05-09-19

STAT1211- Statistics for Engineers.

Def:

Statistics: is the science of Collection, organization presentation, analyze and interpretation of Sample data. (comment)

Age - 18, 19, 21, 20, 16, 24, 25 Collection.

Class (Collection)	Tally (Organization)	Frequ.
16-20	IIII	4
20-24	II	2
24-28	I	1
		7

analyze

Presentation

Book -> ① Introduction to Statistics → Nurul Islam.

② Business Statistics → Md. Abdul Aziz.

Statistics is the science of collection, organization, presentation, analysis and interpretation of sample data.

Application/ Uses of Statistics:-

- Weather forcast
- State Craft
- Economy
- Business

Population : Entire set of element/ Observation

Sample : ~~On~~ only one among population / part or portion of population

M → 80%
P →

Q: Difference between population & sample.

21, 19, 20, 15

Outliers or ~~es~~ extreme value (the odd one)

variable = not fix

ID.	CT	Atten
1938520101	10	90%
03	10	50%
05	10	80%

Data Observation

XI · 2023 day

(গুণবাচক)
Variable : 2 Types → Qualitative Variable
→ Ex. Gender [Can't be defined numerically.]

→ Quantative "
(সংখ্যাবাচক)
C → Ex. Age
D → Roll
2 types [Can be expressed with numeric value]
→ Discrete
→ Continuous

Variable → Constant.

Scale of Measurement :-

1. Nominal
→ Name
→ Eye Color
→ Gender
✗ Rank

2. Ordinal
→ Current Class
→ Teacher Rank
→ Rank in Brotherhood (ভাই/বোন)
✓ Rank

3. Ratio
→ height
→ Weight
[Zero meaningless] (non-negative)

4. Interval
→ Temperature
→ Zero. meaningful.

Nominal No Rank
Ordinal Rank
Ratio & Interval

Nominal, Ordinal, Ratio or Interval.

1. Religion – Nominal

2. Jersy Number – ~~N~~ "

CW. ~~Date~~ ~~02~~09.09.19

Variable: Not known

1. Qualitative
2. Quantative

Data
1. Qualitative
2. Quantative.

Age → Quantitative variable.

23 → Quantative data

24 → "

Data Source

a) Primary
- direct Observation
- Takes time
- Costly

Methods
→ Mail Questionnine
→ Observation
→ Interviewing

b) Secondary
→ Published
→ Un. "

Two-Way Calculation

Marks of 30 Students.

Data:

34	36	31	46	76	86	42
44	32	46	40	54	66	56
50	42	33	80	77	81	46
40	60	63	69	76	56	57
57	70					

→ Construction of Continuous Frequency distribution

i) Range = Highest Value − lowest Value = 86 − 31 = 55

ii) Determination of no. of Class. N=30

$k = 1 + 3.322 \log_{10} N$

$= 5.906$

$\approx 6.$

Assignment ①

(iii) Class interval = $\frac{\text{Range}}{\text{Number of Classes}} = \frac{55}{6} = 9.1667$

≈ 10

Class	Tally Mark	Frequency	Cumulative Frequency
30 − 40 (39)	~~IIII~~	5	5
40 − 50 (49)	~~IIII~~ III	8	13
50 − 60 (59)	~~IIII~~ I	6	19
60 − 70 (69)	IIII	4	23
70 − 80 (79)	IIII	4	27
80 − 90 (86)	III	3	30

N = 30

CamScanner

Stem & leaf display

Steam	leaf
3	1 2 3 4 6
4	~~0~~ 0 2 2 4 4 6 6 6
5	0 4 6 6 7 7
6	0 3 4 6
7	0 6 6 7
8	0 1 6

key 3|1 means 31

— X —

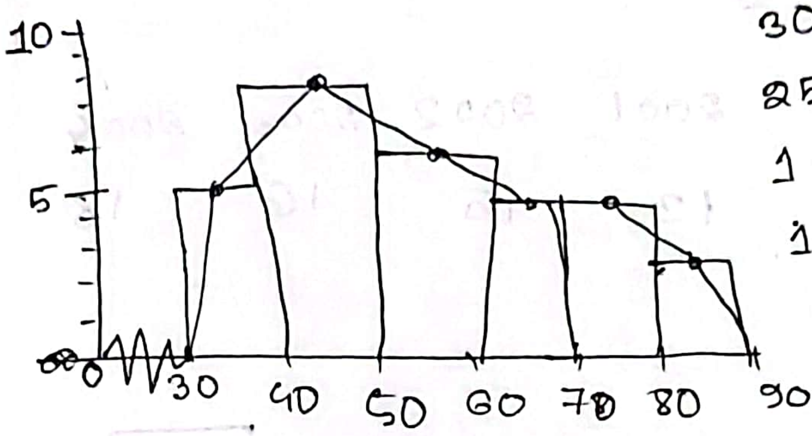
11.09.19

Graph

→ Frequency distribution graph
- → Histrogram
- → Frequency polygon
- → Frequency curve
- → Ogive Curve

→ Time series graph

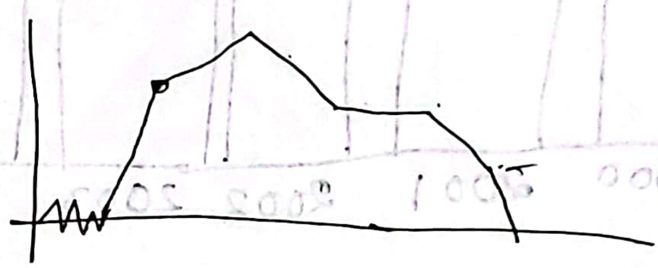
① Histogram



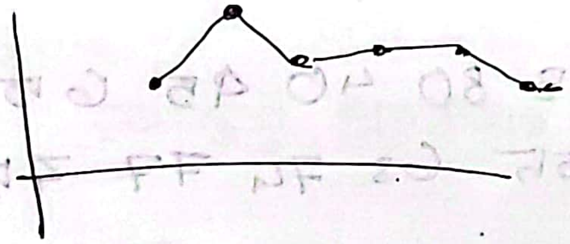
Decreasing C. Freq.	class	Freq	Increasing C.Fq
30	30-40	5	5
25	40-50	8	13
17	50-60	6	19
11	60-70	4	23
7	70-80	4	27
3	80-90	3	30

② F. Poly: Take mid-point & join those.

Mid point
35
45
55
65
75
85

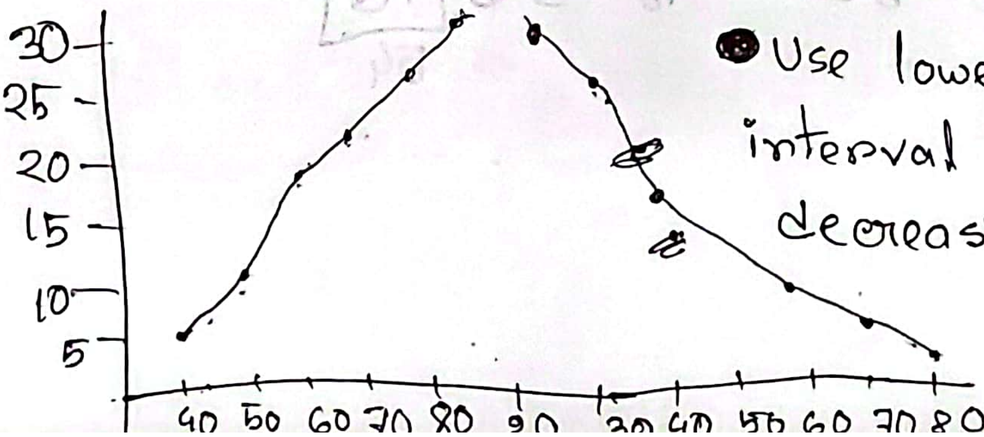


③ Frequency. Curve (Not joined with x-axis)



④ Ogive Curve.

- Use upperlimit of class interval then use increasing C.frequency
- Use lower ~~level~~ limit of class interval then use decreasing C. F.



40 43 1 7 10 13 17 24 31 38 42
41 44 3 8 11 14 20 26 33 39
42 5 9 12 16 22 27 34 41

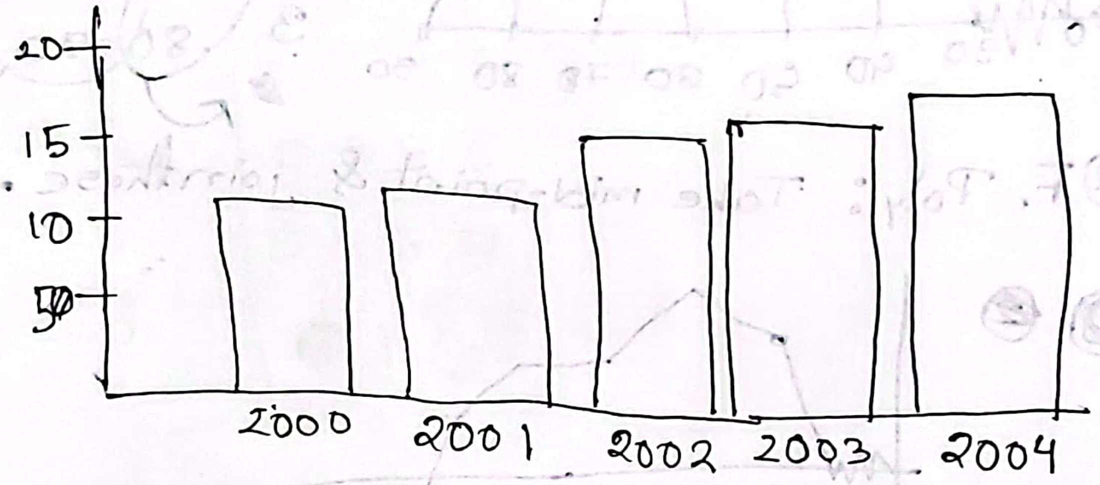
50 53
51 54
52

Time Series graph:

60 63
61 64
62

Year	2000	2001	2002	2003	2004
Pop^n (c)	12	13	15	16	18

70 73
71 74
72
80 83
81 84
82



⊛ The following data represent the daily income (in Tk) of 30 workers.

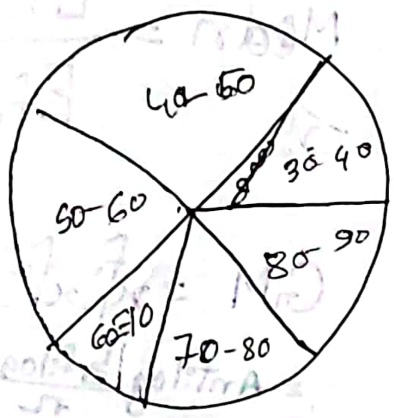
H.W.

50 70 55 75 80 40 45 65
46 35 30 55 68 74 77 45
58 44 36 60 62 64 48 52
50 50 52 48 35 $\boxed{\text{id}}$
54

PIE CHART

class	F		$\frac{360^\circ \times n}{N}$
30-40	5	60	
40-50	8	96	
50-60	6	72	
60-70	4	48	
70-80	4	48	
80-90	3	36	
	N=30	360°	

$$\begin{array}{r} 60 \\ +\,96 \\ \hline 156 \\ +\,72 \\ \hline 228 \end{array}$$



Measure of Centeral Tendency : Measures of
কেন্দ্রিয় প্রবণতার পরিমাপক central tendency in a single value that summaries a set of data. It locates the center of the value. —

Ex : Mean, Median, mode.

-> Mean গড়

-> Median মধ্যক

-> Mode প্রচুরক

Central Tendency

- Mean
  - Arithmetic
  - Geometric
  - Harmonic
- Median
- Mode.

34 - - . 70

Raw data : Ungrouped data

1. Mean, $\boxed{\bar{x} = \frac{\sum x_i}{n}}$

$$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$$

1 is ⓐ Arithmetic mean

ⓑ Geometric (GM)

$$GM = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

ⓒ Harmonic.

$$HM = \frac{n}{\sum \frac{1}{x_i}}$$

Grouped data

	f_i	mid x_i
30 - 40	5	35
		45
80 - 90	3	85

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i / N}$$

$x_i \rightarrow$ mid

$$GM = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_6^{f_6}\right)^{1/N}$$

$$= \text{Antilog} \frac{\sum f_i \log x_i}{n}$$

$$H.M = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots}$$

AM ≥ GM ≥ HM : Theorem

ungrouped

5, 2, 3, 4, 6, 10

$A.M, \bar{x} = \frac{\Sigma x_i}{n}$

$N = 6$

$GM., = \text{Antilog } \frac{\Sigma \log x_i}{n}$

$H.M = \frac{n}{\Sigma \frac{1}{x_i}}$

(a) 5.

(b) 4.3942

(c) $\frac{6}{\frac{31}{20}} = 3.87$

Class	f_i	x_i	$f_i x_i$
30-40	5	35	175
40-50	8	45	360
50-60	6	55	330
60-70	4	65	260
70-80	4	75	300
80-90	3	85	255
	30 N		$\Sigma f_i x_i = 1680$

$GM = (35^5 \cdot 45^8 \cdot 55^6 \cdot 65^4 \cdot 75^4 \cdot 85^3)^{1/30}$

$= 53.82$ (b)

$\therefore \text{Mean} = \frac{1680}{30} = 56$ (a)

$$H.M = \frac{30}{\frac{5}{35} + \frac{8}{45} + \frac{6}{55} + \frac{4}{65} + \frac{4}{75} + \frac{3}{85}}$$

$= 51.7337$ (c)

** Uses of Mean, Median, Mode.

22, 23, 24, 22, 21, 20 → Mean

21 22 22 23 24

22, 23, 24, 22, 21, 121 → Median

22, 23, 24, 22, 21, 22 → Mode

~~10~~.16.09.19

H.W.

Graph

The following data represent the daily income (in Tk) of 30 works.

50	~~40~~ 57	55	75	80	40	45	65	46	35
30	55	68	74	77	45	58	44	36	60
62	64	48	52	50	50	52	48	35	$\boxed{\text{id}}$ 54

(i) Range = Highest value - Lowest value $= 80 - 30$

$= 50$

(ii) Determination of no. of Class -

$K = 1 + 3.322 \log_{10} N$

$= 1 + 3.322 \log_{10} 30$

$= 5.9$ ~~≈6~~

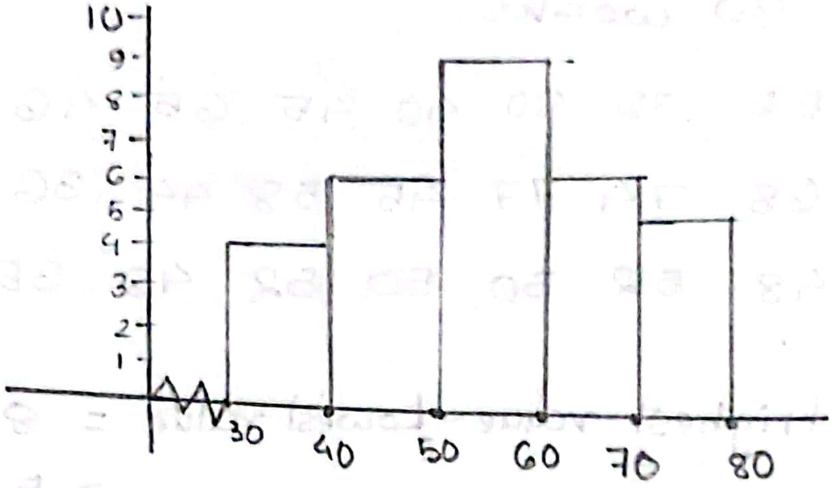
≈ 5

(iii) Class interval $= \dfrac{50}{5} = 10$.

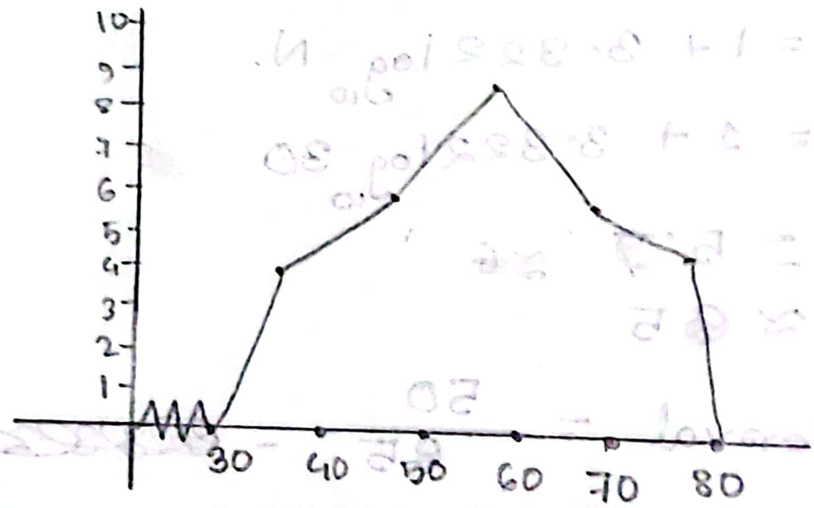
Class Interval	Tally	Frequency	Mid Value	Cumulative Frequency
30 - 40	||||	4	35	4
40 - 50	~~||||~~ |	6	45	10
50 - 60	~~||||~~ ||||	9	55	19
60 - 70	~~||||~~ |	6	65	25
70 - 80	~~||||~~	5	75	30

lower limit is counted

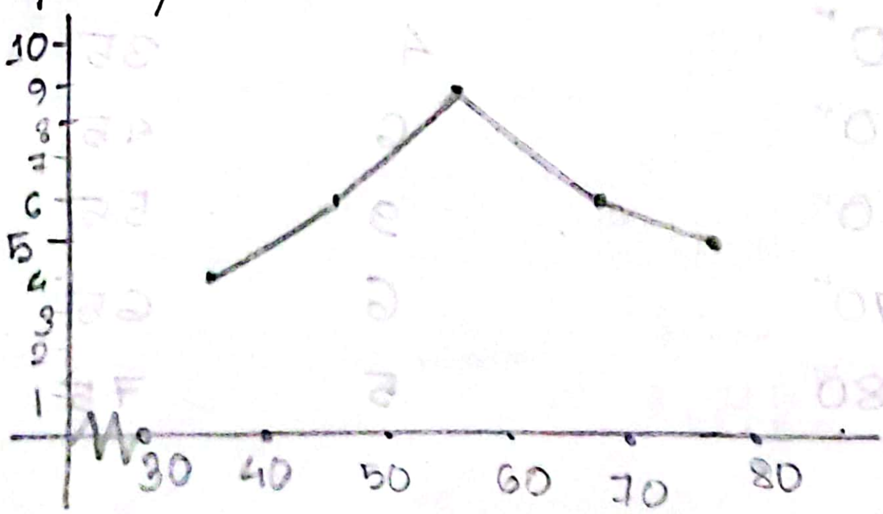
① Histogram :-



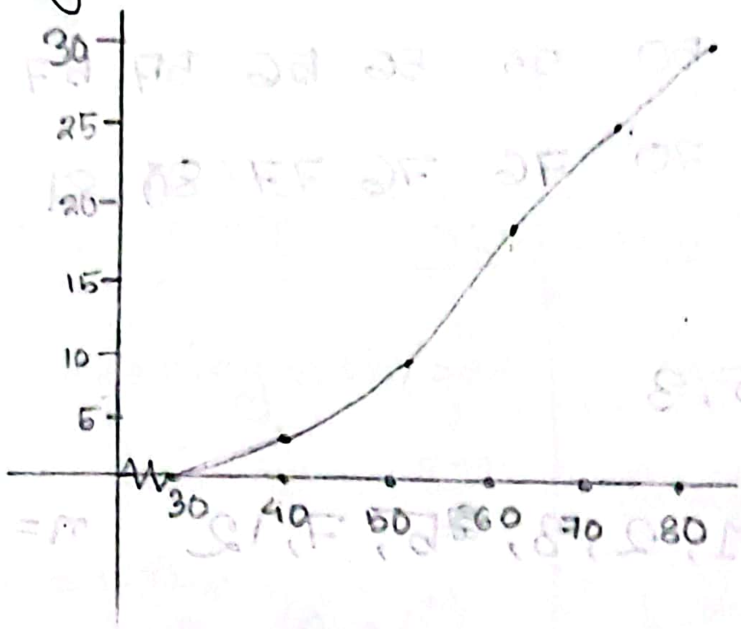
② Frequency Polygon:-



③ Frequency curve :-



③ Ogive Curve :-



④ Pie Chart :-

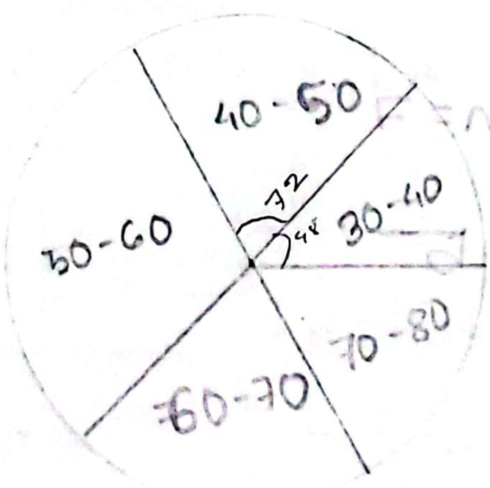
For Class interval 30-40, $\frac{360\times4}{30} = 48°$

" " " 40-50, $\frac{360\times6}{30} = 72°$

" " " 50-60, $\frac{360\times9}{30} = 108°$

" " " 60-70, $\frac{360\times6}{30} = 72°$

" " " 70-80, $\frac{360\times5}{30} = 60°$



Median

16.09.19

C.W.

$n=30$

31	32	33	34	36	40	40	42	42	44
46	46	46	50	54	56	56	57	57	60
63	64	66	70	76	76	77	80	81	86

Q: 1, 7, 12, 2, 5, 3

Ascending: 1, 2, 3, 5, 7, 12 $\quad n=6$

$$= \left(\frac{6}{2}\right)^{th} + \left(\frac{6}{2}+1\right)^{th}$$

$$= \frac{3^{th} + 4^{th}}{2}$$

$$= \frac{3+5}{2}$$

$$= 4$$

1, 2, 3, 5, 7, 10, 12 $\quad n=7$

$$\frac{7+1}{2} = 4^{th} \text{ term} = 5$$

Median

34, 31, . . .

