(- 3(W-2)

17/4/24

[Late Collect Prev Part]

Statistics

Statistics is a branch of applied mathematics which is applied to observational data. It deals with data collection, Presentation, organization and summarizing data, decision making etc.

Types:

i) Descriptive, [Everything upto Probability]

ii) Inferential; [Af Theory of Probability && TOPIC]

↳ Theory of Probability, Sampling, Test of Significance, Test of hypothesis etc → based on sample → if random is involved

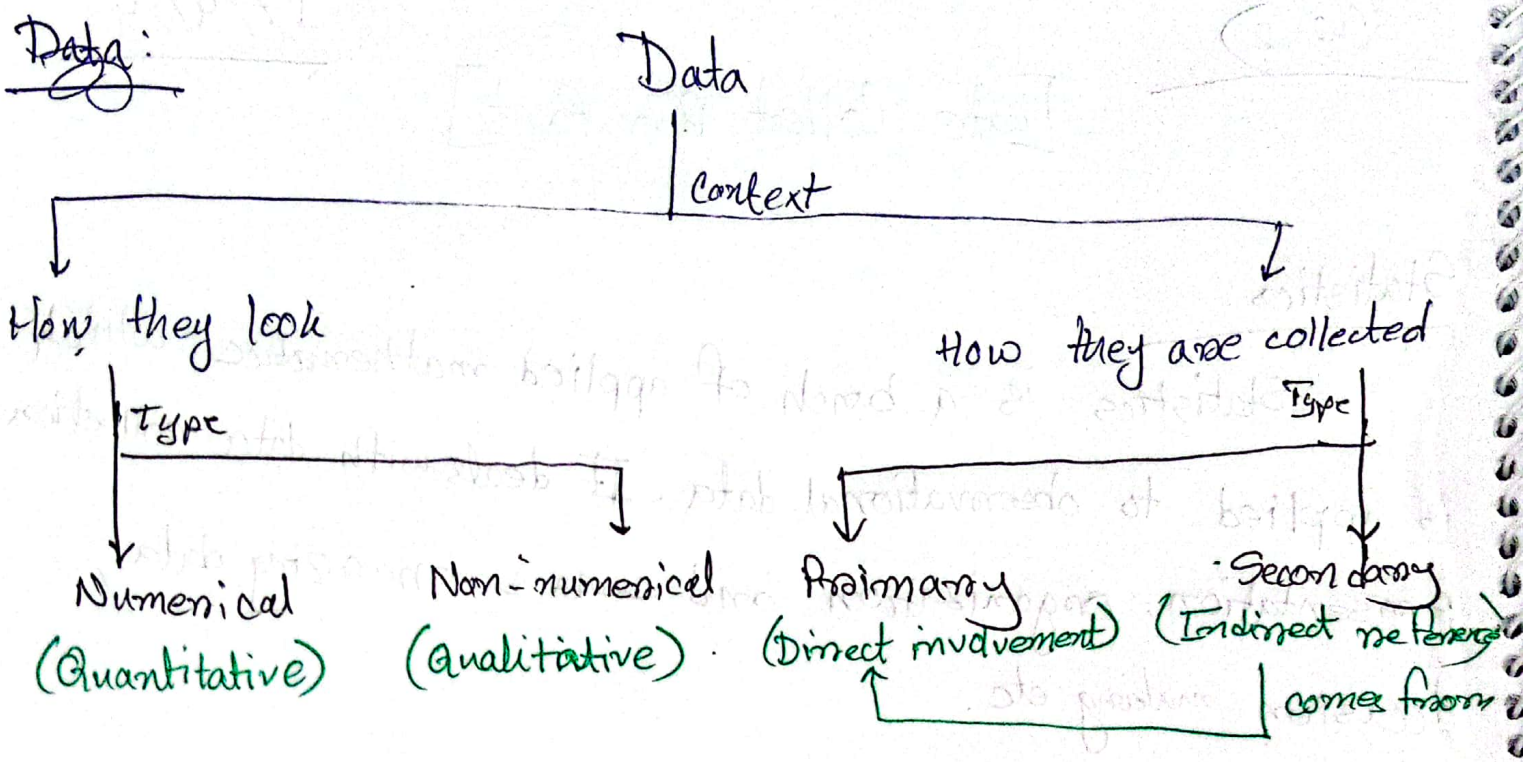
↳ Data Presentation, Organization and summarization of data, measures of location, dispersion, shape characteristics etc (central tendency things)

Population The entire set of individuals or objects of interest or the measurements obtained from all individuals or object of interest.

