties \rightarrow 5.2.1

1. Errors ε_i are uncorrelated

ε_i independent

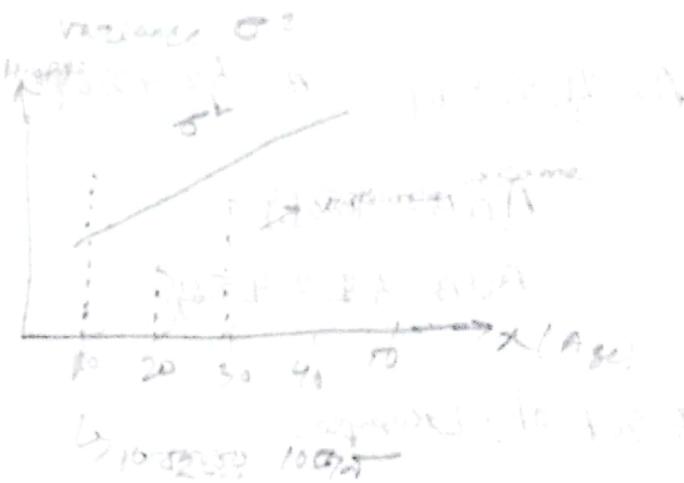
2. Constant variance \rightarrow homoscedasticity of errors in the regression function

3. Y is Distribution Normal

1. $Y \sim N(\mu, \sigma^2)$

2. $\text{Corr}(\varepsilon_i, \varepsilon_j) = 0$

3. $\varepsilon_i \sim N(0, \sigma^2)$



Example with a simple LR

Assume model \rightarrow regression

Explanatory

Y

Random error term

Explanatory variable

Response variable

Random error term

Explanatory variable

Response variable

Random error term

Explanatory variable

Response variable

Large slope

Small slope

Large error

Small error

Probability:

Every event outcome has its own chance
↳ Exact event value.

Possibility \rightarrow Yes/No

↳ True

Set \rightarrow elements / Members

↳ Discrete value

Subset, Intersection, Union, Complementation, Power

$$A = \{1, 3, 5, 6\} \quad B = \{2, 4, 5, 6\}$$

$$A \cap B = \{5, 6\}$$

$$A \cup B = \{1, 2, 4, 5, 6\}$$

* Set \emptyset ? Example,

Experiment, Random Experiment

↳ Every event
outcome has
chance

↳ Random variable

↳ Possible sample states for randomly
chance

Sample space:

Event

Discrete & continuous Sample space:

↳ Categorical

↳ Rainfall range answer

Classical Approach, Relative Frequency Approach, Subjective Approach

$$P(A) = \frac{n(A)}{n(S)}$$

→ If event A is favorable
to event S

Sample set

event S

number of favorable outcomes

Ex-22: 10 balls \rightarrow 4 white, 6 red, find Possibility?

$$4/10 \rightarrow \frac{4}{10}, \text{ red} = \frac{6}{10}$$

Conditional Probability

$$A \cap B \rightarrow P(A) \cdot P(A|B), P(B)$$

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

↳ Dependent

$$\text{Independent} \rightarrow P(A|B) = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

7.37

$$P(A|B) = ?$$

$$A = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}$$

$$B = \{(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{6/36} = \frac{1}{3}$$

F.38: Bayes Theorem

↳ Prior, posterior

$$\text{Posterior} = \text{Data} \times \text{Prior}$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) P(A)}{P(B)}$$

* Machine A & B, output \rightarrow 100% 20% 40% B, 60% A.

1. A 25% produce 3% defective

2. B 40% produce 5% defective

$$P(A) = 0.60$$

$$P(B) = 0.40$$

$$P(D|A) = 0.03$$

$$P(D|B) = 0.05 \rightarrow P(D) = \frac{P(D|A)P(A) + P(D|B)P(B)}{P(A) + P(B)} = 0.045 \rightarrow \text{Probability of D}$$

↳ 1. Prior, Posterior

2. Marginal Probability

3. Posterior

$$P(A|D) = ?$$

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.03 \times 0.60}{0.045} = 0.40$$

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \frac{0.05 \times 0.40}{0.045} = 0.44$$

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.03 \times 0.60}{0.045} = 0.40$$