28-2-24

Lec-1

Introduction of Statistics & Probability

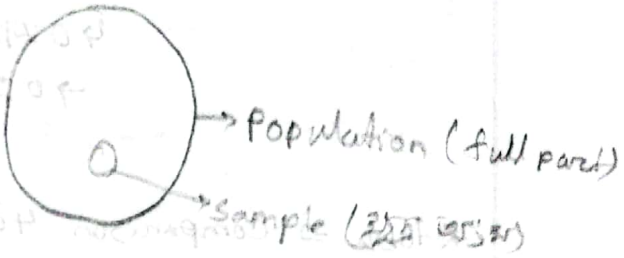
→ M. Nurul Islam

Ch-2

Statistics

↳ Senses, Sample

↳ Population (all possible sample)



→ Data

1. Quantitive (Hight, weight)
2. ~~Qualitive~~ → Numeric, contigous
2. Qualatitive (catagorical)

↳ gender → male, female

↳ status → yes, no

↳ economic status

MACS, Stapo scores

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

* Level of Measurement:

↳ Neumerical

* Nominal Level of Measurement: Color,

↳ যেগুলো order করা যায় না, শুধু নাম দ্বারা expression করে,
যেমন: বিদ্যালয় " " "

* Ordinal Level of Measurement: Health status, job rank

↳ Order/Rank করা যায়

$>, <, =, \leq, \geq$ " "

* Interval Level of Measurement: IQ test score.

↳ Interval ↳ আগানো যায় → 2, 4, 6, 8 ...

↳ Multiply / divide করা যায় না → 0 means 0 না

↳ temparature ↗ IQ = 0 Ex

* Ratio Level of Measurement : Hight, weight

↳ Interval ↳ ছিল না.

↳ 0 মানে 0 ✗

→ 0 ✓ Meaning আছে

* table ↳ comparison এর Level ↳ → ↳

* Variable & Attribute:

↳ যেটা vary করতে পারে

↳ Quantitive → Neumerical

Qualatitive

*Attribute : সেটা Qualative variable ↳ distinct category

*Discreate

* Contigious

* Constant

Ordinal
Categorical
Nominal
Interval

* Summarizing and ~~Com~~ presenting dedta:

1. Mean
2. Median → Main influenced by extreme value
3. Mood
Mode → Middle point নিয়ে কাজ করে
↳ কোনটা বেশি পরিমাণে আছে

Ques → 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

⇒

Interval	Frequency	Tally	Cumulative Frequency	Probability/Perce
5.00 - 5.5	15		15	15/25 =
5.6 - 5.10	8		23	8/25 =
5.11 - 6.00	2		25	2/25 =
	25			

$P(X \approx 5.62) = \frac{8}{25}$

Lower class Limit

Upper " "

* Class Mark = Mid point
* Instant steps

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

Frequency Distribution

L = Large value

S = Small "

d = difference

Data
- Quantitative
  - Age, Height, weight → contiguous
  - Number, CGPA → Discrete
    - → Binary
- Qualitative
  - gender, name, categorical → Nominal

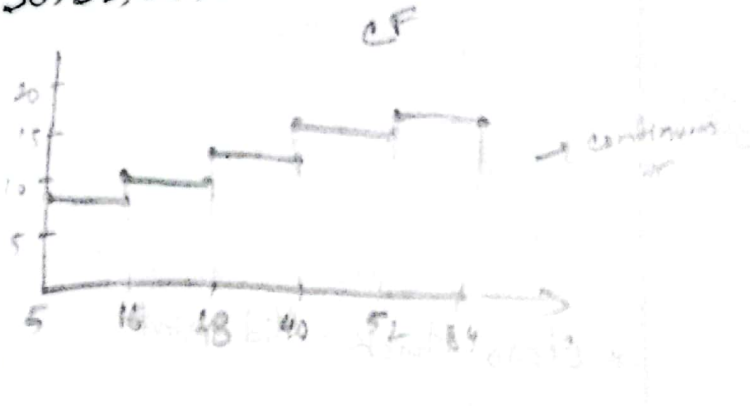
Frequency Distribution:

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

→

$R = 60 - 5 = 55$

$\frac{55}{5} = 11$ → Class interval

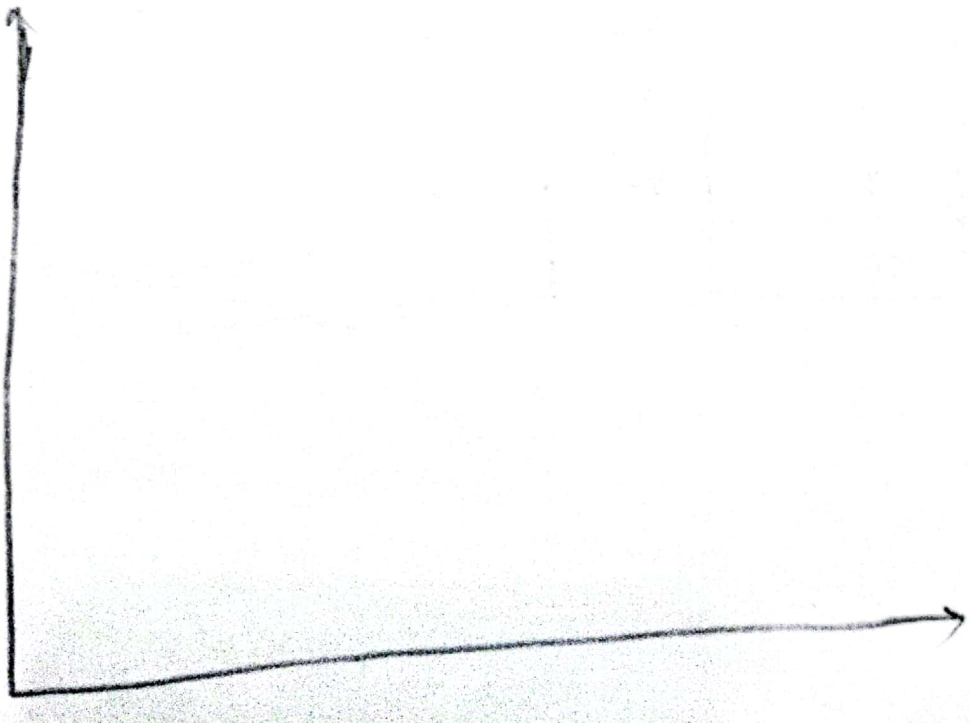


Class Interval	Frequency	Frequency	CF	Percentage (%)
5 - 16	𝍸 III	8	8	8/20 =
17 - 28	II	2	10	2/20 =
29 - 40	𝍸	4	14	4/20 =
41 - 52	𝍸 I	4	18	4/20 =
53 - 64	II	2	20	2/20 =

Qualitative :

Cat, Cat, Cat, Cat, Cat, Hen, Hen, Hen, Hen, Hen, Hen, Hen, Hen, Hen, Hen, Hen, Dog, Dog, Dog, Dog, Dog, Large, Large, Large, Large, Large, Small, Small, Small

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



Cumulative Frequency · 2 types : ↗, ↘

≤16 → 8 জন
≤28 → 10 জন
≤40 → 14 জন
→ Less than C Graph
↳ upper Limit

≥5 → 20
≥17 → 12
≥29 → 10
→ More than C Graph
↳ Lower Limit

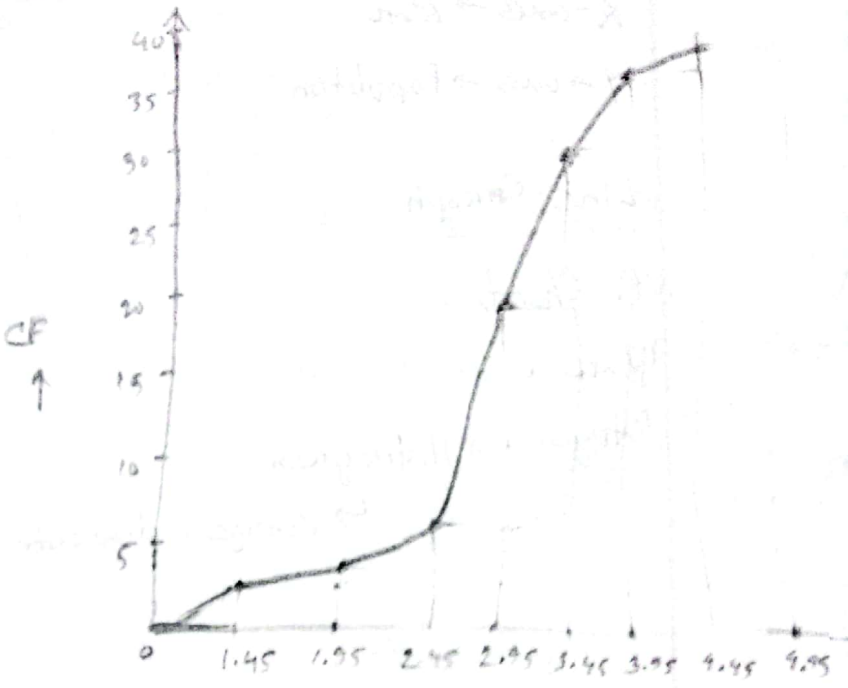
↗ Descending

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	33
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
	40		

Less than CF:

(a)

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

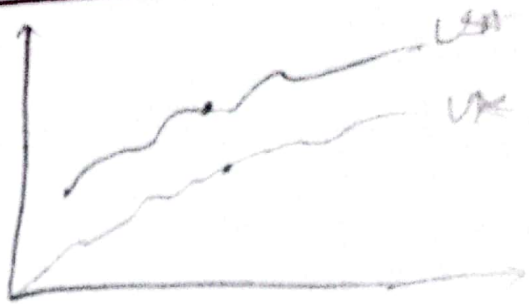


Cont. of Adv.