Sample: A [portion (representative), or part, of the population of interest]

↳ MUST Represent the popopulation

Data:



Types of Levels of Data:

i)

ii) Interval Level Data → 0 does not means any non-existent

iii) Ratio Level Data → 0 means non existent and a state

C-5(w-3) Statistics 24/04/24

(Lee Marshall)

Describing Data:

* Frequency Distribution Table:

Construct a FDT / Arrange the data / Make a table...

k = smallest number of classes

$2^k \geq$ no. of data ; For what minimum value of k, 2^k will be greater than number of data

For 180,

$k = 8$

$\therefore 2^8 = 256 > 180$

$\therefore$ Min class number = 256

i = class width, H = highest value, $?$ = lowest value

$i \geq \frac{H - ?}{k}$; always $\lceil i \rceil$

* If gap is high, you can adjust data

H = 3904
L = 204
i = 375

→ H = 3900, L = 200, i = 400

Here, the observation number is,

$n = 180$ [Check Q from Book]

The smallest number of classes(k) is,

$2^k \geq 180$

or, $2^8 \geq 180$ // find it in rough

which means, number of class, $k = 8$.

And, the class width (i) is,

$i \geq \frac{H - L}{k}$; { Here, highest value, $H = 3292$, lowest " , $L = 294$

or, $i \geq \frac{3292 - 294}{8}$

≥ 374.75

* N → usually population size
* but in this case, N = total frequency
* n = sample size