Class	Fre	C.F	class	F	CF.
30-40	5	5	60-70	4	23
40-50	8	13	70-80	4	27
Median class 50-60	6	19	80-90	3	30

UNGROUPED | **Median** | GROUPED

1. Checking even/odd

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

If even,

$$M_e = \frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$$

$$= \frac{\left(\frac{30}{2}\right)^{th} + \left(\frac{30}{2}+1\right)^{th}}{2}$$

$$= \frac{15^{th} + 16^{th}}{2}$$

$$= \frac{54+56}{2}$$

$$= 55$$

$$\text{Median} = L + \frac{\frac{n}{2} - f_c}{f_m} \times h$$

$\frac{n}{2} = 15$

Lower limit of that class = 50

C.F of previous class = 13 = F_c

F of Median class = 6 = F_m

h = class interval = 10

$$M = 50 + \frac{15-13}{6} * 10$$

$$= 53.33$$

In odd,

$$\frac{29+1}{2} = \frac{30}{2} = 15^{th} = 54$$

$$L = \frac{\frac{n}{2} - f_c}{f_m} \times h$$

$\boxed{\text{Mode}}$

Very simple for identifying mode from ungrouped data. M_0 = The observation having repeated mostly / frequency

$$\text{Mode} = L + \frac{f_1}{f_1 + f_2} \times h$$

(max f class lower limit) $L = 40$ $\qquad = 40 + \frac{3}{3+2} \times 10$

$f_1 = 8 - 5 = 3$

$f_2 = 8 - 6 = 2$ $\qquad = 46$

$h = 10$

$AM \geq GM \geq HM$ (two observation)

$$AM = \frac{x_1 + x_2}{2}$$

$$GM = (x_1 . x_2)^{\frac{1}{2}} = \sqrt{x_1 . x_2}$$

$$H.M = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Now, for any two positive quantity,

$$(\sqrt{x_1} - \sqrt{x_2})^2 \geq 0$$

$$= (\sqrt{x_1})^2 - 2.\sqrt{x_1 x_2} + (\sqrt{x_2})^2 \geq 0$$

$$= x_1 - 2\sqrt{x_1 . x_2} + x_2 \geq 0$$

$$= x_1 + x_2 - 2\sqrt{x_1 . x_2} \geq 0$$

or, $x_1 + x_2 \geq 2\sqrt{x_1 x_2}$

or, $\frac{x_1 + x_2}{2} \geq \sqrt{x_1 x_2}$

$AM \geq GM$ — ①

Now for any two positive quantity,

$$\left(\frac{1}{\sqrt{x_1}} - \frac{1}{\sqrt{x_2}}\right)^2 \geq 0$$

or, $\left(\frac{1}{\sqrt{x_1}}\right)^2 - 2 \cdot \frac{1}{\sqrt{x_1}} \cdot \frac{1}{\sqrt{x_2}} + \left(\frac{1}{\sqrt{x_2}}\right)^2 \geq 0$

or, $\frac{1}{x_1} + \frac{1}{x_2} - 2 \cdot \frac{1}{\sqrt{x_1 \cdot x_2}} \geq 0$

or, $\frac{1}{x_1} + \frac{1}{x_2} \geq \frac{2}{\sqrt{x_1 \cdot x_2}}$

or, $\frac{1}{\frac{1}{x_1} + \frac{1}{x_2}} \leq \frac{\sqrt{x_1 \cdot x_2}}{2}$

or, $\frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} \leq \sqrt{x_1 \cdot x_2}$

$\therefore \sqrt{x_1 \cdot x_2} \geq \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$

$\Rightarrow \sqrt{x_1 \cdot x_2} \geq \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$

$GM \geq H.M$ — ②

From ① & ②

$AM \geq GM \geq H.M$

$$\frac{1}{\frac{2}{3}}$$

Mean = Median = Mode

AM = GM = HM $\qquad 3 = 3 = 3 = 3 = 3$

For similar number.

<u>Proof:</u> AM = GM = HM (n obn)

Let, us consider a set of number is

$x_1 = x_2 = \cdots = x_n = k$

Again,

$$AM = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{k + k + \cdots + k}{n}$$

$$= \frac{nk}{n} = k$$

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} = \sqrt[n]{k \cdot k \cdot k \cdots k}$$

$$= \sqrt[n]{k^n}$$

$$= k^{n \cdot \frac{1}{n}}$$

$$H.M = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n}}$$

$$= \frac{n}{\frac{1}{k} + \frac{1}{k} + \cdots + \frac{1}{k}}$$

$$= \frac{n}{n \cdot \frac{1}{k}}$$

$$= k$$

১৫/৭/২২

1, 5, 7, 9, 10, 12, 20, 35, 26

কত বেশি তার measure.

১/ Range (H − L)

2. Mean deviation, $M_{\dot{e}}(\bar{x}) = \frac{\Sigma|x_i - \bar{x}|}{n}$

3. Quartile " $\frac{Q_3 - Q_1}{2}$

* 4. Variance " $S^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$

* 5. Standard " $S_d = \sqrt{S^2}$

19-09-19

Measure of Dispersion

Absolute M.D

1. Range
2. Mean deviation
3. Quartile "
4. Variance "
5. Standard daviation

Relative M.D

1. Co-efficient of Range
2. " " Me D
3. " " Q: D
4. " " Variation (CV).

30, 40, ... 70.

30-40	5	35
40-50	8	45
50-60	6	55
60-70	4	
70-80	4	
80-90	3	
	30	

Absolute — Un G — G

1. Range — $H-L = 86-31 = 55$ — $H-L = 90-30 = 60$

2. M.D

$MD(\bar{x}) = \frac{\Sigma|x_i-\bar{x}|}{n}$ — $\frac{\Sigma f_i|x_i-\bar{x}|}{n}$

$MD(M_e) = \frac{\Sigma|x_i-Me|}{n}$ — $\frac{\Sigma f_i|x_i-Me|}{n}$

$MD(M_o) = \frac{\Sigma|x_i-M_o|}{n}$ — $\frac{\Sigma f_i|X_i-M_o|}{n}$

Un G G

4. Varience, $s^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$ $s^2 = \frac{1}{n}\sum F_i(x_i-\bar{x})^2$

5 5 8 9 10

5, 10, 9, 8, 5

x_1 $\bar{x}$

$|5-7.4|$

$1-2.4|$

$\frac{|2.4|+}{5}$

$\bar{x} = \frac{5+10+9+8+5}{5}$

$= 7.4$

$f_i($

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

$\frac{}{30}$

$s^2 = \frac{\sum(x_i-\bar{x})^2}{5}$

$= \frac{(5-7.4)^2+(10-7.4)^2+(9-7.4)^2+(8-7.4)^2+(5-7.4)^2}{5}$

$\bar{x}$

$= 4.24$

$s^2 = \frac{\sum f_i(x_i-\bar{x})^2}{n} = \frac{5(35-51)^2+8(45-51)^2+6(55-51)^2+4(65-51)^2+4(75-51)^2+3(85-51)^2}{30}$

$= 249$

1 2 3 3 4

$|1-3|+|2-3|$

(কম হলে ভাল)

Variance' is the sum of square of mean deviation.

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

S.D: SD is positive square root of variance.

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

T20

Criketer-A

24, 23, 20, 21, 19

$$\bar{x} = \frac{107}{5} = 21.4$$

$$S^2 = 3.44$$

Criketer-B

24, 0, 15, 50, 5

$$\bar{x} = \frac{94}{5} = 18.8$$

$$\frac{(24-18.8)^2 + (-18.8)^2 + (15-18.8)^2 + (50-18.8)^2 + (5-18.8)^2}{5}$$

$$S^2 = 311.74$$

G-A

Mean = 3·24

Mode = 10

σ/Sd = 2·25

$$CV = \frac{\sigma}{x} \times 100\%$$

$$= \frac{2{\cdot}25}{3{\cdot}24} \times 100\%$$

$$= 68\%$$

G-B

Mean = 30·24

Mode = 10

σ/Sd = 2·50

$$= \frac{2{\cdot}5}{30{\cdot}24} \times 100\%$$

$$= 8{\cdot}26\%$$

Correlation

$r =$

19.09.19

H.W.

FIND VARIENCE:

Cricketer A: 24, 23, 20, 21, 19

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{24+23+20+21+19}{5} = 21.4 \text{ (average)}$$

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{(24-21.4)^2 + (23-21.4)^2 + (20-21.4)^2 + (21-21.4)^2 + (19-21.4)^2}{5}$$

$$\sigma = 3.44 \text{ (variance)}$$

$$\sigma = \sqrt{S^2} = 1.85 \text{ (standard daviation)}$$

Cricketer B: 24, 0, 15, 50, 5

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{24+0+15+50+5}{5} = 18.8$$

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{(24-18.8)^2 + (0-18.8)^2 + (15-18.8)^2 + (50-18.8)^2 + (5-18.8)^2}{5}$$

$$= 311.76$$

$$\sigma = \sqrt{S^2} = 17.65$$

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

C.W. - 23.09.19

Correlation

(Y) - Dependent
(X) - Independent

(X) Yield (kg)	Fertilizer (kg)
50	1
70	2
100	4

$-1 \leq r \leq 1$

$$r = \frac{\sum x_1 y_1 - \frac{\sum x_1 \cdot \sum y_1}{n}}{\sqrt{\sum x_1^2 - \frac{(\sum x_1)^2}{n}} \sqrt{\sum y_1^2 - \frac{(\sum y_1)^2}{n}}}$$

y	x	xy	x^2	y^2
50	1	50	1	2500
70	2	140	4	4900
100	4	400	16	10000
$\sum y = 220$	$\sum x = 7$	$\sum xy = 590$	$\sum x^2 = 21$	$\sum y^2 = 17400$

$$r = \frac{590 - \frac{7 \times 220}{3}}{\sqrt{21 - \frac{7^2}{3}} \sqrt{17400 - \frac{(220)^2}{3}}}$$

$= 0.997$

On 1st day after Puja vacation Statistics.

(CH-1 & Correlation).

Correlation: It Measures the relationship between two variable i.e., independent & dependent variable.

Example: ① Yield increase with the uses of fertilizer

② Reading time & CGPA

26.08.19
Thursday

Correlation

Correlation:- Find the relationship betw two variables.

~~Collo~~ Correlation Co-efficient: (r) : measure the strength of linear relationship betw two variables.

Co-effecient of determination (r^2): Measure the total variation of dependent variable explained by independent variable.

$r = 0.98$ linear relationship

$r^2 = 0.96$ ~~linear re~~ : fertilizer 96%

Types of correlation

Simple Corr (5 types)

Example: Dependent-1
Inde -1

Mulltiple Corr.

Dependent - 1
independent ->1
(more than one)

$-1 \le r \le 1$

① Simple Correlation

→ Perfect Positive sc $(r=1)$

→ Partial Positive sc $(0<r<1)$

→ Perfect negative sc $(r=-1)$

→ Partial " sc $(-1<r<0)$

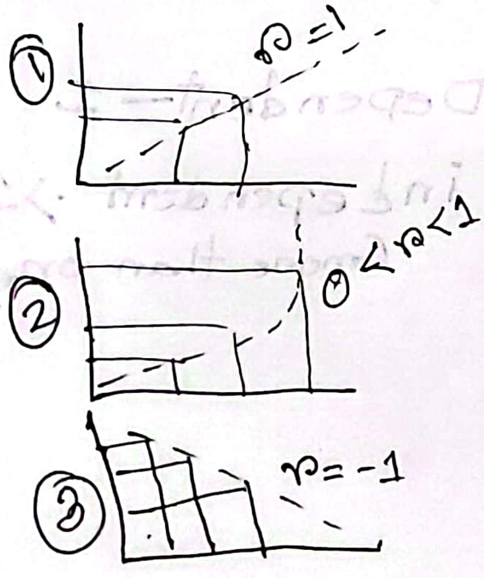
→ Zero correlation $(r=0)$

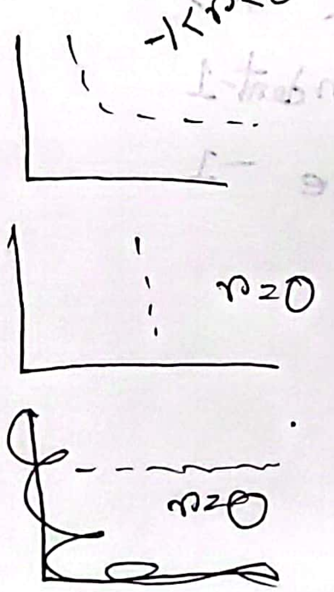
3 formula for finding co-efficient of correlation

1. Scatter diagram

2. Karl pearson Co-efficient of correlation.

3. Rank Correlation





independent = dependent

increase = increase

increase ≠ increase

increase = decrease

2. KPCC

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

3. Rank Correlation:- $r_{xy} = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$

$-1 \leq r \leq 1$

We know the sum of square is greater than or equal to zero.

$$\sum\left(\frac{x_i - \bar{x}}{\sqrt{\sum(x_i-\bar{x})^2}} \pm \frac{y_i - \bar{y}}{\sqrt{\sum(y_i-\bar{y})^2}}\right)^2 \geq 0$$

$$\text{or, } \sum \frac{(x_i-\bar{x})^2}{\left(\sqrt{\sum(x_i-\bar{x})^2}\right)^2} \pm 2\frac{x_i-\bar{x}}{\sqrt{\sum(x_i-\bar{x})^2}} \cdot \frac{y_i-\bar{y}}{\sqrt{\sum(y_i-\bar{y})}} + \frac{(y_i-\bar{y})^2}{\left(\sqrt{\sum(y_i-\bar{y})^2}\right)^2} \geq 0$$

$$\text{or, } \frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2} \pm \frac{2\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})}} + \frac{\sum(y_i-\bar{y})^2}{\left(\sqrt{\sum(y_i-\bar{y})^2}\right)^2} \geq 0$$

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

$\Rightarrow 1 \pm 2r + 1 \geq 0$

$\Rightarrow 2 \pm 2r \geq 0$

$2(1 \pm r) \geq 0$

$(1 \pm r) \geq 0$

$1 + r \geq 0$ $\qquad$ $1 - r \geq 0$

$r \geq -1$ $\qquad$ $-r \geq -1$

$r \leq 1$

Combining those: $-1 \leq r \leq 1$

To be Sumitted in the next Class.

Height of Father (in inches): 58 60 61 65 55

Height of Son (in inches): 59 59 62 63 60

① Find the karl pearson co-efficent of correlation

x $\quad$ y $\quad$ xy $\quad$ x^2 $\quad$ y^2

~~C.H.W.~~ ~~2020.09.19~~
~~Monday~~

CH-1: Introduction to Statistics.

Definiton: The science of Collection, organization, presentation, analyze and interpretation of Sample data.

Fields of Application:

1. State craft
2. Economies
3. Trade & Commerce
4. Medical Science
5. Demography (জনসংখ্যা)
6. Weather forecasting
7. Insurance Company
8. Law-reformation and human rights
9. Science.

(6).

* Difference betw Population & Sample.

Population	Sample
① In statistics, population refers to the totality of all the items or individuals having some specific characters.	① A representative and considerably small part of a population is known as ~~sa~~ sample of that population.
② Example: ⓐ All the Students of an University.	② ⓐ A spoonful boiled rice from a pot of boiled rice
ⓑ All possible outcomes in successive tossing of a coin	ⓑ CSE dept. of ICE

Constant: A numerical characteristic which does never change or vary it's value in termed as constant. Example: pi(π) = 3.1416 .

Variable:- The characteristics which varies over the units (or from unit to unit) is called variable. Example:

Father's Occupation of students of ~~IC~~ RU.

Here, the unit or population-unit is a student and their father's occupation varies. Say, one's father is teacher, other's is bussinessman and so on.

$1 - \frac{6\sum d^2}{n(n^2-1)}$

Classification :

ⓐ Qualitative: The characters which are not expressed in any numerical form ~~(i.e, color, gender,~~ occupation ~~etc)~~ but we can arrange them according to their quality or attribute.

Example: color of eye, gender, occupation etc.

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

ⓑ Quantitative: The characteristics of an unit or item that are expressed in numerical form or in numbers are quantitative variable.

Example: Family size, height etc.

Quantitative is has 2 sub-group:

① Discrete : Which possesses isolated or integral value (a whole ~~num~~ non-fractional number) is called ... dis--- . Example: Family size, population of ea country etc.

(ii) Continuous: Which takes values within a range or limit (allows fractional value) is called cont--- ---

Example: height, weight etc.

Scale of measurement:

ⓐ Nominal : A scale that measures a variable nominally (or by name without any order) is called Nominal. These data are:

① Mutually exclusive and exhaustive.
(ii) Have no logical order

Example: Color, gender, religion etc.

ⓑ Ordinal : A scale that identifies the values of a variable and arranges the values meaningfully in order of magnitude is. The properties

① Data classification are mutually exclusive and exhaustive
(ii) " " are Ranked or ordered according to a particular trait they possess

Example: Teacher's rank:- Teacher, Lecturer. Asst. Professor, Professor.

(c) Interval Scale: A scale which includes a definition of distance betw. the categories in terms of fixed and equal units is called interval scale. The concept of zero (0) is not included. So, 0 is meaningful.

Example: Temperature (−30°C, 0°F), I.Q etc.

(d) Ratio Sacle: The scale of measurment includes all properties of an interval scale ~~e~~xcluding concept of (0). Here we can perform arithmetic operations.

Example: height, weight etc.

→ **Data:** Data is the plural form of datum. Data are the collection of raw facts and figures from any sectors of inquiry for the purpose of statistical analysis.

(i) Qualitation data: The data which can not be measure by the numerical form is called qualitative. Such as: religion, economical condition etc.

(ii) Quantitative: The data which are expressed in numerical form. Exam; Population, height etc

Sources of data:

(a) Primary Data: Data which are obtained by direct observation from population or sample is called primary. The primary data are original in character and not well-organized. They are also called raw data or original data. The collection is highly expensive in respect to money, time and labour. ~~Metho~~

Methods of P.D. Collection:-

(a) Observation: The method where we have to use eyes rather than ears or voice. This is also called the classic method of scientific enquiry.

Advantage

(i) ~~If~~ In the cases where there lies greater chance of getting wrong information, we may use this method.

(ii) Useful in natural science

Disadvantage

(i) The scope of observation method is so limited.

(2) The method is very time consuming.

(b) Mail Questionnaire:- Method where a set of question is sent by mail to the respondent.

Advantage

(i) Less expensive

(ii) Free from problem associated with the use of interviewers

(iii)

Disadvantage

(i) Only considerable when questions are simple and easy to understand

Advantage	Dis Advantage
(iii) Questions which are embrassing in nature is suitable to be asked by mail questionnine.	(ii) Inappropiate when spontaneous answers are needed
	(iii) The surveyor can not be sure that the right person completed the questionnaire.

(C) Interviewing: The method of achieving information from the respondents, in which the information is gained by a conversation between interviewer and respondent.

There are three ways.

(i) Telephone Interview

(ii) Indirect Interview

(iii) Personal Interview

(1) Converwation over telephone to attain data.

Advantage	Disadvantage
(1) It's a easier method	(1) It is necessary for every respondent to have telephone.
(2) This method is cheap in terms of time and money.	(2) Question should be simple and less in number.
	(3) No chance to ask any in depth question

(ii) Indirect : The interviewer ~~d~~ asks the neighbours of respondent rather than that person.

Example : To know about terrorists in Dhaka then we use this method.

Advantage

(i) This is a cheap method in terms of time and money

(ii) For qualitative data collection this is a suitable method

DisAdvantage

(i) The respondent may answer incorrectly, or he/she be unable to response correctly for lack of knowledge.

(iii) Personal Interview : The interviewer ask questions from a prior printed schedule or the interviewer can offer the respondent a questionnaire that should be filled by the respondent.

Merits

(i) The chances of having wrong information is less.

(ii) The scope of enquiry can be widely spreaded in personal interview.

(iii) Here is hardly any chance of non-response.

Demerits

(i) It's expensive method

(ii) If the interviewer is not well-trained we may have wrong info.

(iii) Largly depends on quality of the interview.

(iv)

(b) Secondary data: Data which are already obtained by some other persons or organizations and are already published or utilized are called secondary data.

Sources of Secondary data:-

(i) Published Sources: There are many publication international publication, official govt & non govt. research institution publication, NGO's publication, Journals and newspaper center for trade and commerce etc.

Dependency

(ii) Unpublished sources: Records of information are stored in many officies, medical institutions hospitals etc.

Expense

Allah

Difference betw Primary & Secondary data.

Basis	Primary Data	Secondary Data
(i) Definition	The data which are obtained by direct observation from the population or sample is primary data.	The data which are already obtained by some other person or organization and are already published or utilized are sondary data
(ii) Originality	It is Original. And are collected from Original sources	It is not always Original. Secondary data is collected from some organization, journals etc.
(iii) Expenses	It involves large expenses in terms of time, energy and money.	It is relatively a less costly method.
(iv) Suitability	If data collected in a systematic manner its suitability is Positive	It may or may not suit the objective of survey.
(v) Reliable	Primary data more reliable that secondary data	Secondary data less reliable than primary data
(vi) Dependency	Completely independent	Dependent on primary data.
(vii) Precaution	No extra precaution needed	It should be used with care.
(viii) Qualified interviews	Any reliable primary data can be obtained only by qualified interview	No need of Q.I.

Correlation: One of the several measures of the linear statistical relationship between two random variables, indicating both strength and direction of relationship.

Example: (i) Yield increase with the uses of fertilizer.

(ii) Reading & CGPA.

Correlation Co-efficient: The measure of strength of the linear relationship between two variables.

The Karl Pearson invented formula of Correlation co-efficient,

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

The range of value of r is $-1 \leq r \leq 1$

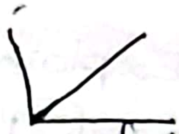
x is independent variable and y is dependent variable.

Correlation is 2 types:

(1) Simple "

(2) Multiple "

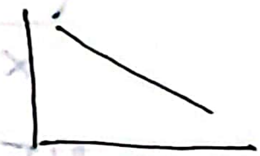
Simple is again 5 types:



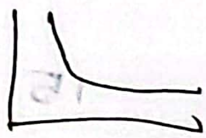
(I) Perfect Positive (direct): If changes of two variable are same direction i.e. if one increase or decrease the correspondent does the same i.e both variables are equal. $r = 1$.

(II) Partial Positive: If both variables increase or decrease ~~but~~ but rate of change is not equal it's partial positive. $0 < r < 1$.

(III) ~~Roe~~ Perfect Negative (inverse): If two variables change values recipocally i.e. one increase then other decrease and vice-versa. $r = -1$. Here rate of change is equal.

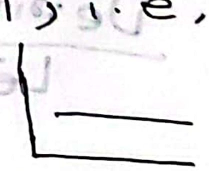


(IV) Partial Negative: If two variables change values recipocally but rate of change are not equal it's partial negative. $0 > r > -1$.



(V) Zero Correlation:

If the change's are independent, i.e. if the increases (or decrease) in one variable results a corresponding no change in othere. $r = 0$



Explain: $r = 0{\cdot}75$: It means by the partial positive correlation between the two variables. In this case, the changes of the two variables in same direction.