(b) More than CF:

X-axis → Year

Y → axis → Population



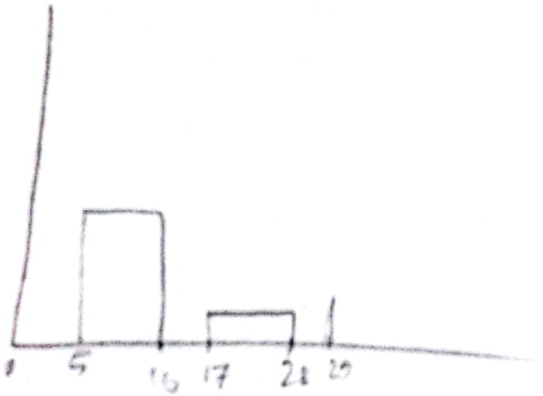
* Line Graph:

* P. Chart,

* Bars → Categorical

* Histogram * Histrogram → Distributed for continuous

↳ Contigous, discrete

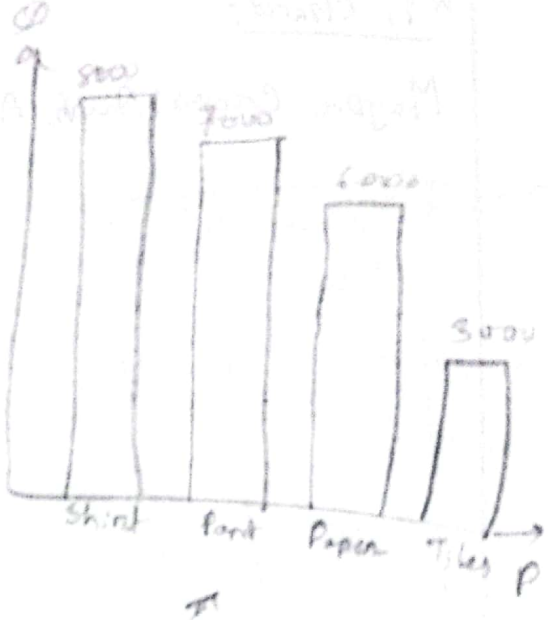


↳ Discrete

Contigrous → Categorical Bar Diagram

→ categorical

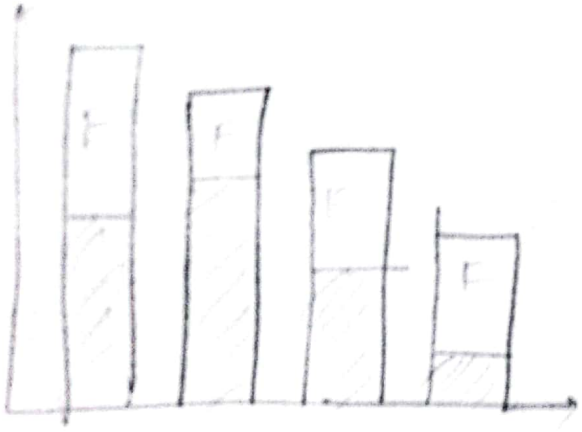
Products	Quantity
Shirt	8000
Pant	7000
Paper	6000
Tiles	3000



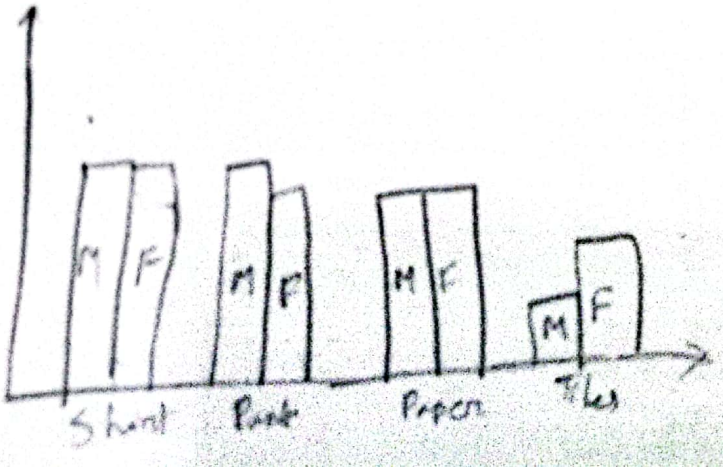
Bar Diagram

↳ Compound Multiple

Compound Bar Diagram:



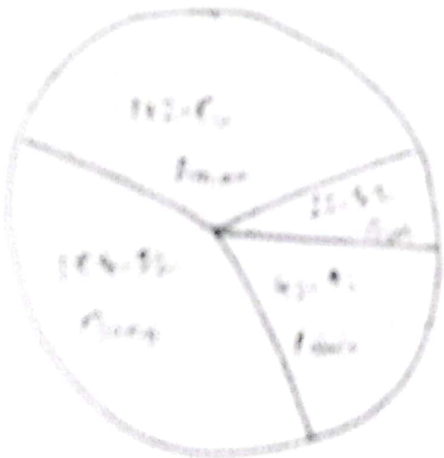
Multiple Bar Diagram:



P. Chart:

Major Crops: Aus, Aman, Boro, Potato

Major Crops	Production (mil t)	R.F (Relative Frequency)	Degree:
Aus	18	18/290 = 0.062	0.062 × 360 = 22.32
Aman	115	115/290 = 0.396	0.396 × 360 = 142.56
Boro	124	124/290 = 0.427	0.427 × 360 = 153.72
Potato	33	33/290 = 0.113	0.113 × 360 = 41.04
	290		



Stem Leaf Plot

Stem and Leaf Plot:

↳ Data nature ကိန်း

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

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

Bivariate table:

↳ 2 ခု variable → Relation ကိန်း → 2 ခုစလုံး categorical

	Gender	Economical Status
1	M	P
2	F	R
3	F	R
4	F	M
5	M	P
6	M	R
7	M	M
8	F	P
9	F	R
10	M	M

Gender	ES			
	R	M	P	
M	2	1	2	5
F	2	2	1	5
	4	3	3	

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

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

* Possible Questions:

1. Data কী? Example, [illegible]
2. Scale, Variable কী. " . Example
3. Frequency, distribution → table draw, histrogram draw.
4. Class কী, Types, frequency, Class Limit / Boundary.
5. কীভাবে Frequency table draw করবে?
   ↳ Cumulative,
6. Bar diagram:
   ↳ Multiple
   ↳ Stack
   ↳ Component
7. Bar diagram & Histrogram diff.

Chisquare:

Association between 2 categorical variable

Central tendency → Ch-3

Measure of central tendency

↳ Mean, Meadian, Mode

↳ Center as [illegible]

* Arithemedic mean কীভাবে বের করে?

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

grouped → class interval ... Ungrouped → ...

Age of Manager (xi)	Frequency (fi)	fixi
11	17	187
12	28	336
13	37	481
14	52	728
20	8	160
	Σfi	

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$ → Ungrouped

Class Interval	Frequency	Mid point (xi)	fixi
17-20		17+20/2	
28-36			
52-60			

→ grouped

$\frac{\Sigma f_i x_i}{\Sigma f_i}$

↑
class interval ka middle point (xi) ans

H.W

* Pooled mean
↳ 2 or group ka Mean ko combine

13-3-24
Lec-4

Pooled Mean

↳ Merge করে

↳ 2টা group এর combined Mean

g^1	g^2
M	F
5.1	5.0
5.6	5.1
5.8	5.9
⋮	⋮
$n_1 = 10$	$n_2 = 15$

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$$

$$\bar{x} = \frac{\Sigma f_i x_i'}{\Sigma f_i}$$

$\bar{x}_1 = 5.5$

$\bar{x}_2 = 5.1$

$$\bar{x}_c = \frac{10\times 5.5 + 5.1\times 15}{10+15}$$

$$= 5.26$$

*Meadian → Ascending order এ সাজাতে হবে

* 34, 21, 19, 17, 12, 7, 2 → ungrouped

→ 2, 7, 12 (17) 19, 21, 34 → Median

n = Odd

$$= \frac{n+1}{2}$$

n = even, Meadian $= \frac{\frac{n}{2} + \left(\frac{n}{2}+1\right)}{2}$

* 1, 2, 7, 12, 17, 17, 21, 34 → ungrouped

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

Grouped Data:

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

l_m = Median class as Lower Limit

h = difference

f_m = Median as Frequency

$F_{(m)-1}$ = Median as 34.50 CF

Age	f_i	CF
24.5-29.5	3	3
29.5-34.5	9	12 ($F_{(m)-1}$)
34.5-39.5 (l_m)	15 (f_m)	27
39.5-44.5	12	39
44.5-49.5	7	46
49.5-54.5	4	50
	n=50	

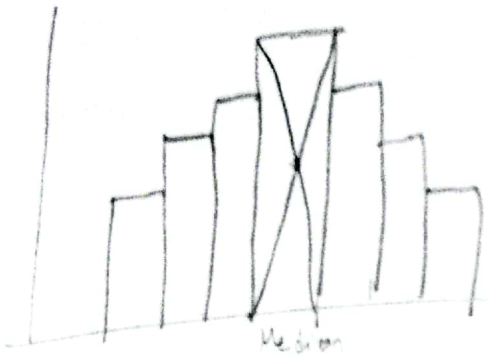
$\frac{50}{2} = 25$

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

$$= 43.33$$

$$= 38.83$$

Histrogram → Normal Distribution :