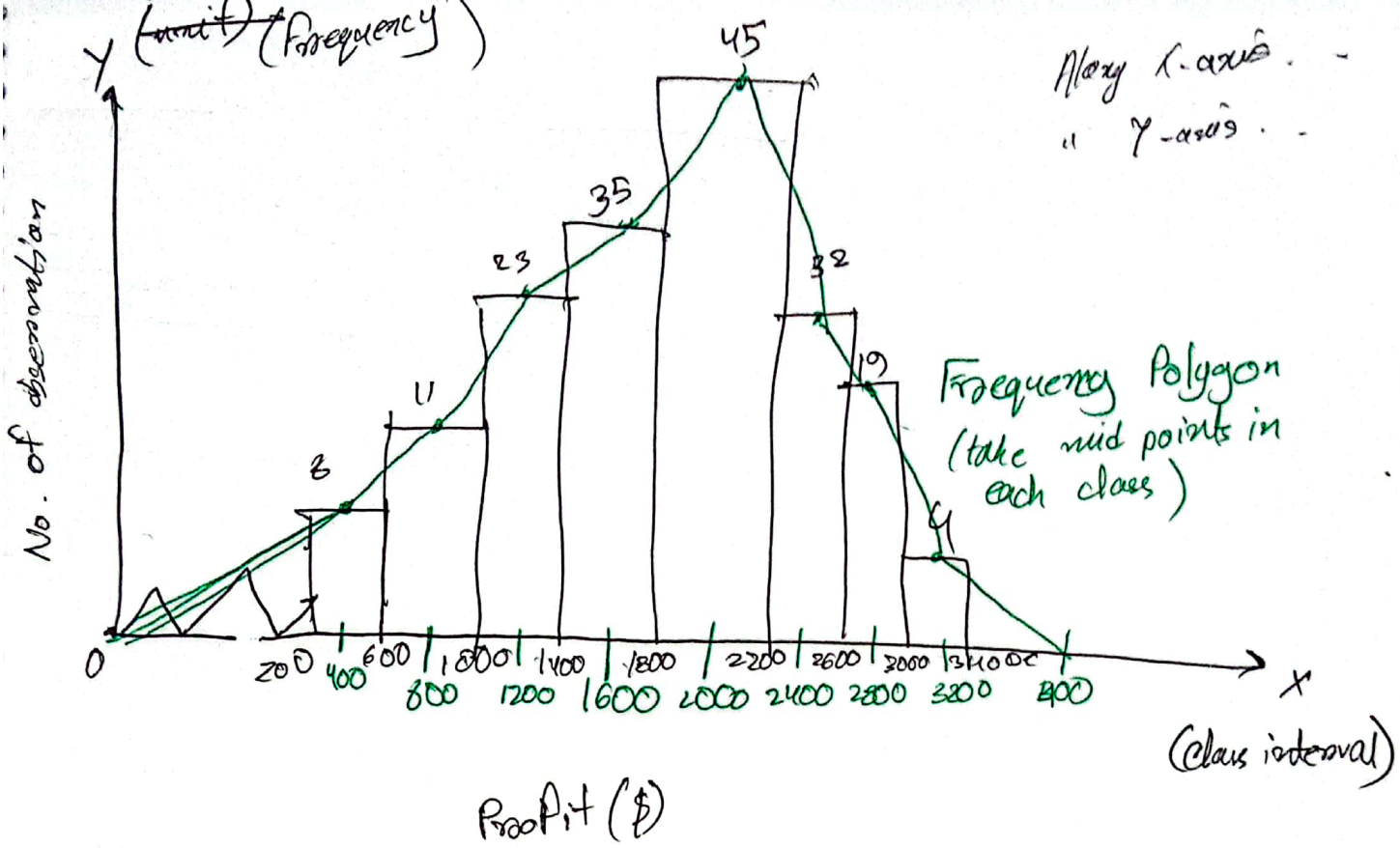
Here,

we consider the class width 400

~~we consider~~ 200 in lieu of 294 and ~~3292 in lieu of 3400~~ 3400 of 3292

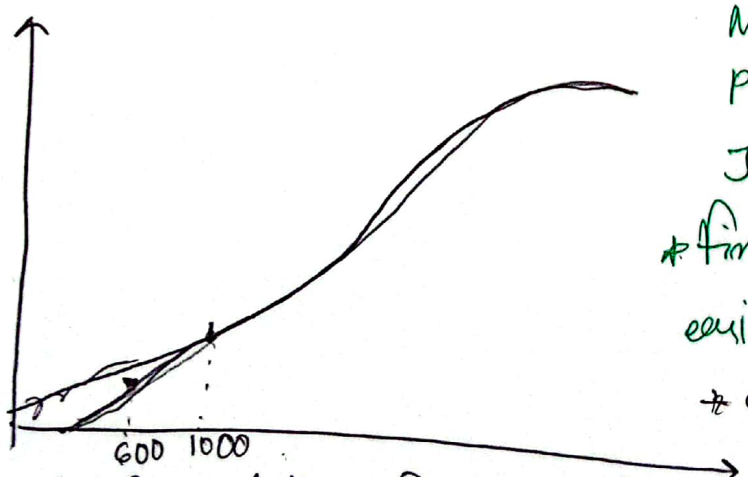
Table: Frequency Distribution Table of Profit ($) on vehicles sold by Applewood Auto // Name is MUST. Write short but MUST

class interval (on Profit)	Tally	Frequency f_i (actual, name less than cumulative frequency)	Cumulative Frequency F_i (used to get a better idea of everything)	Relative Frequency $\left(\frac{f_i}{N}\right)$ (proportion)
200 - 600	HH II	8	8	$\frac{8}{180} = 0.04$
600 - - -	II - -	11	19	$\frac{11}{180} = 0.06$
	$N = \Sigma f_i =$	180		



Many x-axis . .
" y-axis . .

Fig: Histogram of the Profit . . .
(Table Name but replace it ~~fig~~ with Histogram)



Cumulative frequency Polygon
or
Ogive Curve

No need to fit all points
Just draw smooth
* finding the slope is easier
* draw points on the highest class.