— x —

C.W. Statistics : Rank Correlation 14.09.19

$$\boxed{r = 1 - \frac{6\Sigma d_i^2}{n}}$$

x	Y	d_i
25 (3)	100 (3)	0
20 (2)	75 (2)	0
15 (1)	50 (1)	0

Derivation of Formula of Rank Correlation:—

Let, (x_i, y_i) be the rank of sample variable x & y.

x & y can take the values

1, 2, ————

Now,

$$\bar{x} = \bar{y} = \frac{1+2+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$$

Varience,

$$\sigma^2_x = \frac{1}{n}\sum(x_i-\bar{x})^2 = \frac{1}{n}\left(\sum x_i^2 - n\bar{x}^2\right)$$

$$\sum x_i^2 = 1^2 + 2^2 + 3^2 + \ldots + n^2$$

$$= \frac{n(n+1)(2n+1)}{6}$$

$$\frac{\sum d_i^2}{n} = \frac{n^2-1}{12}$$

$$\therefore \sigma^2_x = \frac{1}{n}\left(\sum x_i^2 - n\bar{x}^2\right)$$

$$= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6} - n\cdot\left(\frac{n+1}{2}\right)^2\right]$$

$$= \frac{n+1}{2}\left[\frac{(2n+1)}{3} - \frac{(n+1)}{2}\right]$$

$$= \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n+1}{2}\cdot\frac{4n+2-3n-3}{6}$$

$$= \frac{n+1}{2}\cdot\frac{n-1}{6} = \frac{n^2-1}{12} \text{ (Ans)}$$

Similarly,

$$\sigma_y^2 = \frac{1}{n}\sum(y_i^2 - \bar{y}^2) = \frac{n^2-1}{12}$$

$$d_i = x_i - y_i$$

$$= x_i - \bar{x} + \bar{x} - y_i$$

$$= x_i - \bar{x} - y_i + \bar{x}$$

$$= x_i - \bar{x} - y_i + \bar{y} \quad [\because \bar{x} = \bar{y}]$$

$$d_i = (x_i - \bar{x}) - (y_i - \bar{y})$$

$$d_i^2 = \{(x_i - \bar{x}) - (y_i - \bar{y})\}^2$$

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

$$- 2\sum (x_i - \bar{x}) \sum (y_i - \bar{y})$$

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

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

CamScanner

$$\sum \frac{d_i^2}{n} = \frac{n^2-1}{12} + \frac{n^2-1}{12} - 2\,\mathrm{Cov}(x, y)$$

$$\Rightarrow 2\,\mathrm{cov}(x,y) = \sigma_x^2 + \sigma_y^2 - \sum \frac{d_i^2}{n}$$

$$\mathrm{cov}(x,y) = \frac{\frac{n^2-1}{12} + \frac{n^2-1}{12}}{2} - \sum \frac{d_i^2}{2n}$$

$$= \frac{n^2-1}{6} \times \frac{1}{2} - \sum \frac{d_i^2}{2n}$$

$$= \frac{n^2-1}{12} - \sum \frac{d_i^2}{2n}$$

$$r_{xy} = \frac{\mathrm{cov}(x,y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\frac{n^2-1}{12} - \sum \frac{d_i^2}{2n}}{\sqrt{\frac{n^2-1}{12}}\sqrt{\frac{n^2-1}{12}}}$$

$$= \frac{\frac{n^2-1}{12} - \sum \frac{d_i^2}{2n}}{\frac{n^2-1}{12}}$$

$$= 1 - \frac{\sum \frac{d_i^2}{2n}}{\frac{n^2-1}{12}}$$

$$= 1 - \frac{\sum d_i^2}{2n} \times \frac{12}{n^2-1}$$

$$= 1 - \frac{\sum d_i^2}{n} \times \frac{6}{n^2-1}$$

$$\therefore r_{xy} = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

(Proved)

Math on Rank Correlation:-

x (dscendin)	y (ascending)	R(x)	R(y)	$d_i = x_i - y_i$	d_i^2
8	430	5.5	4	1.5	2.25
6	40	8	5	3	9
5	53	9	2	7	49
12	31	3.5	8	-4.5	20.25
8	12	5.5	10	-4.5	20.25
13	39	2	6	-4	16
7	51	7	3	4	16
12	59	3.5	1	2.5	6.25
29	19	1	9	-8	64
2	32	10	7	3	9

$\sum d_i^2 = 212$

$$1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

$$1 - \frac{6 \cdot 212}{10(10^2-1)}$$

$$= -0.2848$$

Ans

Comment: -0.29 is the value of co-efficient of correlation. The relation have ~~lower~~ partial negative correlation.

Two variable of independent and dependent variable.

—x—

C.W: **Regression Analysis** 21.10.19 Monday

Yield = □ + Fartilizer + Water + Soil + Weather

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots x + \varepsilon$$

$$Y = \alpha + \beta X + \varepsilon \rightarrow \text{parameter } \alpha, \beta$$

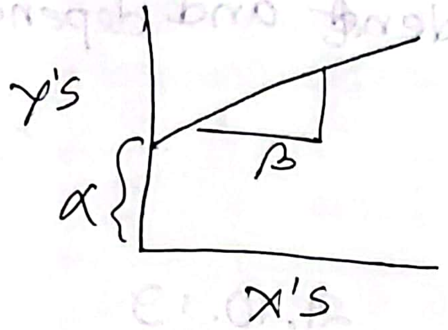
dependent = Constant + slope × independent variable + Error

Regression: ~~Regr~~ Is a mathematical measure of the average relationship between two or more variables interms of the original unit of the data.

1. X's are fixed
2. $\varepsilon_i \sim N(0, \sigma^2)$
3. Y is linear in x.
4. $Y_i \sim N(\mu, \sigma^2)$

} Assumption of RA.

Regression line.



$\hat{y} = \alpha + \beta x$

Regression line:

$\alpha = y - \beta x$

$\beta = \frac{y - \alpha}{x}$

Regression Analysis: - Reg^n measures the probable movement of one variable interms of the others.

Example : Yield vs Fertilizer

- Correlation vs Regression - 7-20

Find (α, β)

We know the regⁿ eqⁿ –

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i = y_i - \alpha - \beta x_i$$

$$\varepsilon_i^2 = (y_i - \alpha - \beta x_i)^2 \text{ [squaring both side]}$$

$$\sum \varepsilon_i^2 = \sum (y_i - \alpha - \beta x_i)^2 \text{ [Summing both side}$$

Differenciating with respect to α & set to zero.

$$\frac{\partial}{\partial \alpha}\left(\sum \varepsilon_i^2\right) \Rightarrow \frac{\partial}{\partial \alpha} \sum (y_i - \alpha - \beta x_i)^2 = 0.$$

$$\Rightarrow 2\sum (y_i - \alpha - \beta x_i) \cdot (0 - 1 - 0) = 0$$

$$\Rightarrow (2y_i - 2\alpha - 2\beta x_i) \cdot (-1) = 0$$

$$\Rightarrow \sum (-2y_i + 2\alpha + 2\beta x_i) = 0$$

$$\Rightarrow \sum (y_i - \alpha - \beta x_i) = 0$$

$$\Rightarrow \sum y_i - \sum \alpha - \sum \beta x_i = 0$$

$$\Rightarrow \sum \alpha = \sum y_i - \sum \beta x_i$$

$$\Rightarrow n\alpha = \sum^{n} - - \sum^{n}$$

$$\alpha = \frac{\sum y_i}{n} - \frac{\sum \beta x_i}{n}$$

$$\Rightarrow \alpha = \frac{\sum y_i}{n} - \beta \frac{\sum x_i}{n}$$

$$\alpha = \bar{y} - \beta \bar{x}$$

Again, differenciating in terms of β and set to 0.

$$\frac{d}{d\beta}\left(\sum \epsilon_i^2\right) \Rightarrow \frac{\partial}{\partial\beta}\sum(y_i - \alpha - \beta x_i)^2 = 0$$

$$\Rightarrow 2\sum(y_i - \alpha - \beta x_i)(-x_i) = 0$$

$$\Rightarrow \sum x_i y_i - \alpha\sum x_i - \beta\sum x_i^2 = 0$$

$$\Rightarrow \beta\sum x_i^2 = \sum x_i y_i - \alpha\sum x_i$$

$$\Rightarrow \beta\sum x_i^2 = \frac{\sum x_i y_i - \alpha\sum x_i}{\cancel{\sum x_i^2}}$$

$$= \frac{\sum x_i y_i - \left(\frac{\sum y_i}{n} - \beta\frac{\sum x_i}{n}\right)\sum x_i}{\cancel{\sum x_i^2}}$$

$$\beta\sum x_i^2 \cancel{\beta} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} + \frac{\beta\left(\sum x_i\right)^2}{n}$$

$$\beta\sum x_i^2 - \beta\frac{\left(\sum x_i\right)^2}{n} = \sum x_i y_i - \frac{\sum x_i y_i}{n}$$

$$\beta\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}.$$

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

$$\hat{y} = 2.02 + 3.03\hat{x}$$

Comment: Here the constant term 2.02 & with the change of one unit of x there will be changed (in an average) 3.03 sloped in dependent variable.

C.W.

Statistics

24-10.19
Thursday.

Regression.

Sales (Tk.lakhs) Promotion Exp (Tk '000)

(x)	(y)
8	2
10	2
9	3
12	4
10	5
11	5
12	5
13	6
14	7
15	8

i) Find the reg^n line for promotion on sales.

ii) If Sales is 20 tk find the promotion Exp.

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

$\Sigma x_i = 114$

$\Sigma x_i^2 = 1344$

$\Sigma y_i = 47$

$\Sigma xy = 572$

$$= \frac{572 - \frac{114 \times 47}{10}}{1344 - \frac{(114)^2}{10}}$$

$= 0.81531$

$y = a + bx$. $a = -4.594594595$

$b = 0.8153153153$

$r = 0.9041978802$

$$\alpha = \bar{y} - \beta\bar{x}$$

$$= \frac{\Sigma y}{n} - \beta\frac{\Sigma x}{n}$$

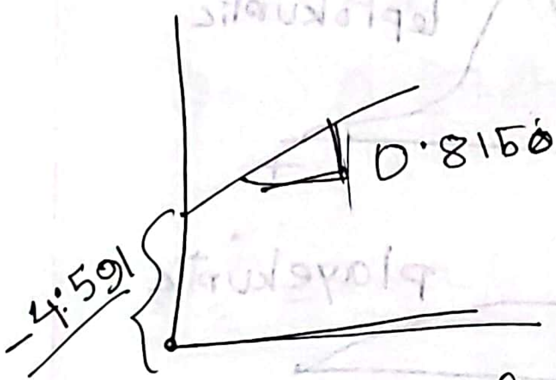
$$= \frac{47}{10} - 0.8153 \times \frac{114}{10}$$

$$= -4.59$$

$y = mx + c$

$$\boxed{\hat{y} = \alpha + \beta\hat{x}}$$

$$= -4.59 + 0.815\hat{x}$$



Comment: With sales of one unit there will be changed 0.815 on the promotion unit $\alpha = -4.591$ says then if there are no sales promotion will be decreased.

$$\hat{y} = -4.59 + 0.815 \times 20$$

$$= 11.71$$

$$\simeq 12$$

Comment: If $x = 20$ there will be changed about 12 (approx) unit in promotion.

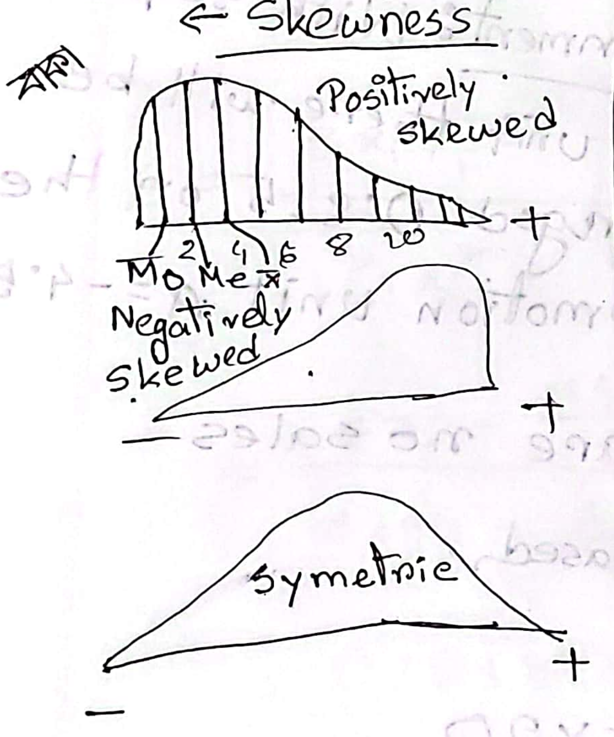
Male's Age :	22	23	23	24	26	27	27	28	30	30
Female's Age :	18	20	21	20	21	22	23	24	25	26

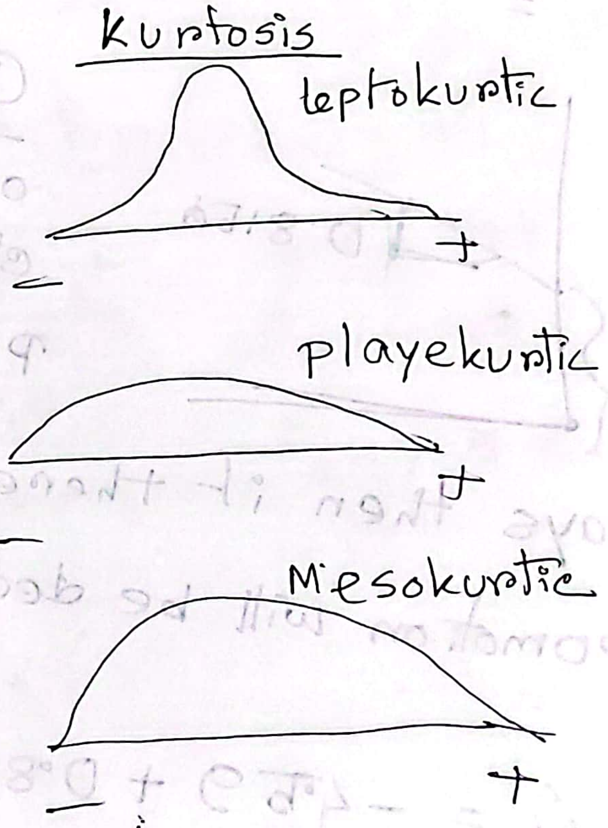
① Fit the regression line for husband (y) on wife (x).
Wife (x) on husband (y).

least square method (determine α, β).

to minimize error we use least square method.

কোনদিকে বেকা বাকানো





Skewness

last method 4. $\beta_1 = \frac{\mu_2^3}{\mu_3^2}$

1. β_1 = Mean – Median (absolute)

2. β_1 = Mean – Mode (absolute)

3. $\beta_1 = \frac{\text{Mean} - \text{Mode}}{\sigma}$ (Pearson)

$\sqrt{\beta_1} < 0$ [Negatively Skewed]

$\sqrt{\beta_1} = 0$ [Symmetric]

$\sqrt{\beta_1} > 0$ [Positively Skewed]

$\mu_2' = \bar{x} - a$

$\sqrt{\beta_1}$ = Co-efficient of skewness

Kurtosis

$\beta_2 = \frac{\mu_4}{\mu_2^2}$

$\beta_2 > 3$ leptokurtic

$\beta_2 = 3$ Mesokurtic

$\beta_2 < 3$ Playekurtic

H.W.

(husband)	(wife)
22	18
23	20
23	21
24	20
~~20~~26	21
27	22
27	23
28	24
30	25
30	26

$n = 10$

① Regression line for husband on wife (x)

② Regression line for wife on husband. (x)

① Husband (Y) and wife (x).

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

$$= \frac{5782 - \frac{220 \times 260}{10}}{4896 - \frac{(220)^2}{10}}$$

$$= \frac{62}{56}$$

$$= 1 \cdot 1071$$

$\sum x = 220$

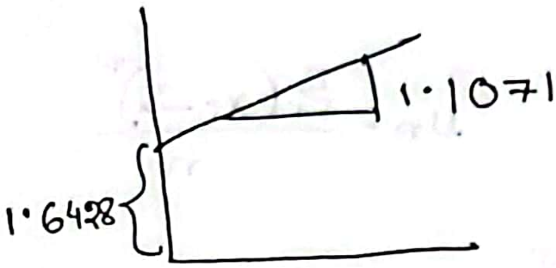
$\sum x^2 = 4896$

$\sum y = 260$

$\sum xy = 5782$

$n = 10$

$$\alpha = \bar{y} - \beta \bar{x}$$
$$= \frac{\Sigma y}{n} - 1.1071 \frac{\Sigma x}{n}$$
$$= \frac{260}{10} - 1.1071 \cdot \frac{220}{10}$$
$$= 1.6428$$



$$\therefore \hat{y} = \alpha + \beta \hat{x}$$
$$= 1.6428 + 1.1071 \hat{x}$$

Comment: So, ~~the~~ with scales of 1 unit of independent variable the dependent variable increasese 1.1071 times with 1.6428 unit as constant increased amount.

(2) Husband (x) & wife (y)

$$\beta = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$= 0.8157$$

$\Sigma x = 260$

$\Sigma y = 220$

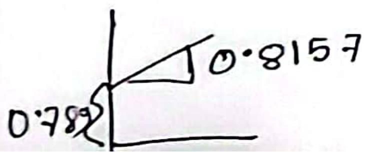
$\Sigma x^2 = 6836$

$\Sigma xy = 5782$

$n = 10$

$$\alpha = \bar{y} - \beta \bar{x} = 0.7894$$

$$\hat{y} = \alpha + \beta \hat{x}$$
$$= 0.7894 + 0.8157 \hat{x}$$



Comment: With scales 1 unit of independent variable the dependend (y) will increases 0.8157 times and $\alpha = 0.7894$ unit increased as constant.

C.W.

Moments

28.10.19
Monday

(i) Central Moment (ii) Raw Moments

$$\mu_r = \frac{\Sigma (x_i - \bar{x})^r}{n} \qquad \mu'_r = \frac{\Sigma (x_i - a)^r}{n}$$

a = arbitrary value

max upto 5.

$$\mu_1 = \frac{\Sigma (x_i - \bar{x})}{n} \qquad \mu_1 = \frac{\Sigma (x_i - a)}{n}$$

$$\mu_2 = \frac{\Sigma (x_i - \bar{x})^2}{n} = \sigma^2 \qquad \mu_2 = \frac{\Sigma (x_i - a)^2}{n}$$

$$\mu_3 = \frac{\Sigma (x_i - \bar{x})^3}{n} \qquad \mu_3 = \frac{\Sigma (x_i - a)^3}{n}$$

Ungrouped 2, 3, 4, 6, 10 $\quad \bar{x} = 5$

$$\mu_1 = \frac{(2-5)+(3-5)+(4-5)+(6-5)+(10-5)}{5}$$

$$= \frac{-3-2-1+1+5}{5}$$

$$= 0$$

$$\mu_2 = \frac{\Sigma (x_i - \bar{x})^2}{n}$$

$$= \frac{(2-5)^2 + (3-5)^2 + (4-5)^2 + (6-5)^2 + (10-5)^2}{5}$$

$$= 8$$

Grouped $\mu_n = \frac{\sum f_i (x_i - \bar{x})^n}{n}$

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3}{n} = 18.$$

$$\sum_{i=1}^{n} a = na$$

Raw

$$\mu_1' = \frac{\sum (x_i - a)}{\cancel{a}\, n}$$

$$\sum_{i=1}^{n} x_i =$$

$$= \frac{\sum x_1 - \sum a}{n}$$

$$= \frac{\sum x_1}{n} - \frac{\sum a}{n}$$

$$= \bar{x} - a$$

Relationship betⁿ Central ~~tendency~~ & Raw moments

$$\mu_1 = 0$$

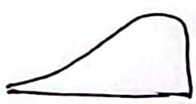
$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

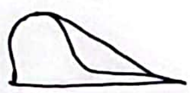
$$\mu_4 = \mu_4' - 4\mu_3'\mu_1 + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$

Skewness- ~~ff~~ যে দিকে ∅ বেশি থাকে.

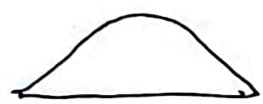
Positive



Negative

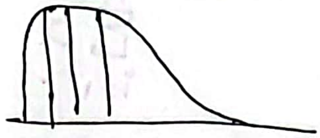


Symmetrical

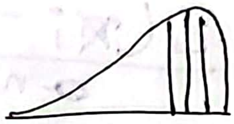


1. Measure of Skewness.

Median

Positive 

mo < me < $\bar{x}$

Negative mo > me > $\bar{x}$ 

Symmetric mo = me = $\bar{x}$

(a) $S_k(P) = \dfrac{\text{Mean} - \text{Mode}}{\sigma} = \dfrac{3(\text{Mean} - \text{Median})}{\sigma}$

$(-3, 3)$

(b) $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

$\sqrt{\beta_1} = \dfrac{\mu_3}{\sqrt{\mu_2^3}}$

$\sqrt{\beta_1} = 0$ Symmetrical

$\sqrt{\beta_1} > 0$ Positive Skewed

$\sqrt{\beta_1} < 0$ Negative "

$$\mu_3 = \frac{\Sigma (x_i - \bar{x})^3}{n}$$

$$\mu_2 = \frac{\Sigma (x_i - \bar{x})^2}{n}$$

2, 4, 6, 12

Moment
Skewness
Kurtosis.

(1) Find β_1 & interpeat your result

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \overline{x} = 6$$

$$\mu_3 = \frac{\Sigma (x_1 - \overline{x})^3}{n} = \frac{(2-6)^3 + (4-6)^3 + (6-6)^3 + (12-6)^3}{4}$$

$$= 36$$

$$\mu_2 = \frac{\Sigma (x_1 - \overline{x})^2}{n} = 14$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{36^2}{14^3} = 0.4723 \text{ (Ans)}$$

$$\sqrt{\beta_1} = 0.687$$

Comment: $\sqrt{\beta_1} = 0.687$. So, it's positively skewed dataset/distribution.

Moments:– A set of constant or descriptive measure which can characterize a set of values or observation uniquely is called moments.

Central moments:

– Computed from Arithmetic mean

$$\mu_r = \frac{\Sigma(x_1 - \bar{x})}{n}$$

Raw moments:

– Computed ~~from~~ using arbitrary/certain values.

$$\mu_r' = \frac{\Sigma(x_i - a)}{n}$$

Skewness:

Skewness is the degree of asymmetry or departure from symmetry of a distribution.

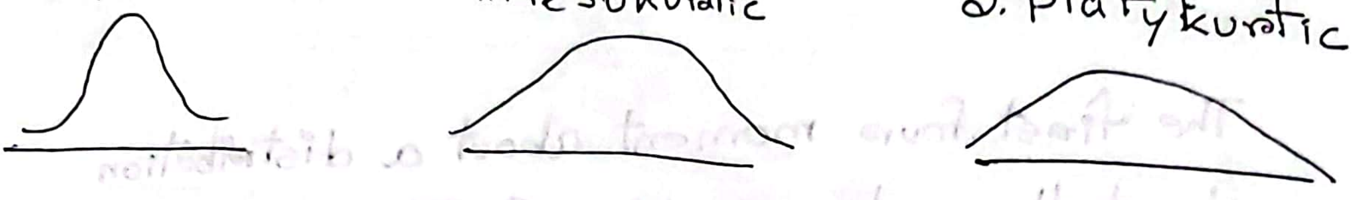
Kurtosis: Kurtosis is the degree of peakedness or flatness of a distribution.

STAT1211 31.10.19

C.W.

Kurtosis

1. Leptokurtic
2. Mesokurtic
3. Platykurtic



Measurment:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$\beta_2 = 3$ mesokurtic

$\beta_2 > 3$ leptokurtic

$\beta_2 < 3$ Platykurtic.

Q: 5, 7, 7, 7, 12 ; Find β_2 & comment.