Quartile:

$\frac{n}{4}$ → n ও 4 দ্বারা Divisible হবে,

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

Q_2 = Median → 50

$\frac{3n}{4}$ → $Q_3 = \frac{\frac{3n}{4}\text{th} + \left(\frac{3n}{4}+1\right)\text{th}}{2}$

n ও 4 দ্বারা Divisible না হবে,

$\frac{15}{4} = 3.75$ Q_1 = Next higher integer

$\approx 4 \rightarrow Q_1$

Q_2 = Median

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

$= \frac{3 \times 15}{4} = 11.325 \approx 12 \leftarrow Q_3$

* 14, 17, 19, 23, 27, 32, 40, 49, 54, 59, 71, 80,

$\rightarrow n = 12$

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

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

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

$Q_4 = 80$

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

$$Q_1 = l_1 + \frac{h}{f_1}\left(\frac{1n}{4} - F_{(1)-1}\right)$$

$$= 34.5 + \frac{5}{15}\left(\frac{1 \times 50}{4} - 12\right)$$

$$= 34.67$$

$\frac{50}{4} = 12.5$ ↳ 3rd class

$$Q_3 = 37.5 + \frac{5}{12}\left(\frac{3 \times 50}{4} - 27\right)$$

$$= 43.875$$

$\frac{3 \times 50}{4} = 37.5$ ↳ 4th class

Percentile:

$$Q_p = l_0 + \frac{h}{f_p}\left(\frac{pn}{100} - F_{(p)-1}\right)$$

$$Q_1 = 29.5 + \frac{5}{9}\left(\frac{12\times 50}{100} - 3\right)$$

$$= 31.16$$

12%. তম শতক,

p

$\frac{pn}{100}$

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

~~12.5~~

75%.

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

≈ 38

↳ 4th class

Deciles:

$$Q_D = l_D + \frac{h}{f_D}\left(\frac{Dn}{10} - F_{(D)-1}\right)$$

#

Mode:

$$M_0 = l_0 + h\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

l_0 = Lower limit of the Modal Class

h = hight of Modal Class

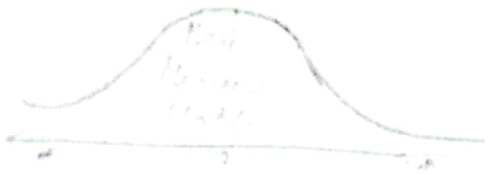
~~Mo = 1~~

~~Mo = 34.5 + 5~~

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

$$= 37.83$$

Symmetric distribution → Mean, Median Mode → same



Abnormal distribution → Mean, Median, Mode same নয়

Right squad:

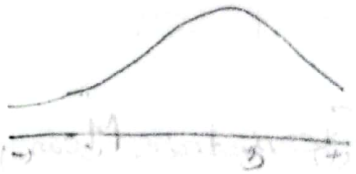
Mean → সবচেয়ে বড়

Median → মাঝে

Mode → Left side এ সবচেয়ে উঁচু point

↳ highest frequency

Left squad



Mean → সবচেয়ে ছোট

Median → মাঝে

Mode → Right side এ সবচেয়ে উঁচু point

↳ highest freque

H.W.

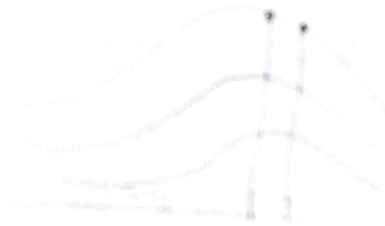
- Arithmetic mean depends on both origin and scale
- Variance " " only scale but not origin