C-7 (W-5) 08/05/24

* When * classes have names, we use bar diagram or pie chart

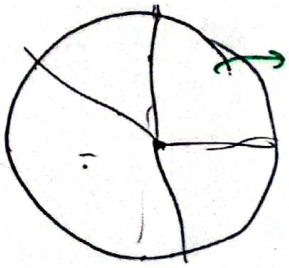
Bar Diagram:

- X-axis: Name ; scaling not needed
  Y-axis: Frequency ; scaling needed
- * Mark the bars : 
- * Classes without names can also be plotted.
  Replace names with class width.
  Y-axis remains same

Pie Chart:

At ~~first~~ we calculate the angles

Class	Frequency	Angle
Total	n	



* No need for protractor
or maintaining serial
or showing calculation

Measures of Central Tendency

1. Mean
   - → Arithmetic mean
   - → Geometric means (%)
   - → Harmonic (x/y units) → rate
2. Median
3. Mode

Formula for Ungrouped Data

* Mean :

Arithmetic mean: $\overline{X} = \frac{\Sigma x_i}{n}$

Geometric mean: $\sqrt[n]{x_0 x_1 x_2 \dots x_n} = \prod(x_i)$

(Lin Marshall - Z2)

(S.P Gupta - Q4 : 99)

Harmonic mean $= \frac{n}{\Sigma \frac{1}{x_i}}$

* Median:

The value which takes the middle most position an ordering arranged data set.

* Mode:

The value with maximum number of occurrence/ frequency in a data set is known as mode.

Formula For Grouped Data:

* Arithmetic Mean: $\bar{X} = \frac{\sum f_i x_i}{N}$

* Geometric Mean: Antilog $\frac{1}{N} \sum f_i \log x_i$; $\log_{10}$ (NOT ln)

Antilog $x = 10^x$

* Harmonic Mean: $\frac{N}{\sum \frac{f_i}{x_i}}$

* Median: $M_e = L + \frac{\frac{N}{2} - c.f}{f_m} \times C$

* Mode: $M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$

L = Lower limit of median class
cf = cumulative frequency of prev. median class
f_m = median class frequency
C = median class width
N = total observation

L = Lower limit
Δ_1 = diff. between frequencies of modal and premodal class
Δ_2 = " " modal and post modal class
C = modal class width

Example of Geo. Mean.

As the data is percentages, we will use geometric mean.

Let, 100 [rate] is invested

(replace with unit mentioned in the question.)

return 30 % means 130 [rate]

" 20 % " 120 [rate]

" 40 % " 60 [rate]

" 200 % " 300 [rate]

we convert the data in smaller form.

That is: 1.3, 1.2, 0.6, 3.0

Then, the (geometric) mean rate of return is:

$$G.M = \sqrt[n]{(x_1)(x_2)\ldots(x_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)}$$

$$= 1.294$$

i.e. 129.4 %

∴ Return is 29.4 %.

Example (S. C. Gupta) :

% Rise	Expenses taking 100 %	log x
32	132	.
40	140	.
:		
.		

* Do : S.C Gupta : Miscellanous Examples

(27, 30, 32, 33,)

(But do all)

Harmonic Mean:

$$H.M = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots}{n}}$$

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

Used Whenever there is a per unit (/unit) involved

* Any rate (km/hr, taka/kg)

x goes 20 km ~~per~~ in 5 hour

$\frac{20 \text{ km}}{5 \text{ hour}} = 4 \text{ km/hr}$; [for finding distance in t time]

$\frac{5 \text{ hour}}{2 \text{ km}} = \frac{1}{4} \text{ hr/km}$; [for finding time required to cross d distance]

Example:

A ⟶ 4 min/work = 0.25

B ⟶ 5 min/work = 0.2

C ⟶ 6 min/work = 0.167

D ⟶ 10 min/work = 0.1

E ⟶ 12 min/work = 0.833

i) Avg work per min = ?