Soln: We know,

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\bar{x} = \frac{5+7+7+7+12}{5} = 7.6$$

$$\mu_4 = \frac{\Sigma(x_i - \bar{x})^4}{n} = \frac{(5-7.6)^4 + (7-7.6)^4 \times 3 + (12-7.6)^4}{5}$$

$$= 84.1792$$

$$\mu_2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$= 5.44$$

$$\therefore \beta_2 = \frac{84.1792}{(5.44)^2} = 2.844 \text{ (Ans)}$$

Comment: Here $\beta_2 = 2.844$. $\therefore \beta_2 < 3$

So, the distribution is platykurtic.

The first four moment about a distribution about the value 5 are 2, 20, 40 and 50 respectfully. Obtain the first four central moments; β_1 & β_2:

$a = 5$

raw moments

$\mu_1' = 2$

$\mu_2' = 20$

$\mu_3' = 40$

$\mu_4' = 50$

(i) $\mu_1 = 0$

(ii) $\mu_2 = \mu_2' - (\mu_1')^2$

$= 20 - 2^2 = 16$

(iii) $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

$= 40 - 3.20.2 + 2(2)^3$

$= -64$

(iv) $\mu_4 = \mu_4' - 4\mu_3'\cdot\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$

$= 50 - 4.40.2 + 6.20.2^2 - 3.2^4$

$= 162.$

(skewness)

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-64)^2}{(16)^3} = 1$$

(co-efficient of skewness)

$$\sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-64}{\sqrt{(16)^3}} = -1$$

$\sqrt{\beta_1}\ (-1) < 0$ ie. The distn is negatively skewed.

(kurtosis) $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{162}{16^2} = 0.6328$

$\because \beta_2\ (0.6328) < 3$. So, the distn is platykurt

Probability 0-1

Possibility 0-100%

Formula for probability:-

$$P = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes.}}$$

If two dice are drawn simultanously make sample space.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Sample Space, $S = \{(1,1), (1,2), \ldots (6,6)\}$

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

A coin tossed 3 times.

	HH	HT	TH	TT
H	HHH	HHT	HTH	HTT
T	THH	THT	TTH	TTT

$S = \{$ HHH, HHT ------ TTT $\}$

Set: A collection of well defined objects is called set.

Example: List of book,

Universal, sub, null, set.

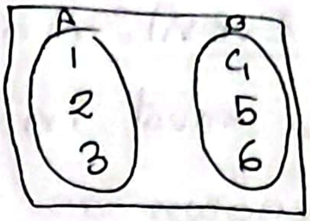
$A' = U - A$

C.W

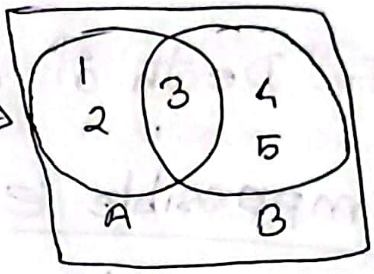
Probability.

04.11.19
Monday

Not Mutually Exclusive set →



~~Not~~ mutually exclusive set. →



Experiment: ~~A~~ Experiment is an act that can be repeated under some specific Condition

Trial: Consider an experiment which though repeated ~~a~~ under essentially identical Conditions does not give unique results but may result in any one of the several possible outcomes.

Ex: Throwing a dice is a trial.

Event: Any possible outcome or a set of possible outcome of a random experiment is called event.

Ex: Sample space $= \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 2, 3\}$

Sure Event: An event whose occurance is a must in any random experiment is known as sure event.

Ex: Death of every living.

Impossible event:-

An event whose occurance is quite impossible in a random experiment is called impossible event.

Ex: To live without breathing.

Equally likely Event:- The outcome of a trial or experiment are said to be equally likely if each of them have equal chance to be occured.

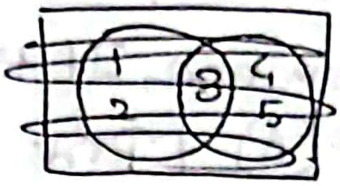
Ex: For a fair unbaised coin $P(H) = P(T) = \frac{1}{2}$

Mutually Exclusive event:- If the happening of any of the event excludes the happening of all the other

Bachelor of Science.

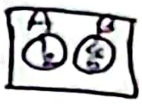
then the events ~~what~~ would be term as mutually exclusive event.

~~Ex~~ Example:- Throwing a coin if ~~co~~ head comes then tail will be excluded



If we know the outcomes it's

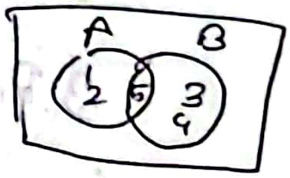
Exhaustive event: If the total number of outcome of an event is known, then it is called exhaustive event. Ex:- Throwing a dice, $S = \{1, 2, 3, 4, 5, 6\}$.



Independent event: If the occurance of a set of a event $\boxed{\text{not}}$ affected by any other set of event in any way, then the event is called ---

Example: Tosing a coin.

Dependent event: " (not হবে না).



Example: If we consider. 4 Red, 3 Black, 3 white

$P(\text{red}) = \frac{4}{10} = \frac{2}{5}$. For second draw the

$P(\text{red}) = \frac{4}{9}$ prob^y of red is affected by any other set of event in any way, then the event is called dependent event.

Complementary event: ~~The probability of occurance,~~ If the proby of ~~P(A) + P(A)~~ happening is substract from 1, then it is called complementary event.

9, 17, 20, 31

$$P(A) + P(A') = 1$$

Mid-term Exam

1. Introduction (chapter-1)
2. Measure of Central Tendency (Chapter-2)
~~3~~. ~~Measure~~ Moments (chapter 3(B))
3. Regression.

4. what is u ? difference betw Correlation & regression.

x				
y				

① Find Regn line —

② If x = —— then y = ?

All book list — Sadia

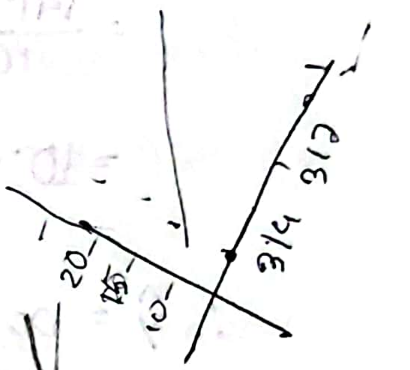
Physics- C to G — S, B

Arraylist to Array — S, B

2016 - 1(d), 2(a)

$y = a + bx$

$10.48 + 0.0109 \times x$



2015 - 5(c)

(ii)

Weight (x)	Distance (y)	x_i^2	$x_i y_i$
314	13.9	98596	4364.6
317	14.0	100489	4438
320	13.9	102400	4448
326	14.1	106276	4596.6
331	14.0	109561	4634
339	14.3	114921	4847.7
346	14.1	119716	4878.6
354	14.5	125316	5133
361	14.5	130321	5234.5
369	14.4	136161	5313.6
$\Sigma x = 3377$	$\Sigma y = 141.7$	$\Sigma x_i^2 = 1143757$	$\Sigma x_i y_i = 47888.6$

$$\beta = \frac{\Sigma x_i y_i - \frac{\Sigma x \Sigma y}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}} = \frac{47888.6 - \frac{3377 \times 141.7}{10}}{1143757 - \frac{(3377)^2}{10}}$$

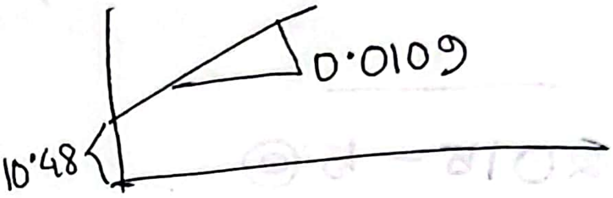
$$= 0.0109$$

$$\alpha = \bar{y} - \beta \bar{x}$$

$$= \frac{\sum y}{n} - \beta \frac{\sum x}{n}$$

$$= \frac{141.7}{10} - 0.0109 \times \frac{3377}{10}$$

$$= 10.48907$$



$$\therefore \hat{y} = \alpha + \beta \hat{x}$$

$$= 10.48907 + 0.0109 \hat{x}$$

Comment: So, with the ~~sace~~ scales of 1 unit of independent variable the dependent variable increases 0.0109 times. on the promotion unit $\alpha = 10.489$.

(iii) If $x = 360$:

$$\hat{y} = \alpha + \beta \hat{x}$$

$$= 10.48907 + 0.0109 \times 360$$

$$= 14.413 \quad \text{Ans.}$$

Central Tendency: In a representative sample, the value of a series of data has always been a tendency to cluster around a certain point ~~or~~ usually at the centre of the series is called central tendency

Regression: Regression is a mathematical measure of the average relationship between two or more variables in terms of original unit of the data.

Linear Regression: The regression between two variables is called linear regression.

The linear regression betw the dependent variable y and independent variable x can be written as equation of a straight line.

$y = a + bx + e$

where, a is the intercept, b is the regression co-efficient or shape and e are random error components.

2018 - 1(c)

Class interval	Frequency (f_i)	Mid value (x_i)	Cumulative frequency f_c	f_ix_i
50 - 60	18	55	18	990
Median 60 - 70	32	65	50	2080
70 - 80	24	75	74	1800
80 - 90	16	85	90	1360
90 - 100	10	95	100	950
				7180

For grouped data,

$$\text{mean}, \bar{x} = \frac{\Sigma f_ix_i}{n} = \frac{7180}{100} = 71.8$$

$$\text{Median}, M_e = L + \frac{\frac{n}{2} - f_c}{f_m} \times c$$

$$= 60 + \frac{50 - 50}{32} \times 10$$

$$= \cancel{51.25}\ \cancel{52.8}\ \cancel{66.8}$$

60 - 3.125

56.875

L = 24 (mode class)

$f_1 = 32 - 18 = 14$

$f_2 = 32 - 24 = 8$

$c = 10$

$$\text{Mode}, M_o = L + \frac{f_1}{f_1 + f_2} \times c$$

$$= 60 + \frac{14}{14 + 8} \times 10$$

$$= 66.36$$

$$s^2 = \frac{\Sigma f_i(x_i - \bar{x})^2}{n} = 149.8 \qquad S.D = 12.24$$

2013

1. ⓒ

Mid value (x_i)	5	10	15	20	25	30	35	40	45	50
frequency (f_i)	1	4	15	31	49	22	25	12	5	3
f_ix_i	5	40	225	620	1225	660	875	480	225	150

$\therefore$ mean, $\bar{x} = \frac{\Sigma f_ix_i}{n}$

$= \frac{5+40+225+620+1225+660+875+480+225+150}{10}$

$= 450.5$

2015-1 ⓐ

The characteristics of a good measure of central tendency are given below:-

1. It should be rigidly (কঠিনভাবে) defined.
2. To facilitate comparisons between data
3. It should be based upon all values of given data
4. It should be capable of further mathematical treatment.
5. It should have sampling stability
6. It should be not be unduly affected by extreme values.

even

2 3 4 5

$$\frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$$

$$= \frac{2nd + 3rd}{2}$$

$$= \frac{3+4}{2}$$

$= 3.5$

1 2 3 4 5

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

$$\frac{5+1}{2}$$

$$\frac{6}{2}$$

3

* Univariate :- When only one variable is studied. Example: Measure of central tendency, Measure of dispersion.

* Bivariate :- When relationship between two variables are studied.

* Multivariate: When relationship between more than two variables are studies.

2011 - 2 (a)

Which measure of central tendency is better than others and why?

Actually it depends on data ~~an~~ or variable.

Type of Variable.	Best measure of central tendency
(name) Nominal	Mode
(ranked ") Ordinal	Median
Interval / Ratio (not skewed)	Mean
Interval / Ratio (skewed)	Median.

2012 - 5 (a)

Why are there two regression lines in a bivariate distribution?

There are always two line of regression, one of x on y and other of y on x. The line of regression of y on x is used to estimate the value of y for any given value of x i.e, when y is dependent variable and x is independent. ~~The~~ .

Comparison of regression and correlation :

Topic	Correlation	Regression : Measures
(1) Definition	Indicates whether there is any relation between the variables	Probable movement of one variable interms of other
(2) Co-efficients	$r_{xy} = r_{yx}$	$b_{xy} \neq b_{yx}$
(3) Limit of Coefficient	$-1 < r < 1$	$-\infty < b < \infty$
(4) Indication	Indicates linear relationship between two variables	Indicates any type of relationship
(5) ~~Measurement~~ Pure number	Correlation co-efficient is a pure number	Regression coefficient is not pure number.
(6) Measurement	Measures degree of association between the variables	Measures form of relationship between one dependent and one or more independent variable.

Defi relation, $b_{xy} \neq b_{yx}$

Coeff. ~~ r_{xy} ~~ $r_{xy} = r_{yx}$ $-\infty < b < \infty$

range $-1 < r < 1$

no

any type

pure no. yes

linear

dependent form & re bet^n one dependent to one or more independent

degree of associ

Meo

$$\beta = \frac{\Sigma x_1 y_1 - \frac{\Sigma x \Sigma y}{n}}{\Sigma_1 x_1^2 - \frac{(\Sigma_1 x)^2}{n}}$$

Prove: Regression co-efficient is independent on origin but depend on scale.

Let, the two variables x and y having n pairs of observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with their respective means $\overline{x}$ and $\overline{y}$.

$\therefore$ regression co-efficient of y on x, $b_{yx} = \frac{(\Sigma_1 x_i - \overline{x})(\Sigma y_1 - \overline{y})}{(\Sigma x_i - \overline{x})^2}$

Let, new variable,

$u_i = \frac{x_i - a}{c}$ $\left[\begin{array}{l} a = \text{origin} \\ c = \text{scale } (c>0) \end{array}\right]$

or, $x_i - a = c u_i$

or, $x_i = a + c u_i$

or, $\frac{\Sigma x_i}{n} = \frac{na}{n} + \frac{c \Sigma u_i}{n}$

or, $\overline{x_i} = a + c\overline{u}$

Another new variable,

$v_i = \frac{y_i - b}{d}$ $\left[\begin{array}{l} b = \text{origin} \\ d = \text{scale } (d>0) \end{array}\right]$

or, $y_i - b = d v_i$

or, $y_i = b + d v_i$

or, $\frac{\Sigma y_i}{n} = \frac{nb}{n} + \frac{d \Sigma v_i}{n}$

$\overline{y} = b + d\overline{v}$

Now, $b_{yx} = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\Sigma(x_i - \overline{x})^2}$

$= \frac{\Sigma(a + cu_i - a - c\overline{u})(b + dv_i - b - d\overline{v})}{\Sigma(a + cu_i - a - c\overline{u})^2}$

$= \frac{cd\,\Sigma(u_i - \overline{u})(v_i - \overline{v})}{c^2\,\Sigma(u_i - \overline{u})^2}$

$= \frac{d}{c} b_{vu}$

$\therefore b_{yx} = \frac{d}{c} b_{vu}$

Similarly,

$b_{xy} = \frac{c}{d} b_{uv}$

$\therefore$ Regression co-efficiend is independent of origin but dependent on variable.

Estimation of regression co-efficient by using the method of least square.

We know the regression equ^n:

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\varepsilon_i = Y_i - \alpha - \beta x_i$$

$$\varepsilon_i^2 = (Y_i - \alpha - \beta x_i)^2 \text{ [squaring both side]}$$

$$\Sigma \varepsilon_i^2 = \Sigma (Y_i - \alpha - \beta x_i)^2 \text{ [summing both side]}$$

Now, differenciating ~~with re~~ $\Sigma \varepsilon_i^2$ with respect to α and β and equating to zero we get,

$$\frac{d}{d\alpha}(\Sigma \varepsilon_i)^2 = 0$$

$$\text{or, } \frac{d}{d\alpha}\Sigma (Y_i - \alpha - \beta x_i)^2 = 0$$

$$\text{or, } 2\Sigma (Y_i - \alpha - \beta x_i)(-1) = 0$$

$$\text{or, } \Sigma (Y_i - \alpha - \beta x_i) = 0$$

$$\text{or, } \Sigma Y_i - \Sigma \alpha - \beta \Sigma x_i = 0$$

$$\text{or, } \Sigma \alpha = \Sigma Y_i - \beta \Sigma x_i$$

$$\text{or, } \frac{n\alpha}{n} = \frac{\Sigma Y_i}{n} - \frac{\beta \Sigma x_i}{n}$$

$$\therefore \alpha = \bar{Y} - \beta \bar{x} \quad \text{--- (i)}$$

$$\beta = \frac{\sum x_i y_1 - \frac{\sum x \sum y}{n}}{\sum x_1^2 - \frac{(\sum x_1)^2}{n}}$$

Similarly,

$$\frac{d}{d\beta}\sum(y_1 - \alpha - \beta x_i)^2 = 0$$

$$\text{or, } 2\sum(y_i - \alpha - \beta x_i)(-x_i) = 0$$

$$\text{or, } \sum(y_i - \alpha - \beta x_i)(x_i) = 0$$

$$\text{or, } \sum x_i y_i - \alpha \sum x_i - \beta \sum x_i^2 = 0$$

$$\text{or, } \sum x_i y_i = (\bar{y} - \beta\bar{x})\sum x_i + \beta \sum x_i^2$$

$$\text{or, } \sum x_i y_i = \bar{y}\sum x_i - \beta\bar{x}\sum x_i + \beta\sum x_i^2$$

$$\text{or, } \sum x_i y_i - \bar{y}\sum x_i = \beta(\sum x_i^2 - \bar{x}\cdot\sum x_i)$$

$$\text{or, } \sum x_i y_i - \frac{\sum y \sum x}{n} = \beta\left(\sum x_i^2 - \frac{\sum x_1 \sum x_i}{n}\right)$$

$$\therefore \beta = \frac{\sum x_i y_i - \frac{\sum x \sum y}{n}}{\sum x_i^2 - \frac{(\sum x_1)^2}{n}} \quad \text{--- (ii)}$$

where α is the intercept of the model and β is the regression co-efficient of y on x.

$$L + \frac{\frac{n}{2} - f_c}{f_m} \times c.$$

$$60 + \frac{\frac{100}{2} - 18}{32} \times 10$$

$$60 + 10$$

$$70$$

$$\frac{\sum x_i y_1 - \frac{\sum x \sum y}{n}}{\sum x_1^2 - \frac{(\sum x)^2}{n}}$$

When ~~an~~ geometric mean is prefered to arithmetic means

(i) The geometric mean is used to calculate average ~~of~~ of ratio and percentage.

(ii) It is used also for computing average rate of increase or decrease

(iii) It is useful for construction of index numbers.

When is median prefered to arithmetic mean?

(i) Median is very useful when our observations are not quantitative.

(ii) It ~~was~~ verily helps us when there exists open interval.

(iii) It is useful for comparing two or more sets of qualitative data.

$AM \geq GM \geq HM$ [For 2 observation]

$$AM = \frac{x_1 + x_2}{2}$$

$$GM = (x_1 + x_2)^{\frac{1}{2}} = \sqrt{x_1 + x_2}$$

$$HM = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Now for any two positive quantity,

$$(\sqrt{x_1} - \sqrt{x_2})^2 \geq 0$$

or, $(\sqrt{x_1})^2 - 2\sqrt{x_1 x_2} + (\sqrt{x_2})^2 \geq 0$

or, $x_1 - 2\sqrt{x_1 x_2} + x_2 \geq 0$

or, $x_1 + x_2 \geq 2\sqrt{x_1 x_2}$

$\therefore \dfrac{x_1 + x_2}{2} \geq \sqrt{x_1 x_2}$

$\therefore$ AM $\geq$ GM

For any two positive quantity,

$$\left(\frac{1}{\sqrt{x_1}} - \frac{1}{\sqrt{x_2}}\right)^2 \geq 0$$

or, $\dfrac{1}{x_1} - \dfrac{2}{\sqrt{x_1 x_2}} + \dfrac{1}{x_2} \geq 0$

or, $\dfrac{1}{x_1} + \dfrac{1}{x_2} \geq \dfrac{2}{\sqrt{x_1 x_2}}$

or $\sqrt{x_1 x_1} \geq \dfrac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$

$\therefore$ GM $\geq$ HM.

$\therefore$ AM $\geq$ GM $\geq$ HM.

$\boxed{\text{Proved}}$

AM=GM=HM

Let, there is n number of variables $x_1, x_2, x_3 \dots x_n$.

$x_1 = x_2 = x_3 = \dots = x_n = k$

$$AM = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$= \frac{k + k + \dots + k}{n}$$

$$= \frac{nk}{n}$$

$$= k.$$

$$GM = (x_1 . x_2 . x_3 \dots x_n)^{\frac{1}{n}}$$

$$= (k.k.k \dots k)^{\frac{1}{n}}$$

$$= (k^n)^{\frac{1}{n}}$$

$$= k$$

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

$$= \frac{n}{\frac{1}{k} + \frac{1}{k} + \dots + \frac{1}{k}}$$

$$= \frac{n}{\frac{n}{k}}$$

$$= n \times \frac{k}{n}$$

$$= k$$

$\therefore$ AM = GM = HM.

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

C.W. $\boxed{-1 \le r \le 1}$ *

14-11-19

Sample Space: The collection of all possible outcomes of an experiment is called S.S.

Example:- For one times tossing a coin the sample space will, $S = \{H, T\}$. For two times, $S = \{HH, TH, HT, TT\}$

~~Statemen~~

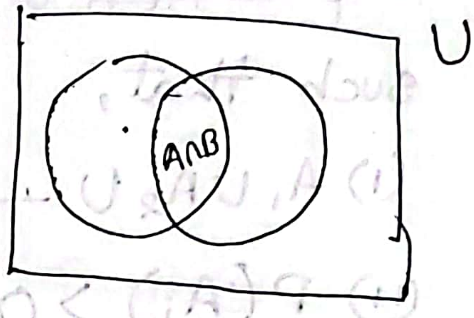
Theorem-1: The additive law of probability for non-mutually exclusive events.

Statement: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Symblically, if A & B are ~~non~~ two non-mutually exclusive event then

Proof: Let A & B are two non-mutually exclusive events of the obtained Sample Space S from a random experiment E.

Then the venn diagram is given by,



Here, $n(S)$, the total number of elements in sample S.

$n(A)$; the total number of element for event A.

$n(B)$; " " " " " B

$n(A\cap B)$; " " " " " compound A & B.

Thus, we know,

$$P(A) = \frac{n(A)}{n(S)} \; ; \; P(B) = \frac{n(B)}{n(S)} \; \& \; P(A\cap B) = \frac{n(A\cap B)}{n(S)}.$$

Now,

$$n(A\cup B) = \{n(A) - n(A\cap B)\} - n(A\cap B) + \{n(B) - n(A\cap B)\}$$

(After 2 pages)

State & Prove Bayes theorem :-

*** **Bayes Theorem:** Let,

$\{A_1, A_2, \ldots A_i, \ldots A_k\}$ be a set of mutually exclusive & exhustive events forming a ~~partion~~ partition of the sample space S such that,

(i) $A_1 \cup A_2 \cup \ldots \cup A_k = S$

(ii) $P(A_i) > 0$; $i = 1 \ldots\ldots k$

~~(iii)~~ Let, B be the event of S such that $P(B) > 0$

then $P(A_i \overset{\text{given}}{|} B) = \dfrac{P(A_i)\,P(B|A_i)}{\sum\limits_{i=1}^{k} P(A_i)P(B|A_i)}$; $i = 1, \ldots k$

which is Bayes theorem.