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

* C.S. 10 12

→ 5 or 9 or 10 or 12 or 4

→ $\frac{5}{4}$ $\frac{9}{4}$ $\frac{10}{4}$ $\frac{12}{4}$



* Geometric Mean, harmonic Mean

* Trimean, Trimmean, Pooled Mean

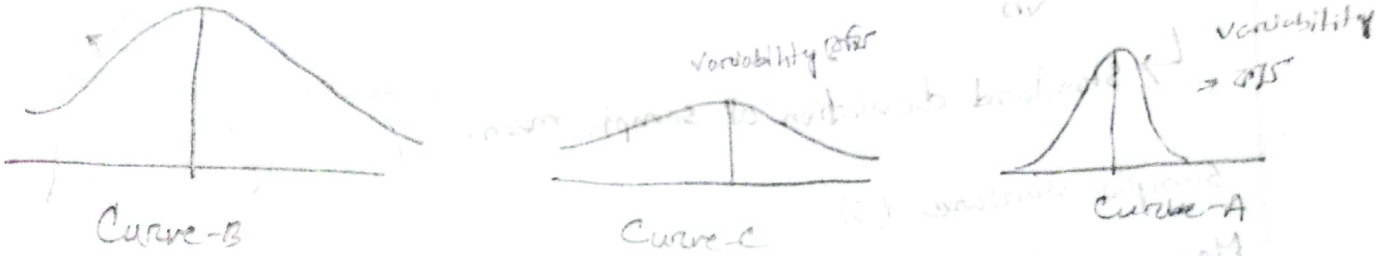
18-3-24
Lec-5

Ch-4

Variance

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

Disparation:



* Mean থেকে যতটার Observation এর দূরত্ব

* Measure of Disparation:

1. Variance
2. Range → upper, lower limit diff.
3. Inter quardile ~~$\phi_3 - \phi_1$~~ (IQR) Range: $\phi_3 - \phi_1$ → (25% - 75%) এর মধ্যে কতগুলো
4. Mean Deviation:
   দূরত্ব

↳ Observation থেকে Mean এর দূরত্ব

$$M_D(a) = \frac{\Sigma|x_i - a|}{n}$$

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

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

Group data → $\frac{\Sigma|f_i x_i - a|}{n}$ → table-4.2

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

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

$SD = \sqrt{Var}$

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

↳ Standard deviation of sample mean.

Sample variance (S)

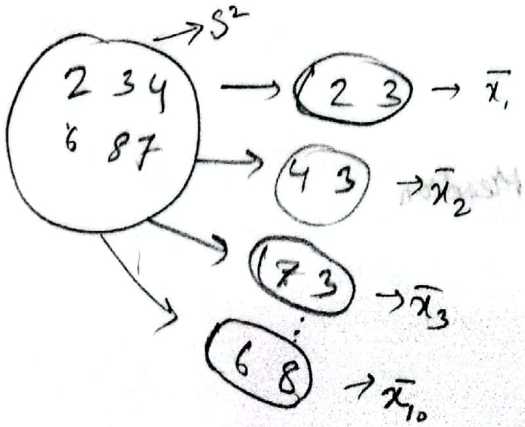
Mean = Mean of Standard deviation

Variance of Mean

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

→ M_1, S_1 } Mean of Mean
→ M_2, S_2 } Mean for grand mean diff. } Mean of Variance

$$\frac{\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{n}$$



* Variance of Property

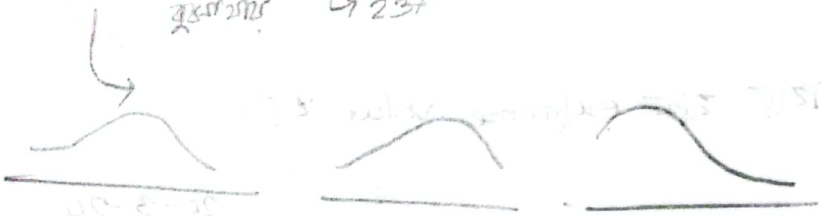
* Co-efficient of variation : $CV = \frac{SD}{\bar{x}}$

↳ Adv. → 2টা variable এর Mean সমান হলেও সহজে পার্থক্য বুঝা যায়

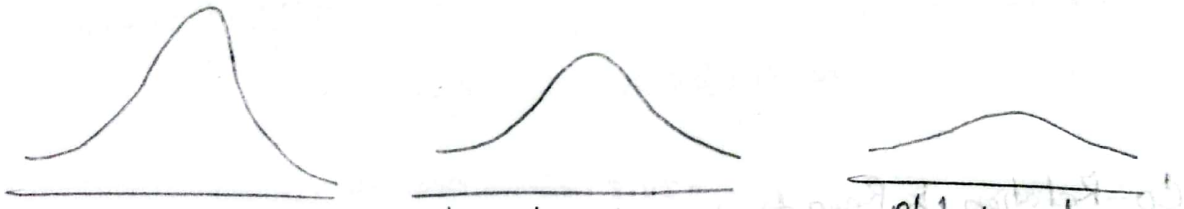
→ 232 page

Squeness, Quartosis

↳ Shape বুঝাবার ↳ 237

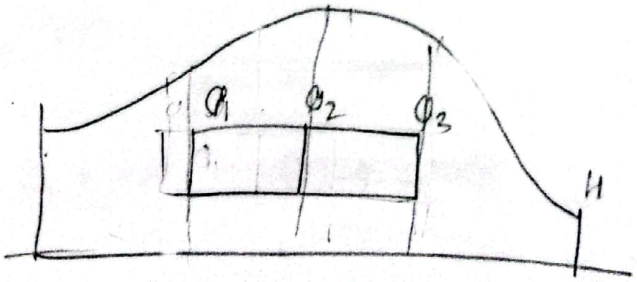


Quartosis: Picked অর্থে ব্যবহার



Box plot: (Math)