Proof :- According to the Bayes theorem A_i & B are dependent.

Then by the formula for multiplication rule, we have

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

$$P(A_i \cap B) = P(B)\, P(A_i|B) \quad \text{--- (i)}$$

$$P(A_i \cap B) = P(A_i) . P(B|A_i) \quad \text{--- (ii)}$$

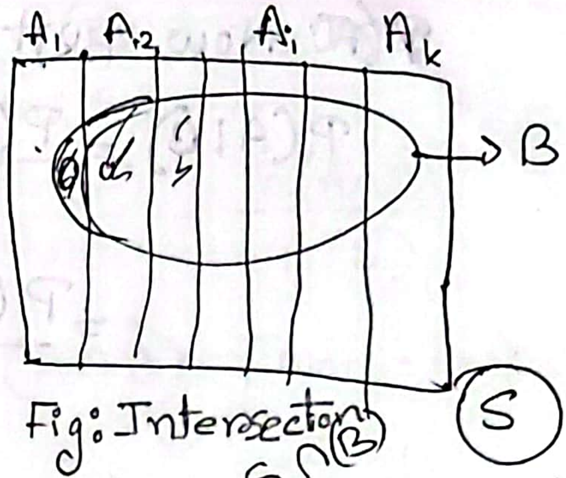


Fig: Intersection of B with $A_1, A_2, \ldots A_k$

$B = S \cap B$

From (i) we have,

$$P(A_i \cap B) = P(B)\, P(A_i|B)$$

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

$$= \frac{P(A_i)\, P(B|A_i)}{P(B)} \quad \text{from (ii)}$$

$$\therefore P(A_i|B) = \frac{P(A_i) P(B|A_i)}{P(B)} \quad \text{--- (iii)}$$

$$S = A_1 \cup A_2 \cup \ldots \cup A_k$$

$$B = S \cap B$$

$$= (A_1 \cup A_2 \cup \ldots \cup A_k) \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \cancel{(A_3 \cap)} \ldots \cup (A_k \cap B).$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_k \cap B)$$

$$= \sum_{i=1}^{k} P(A_i \cap B)$$

$$\therefore P(B) = \sum_{i=1}^{k} . P(A_i)\, P(B|A_i) \quad \text{[From (ii)]}$$

~~P(A)~~ Now putting the value of B in (iii),

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

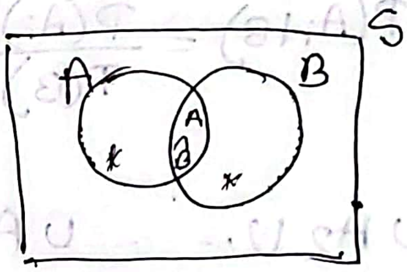
$$= \frac{P(A_i)\,P(B|A_i)}{\sum P(A_i)\,P(B|A_i)}$$

[Proved]

C.W. Probability 18.11.19

~~The~~

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



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

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

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

$$n(A\cup B) = \{n(A) - n(A\cap B)\} + n(A\cap B) + \{n(B) - n(A\cap B)\}$$

$$= n(A) + n(B) - n(A\cap B)$$

18-11-19 16

Jefa Jeba, Jannat, Jnc. Sadia, Tasfi, Prema, Abir, Nahid, Green, Fahim, Mithun, Riod, sabuj, Dip, Nadim, ~~Mahfujur~~, Mujahid

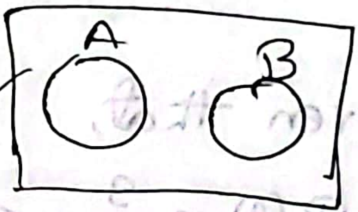
$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A) + n(B) - n(A \cap B)}{n(S)}$$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad [\text{Non-mutual}]$$

Additive law of Prob^y for two mutually exclusive events.

$$P(A \cup B) = P(A) + P(B)$$



Statements: Multiplicative law of prob. for two $\boxed{\text{dependent}}$ events.

For two event A & B the multiplicative law says that,

$$* \; P(A \cap B) = \underbrace{P(B)}_{\text{(uncondi)}} \underbrace{P(A \setminus B)}_{\text{(conditional)}}$$

$$* \; P(A \cap B) = P(A)\, P(B \setminus A)$$

the joint prob^y will be unconditional event will conditional event in term of probability.

Proof: For two event we know for conditional probability,

$$P(A \setminus B) = \frac{\cancel{P}n(A \cap B)}{n\cancel{P}(B)} \quad \text{--- (i)}$$

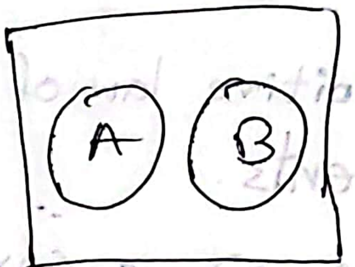
$$P(B \setminus A) = \frac{n\cancel{P}(A \cap B)}{n\cancel{P}(A)} \quad \text{--- (ii)}$$

Now, $P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{n(A \cap B)}{n(A)} \times \frac{n(A)}{n(S)} = P(B \backslash A)\, P(A)$

$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{n(A \cap B)}{n(B)} \times \frac{n(B)}{n(S)} = P(A \backslash B)\, P(B)$

Multiplicative law of two independent events:

$P(A \cap B) = P(A) P(B)$



Given that,

$P(A) = \frac{3}{8}$, $P(B) = \frac{5}{8}$ & $P(A \cup B) = \frac{3}{4}$

Find $P(A|B)$, $P(B \backslash A)$ & A & B are independent.

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{5/8} = \frac{1}{\cancel{4}} \times \frac{\cancel{8}^{2}}{5} = \frac{2}{5}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cap B) = \frac{3}{8} + \frac{5}{8} - \frac{3}{4}$

$= \frac{3+5-6}{8}$

$= \frac{2}{8}$

$= \frac{1}{4}$

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

$= \frac{1}{\cancel{4}} \times \frac{\cancel{8}^{2}}{3}$

$= \frac{2}{3}$

For independent

$P(A\cap B) = \frac{1}{4}$

$P(A)P(B) = \frac{3}{8} \times \frac{5}{8} = \frac{15}{64}$

$\therefore P(A\cap B) \neq P(A)P(B)$ $\therefore$ A & B are ~~det~~ not independent.

At least/any one/or/ $\rightarrow P(A\cup B)$

Both A & B $\longrightarrow P(A\cap B)$.

In a survey of 100 ~~ro~~ readers, it was found that 40 read the daily Ittefaq & ~~10~~ 15 read the daily Star and 10 are both. What is the proby of a person reading at least one of ~~nea~~ newspaper.

$P(A\cup B) = P(I) + P(S) - P(I\cap S)$

$= \frac{40}{100} + \frac{15}{100} - \frac{10}{100}$

$= \cancel{0.45}$

$= \frac{45}{100} = \frac{9}{20}$

$\boxed{\text{Hypothesis Testing}}$

Q.9

21.11.19

Hypothesis testing

$\boxed{\bar{x} = 22 \quad s = 2.50 \quad n = 25}$

20.25 16.16
2625. 17.25

Sample mean

Test the hypothesis that the ~~population~~ sample average age is 30.

(one tail > / <)

Sol^n

(~~average~~/mean) population

step-1 $\begin{cases} H_0 : \mu = 30 & \text{[null hypothesis]} \\ H_1 : \mu_1 \neq 30 & \text{[two tail]} \end{cases}$

step-2 { level of significant $\alpha = 0.05$ [কিছু না দেওয়া থাকলে 0.05 5%]

step-3 { $t_{cal} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ (statistic).

Step-4 { Determine the critical region

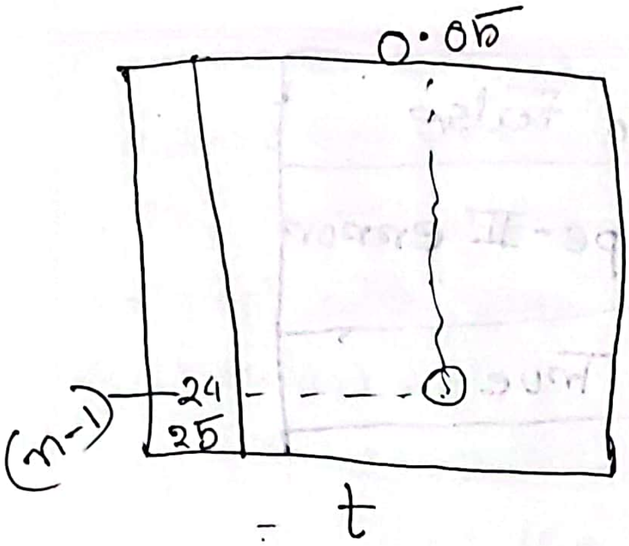
t max value 30

Step-5 { Calculating test statistics

Step 6 { Taking decision

H_0 accept

H_1 accept



1% 0.01	1.64
5% 0.05	1.96

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(step-3)

If, calculated value (t_{cal}) > then value of t (t_{tab}) ~~then~~ then the null hypothesis rejected

Alternative ~~w~~ $t_{tab} > t_{cal}$ = H_0 fail to reject

Defi: ~~Alt~~ null, two $t_{tab} < t_{cal}$ = H_0 accept.

($df = 24$, $t_{tab} = 1.710$)

t_{tab}

Error

Type-I → কম error

Type-II → বেশি "

	H_0 True	H_0 False
H_0 Accept	True	Type-II error
H_0 Reject	α Type-I error	True

test

Statistics

C.W. Hypothesis testing 25.11.19 Monday

Step 1: Setting up of hypothesis
" 2: Set up a suitable level of significance
" 3: Computation of test statistic.
" 4: Determine the critical region.
" 5: Calculating the test Statistics
" 6: Taking Decision

Accept H_0 (Null hypotesis)

Accept H_1. (Alternate hypothesis)

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \quad t_{tab} = t_{(n-1), 0.05 / \text{if given}}$$

When $n > 30$ we put z & use data for z-table

0.01 1%	1.64
0.05 5%	1.96

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

A sample of 25 items is taken from a popu^n whose standard deviation is 1·5. The mean of the sample is 25. test whether the sample has come from the pop^n with mean 26·8 (Test at 5% level of significance).

Step 1 : $H_0 : \mu = 26{\cdot}8$ (≠ then two tailed)

$H_1 : \mu \neq 26{\cdot}8$ (> / < → then One tailed).

Step 2 : $\alpha = 0{\cdot}05$

Step 3 : $|t_{cal}| = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| = \left|\frac{25-26{\cdot}8}{1{\cdot}5/\sqrt{25}}\right| = |-6| = 6$

Step 4 : $t_{tab} = t_{(25-1),0{\cdot}05} = 2{\cdot}33$

Step 5 : $t_{tab} < t_{cal}$

Step 6 : H_0 is ~~reg~~ rejected. H_1 (Alternate) is accepted.

$t_{tab} > t_{cal} \rightarrow H_0$ is accepted

$t_{tab} < t_{cal} \rightarrow H_0$ is rejected, H_1 accepted.

~~Group~~

$\underline{H_0}$ table > cal
table accept

table < Cal
reject

Two mean test

Group-1	Group-2
$\bar{x}_1 = 20.5$	$\bar{x}_2 = 25.5$
$s_1 = 2.5$	$s_2 = 1.5$
$n_1 = 20$	$n_2 = 22$

$$z_{tab} = z_{(n_1+n_2-2),0.05}$$
$$= z_{(20+22-2),0.05}$$
$$= z_{40,0.05}$$
$$= 1.96.$$

Test whether there are any significant mean difference.

$$z_{cal} = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| = \left|\frac{20.5 - 25.5}{\sqrt{\frac{2.5^2}{20} + \frac{1.5^2}{20}}}\right| = |-7.76| = 7.76$$

$z_{cal} > z_{tab}$

$\therefore H_0$ is rejected. Alternate accepted.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

\# 20, 21, 22, 20, 24   22, 23, 25, 26, 29

$\bar{x}, s, n_1$ $\bar{x}_2, s_2, n_2$

The average monthly income of 80 teachers in Dhaka city college is taka 20 thousand with standard deviation tk 3000 per month. While, for 60 teachers in Dhaka Commerce college the corresponding quantities are tk 22.5 thousand and 5000 respectively. Do, the above data indicate any real difference betn the average monthly income between two college.

$n_1 = 80$ $\quad n_2 = 60 \quad \alpha = 0.05$

$\bar{x}_1 = 20 \quad \bar{x}_2 = 22.5$

$S_1 = 3 \quad S_2 = 5.$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$|Z_{cal}| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{n}}} \right| = \left| \frac{20 - 22.5}{\sqrt{\frac{3^2}{80} + \frac{5^2}{60}}} \right| = |-3.436| = 3.436$$

$$Z_{tab} = Z_{(80+60-2),\, 0.05} = 1.96$$

$Z_{tab} < Z_{cal}$

$\therefore$ H_0 not accepted, i.e Alternate H_1 is accepted.

Hence there is significant difference between the two college regarding monthly income

28-11-19
Thursday

~~norta~~ normal test → z test	t test
$n > 30$	$n < 30$
$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
$z = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
If variace is known but sample $n<30$	If variance ~~kno~~ unknown

9.2, 8.7, 8.6, 8.8, 8.5, 8.7, 9.0

At the 0.01 significance level can we conclude that the mean weight is less than 9.0?

Step-1 $H_0 = 9.0$

$H_1 < 9.0$

Step-2 level of significance $\alpha = 0.01$

Step-3 $|t_{cal}| = \frac{\bar{x} - \mu}{s/\sqrt{n}} = |-2.53| = 2.53$

$\bar{x} = 8.78$

$n = 7$

$s = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2} = 0.23$

μ

$t_{tab} = 3.143 > t_{cal}$

$\therefore H_0$ is accept

$a - b = 4$

$ab = 60$

$ab = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$

$60 + \left(\frac{4}{2}\right)^2 = \left(\frac{a+b}{2}\right)^2$

$\sqrt{60+4} \times 2 = a+b$

$a+b = 16$

iii| Mean |

$\boxed{\text{Variance}}$ σ

$H_0 : \sigma = \sigma_0$

$H_1 : \sigma \neq \sigma_0$ or $\sigma > \sigma_0$ or $\sigma < \sigma_0$

test Statistic $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$ (χ^2 - chi-square বর্গ)

s^2 = sample variance

σ^2 = population

60, 67, 71, 55, 66, 74, 81, 80, 77, 85

Do this data singnificatly satistied that the population variance is 27. [$\alpha = 0.05$]

Step 1 : $H_0 : \sigma = 27$

$H_1 : \sigma \neq 27$

$s = 92.93$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$= \frac{(10-1)\,92.93}{27}$$

$$= 30.97$$

$\chi^2_{cal} = 30.977$

$\chi^2_{tab} = \chi^2_{(\frac{\alpha}{2}),(n-1)} = \chi^2_{0.025,9} = 19.822$.

$\chi^2_{cal} > \chi^2_{tab}$

H_0 reject.

02.12.19
Monday.

Hypothesis testing

Null Hypothesis: The statistical hypothesis that is set up for testing a hypothesis is known as null hypothesis.

The hypothesis that we assume is said to be null hypothesis. A null hypothesis states that there is no difference between a sample estimate and the true population value. The null hypothesis is generally denoted by H_0.

Alternate hypothesis: The negative of null hypothesis is called alternate hypothesis. In other words, any hypothesis which is not a null hypothesis is called an alternate hypothesis. Alternate hypothesis is denoted by H_1 or H_a

Example: If $H_0: \mu = 0$ then its alternate hypothesis may be $H_1: \mu \neq 0$, $H_1: \mu > 0$ or $H_1: \mu < 0$

If $H_0: \mu_1 = \mu_2$ then

$H_1: \mu_1 \neq \mu_2$; $H_1: \mu_1 > \mu_2$ or, $H_1: \mu_1 < \mu_2$ etc.

Type-I error: In a test of hypothesis, the ~~type-I~~ type-I error occurs when the null hypothesis H_0 is rejected although it was true.

The probability of a type-I error is denoted by α, α is also known as the level of significance.

Type-II error: In a test of hypothesis, the typ-II error occurs when the null hypothesis H_0 is not rejected (i.e accepted) although it was false.

The probability of typ-II error is denoted by β.

Level of Significance: The maximum probability of making a type I error, generally specified in a test. If there rests no specification then we consider the level of significance $\alpha = 0.05$. That implies we are 95% (0.95) confident about the significance of our decision & we have 5% (0.05) chance of occurring type-I error.

Degree of Freedom (d.f): The number of cases to which the values can be assigned arbitrarily without violating the restrictions placed.

For example: If we are to choose 5 numbers with their total 120, then number restriction impose is 1. and that is, their total 120. Here, although we are to choose 5 but we can choose 4 numbers arbitrarily. Here, the degree of freedom, $v = n - r = 5 - 1 = 4$, where r refers to number of independent constraints.

Acceptance Region: An acceptance region is a set of possible values of the test statistic (in a test of hypothesis) that leads the null hypothesis to be accepted.

~~Acer~~ Rejection region: (or Critical region): A rejection region is a set of possible values of the test statistic (in a test hypothesis) that leads the null hypothesis to be rejected.

~~Simple~~ Parameter Hypothesis: Any hypothesis about the parameter of a population distribution is known as parametric hypothesis.

For example: Suppose, the average age of all students in college is 20 years. The hypothesis about their mean value is called parametric hypothesis.

Statistical hypothesis: It is some assumption or statement, which may or may not be true, about a population or about the probability of given population, which we test on basis of evidence from random sample.

02.12.19
Monday

C.W. A company two types of bulbs, A & B 200 bulbs of each type were tested and it was found that type A has mean life of 2560 hours and ~~SD~~ 90 hours, where as type B had 2650 hours mean life with SD 75 hours. Is there a significant difference between the mean life of types.

~~$n_1 = 200$~~ ~~$n_2 = 2$~~

$n_1 = n_2 = 200$

$S_1 = 90$, $S_2 = 75$

$\bar{x}_1 = 2560$, $\bar{x}_2 = 2650$

~~Sol~~ Step 1: $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Step 2: $\alpha = 0.05$

Step - 3: $|z_{cal}| = \left\{ \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} \right\}$

$= \left\{ \frac{2560 - 2650}{\sqrt{\frac{90^2}{200} + \frac{75^2}{200}}} \right\}$

$= \{-10.86\}$

~~$= 10.86$~~ $|z_{cal}| = 10.86$

Step-4:

$z_{tab} = z_{0.05, 398} = 1.96$

$z_{tab} = z_{(n_1+n_2-2), 0.05}$ / if any

Step-5:

$z_{tab} < z_{cal}$

$\therefore H_0$ is ~~neg~~ rejected.

H_1 i.e. alternate hypothesis is accepted.

~~$\therefore$ There~~

0.01α $\boxed{1.64}$

$z_{(n_1+n_2-2)}$ 0.05α 1.96

∴ There is a significant difference between the mean lifes of the type.

$\boxed{\text{Two Varience test (F Test)}}$

$n_1 = 15$ $\qquad$ $n_2 = 10$

$\bar{x}_1 = 24.25$ $\qquad$ $\bar{x}_2 = 20.25$

$S_1 = 2.50$ $\qquad$ $S_2 = 1.50$

---

for previous math

Comment: $z_{cal} > z_{tab}$; So null hypothesis will be rjected. That means there are significanc difference betn two groups A & B in mean life time

---

Is there any significance Variation betw two groups.

$$F = \frac{S_1^2}{S_2^2}$$

$$F_{tab} = F_{(n_1-1),(n_2-1),\alpha}$$

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 \neq \sigma_2$

$$\boxed{F_{cal} = \frac{2.5^2}{1.50} = 2.78}$$

$$F_{tab} = F_{(n_1-1),(n_2-1),\alpha}$$

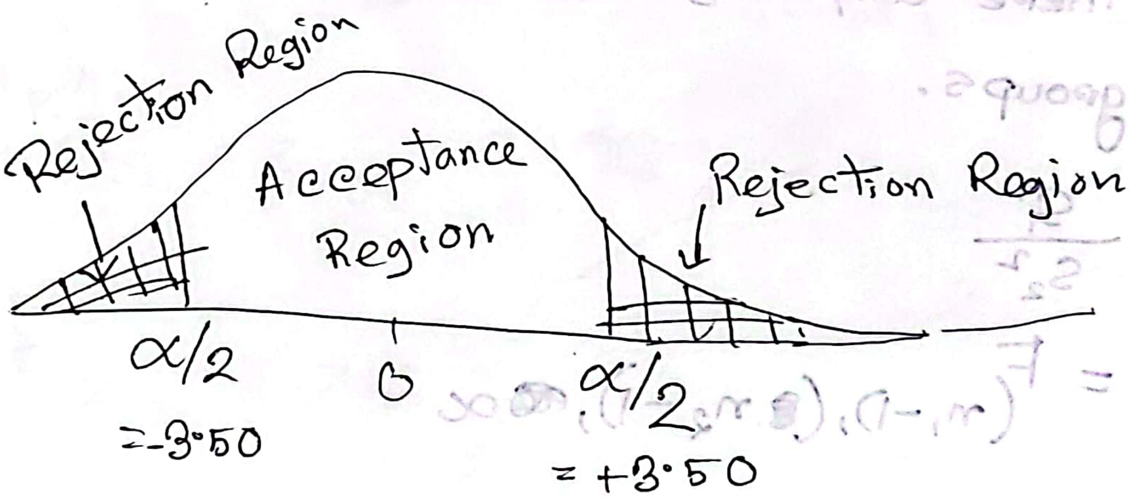
$$= F_{4,9,0.05}$$

$$= 3.6331$$

$F_{tab} > F_{cal}$.

Comment: $\therefore H_0$ is accepted.

So, no variation between two groups.



05-12-19
Thursday

C.W.

2x2 Contigency Table

Do these figures support the hypothesis that skilled fathers have intelligent boys?

Fathers \ Boys	Intelligent	Unintelligent	
Skilled	40 (39.67)	30 (30.30)	70
Unskilled	70 (70.30)	54 (53.7)	124
	110	84	194

(grand total).

observed

$\frac{70 \times 110}{194}$ Expected

$\frac{70 \times 84}{194}$ Ex

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O	E	$(O-E)^2$	$(O-E)^2/E$
40	39.7	0.090	0.00226
30	30.3	0.090	0.0029
70	70.3	0.090	0.0012
54	53.7	0.090	0.0016

$$\chi^2 = \frac{\Sigma (O-E)^2}{E} = 0.00796$$

(Always try that alternative is accepted).

H_0 : No association

H_1 : Association

$$\chi^2_{tab} = \chi^2_{(R-1),(c-1),\ \alpha}$$

$$= \chi^2_{1,1,0.05}$$

$$= 3.84$$

$\therefore$ There is no association ~~of~~ between father's skill and boy's intellegence.

$\therefore H_0$ ~~fal~~ fail to reject.

$$E_{ij} = \frac{R_i \times C_j}{G}$$

$$E_{11} = \frac{R_1 \times C_1}{G}$$

$$E_{12} = \frac{R_1 C_2}{G}$$

$$E_{21} = \frac{R_2 C_1}{G}$$

$$E_{22} = \frac{R_2 C_2}{G}$$

	a	b	$a+b=R_1$
	c	d	$c+d=R_2$
	$a+c=C_1$	$b+d=C_2$	$R_1+R_2=C_1+C_2=G$

How do you test for 2×2 contingency table?