Smallest measurement, Q_1, Q_2, Q_3, highest measurement নিয়ে যে Box plot হয় তা।



Inner Fench:

$Q_1 \pm 1.5 (IQR)$

Inner Fench এর বাইরে থাকলে Out layer

Outer Fench:

$Q_1 \pm 3 (IQR)$

Outer Fench এর বাইরে হলে extreme value হবে।

20-3-24
Lec-

Ch-5

* কোন একটা Data র Centre নিয়ে কাজ করে → Central tendency
* Data variability measure করে → Dispersion

Co-Relation & Regression: → 2টার relation বোঝায়

Relation জানা দিবে

Co-Relation:

Cor(x, y)

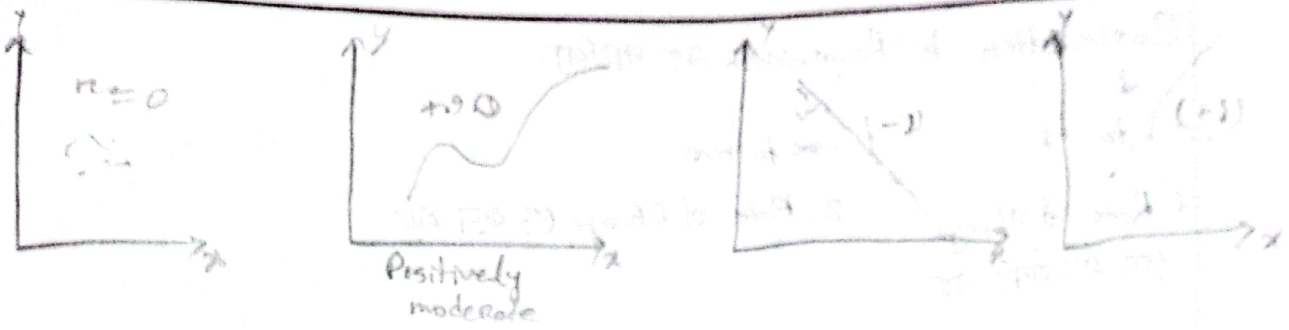
x	y

$-1 \le Cor(x,y) \le 1$

Cor → +ve → 1 টির সাথে অপরটি বাড়বে
→ -ve হলে → 1 টি বাড়লে অপরটি কমবে
= 0 → কোন Relation নাই

x (Study hours)	y (cgpa)

Range (-1, 1)
Middle → 0



285 page

287 ↓
5.7
Sales → Math

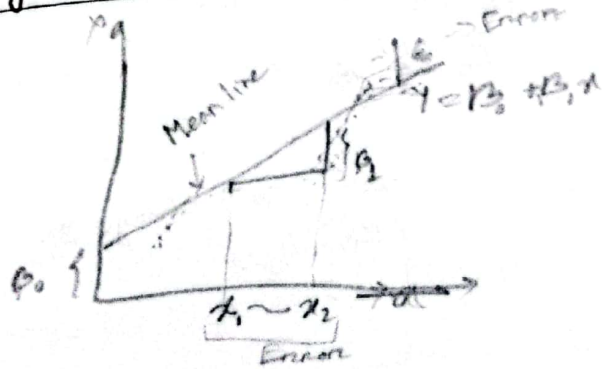
$$Corr_{(x,y)} = \frac{Cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$$

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

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

∴

Regression: → Relation এর সাধারণ → $-\infty$ to $+\infty$; B_1 = rate of change



$Y = mx + C$

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

m = ঢাল
C = constant

$B_1 \rightarrow +ve \rightarrow$ x বাড়লে y বাড়বে

$-ve \rightarrow$ x বাড়লে y কমবে

$x = 0 \rightarrow$ intercept

BMI
$Y = a - Bx$
Health condition
Overall
OR CGPA

Statistics → Approximate

Correlation & Regression এর পার্থক্য:

Correlation

1. −1 to +1
2. Rate of change বের করা যায় না

Regression

1. −∞ to +∞
2. Rate of change বের করা যায়

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

$Y = \beta_0 + \beta_1 x + \epsilon$ → Regression Line

↳ ϵ → Error

$E(Y|x) = \hat{\beta}_0 + \hat{\beta}_1 x$ → ; $E(\epsilon|x) = 0$

↳ Mean এর চেয়ে কিছু কম বা বেশি Error থাকবে

O → উপরে Error −ve

O − নিচে " = +ve

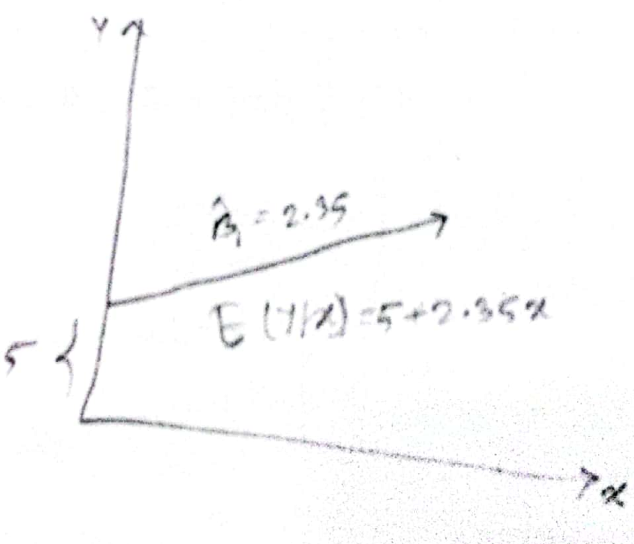
Error = 0

x (Hight) cm	y (Age) Year

Mean age

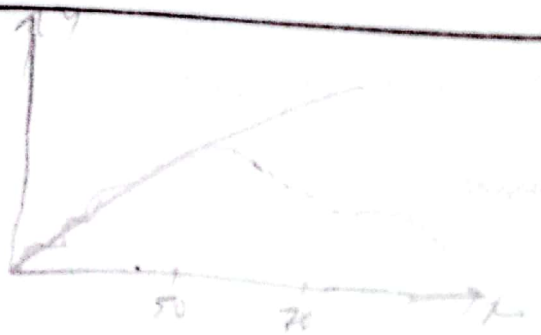
$\hat{\beta}_0 = 5$ → $x = 0$;

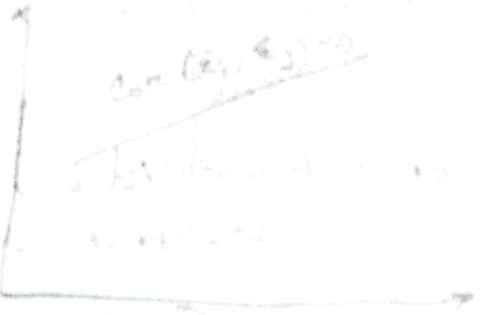
$\hat{\beta}_1 = 2.35$



অর্থাৎ 1 unit change হবে

Mean Hight 2.35 বাড়বে যদি 1 year age বাড়ে



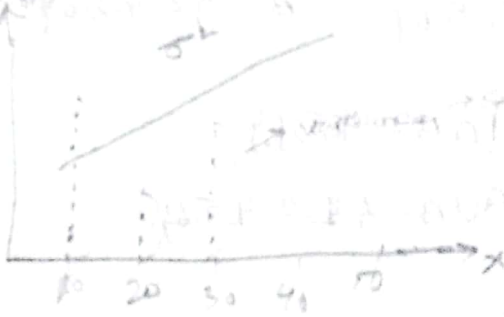


* Regression Properties → 5.2.1

1. Error term uncorrelated

2. Constant variance

3. Y is Distribution Normal
   1. $Y \sim N(\mu, \sigma_y)$
   2. $Cor(\varepsilon_i, \varepsilon_j) = 0$
   3. $\varepsilon_i \sim N(0, \sigma_2)$



Probability :

কোন একটা outcome হওয়ার chance

↳ Exact একটা value বলা

Possibility → Yes/No

↳ হতে

Set → elements/ Members

↳ Discrete value

Subset, Intersection, Union, complementation, power

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

$A \cap B = \{2, 5, 6\}$

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

* set কী? Example,

Experiment, Random Experiment

↳ কোনকিছুকে observe করা

↳ Random variable

↳ Possible sample statas (যেটা randomly আসতে chance

Sample space:

Event

Discreate & continuous Sample space:

↓

coin toss

↳ rainfall মাপার একক

Classical Approach, Relative Frequency Approach, Subjective Appro

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

↳ নির্ভুল হবে sample se

Ex-22: 10 balls → 4 white, 6 red, red হওয়ার possibility

white → $\frac{4}{10}$, red = $\frac{6}{10}$

Conditional Probability Probability:

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

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

↳ dependent

7.37

7.38, 7.3

Independent → $P(A|B) = \frac{P(A) * P(B)}{P(B)} = P(A)$

7.37

$P(A|B) = ?$

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

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

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

event / sample space

~~7.38~~ Bayes Theorem

↳ Prior, posterior

Posterior = Data * Prior → Data → Prior

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

* Machine A & B, output → 100%. 60% B, 40% A

1 " A থেকে produce defective 3%.

" B " " " 5% "

$P(A) = .60$

$P(B) = .40$

$P(D|A) = 0.03$