~~step~~ Setting up hypothesis –

H_0 : No Assocn ..

H_1 : Associa

Test ~~bgg~~

		College	
		Private	Government	
Income	Low	200 (360)	400 (240)	600
	High	1000 (840)	400 (560)	1400
		1200	800	2000

Test any association betn college ~~&~~ type & ~~inc~~ income. $\boxed{\chi^2_{1, 0.05} = 3.84}$

H_0 : No association

H_1 : Association.

O	E	$(O-E)^2$	$(O-E)^2/E$
200	360	25600	71·11
400	240	25600	106·66
1000	840	25600	30·47
400	560	25600	45·71

chi-square (χ^2) কাই

$\chi^2_{cal} = 253·95$

$\chi^2_{tab} = 3·84$

H_0: ~~fail to~~ reject.

~~There is no association betw college & income.~~

H_1 : There is association betw college & income.

C.W

09.12.19
Monday.

Marriage adjustment Score.

	Very low	Low	High	Very high	
College	24	97	62	58	241
High School	22	28	30	41	121
Middle School	32	10	11	20	73
	78	135	103	119	485

Level of education

(O) Original	(E) Expected	$(O-E)^2$	$(O-E)^2/E$
24	43.21	369.02	8.54
97	74.79	493.28	6.59
62	57.06	24.40	0.42
58	65.92	62.72	0.95
22	21.69	0.096	0.00442
28	37.55	91.20	2.42
30	28.65	1.82	0.06
41	33.10	62.41	1.885
32	13.08	357.96	27.36
10	22.65	160.02	7.06
11	17.28	39.43	2.28
20	19.97	0.0009	4.5×10^{-5}

$\chi^2_{cal} = \sum \frac{(O-E)^2}{E} = 57.56$

$\chi^2_{tab} = \chi^2_{(R-1)(C-1),\alpha} = 2, 3, 0.05 = 12.59$

$\chi^2_{cal} > \chi^2_{tab}$

H_0 = no asso.

$\therefore H_1$ is accepted.

Comment: H_0 is rejected ~~Ho~~ $\chi^2_{cal} > \chi^2_{tab}$.

$\therefore$ Alternate accepted

Hypothesis :- Any statement about the ~~poor~~ population is called hypothesis.

Hypothesis → Null, Alternate.

~~Alter~~ Null : The statistical hypothesis that is set up for testing ~~a~~ a hypothesis is known as null hypothesis.

Alternate : A hypothesis that is not null hypothesis called alternate or opposite of null hypthesis hypothesis.

Example: Null hypothesis : $H_0 : \mu = \mu_0$

Alternate hypothesis : H_1 or $H_A : \mu \neq \mu_0$

Difference between Null & alter hy... (Page 6-28)

Type I error: In a test of hypothesis, the type I error occurs when the null hypothesis Ho is ~~not~~ rejected although it was true (false) — type II

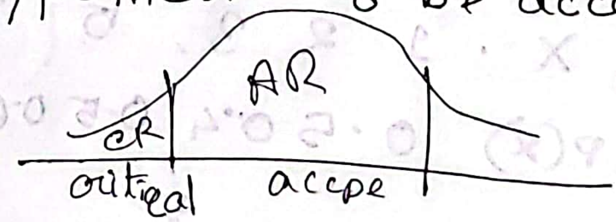
Level of significance (α):- The level of significance is the ~~min~~ maximum prob^y of making a type-I error.

$\alpha = 0.05$ means
$\{H_0 ; \mu = \mu_0$ it would be occured 5%$\}$

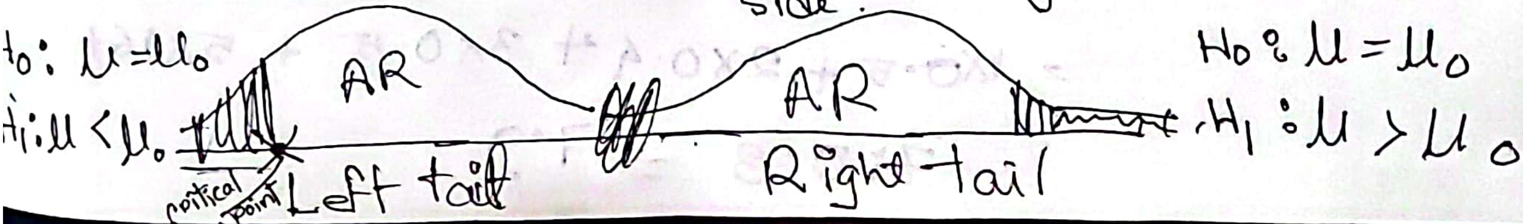
Test Statistics: ~~The~~ Test statistic is a function of sample observation.

Example: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ which is the f^n of sample.

Acceptance region: An acceptance region is a set of possible values of the test statistic that leads the ~~null~~ null hypothesis to be accepted.



One-tail test: — Rejection region consist one side.



Two tail test:
— rejection region consist two side.

Degrees of freedom:-
— No of independent variable.

$(n-1)$ d f

$(n-1)$ ~~dot~~ df ~~degree~~

C.W. 12-12-19

Mathematical Expectation

Sampling 1, 2, 3, 5, 7

$\mu = \bar{x} = \frac{1+2+3+5+7}{5}$ = value = 3.6

Probability distribution.

X	1	2	3	5	7
P(x)	0.5	0.4	0.5	0.6	0.3

$\mu = E(x) = \sum_{i=1}^{5} x\,p(x)$

$= 1\times 0.5 + 2\times 0.4 + 3\times 0.5 + 5\times 0.6 + 7\times 0.3 = 7.9$

If a dice is thrown one time. Find the probability distⁿ.

X :	1	2	3	4	5	6
$P(x)$:	1/6	1/6	1/6	1/6	1/6	1/6

$\mu = E(x) = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + \frac{6}{6}$

$= 3.5$

$E(x) = \sum x P(x)$

$E(x^2) = \sum x^2 P(x)$

$V(x) = E\{X - E(x)\}^2$

$= E(x^2) - (E(x))^2$

X :	0	1	2
$P(x)$:	1/4	1/2	1/4

Find $E(x)$; $E(x^2)$ & $V(x)$

$E(x) = 0\times\frac{1}{4} + 1\times\frac{1}{2} + 2\frac{1}{4} = 1$

$E(x^2) = 0^2\times\frac{1}{4} + 1^2\times\frac{1}{2} + 2^2\times\frac{1}{4} = \frac{1}{2} + 1 = 1.5$

$V(x) = E(x^2) - (E(x))^2 = 1.5 - 1 = \frac{1}{2}$

Rules and laws of Expectation:

(i) $E(c) = c$ $\quad E(3) = 3$

(ii) $E(ax \pm b) = aE(x) \pm b$ $\quad E(3x+2) = 3E(x) + 2$

(iii) $E(x \pm y) = E(x) \pm E(y)$ $\quad E(x) \pm E(y)$

(iv) $E(xy) = E(x)\,E(y)$ $\quad E(xy) = E(x)\,E(y)$.

$E(2x) = 2E(x) = 2\sum x\,p(x) = 2 \times 1 = 2$

(i) $v(c) = 0$ $\quad v(ax) = a^2 v(x)$.

(ii) $v(ax \pm b) = a^2 v(x)$

(iii) $v(x \pm y) = v(x) \pm v(y) \pm \cancel{Cov}(x, y)$ (coefficient)

(iv) $\cancel{v(xy)} = Cov(x, y) = E\{(x - E(x))(y - E(y))\}$

$CoV(x, y) = x_i y_i -$

Random Variable:-

When a variable may have the values which can not be known in advance; rather the values depend on chance then the variable might be termed as RV.

RV

- Discrete: # of heads in two successive tosses of a fair coin
- Continuous: height of ICE students.

Write down the probability distribution of the number of heads in two successive ~~two~~ tosses of a fair coin. $\{HH, HT, TH, TT\}$

$P(H) = P(T) = \frac{1}{2}$

x	0	1	2
$P(x)$	1/4	1/2	1/4

When head $= 0$

Sample Space $= \{TT\}$

$P(\underset{=0}{H}) = P\{TT\} = P\{T\}P\{T\}$

$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Head $= 1$,

$S.S = \{HT, TH\}$.

$P(H=1) = P\{HT\} + P\{TH\}$

$= P\{H\}P\{T\} + P\{T\}P\{H\}$

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

$= \frac{1}{4} + \frac{1}{4}$

$= \frac{1}{2}$

Head $= 2$

$S.S = \{HH\}$

$P(H=2) = P\{HH\} = P\{H\}P\{H\}$

$= \frac{1}{2} \times \frac{1}{2}$

$= \frac{1}{4}$

$f(x) = \frac{1}{12}(x+2)$; $1 < x < 3$ then find

$P(x > 0)$.

<u>Solⁿ:-</u> $P(x>2) = \int_2^3 f(x)\,dx$

$$= \int_2^3 \frac{1}{12}(x+2)\,dx$$

$$= \int_2^3 \left(\frac{x}{12} + \frac{2}{16}\right) dx$$

$$= \int_2^3 \frac{x}{12}\,dx + \frac{2}{16}\int_2^3 dx$$

$$= \frac{1}{12}\left[\frac{x^2}{2}\right]_2^3 + \cancel{\frac{2}{16}}\ \frac{1}{6}\,[x]_2^3$$

$$= \frac{1}{24}(9-4) + \frac{1}{6}$$

8, 24
2 | 4, 12
2 | 4, 6
2, 3

$$= \frac{5}{24} + \frac{1}{6}$$

$$= \frac{3}{8}\ \text{(Ans)}$$

1.

$P(2 < x < 3) = \int_2^3 f(x)\,dx = 3/8$

Distribution – Binomial
– Poisson
– Normal
– Bivariate

19-12-19
Thursday

C.W

Binomial distribution:-

$P(x) = f(x; n, p) = {}^{n}c_x \, p^x \underbrace{(1-p)}_{q}^{n-x}$ $x = 0, 1, 2, 3, \dots, n$

$\boxed{p + q = 1}$

x = number of ~~trail~~ success

n = ~~repeated~~ " ~~success~~ trial.

P = Probability of success

* Find the mean & variance of Binomial dist^n

$E(x) = \sum x \, P(x)$

$= \sum_{x=0}^{n} x \; {}^{n}c_x \, p^x q^{n-x}$

$= 0 \; {}^{n}c_0 \, p^0 q^{n-0} + {}^{n}c_1 \, p q^{n-1} + 2 \, {}^{n}c_2 \, p^2 q^{n-2}$

$+ \dots + n \, {}^{n}c_n \, p^n q^{n-n}$

$= 0 + n p q^{n-1} + \cancel{2 \, {}^{n}c_2 \, p^2 q^{n-2}} +$

$2 \cdot \frac{n!}{2!\,(n-2)!} \cdot p^2 \cdot q^{n-2} + \dots + n \cdot 1 \cdot p^n q^0$

${}^{n}p_r = \frac{n!}{(n-r)!}$

${}^{n}c_r = \frac{n!}{r!\,(n-r)!}$

$= n p q^{n-1} + \frac{n(n-1)(n-2)!}{(n-2)!} p^2 q^{n-2} + \dots + n p^n$

$= n p q^{n-1} + n(n-1) p^2 q^{n-2} + n p^n$

$= np \left(q^{n-1} + (n-1) p q^{n-2} + \dots + p^{n-1} \right)$

$= np \, (q + p)^{n-1}$

$= np \, (1)^{n-1}$

$= np$

$$V(x) = E(x^2) - (E(x))^2$$

$$E(x^2) = E\{x(x-1)+x\}$$

$$= E[x(x-1)] + E(x)$$

$$= \sum x(x-1)\, P(x) + np$$

$$= \sum_{x=0}^{n} x(x-1)\, {}^nC_x\, p^x q^{n-x} + np$$

$$= 0 + 0 + 2(2-1)\, {}^nC_2\, p^2 q^{n-2} +$$

$$3(3-1)\, {}^nC_3\, p^3 q^{n-3} + \ldots$$

$$n(n-1)\, {}^nC_n\, p^n q^{n-n} + np$$

$$= 2\, \frac{n(n-1)(n-2)!}{2!\,(n-2)!}\, p^2 q^{n-2} + 6\, \frac{n(n-1)(n-2)(n-3)!}{3!\,(n-3)!}\, p^3 q^{n-3}$$

$$+ \ldots + n(n-1)p^n + np$$

$$= n(n-1)\, p^2 q^{n-2} + n(n-1)(n-2)\, p^3 q^{n-3} +$$

$$\ldots + n(n-1)\, p^n + np$$

$$= n(n-1)p^2 \left[q^{n-2} + (n-2)\, p\, q^{n-3} + \ldots + p^{n-2} \right] + np$$

$= n(n-1)p^2 (q+p)^{n-2} + np$

$= n(n-1)p^2 + np$

$= n^2p^2 - np^2 + np$

$V(x) = E(x^2) - (E(x))^2$

$= n^2p^2 - np^2 + np - (np)^2$

$= n^2p^2 - np^2 + np + n^2p^2$

$= np - np^2$

$= np(1-p)$

$= npq$

$E(x) > V(x)$

~~Poisson Distn~~ <u>Binomial distn</u> :- If x is a random variable then the binomial distribution can be defined & denoted by - $f(x; n, p) = {}^nc_x p^x q^{n-x}$; $x = 0, 1, 2, 3 \ldots n$

Uses:-

① Binomial $dist^n$ can be used to derive poissons $dist^n$

② Test ratio of small sample

③ Frequency distribution

Poisson's

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} ; x = 1, 2, \ldots, \infty$$

$$E(x) = \sum x P(x)$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{x\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{x\lambda^x}{x(x-1)!}$$

$$= \cancel{e^{\lambda} = 0 \cdot e^{-\lambda} = 1 \cdot \frac{e^{-\lambda}\lambda^1}{1!}} + \cancel{= e^{-\lambda}\left(\frac{0\cdot\lambda^0}{0!} + \frac{1\cdot\lambda^1}{1!} + \frac{2\lambda^2}{2!} + \ldots\right)}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{(x-1)!}$$

$$\cancel{= e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\ldots\right)} = 0\, P(0) + 1\, P(1) + 2\, P(2) + \ldots$$

$$\cancel{= e^{-\lambda}\lambda \cdot e^{\lambda} = e^{-\lambda}}$$

$$\cancel{= \lambda e^{\lambda+\lambda}} = P(1) + 2 \cdot P(2) + \ldots$$

$$= \frac{e^{-\lambda}\cdot\lambda^1}{1!} + 2x.$$

$$\cancel{= \lambda}$$

$$= e^{-\lambda} \lambda \left(1+\lambda+\lambda^2+\ldots\right)$$

$$= e^{-\lambda} \lambda \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \ldots\right)$$

$$= e^{-\lambda} \lambda e^{\lambda} = \lambda$$

$V(x) = E(x^2) - \{E(x)\}^2 = E[x(x-1) + x] - \{E(x)\}^2$

$= E\{x(x-1)\} + E(x) - \{E(x)\}^2$

$E(x^2) = E\{x(x-1) + x\}$

$= E\{x(x-1)\} + E(x)$

$P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$E\{x(x-1)\}$

$= \sum_{x=0}^{\infty} x(x-1)P(x)$

$= 0(0-1)P(0) + 1(1-1)P(1) + 2(2-1)P(2) + 3(3-1)P(3) + 4(4-1)P(4) + \cdots + \infty$

$= 2\dfrac{e^{-\lambda}\lambda^2}{2!} + 3.2\dfrac{e^{-\lambda}\lambda^3}{3!} + 4.3\dfrac{e^{-\lambda}\lambda^4}{4!} + \cdots \infty$

$= e^{-\lambda}\lambda^2 + e^{-\lambda}\lambda^3 + \dfrac{e^{-\lambda}\lambda^4}{2!} + \cdots + \infty$

$= e^{-\lambda}\lambda^2\left(1 + \lambda + \dfrac{\lambda^2}{2!} + \cdots \infty\right)$

$= e^{-\lambda}\lambda^2 . e^{\lambda} = \lambda^2 e^{-\lambda+\lambda} = \lambda^2.$

$\therefore E(x^2) = E\{x(x-1)\} + E(x) = \lambda^2 + \lambda$

$\therefore V(x) = E\{x(x-1)\} + E(x) - \{E(x)\}^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

C.W. **Poissions**

26.12.19
Thursday

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \; ; \; x = 0, 1, 2 \ldots \infty$$

$$E(x) = \sum_{x=0}^{\infty} x\,P(x)$$

$$= 0\,P(0) + 1\,P(1) + 2\,P(2) + \ldots + \infty$$

$$= 0\,\frac{e^{-\lambda}\lambda^{0}}{0!} + 1 \cdot \frac{e^{-\lambda}\lambda^{1}}{1!} + 2\,\frac{e^{-\lambda}\lambda^{2}}{2!} + 3\,\frac{e^{-\lambda}\lambda^{3}}{3!} + \ldots$$

$$= 0 + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^{2}}{1!} + \frac{3\,e^{-\lambda}\lambda^{3}}{3 \times 2!} + \ldots$$

$$= e^{-\lambda}\lambda + e^{-\lambda}\lambda^{2} + \frac{e^{-\lambda}\lambda^{3}}{2!} + \ldots$$

$$= e^{-\lambda}\lambda\left(1 + \frac{\lambda}{1!} + \frac{\lambda^{2}}{2!} + \ldots\right)$$

$$= e^{-\lambda} \cdot \lambda\, e^{\lambda}$$

$$= \lambda$$

$$V(x) = E(x^{2}) - \{E(x)\}^{2}$$

$$= E\{x(x-1) + E(x)\} - \{E(x)\}^{2}$$

$$= E(x(x-1)) + E(x) - \{E(x)\}^{2}$$

$$E(x(x-1)) = \sum_{x=0}^{\infty} x(x-1)\,P(x)$$

$$= 0(0-1)P(0) + 1(1-1)P(1) + 2(2-1)P(2) + 3(3-1)P(3)$$
$$4(4-1)P(4) + \ldots$$

$$= 0 + 0 + 2\,\frac{e^{-\lambda}\lambda^2}{2!} + 3.2\,\frac{e^{-\lambda}\lambda^3}{3!} + 4.3\,\frac{e^{-\lambda}\lambda^4}{4!}$$

$$= e^{-\lambda}\lambda^2 + \frac{e^{-\lambda}\lambda^3}{1!} + \frac{e^{-\lambda}\lambda^4}{2!}$$

$$= e^{-\lambda}\lambda^2 \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right)$$

$$= e^{-\lambda}\lambda^2 e^{\lambda}$$

$$= \lambda^2$$

$$\therefore V(x) = E(x(x-1)) + E(x) - \{E(x)\}^2$$

~~Propo~~

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

(Rare cases) Road accident, Printing mistake

Properties of poisson distⁿ :-

1) The mean & variance are same.

2) If the value of parameter of poisson distⁿ is very large then poisson tends to normal distⁿ

3) Poisson distⁿ is a negatively skewed & platy kurtic.

Applications:

1) Number of Road accidents in a particular road within one year.

2) Number of Printing mistake in a book.

For small probability

3. To Single out. the faulty products of a factory

bivariante / Normal distⁿ:

If x be a random variable with mean μ and variance σ^2 then the normal distⁿ can be defined & denoted by,

$$f(x;\ \underset{\text{mean}}{\mu},\ \underset{\text{variance}}{\sigma^2}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\left[\begin{array}{c} -\infty < x < \infty \\ \sigma^2 > 0 \\ -\infty < \mu < \infty \end{array}\right]$$

* Wh[illegible] binomial distⁿ tends to normal distⁿ?

1) The number of trials are large.
2) The prob^y of success & failure are approximately equal.

bernolli is once tossed. $p^x q^{-x}$

ber (P) =

b (n, P) =

P (λ) =

N (μ, σ^2) =

Properties of Normal;

1. The mean and variance of normal $dist^n$ are μ & σ^2 resp.

2. Skewness ; $\beta_1 = 0$

3. Mean = Median = Mode.

4. The $dist^n$ is called bell shaped symmetric at point $x = \mu$.

5. The total area under the curve is 1.

Uses of normal $dist^n$:

1. Normal $dist^n$ can be used in quality control theory.

2. Binomial poisson can be approximately by normal $dist^n$.

3. Large sample sampling $dist^n$:
(t - $dist^n$, F - $dist^n$)

06.01.20

C.W.

~~$^nC_r\, p^r\, q^{n-r}$~~ $\qquad q = 1-p$

(density function)

$f(x; n, p) = {}^nC_x\, p^x\, (1-p)^{n-x}$

The prob^y that a patient survives from neurosurgery is 0·70. If 10 patients have the surgery, then what is the prob^y that. (i) None will survive

(ii) At least one will survive

(iii) At most 3 will survive.
/at best

(iv) bet^w 2 & 4

$P(x; n, p) = {}^nC_x\, p^x\, q^{n-x}$

Here, $n = 10$

$p = 0·70$

$q = 1 - p = 1 - 0·70 = 0·30$

So, the prob^y f^n –

$P(x; 10, 0·70) = {}^{10}C_x\, 0·7^x\, 0·3^{10-x}$; $x = 0, 1, 2 \ldots 10$

(i) $P(x=0) = {}^{10}C_0\, 0·7^0\, 0·3^{10-0} = 0·0000059$

(ii) $P(x \geq 1) = 1 - P(x<1) = 1 - P(x \leq 0)$

$= 1 - P(x=0)$

$= 1 - {}^{10}C_0\, 0·7^0\, 0·3^{10-0} = 0·99999$

(iii) $P(x \le 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$

$= {}^{10}C_0 (0.7)^0 (0.3)^{10-0} + {}^{10}C_1 (0.7)^1 (0.3)^{10-1} + {}^{10}C_2 (0.7)^2 (0.3)^{10-2}$

$+ {}^{10}C_3 (0.7)^3 (0.3)^{10-3}$

$= 0.0106$

(iv) $P(2 \le x \le 4) = P(x=2) + P(x=3) + P(x=4)$

$= {}^{10}C_2 (0.7)^2 (0.3)^{10-2} + {}^{10}C_3 (0.7)^3 (0.3)^{10-3} + {}^{10}C_4 (0.7)^4 (0.3)^{10-4}$

$= 0.0472$

A fair coin is tossed 12 times what is prob^y that - (i) No head will appear

(ii) At least two head will appear.

$P(x; n, p) = {}^nC_x \, p^x \, q^{n-x}$ $\quad n = 12$

$p = 0.5$

$q = 0.5$

$P(x; 12, 0.5) = {}^{12}C_x (0.5)^x (0.5)^{12-x}$

(i) $p(x=0) = {}^{12}C_0 (0.5)^0 (0.5)^{12-0} = 2.4 \times 10^{-4}$

(ii) $P(x \ge 2) = 1 - P(x \le 1)$

$= 1 - P(x=0) - P(x=1)$

$= 1 - 2.4 \times 10^{-4} - {}^{12}C_1 (0.5)^1 (0.5)^{11} = 0.9968$

λ of poisson indicates average.

Poisson dist^n –

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} ; x = 0, 1, 2, \ldots \infty$$

If electricity power failures occur according to a poisson dist^n with an average of 3 failures every twenty weeks. Calculate the prob^y that there will not be more than one failure during a particular week.

$$f(x;\tfrac{1}{3}) = \frac{e^{-\frac{1}{3}}\cancel{\lambda^x}\left(\frac{1}{3}\right)^x}{x!} \qquad \boxed{\lambda = \frac{1}{3}}$$

$$P(x \le 1) = P(x=0) + P(x=1)$$

$$= \frac{e^{-\frac{1}{3}}\frac{1}{3}^0}{0!} + \frac{e^{-\frac{1}{3}}\frac{1}{3}^1}{1!}$$

$$= \cancel{0.1099}\ 0.9553$$

$2\times2 \quad \chi^2 = \frac{N(ad-bc)^2}{(a+b)(b+c)(c+d)(b+d)}$

20.01.20 : Hypothesis testing.

Statistics

20.01.20
Monday

C.W.

Telephone calls arrive at a switchboard at a mean rate of 0·5 calls per minute. Calculate the proby. that two calls will arrive in a particular five-minute period.

mean/average poission

Soln:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \dots \infty$$

$$\lambda = np = 5 \times 0{\cdot}5 = 2{\cdot}5$$

$$P(x=2) = \frac{e^{-\lambda}\lambda^2}{2!}$$

$$= \frac{e^{-2\cdot5} \cdot 2{\cdot}5^2}{2!}$$

$$= 0{\cdot}257.$$

Poission	Binomial
λ = mean = variance = np	mean = np
	variance = npq

*The average number of calls received by a telephone operator during a time interval of 10 minutes during 5 pm to 5:10 pm daily is 3. What is the proby that operator will receive (i) no call (ii) exeactly one call

Soln: $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\lambda = 3$

(i) $P(x=0) = \frac{e^{-\lambda}\lambda^0}{0!} = \frac{e^{-3}\,3^0}{0!} = 0{\cdot}0497 = 0{\cdot}05$

(ii) $P(x=1) = \frac{e^{-\lambda}\lambda^1}{1!} = \frac{e^{-3}\,3^1}{1!} = 0{\cdot}15$

(iii) At least ~~3~~ 2 calls

$$P(x \geq 2) = 1 - P(x \leq 1)$$

$$= 1 - \{P(x=0) + P(x=1)\}$$

$$= 1 - (0{\cdot}05 + 0{\cdot}15)$$

$$= 1 - 0{\cdot}2 = 0{\cdot}8 \text{ (Ans)}$$

20 students marks:-

31, 34, 14, 26, 33, 13, 27, 23, 16, 16, 19, 23, $\cancel{38}$ 37, 39, 29, 15, 17, 17, 18, 30.

(1) Find mean, median, mode, variance & standard deviation using raw data.

(2) " " " ----- --- after ~~using~~ constructing frequency table.

(3) Make histogram, Pie-chart from frequency table

(4.) Find 1^{st}, 2^{nd}, 3^{rd} raw & central moment as well as Skewness & kurtosis.

arithmetic mean, $\bar{x} = \frac{\Sigma x_i}{n} = \frac{478}{20} = 23.9 \approx 24$

$G.M = \text{Antilog} \frac{\Sigma \log x_i}{n} = 22.544 \approx 23$

$H.M = \frac{n}{\Sigma \frac{1}{x_i}} = 21.27 \approx 21.3$

A.O = 13, 14, 15, 16, 16, 17, 17, 18, 19, 23, 23, 26, 27, 29, 30, 31, 33, 34, $\cancel{38}$ 37, 39.

$$\text{Median, } M_e = \frac{1}{2}\left[\frac{n}{2}^{th} + \left(\frac{n}{2}+1\right)^{th}\right]$$

$$= \frac{1}{2}\left[\frac{20}{2}^{th} + \left(\frac{20}{2}+1\right)^{th}\right]$$

$$= \frac{1}{2}(10^{th} + 11^{th})$$

$$= \frac{1}{2} \times (23+23)$$

$$= 23.$$

Mode, M_o = 16, 17, 23,

Variance, $\delta^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 = 65.79$

Standard Deviation, $\delta = \sqrt{\text{variance}} = 8.111$

2. L.L = 13 Range = 39 - 13 = 26
V.L = 39

~~Range Class = 1 + 3.322 log₁₀ N = 1 + 3.322 log₁₀ 26 = 11.82 ≈ 11~~

Let, Class no. = 5

∴ Class interval $= \frac{\text{Range}}{\text{No. of Class}} = \frac{26}{5} = 5.2 \approx 5$

Class interval	Tally	Frequency f_i	Mid Value x_i	Cumulative freq	$f_i x_i$
13 - 18	𝍸 II	7	15.5	7	108.5
18 - 23	II	2	20.5	9	41
23 - 28	IIII	4	25.5	13	102
28 - 33	III	3	30.5	16	91.5
33 - 38	III	3	35.5	19	106.5
38 - 43	I	1	40.5	20	40.5

$\Sigma f_i x_i = 490$

A. mean, $\bar{x} = \frac{\Sigma f_i x_i}{n} = \frac{490}{20} = 24.5$.

G.m $= \text{Antilog} \frac{\Sigma f_i \log x_i}{N} = \text{Antilog} \frac{62.84}{20} = 23.15$

H.M $= \frac{N}{\Sigma \frac{f_i}{x_i}} = \frac{20}{0.9135} = 21.89$

$$\text{Median} = L + \frac{\frac{n}{2} - f_c}{f_m} \times c$$

$$= 13 + \frac{\frac{20}{2} - 0}{7} \times 6$$

$$= 13 + \frac{10}{7} \times 6$$

$$= 21.57.$$

$$\text{Mode}, M_0 = L + \frac{f_1}{f_1 + f_2} \times h$$

$$= 13 + \frac{7}{7+5} \times 5$$

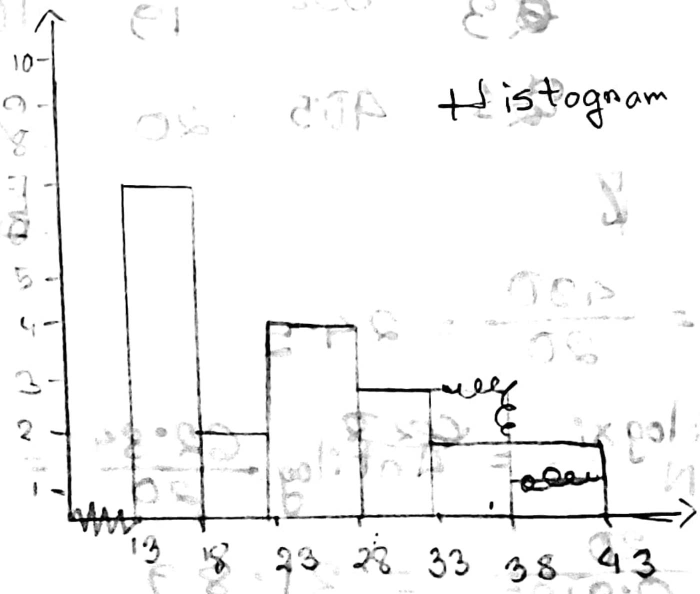
$$= 15.91$$

$$\approx 16$$

Variance, $\delta^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 = \frac{511}{20} = 25.55$

Standard deviation, $S_d = \sqrt{\delta^2} = 5.05$

3. Histogram



Pie-chart.

For class interval 13-18, $\frac{360\times 7}{20} = 126^\circ$

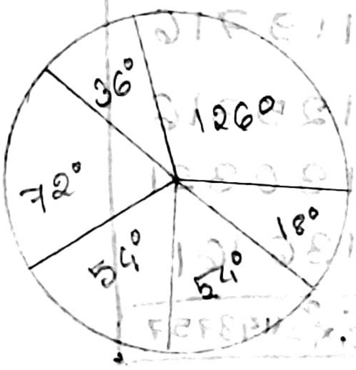
" " " 18-23, $\frac{360\times 2}{20} = 36^\circ$

" " " 23-28, $\frac{360\times 4}{20} = 72^\circ$

28-33, $\frac{360\times 3}{20} = 54^\circ$

33-38, $\frac{360\times 2}{20} = \cancel{54^\circ}\ 36^\circ$

38-43, $\frac{360\times 2}{20} = \cancel{18^\circ}\ 36^\circ$



$a = 10.483$
$b = 0.0109$

2015

5. (c) (ii)

Distance (Y)	Weight (X)	$(x_i y_i)$	x_i^2
13.9	314	4364.6	98596
14.0	317	4438	100489
13.9	320	4448	102400
14.1	326	4596.6	106276
14.0	331	4634	109561
14.3	339	4847.7	114921
14.1	346	4878.6	119716
14.5	354	5133	125316
14.5	361	5234.5	130321
14.4	369	5313.6	136161
$\Sigma y_i = 141.7$	$\Sigma x_i = 3377$	$\Sigma x_i y_i = 47888.6$	$\Sigma x_i^2 = 1143757$

We know,

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

$$= \frac{47888.6 - \frac{141.7 \times 3377}{10}}{1143757 - \frac{(3377)^2}{10}}$$

$$= 0.0109$$

$$\alpha = \bar{y} - \beta \bar{x}$$

$$= \frac{\Sigma y}{n} - 0.0109 \times \frac{\Sigma x}{n}$$

$$= \frac{141.7}{10} - 0.0109 \times \frac{3377}{10}$$

$$= 10.48$$

$$\therefore \hat{y} = \alpha + \beta \hat{x}$$

$$= 10.48 + 0.019 \hat{x}$$

(iii) Given, $x = 360$

$$\bar{Y} = 10.48 + 0.019 \times 360$$

$$= 17.32 \text{ miles/hour.}$$

~~a+b~~

Q Let for raw moment arbitoray value = 23

Marks (X)	$d = X - A$, $A = 23$	d^2	d^3
13	-10	100	-1000
14	-9	81	-729
15	-8	64	-512
16	-7	49	-343
16	-7	49	-343
17	-6	36	-216
17	-6	36	-216
18	-5	25	-125
19	-4	16	-64
23	0	0	0
23	0	0	0
26	3	9	27
27	5	25	125
29	6	36	216
30	7	49	343
31	8	64	512
33	10	100	1000
34	11	121	1331
38	15	225	3375
39	16	256	4096
	$\Sigma d = 19$	$\Sigma d^2 = 1341$	$\Sigma d^3 = 7477$

1st raw moment, $\mu_1' = \frac{\Sigma d}{n} = \frac{19}{20} = 0.95$

2nd raw ", $\mu_2' = \frac{\Sigma d^2}{n} = \frac{1341}{20} = 67.05$

3rd raw ", $\mu_3' = \frac{\Sigma d^3}{n} = \frac{7477}{20} = 373.85$

1st central moment, $\mu_1 = 0$

2nd central moment, $\mu_2 = \mu_2' - (\mu_1')^2$

$= 67.05 - (0.95)^2$

$= 66.1475$

3rd central moment, $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$= 373.85 - 3 \times 67.05 \times 0.95 + 2(373.85)^3$

$= 104501592.8$

Bayer's theorem:- Let, $\{A_1, A_2, A_3 \ldots A_k\}$ ~~are~~ be a set of mutually exclusive events forming a partition of sample space S. such that,

i) $A_1 \cup A_2 \cup A_3 \ldots \cup A_k = S$

ii) $P(A_i) > 0 \;;\; i = 1, 2, \ldots k$

Let, B be any ~~an~~ event of S such that $P(B) > 0$.

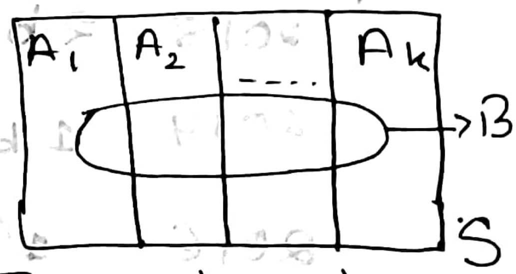
$$\text{Then } P(A_i/B) = \frac{P(A_i)\,P(B/A_i)}{\sum_{i=1}^{k} P(A_i)\,P(B|A_i)} \;;\; i = 1, \ldots k.$$

which is the Bayer's theorem.

<u>Proof:</u>- According to Bayer's theorem A_i & B are dependent. By rule of multiplication,

$$P(A_i \cap B) = P(B)\,P(A_i|B) \quad \text{--- (i)}$$

$$P(A_i \cap B) = P(A_i)\,P(B|A_i) \quad \text{--- (ii)}$$



<u>Figure:</u> Intersection of B with $A_1, A_2, \ldots A_k$

From (i) we get,

$$P(A_i \cap B) = P(B)\,P(A_i|B)$$

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} \quad \text{---- (iii)}$$

Here, $S = A_1 \cup A_2 \cup A_3 \ldots \cup A_k$

$$B = S \cap B$$

$$= (A_1 \cup A_2 \ldots \cup A_k) \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) \ldots \cup (A_k \cap B).$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) \dots + P(A_k \cap B)$$

$$= \sum_{i=1}^{k} P(A_i \cap B)$$

$$= \sum_{i=1}^{k} P(A_i)\, P(B|A_i) \quad [\text{from (ii)}]$$

$\therefore$ In (iii),

$$P(A_i|B) = \frac{P(A_i)\, P(B|A_i)}{\sum_{i=1}^{k} P(A_i)\, P(B|A_i)}$$

This proves the theorem.

$\boxed{\text{Google:}}$

2018	2 b, c | 4 b | 5 b, c
2017	1 b, c | 4 b | 5 b, c
2016	1 (a)
2015	4 b
2013	3, 4 b, 6 b

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

2018

2(c) Show that, $\boxed{\text{variance}}$ is independent of origin but not of scale.

Let, $x_1, x_2, \ldots x_n$ be a set of n values of a variable x. If these values are transformed to a new set of values $y_1, y_2, \ldots y_n$ of a variable y, such that,

$$y = \frac{x-a}{h} \quad \text{---- (i)}$$

where a and h are two constant and $h>0$, then the theorem asserts that

$$S_x^2 = h^2 s_y^2 \quad \text{---- (ii)}$$

To prove this, consider the ith value of variable y defined in (a) above,

$$y_i = \frac{x_i - a}{h}$$

giving $\quad x_i = a + h y_i$ ——— (iii)

From which $\quad \bar{x} = a + h\bar{y}$ ——— (iv)

Hence from (iii) & (iv)

$$x_i - \bar{x} = a + hy_i - (a + h\bar{y}) = h(y_i - \bar{y})$$

Squaring, summing and dividing both sides of above equation by $n-1$, we arrive at following expression,

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

$$S_x^2 = h^2 S_y^2 \quad \text{(Proved)}$$

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$n(S) = 36$

$n(\text{Total } 6) = 5$

$n(\text{even no}) = 9$

$$P(\text{Total } 6) = \frac{n(\text{Total } 6)}{n(S)} = \frac{5}{36}$$

$$P(\text{even no}) = \frac{n(\text{even no})}{n(S)} = \frac{9}{36} = \frac{3}{12} = \frac{1}{4}$$

$$\therefore P(\text{Total } 6 \text{ or even no}) = \frac{5}{36} + \frac{1}{4}$$

$$= \frac{17}{18}$$

2017 2(a)

7

$P(A_1) = \boxed{\frac{3}{2}} P(A_2)$.

$P(A_2) = 2P(A_3)$ ∴

$P(A_1) = \frac{3}{\cancel{2}} \cdot \cancel{2} P(A_3) = 3P(A_3)$ ∴ $\frac{1}{3} P(A_1) = P(A_3)$

$P(A_1) = x$.

$P(A_2) = \frac{2x}{3}$ ∴

$P(A_3) = \frac{1}{3}x$

∴ $x + \frac{2x}{3} + \frac{x}{3} = 1$

or, $\frac{3x+2x+x}{3} = 1$.

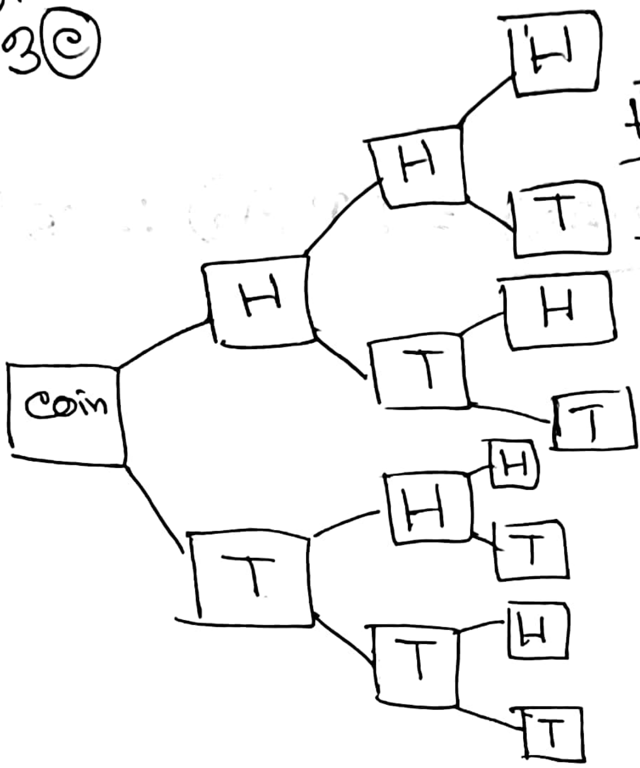
or, $6x = 3$

or, $2x = 1$

or, $x = \frac{1}{2}$

∴ $P(A_1) = \frac{1}{2}$, $P(A_2) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$, $P(A_3) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$

2017
3(c)



	HH	HT	TH	TT
H	HHH	HHT	HTH	HTT
T	THH	THT	TTH	TTT

(i) $\frac{1}{8}$, (ii) {HHH, HHT, HTH, THH} ; $\frac{4}{8} = \frac{1}{2}$

3 (b) Let, x be a binomial variate with parameter n and p, such that, $p + q = 1$.

We know, mean $E(x) = np$ ---- (i)

and variance, $V(x) = npq$

$$\Rightarrow V(x) = np(1-p)$$

$$\Rightarrow V(x) = np - np^2$$

$$\Rightarrow V(x) = E(x) - np^2$$

$$\Rightarrow E(x) - np^2 = V(x)$$

$$\Rightarrow E(x) = V(x) + np^2$$

Since, $n \geq 1$ and $p > 0$, So, $np^2 > 0$

$E(x) = V(x) + \text{positive value}$

$\therefore E(x) > V(x)$.

When we have positive values $= 0$ then

$E(x) = V(x)$

$\therefore E(x) \geq V(x)$

$1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$

X	Y	R(x)	R(y)	d = R(x) - R(y)	d^2
50	11	19.5	23	-3.5	12.25
50	13	~~18~~ 19.5	7.3	12.2	148.84
55	14	18	5.5	12.5	156.25
60	16	4.25	1.5	2.75	7.5625
65	16	1.33	1.5	-0.17	0.0289
65	15	1.33	3.5	-2.17	4.7089
65	15	1.33	3.5	-2.17	4.7089
60	14	4.25	5.5	-1.25	1.5625
60	13	4.25	7.3	-3.05	9.3025
60	13	4.25	7.3	-3.05	9.3021

$\Sigma d^2 = 349.8078$

$= 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)} =$

$= -1.19$

2017

1. (c) First n natural numbers $= 1, 2, 3, 4 \ldots n$

$$\text{Mean} = \frac{1+2+3+4+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n(n+1)}{2} \times \frac{1}{n}$$

$$= \frac{n+1}{2} \text{ (Ans)}$$

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

As the data is very large $(x_i - \bar{x})^2$ calculation is difficult. Hence, we use formula,

$$\text{Variance}(\sigma^2) = \frac{\Sigma (x_i)^2}{n} - (\text{Mean})^2$$

$$= \frac{1^2 + 2^2 + \ldots n^2}{n} - \left(\frac{n+1}{2}\right)^2$$

Since,

$$1^2 + 2^2 + 3^2 + \ldots n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\therefore \text{Variance}(\sigma^2) = \frac{\cancel{n}(n+1)(2n+1)}{6\cancel{n}} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n+1}{2}\left\{\frac{2n+1}{3} - \frac{n+1}{2}\right\}$$

$$= \frac{n+1}{2} \cdot \frac{4n+2-3n-3}{6}$$

$$= \frac{n+1}{2} \cdot \frac{n-1}{6}$$

$$= \frac{n^2-1}{12} \text{ (Am)}$$

2012

3.a Distinguish betw discrete & continuous random variable.

Discrete		Continuous
Finite number of isolated values	1.	Infinite number of different values
Complete range of specific number	2.	Incomplete range of specific number
Values are obtained by counting	3	Values are obtained by measuring.
Non-overlapping	4.	Overlapping
Assumed values are distinct or separate values	5	Any value between the two values
Represented by isolated points	6	Represented by connected points

Probability density fn: It is a non-negative function and is constructed so that the area under it's curve bounded by the x-axis is equal to unity when computed over the range of x, for which $f(x)$ is defined.

~~2012~~ 2 a,b

Cumulative distribution func: The distⁿ func $F(x)$ of a discrete random variable X with probability function $f(x)$ defined over all real number x is the cumulative probability upto and including point x.

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

(b)

x	1	2	3	4
$P(x)$	0·1	0·3	0·4	0·2

2012

$P(x \geq 2)$, $P(x < 3)$, $E(x)$, $V(x)$

$$P(x \geq 2) = 1 - P(x < 2)$$
$$= 1 - P(x = 1)$$
$$= 1 - 0·1$$
$$= 0·9 \text{ (Ans)}$$

$$P(x < 3) = P(x=1) + P(x=2)$$
$$= 0·1 + 0·3$$
$$= 0·4 \text{ (Ans)}$$

or, $P(x < 3) = 1 - P(x \geq 3)$
$$= 1 - P(x=3) - P(x=4)$$
$$= 1 - 0·4 - 0·2$$
$$= 0·4$$

$$E(x) = 1 \times 0·1 + 2 \times 0·3 + 3 \times 0·4 + 4 \times 0·2 = 2·7 \text{ (Ans)}.$$

$$E(x^2) = 1^2 \times 0·1 + 2^2 + 0·3 + 3^2 \times 0·4 + 4^2 \times 0·2 = 8$$

$$\therefore V(x) = E(x^2) - \{E(x)\}^2$$
$$= 8 - (2·7)^2$$
$$= 0·71 \text{ (Ans)}.$$

CamScanner

3(e) What is Poisson distribution? Find mean & variance of this distribution and show that mean & variance of distribution are equal.

Poisson : ~~Let, x~~ The probability distribution of a random variable x is said to have poisson distribution if its probability function is give by

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\text{Mean, } E(x) = \sum_{x=0}^{\infty} x P(x)$$

$$= 0\,P(0) + 1P(1) + 2.P(2) + 3\,P(3) + \cdots$$

$$= P(1) + 2P(2) + 3P(3) + \cdots$$

$$= \frac{e^{-\lambda}\lambda}{1!} + 2.\frac{e^{-\lambda}\lambda^{2}}{2!} + 3\frac{e^{-\lambda}\lambda^{3}}{3!} + \cdots$$

$$= e^{-\lambda}\lambda + e^{-\lambda}\lambda^{2} + \frac{1}{2}e^{-\lambda}\lambda^{3} + \cdots$$

$$= e^{-\lambda}\lambda\left(1 + \lambda + \frac{\lambda}{2} + \cdots\right)$$

$$= e^{-\lambda}\lambda\left(1 + \frac{\lambda}{1!} + \frac{\lambda}{2!} + \cdots\right)$$

$$= e^{-\lambda}\lambda\, e^{\lambda}$$

$$= \lambda$$

Variance, $V(x) = E(x^2) - \{E(x)\}^2$

$= E[x(x-1)+x] - \{E(x)\}^2$

$= E\{x(x-1)\} + E(x) - \{E(x)\}^2$

Here, $E\{x(x-1)\} = \sum_{x=0}^{\infty} x(x-1)P(x)$

$= 0(0-1)P(0) + 1(1-1)P(1) + 2(2-1)P(2) + 3(3-1)P(3)$
$+ 4(4-1)P(4) + \ldots\ldots$

$= 0 + 0 + 2.1\,P(2) + 3.2\,P(3) + 4.3\,P(4) + \ldots$

$= 2\dfrac{e^{-\lambda}\lambda^2}{2!} + 3.2\cdot\dfrac{e^{-\lambda}\lambda^3}{3!} + 4.3\dfrac{e^{-\lambda}\lambda^4}{4!} + \ldots\ldots$

$= e^{-\lambda}\lambda^2 + e^{-\lambda}\lambda^3 + \dfrac{e^{-\lambda}\lambda^4}{2!} + \ldots\ldots$

$= e^{-\lambda}\lambda^2\left\{1 + \lambda + \dfrac{\lambda^2}{2!} + \ldots\right\}$

$= e^{-\lambda}\lambda^2\left\{1 + \dfrac{\lambda}{1!} + \dfrac{\lambda^2}{2!} + \ldots\right\}$

$= e^{-\lambda}\lambda^2 \cdot e^{\lambda}$

$= \lambda^2$

$\therefore V(x) = E\{x(x-1)\} + E(x) - \{E(x)\}^2$

$= \lambda^2 + \lambda - \lambda^2$

$= \lambda$

$\therefore$ Mean, $E(x)$ = Variance, $V(x)$. [Proved]

That is mean & variance of poisson distribution is same

2012

2 (a) Equally Likely event: ~~If~~ The outcomes of a trial or experiment are said to be equally likely if ~~every~~ each of them have equal chances to be occured.

Example: In case of tossing a fair coin head and tail are equally likely event.

Complementary event: A event "A occurs" and event "A does not occur" are called complementary events.

The "event A does not occur" is denoted by $\bar{A}$

$\therefore P(A) + P(\bar{A}) = 1.$

$P(A) = 0.35 \qquad P(A \cap B) = 0.2$

$P(B) = 0.75$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.35 + 0.75 - 0.2$

$= 0.9$

~~$P(B^c) = 1 - P(B) = 1 - 0.75 = 0.25$~~

$P(A \cap B^c) = \cancel{P(A)\,P(B^c) = 0.35 \times 0.25 = 0.0875}$

$= P(A) - P(A \cap B)$

$= 0.35 - 0.2$

$= 0.15$

b) Conditional Probability: ~~When~~ When A & B are two dependent events & occurance of B depends on A and vice verse then it's C.P.

$P(B|A) = \frac{P(AB)}{P(A)}$; $P(A) > 0$.

Multiplicative law of Probability:

(i) For two dependent event:

For two events A and B the multiplicative law says,

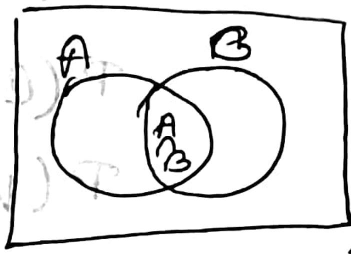
(i) $P(A\cap B) = P(A).P(B|A)$

(ii) $P(A\cap B) = P(B)P(A|B)$

Proof:- If A & B are two events which are dependent on each other than, according to conditional probability,

$$P(A|B) = \frac{n(A\cap B)}{n(B)} \quad \text{--- (i)}$$

$$P(B|A) = \frac{n(A\cap B)}{n(A)} \quad \text{--- (ii)}$$



S

Now, $P(A\cap B) = \frac{n(A\cap B)}{n(S)} = \frac{n(A\cap B)}{n(A)} \times \frac{n(A)}{n(S)}$

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

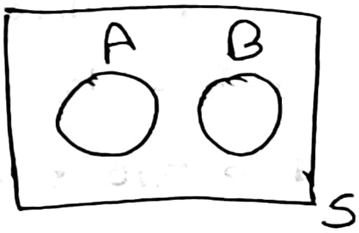
Again, $P(A\cap B) = \frac{n(A\cap B)}{n(S)} = \frac{n(A\cap B)}{n(B)} \times \frac{n(B)}{n(S)}$

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

(ii) Multiplicative law of 2 independent event,

The probability of joint occurrence of two independent event is equal to the product of individual probability. $P(A\cap B) = P(A)\cdot P(B)$

Let, $n(S_1)$ & $n(S_2)$ are sample space for A & B events.



$$\therefore P(A) = \frac{n(A)}{n(S_1)} \quad \& \quad P(B) = \frac{n(B)}{n(S_2)}$$

$$\therefore P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{n(A)\, n(B)}{n(S_1)\, n(S_2)} = P(A)\, P(B)$$

[Proved]

(c) $P(\text{Diagnosed Cancer}) = \frac{65}{100}$

$P(\text{Correct Diagnosis death}) = \frac{30}{100}$

$P(\text{Wrong Diagnosis death}) = \frac{75}{100}$.

$$P(\text{death even for correct diagnosis}) = \frac{65}{100} \times \frac{30}{100}$$

$$= \frac{39}{200}$$

Measures of location: The position at which the central tendency of collective data is described is location.

The measures of

2017 6(b)

Given. Samples = 66, 65, 69, 68, 70, 71, 63, 64, 68, 69

$$\bar{x} = \frac{\Sigma x_i}{n} = 67.3$$

$$S = \cancel{\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}} \sqrt{\frac{1}{n-1}\left\{\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right\}}.$$

$$= \cancel{2.8583}\ 2.66$$

Let, $H_0 : \mu = 65$

$H_1 : \mu \neq 65$

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = 2.73$$

$$t_{tab} = t_{0.05,9} = 2.26$$

$t_{tab} < t_{cal}$

$\therefore H_0$ is rejected.

H_1 i.e alternate hypothesis accepted.

2017.

6(a)

	Set of 40 plot	Set of 60 plots
Mean Yeild / plot	1254	1243
Standard deviation	34	28

Here,

$n_1 = 40 \quad \bar{x}_1 = 1254 \quad s_1 = 34$

$n_2 = 60 \quad \bar{x}_2 = 1243 \quad s_2 = 28$

We set the test,

$H_0 : \mu_1 = \mu_2$ (i.e no significant difference)

$H_1 : \mu_1 \neq \mu_2$ (i.e significant difference)

the test statistics, $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \dfrac{1254 - 1243}{\sqrt{\frac{34^2}{40} + \frac{28^2}{60}}}$

$= 1.698$

The critical value of z at 5% $= Z_{0.05, (n_1+n_2-2)}$

$= Z_{0.05, 98}$

$= 1.96$

$Z_{tab} > Z_{cal}$

$\therefore$ Ho is not rejected . ie there is no significate difference between the two set of plots.

2016

2. (a) Probability of an event: Suppose an event E can happen r ways out of a total of n possible equally likely ways.

Then Probability of occurrence of event (called it success) is denoted by, $P(E) = \frac{r}{n}$

The probability of event (called it failure) is denoted by, $P(\bar{E}) = \frac{n-r}{n} = 1 - \frac{r}{n}$

Thus, $P(E) + P(\bar{E}) = 1$

In words, ~~there~~ this means that sum of probability in experiment of event is 1.

6 (a) ~~(a)~~ The table where the degree of freedom are calculated by using marginal total place restricting in selecting cell frequency is contingency table.

For example,

	Red Dress	Black Dress	Marginal Cost
Boys	a	b	a+b
Girls	c	d	c+d
Marginal Cost	a+c	b+d	N=a+b+c+d Grand total

$\underbrace{(c-1)}_{\text{no. of } \cancel{\text{Row}} \text{ column}}\ \underbrace{(r-1)}_{\text{no of rows}}$

$= (2-1)(2-1)$

$= 1 \times 1$

Thus for 2x2 table, the ~~nub~~ no. of cell frequencies assigned arbitrarily

In this case, chi-square statistics,

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2_{(r-1)(c-1)}$$

where $(r-1)(c-1)$ is degree of freedom

2016

5 (c) Between type-I & type-II error, type-II error is more serious. We know, when the null hypothesis H_0 is true but still we reject it then it's type-I error, which means we are rejecting the good thing. And when the null hypothesis is false but if we still accept it even though we know it's false then it's called type-II error. So, we know the result is false but still we except it. So, it's type-II error will cause more problem. And so, type-II is more serious than type-I error.

2016

~~2015~~

6. (b)

O	E	$(O-E)^2$	$(O-E)^2/E$
40	45	25	0.55
60	55	25	0.45
50	45	25	0.55
50	55	25	0.45

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2$$

H_0: No association

H_1: Association.

$$\chi^2_{0.05} = 3.84 > \chi^2 = 2.$$

Comment: H_0 is not rejected. So, there is no association between hair color & eye color. (Null hypotesis is accepted)

2016 3(b)

$$\left.\begin{aligned} E(aX+b) &= \sum_x (ax+b)\,P(x) \\ &= \sum_x (ax.P(x) + b.P(X)) \\ &= \sum_x ax.p(x) + \sum_x b\,p(x) \\ &= a\sum_x x\,p(x) + b\sum_x p(x) \\ &= a\,E(X) + b \end{aligned}\right\}$$

Similarly, the result can be obtained when X is a continuous random variable.

$$t_{cal} = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

2 mean test

$$t_{cal} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}}$$

Coefficent of variation (CV): The ratio of standard deviation of the distribution to the mean of same distribution expressed in percentage form,

$$CV = \frac{S_x}{\overline{x}} \times 100$$

2017 Q 1 (b):

Let, $x_1, x_2, \ldots x_n$ be a set of n non-negative values. The square of the sum of these values can be expressed as follows,

$$(x_1 + x_2 + \ldots x_n)^2 = x_1^2 + x_2^2 + \ldots + x_n^2 + \sum_{i \neq j} x_i x_j$$

That is

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i \neq j} x_i x_j \text{ ---- } (i)$$

Since x_i's are non-negative, $\sum_{i \neq j} x_i x_j \geq 0$. Hence (i) can be expressed as

$$\left(\sum_{i=1}^{n} x_i\right)^2 \geq \sum_{i=1}^{n} x_i^2 \text{ ---- } (ii)$$

Substracting $(\Sigma x_i)^2/n$ from both sides of (ii)

$$(\Sigma x_i)^2 - \frac{(\Sigma x_i)^2}{n} \geq \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

or, $(\Sigma x_i)^2 \left(1 - \frac{1}{n}\right) \geq n s^2$

or, $\frac{n-1}{n}(n\bar{x})^2 \geq n s^2$

or, $\bar{x}^2 (n-1) \geq s^2$

or, $\bar{x}\sqrt{n-1} \geq s$

[Showed]

2016 6 (a)

$H_0 : \mu = 0.2545$

$\mu \neq 0.2545$

$\bar{x} = 0.255$

$S = 0.0001$

$n = 10$

$\mu = 0.2545$

$Z_{0.025} = 1.96$

$$Z_{cal} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{0.255 - 0.2545}{0.0001/\sqrt{10}}$$

$$= 15.88.$$

$Z_{0.05} < Z_{cal}$

Comment:

$\therefore$ Null hypothesis rejected.

Alternate " accepted.

$$F(\lambda; x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

Theorem: Let X be a discrete random variable with parameter λ.
Then skewness γ_1 of X is given by

$$\gamma_1 = \frac{1}{\sqrt{\lambda}}$$

From Skewness in terms of Non-Central Moment

$$\gamma_1 = \frac{E(x^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

where μ is mean of X, σ is standard deviation

We have, by Expectation of Poisson distribution:

$$E(X) = \lambda$$

By variance of Poisson Distribution: $var(X) = \sigma^2 = \lambda$

So, $\sigma = \sqrt{\lambda}$

Now to calculate γ_1, we must calculate $E(x^3)$

We find this using moment generating function of X, M_x.

By Moment generating fun^c of Poisson distribution

$$M_X(t) = e^{\lambda(e^{t}-1)}$$

$$E(x^3) = M_x'''(0)$$

$$M_x''' = \lambda(\lambda e^t + 1) e^{\lambda(e^t-1)+t}$$

$$\Sigma\left(\frac{x_i-\bar{x}}{\sqrt{\Sigma(x_i-\bar{x})^2}} \pm \frac{y_i-\bar{y}}{\sqrt{\Sigma(y_i-\bar{y})^2}}\right)^2 \geq 0$$

$$\frac{\Sigma(x_i-\bar{x})^2}{\Sigma(x_i-\bar{x})^2} \pm 2\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\cdot\Sigma(y_i-\bar{y})^2}} + \frac{\Sigma(y_i-\bar{y})^2}{\Sigma(y_i-\bar{y})^2} \geq 0$$

We know,

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

$\therefore 1 \pm 2r + 1 \geq 0$

$2 \pm 2r \geq 0$

$1 \pm r \geq 0$

$\therefore 1 + r \geq 0$

$r \geq -1$

$\therefore 1 - r \geq 0$

$1 \geq r$

$\therefore 1 \geq r \geq -1$

Spearman's Rank Correlation: To measure the association between two variables ~~without making any~~ we rank each individual with respect to interest. Then the co-efficient of correlation betn two sets of rank is called rank correlation.

Let (x_i, y_i) be the rank correlation of sample variable x & y. x & y can take the values $1, 2, 3 \ldots, n$.

Here, $\bar{x} = \bar{y} = \frac{1+2+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$

Variance of variable x,

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

$$\sum x_i^2 = 1^2 + 2^2 + 3^2 + \ldots n^2 = \frac{n(n+1)(2n+1)}{6}$$

$\therefore$ Variance,

$$\sigma_x^2 = \frac{1}{\cancel{n}} \cdot \frac{\cancel{n}(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n+1}{2}\left\{\frac{2n+1}{3} - \frac{n+1}{2}\right\}$$

$$= \frac{n+1}{2} \cdot \frac{4n+2-3n-3}{6}$$

$$= \frac{n+1}{2} \cdot \frac{n-1}{6} = \frac{n^2-1}{12}$$

Similarly, $\sigma_y^2 = \frac{1}{n}\sum(y_i - \bar{y})^2 = \frac{n^2-1}{12}$

Let's consider,

$$d_i = x_i - y_i = x_i - \bar{x} + \bar{x} - y_i = x_i - \bar{x} - (y_i - \bar{y})$$

$[\because \bar{x} = \bar{y}]$

CamScanner

or, $d_i^2 = \{(x_i - \bar{x}) - (y_i - \bar{y})\}^2$

or, $d_i^2 = (x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2$

or, $\frac{\Sigma d_i^2}{n} = \frac{\Sigma (x_i - \bar{x})^2}{n} - \frac{2(x_i - \bar{x})(y_i - \bar{y})}{n} + \frac{\Sigma (y_i - \bar{y})^2}{n}$

or, $\frac{\Sigma d_i^2}{n} = \sigma_x^2 - 2\,\mathrm{Cov}(x, y) + \sigma_y^2$

or, $2\,\mathrm{Cov}(x, y) = \sigma_x^2 + \sigma_y^2 - \frac{\Sigma d_i^2}{n}$

or, $\mathrm{Cov}(x, y) = \frac{\sigma_x^2 + \sigma_y^2}{2} - \frac{\Sigma d_i^2}{2n}$

$$= \left(\frac{n^2 - 1}{12} + \frac{n^2 - 1}{12}\right) \times \frac{1}{2} - \frac{\Sigma d_i^2}{2n}$$

$$= \frac{n^2 - 1}{12} - \frac{\Sigma d_i^2}{2n}$$

$$r_{xy} = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

$$= \frac{\frac{n^2-1}{12} - \frac{d_i^2}{2n}}{\sqrt{\frac{n^2-1}{12}}\sqrt{\frac{n^2-1}{12}}}$$

$$= \frac{\frac{n^2-1}{12} - \frac{d_i^2}{2n}}{\frac{n^2-1}{12}}$$

$$= 1 - \frac{\Sigma d_i^2}{2n} \times \frac{12}{n^2-1}$$

$$= 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$$

Rank correlation co-efficient of Spearman is denoted by ρ (rho) then. $\rho_{xy} = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$

2016 2(c)

~~P(Manufactured by machine A,B&C) = 25/100~~

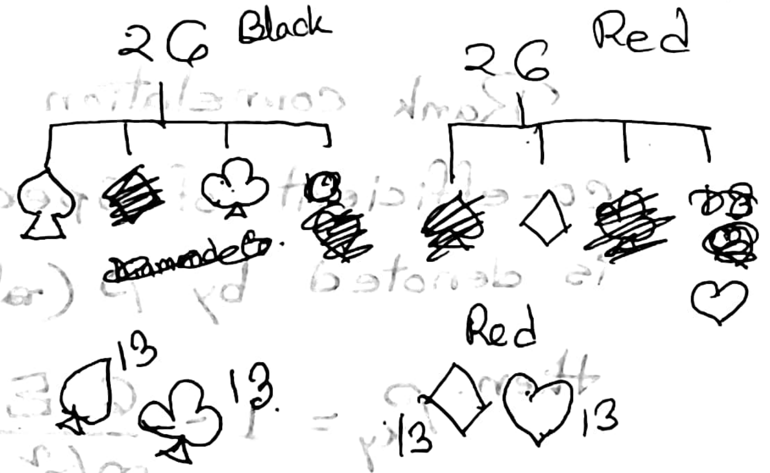
$$P(A) = \frac{25}{100}, \quad P(B) = \frac{35}{100}, \quad P(C) = \frac{40}{100}$$

$$P(A_0) = \frac{5}{100}, \quad P(B_0) = \frac{4}{100}, \quad P(C_0) = \frac{2}{100}$$

From A, $P(A_1) = \frac{5}{100} \times \frac{100}{25} = \frac{1}{5}$

From, B, $P(B_1) = \frac{P(B_0)}{P(B)} = \frac{4}{100} \times \frac{100}{35} = \frac{4}{35}$

From, C, $P(C_1) = \frac{P(C_0)}{P(C)} = \frac{2}{100} \times \frac{100}{40} = \frac{1}{20}$



$$P(\text{Spade}) = \frac{13}{52}$$

$$P(\text{King}) = \frac{4}{52}$$

$$P(\text{spade or king}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$= \frac{17}{52} - \frac{1}{52}$$

$$= \frac{16}{52}$$

Black

I

10

K

Q

J

A

2017 4 (b)

Correlation co-efficient is independent of origin and scale.

Let x, y are two variable

Now, $u = \frac{x-a}{h}$; $h > 0$

$v = \frac{y-b}{g}$; $g > 0$

Now, we can write,

$x = a + uh$

$\bar{x} = a + h\bar{u}$

$(x-\bar{x}) = h(u-\bar{u})$

$y = b + gv$

$\bar{y} = b + g\bar{v}$

$(y-\bar{y}) = g(v-\bar{v})$

$$\text{Correlation, } r_{xy} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sigma_x \cdot \sigma_y}$$

$$= \frac{\Sigma(u_i-u)(v_i-v)}{\sigma_u \cdot \sigma_v}$$

$$= r_{uv}$$

Alternate

We know,

$$r_{xy} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\left(\sqrt{\Sigma x_i^2 - \frac{(\Sigma x)^2}{n}}\right)\left(\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}\right)} = \frac{\Sigma y_i x_i - \frac{\Sigma y_i \Sigma x_i}{n}}{\left(\sqrt{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}\right)\left(\sqrt{\Sigma x_i^2 - \frac{(\Sigma x)^2}{n}}\right)}$$

$$= r_{yx}$$

x & y are independent

2018 2(b)

Let, x_1 & x_2 are two observations.

$$\therefore S^2 = \frac{1}{n}\left\{(x_1-\bar{x}_1)^2 + (x_2-\bar{x}_2)^2\right\}$$

$$= \frac{1}{n}\left\{\left(x_1 - \frac{x_1+x_2}{2}\right)^2 + \left(x_2 - \frac{x_1+x_2}{2}\right)^2\right\}$$

$$= \frac{1}{2}\left\{\left(\frac{2x_1-x_1-x_2}{2}\right)^2 + \left(\frac{2x_2-x_1-x_2}{2}\right)^2\right\}$$

$$= \frac{1}{2}\left\{\left(\frac{x_1-x_2}{2}\right)^2 + \left(\frac{x_2-x_1}{2}\right)^2\right\}$$

$$= \frac{1}{2}\left\{\left(\frac{x_1-x_2}{2}\right)^2 + \left(\frac{x_1-x_2}{2}\right)^2\right\}$$

$$= \frac{1}{\cancel{2}}\left\{\left(\frac{x_1-x_2}{2}\right)^2 \times \cancel{2}\right\}$$

$$= \left(\frac{x_1-x_2}{2}\right)^2$$

$$\therefore S^2 = \left(\frac{x_1-x_2}{2}\right)^2$$

$$\Rightarrow S = \frac{x_1-x_2}{2}$$

i.e, Standard deviation is half of range.

৪২,৪৭.

9.2, 8.7, 8.6, 8.8, 8.5, 8.7, 9

$$\bar{x} = \frac{9.2 + 8.7 + 8.6 + 8.8 + 8.5 + 8.7 + 9}{7}$$

$$= 8.78$$

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

$$= \sqrt{\frac{1}{7}\left\{(9.2-8.7)^2 + \cancel{(8.7-8.7)^2} + (8.6-8.7)^2 + (8.8-8.7)^2 + (9-8.7)^2 + (8.5-8.7)^2\right\}}$$

$$= \sqrt{\frac{1}{7} \times \frac{2}{5}}$$

$$= 0.23$$

z test (Normal test)	t - test
1. Used for large sample ($n \geq 30$) test	1. Used for small sample ($n < 30$) test
2. AKA large sample test	2. AKA small sample test
3. There is no degree of freedom in z-test	3. There is degree of freedom in t-test.
4. The critical value of z-test taken from table of standard ~~norta~~ normal distribution	4. The critical value of t-test can be taken from table of t-distribution.
5. If population variance (σ^2) is known, then normal test can be used.	5. If population variance (σ^2) is known, then t-test can not be used.

Class Test -1

Imperial College of Engineering, Khulna

Department of CSE

Stat-1211

Time: 50 minutes [Must answer **question no 4** and **TWO** from others] Marks: 10

1. Define Statistics. Point out the importance and scope of statistics. Write down two basic differences between primary data and secondary data. [3.33]

2. What are the scale of measurements? Find the different types of scale of measurement from the examples: family size, religion, temperature, weight and economic status. What are the sources of data? [3.33]

3. Define correlation and coefficient of correlation. Identify different types of simple correlation when r = -1, 0 < r < 1. Also figured out the scatter diagram for identifying different types of correlation. [3.33]

4. Calculate the correlation from the following data and comment in your result. [3.34]

X	8	6	5	12	8	13	7	12	29	2
Y	43	40	53	31	12	39	51	59	19	32