ii) How much work in 6 hours per day

∵ The work is done per unit, we ~~do~~ will use harmonic mean

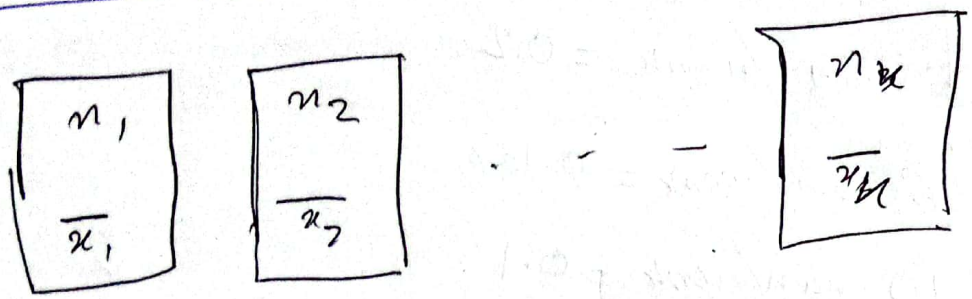
$$\therefore H.M = \frac{5}{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}+\frac{1}{12}} = 0.25 \text{ min/work}$$

$$= 0.16 \text{ work/min} \quad (\text{Ans.})$$

∴ work in 6 hours = $0.16 \times 60 \times 60$

$=$.

Combined Mean



k number of groups containing n_i elements with $\bar{x}_i$ mean

~~∴ They will have Com~~

∴ Combined Mean, $\bar{X} = \frac{\sum n_i x_i}{\sum n_i}$

Example (6(a)) :

Given that,

$n_1 = 100$

$n_2 = 80$

$\bar{x}_1 = 4570\$$

$\bar{x}_2 = 750\$$

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

Given: 6(b)

$\overline{x_c} = 72$

$n_1 + n_2 = 100$

$n_1 = 70$

$n_2 = 30$

$\overline{x_1}$ = something

$\overline{x_2}$ = ?

$$\overline{x_c} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2}$$

$\therefore \overline{x_2} =$.

Weighted Mean:

$$\overline{x_w} = \frac{\sum w_i x_i}{\sum w_i}$$

* Arithmetic mean but * FANCY * ~

@

Grouped Data

Example: Random sample of 25 from 2 brands. Decide which brand performs better in average

Hour	1-2	2-3	3-4	4-5	5-6
Type-I	2	4	8	7	4
Type-II	3	6	8	6	2

* Anything that has less deviation is more consistent

∴ DO NOT use this method.

Class	Frequency (f_i)	x_i	$f_i x_i$ (for arith mean)	$f_i \log x_i$ (for geo mean)	$\frac{f_i}{x_i}$ (harmonic mean)	Cumulative frequency c.f. (median)
				No approximation is allowed. BUT keep upto 3 digits. i.e. 1.99999 ≈ 1.999		
Total	$N =$		$\Sigma f_i x_i$	$\Sigma f_i \log x_i$	$\Sigma \frac{f_i}{x_i}$	

* Always ALWAYS,

$$A.M > G.M > H.M$$

* Median class is the class with half of the total frequency

Consumption of fuel (in lit/hr.)	Number of Generators (f_i)	x_i	$f_i x_i$	$f_i \log x_i$	$\frac{f_i}{x_i}$	Cumulative Frequency $c.f_i$
1-2	2	1.5	3	0.352	1.33	2
2-3	4	2.5	10	1.592	1.6	6
3-4	8	3.5	28	4.352	2.285	14
4-5	7	4.5	31.5	4.572	1.555	21
5-6	4	5.5	22	2.961	0.727	25
Total	$N = 25$		$\Sigma f_i x_i = 94.5$	$\Sigma f_i \log x_i = 13.829$	$\Sigma \frac{f_i}{x_i} = 7.497$	

$\therefore$ Arithmetic mean, $\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{94.5}{25}$

$= 3.78$ lit/hr

$\therefore$ Geo. Mean $=$ Antilog $\frac{1}{N} \Sigma f_i \log x_i$

$=$ Antilog $\{\frac{1}{25} \times 13.829\}$

$= 3.574$ lit/hr

Harmonic mean, $H.M = \frac{N}{\sum \frac{f_i}{x_i}} = \frac{25}{7.497}$

$= 3.334$ lit/hr.

Here, we have $N = 25$ observations and $N/2$

$= 25/2 = 12.5$

Thus, from the table we may decide median class is 3-4.

Therefore, the median is, $M_e = L + \frac{\frac{N}{2} - C.f}{f_m} \times C$

$= 3 + \frac{12.5 - 6}{8} \times 1$

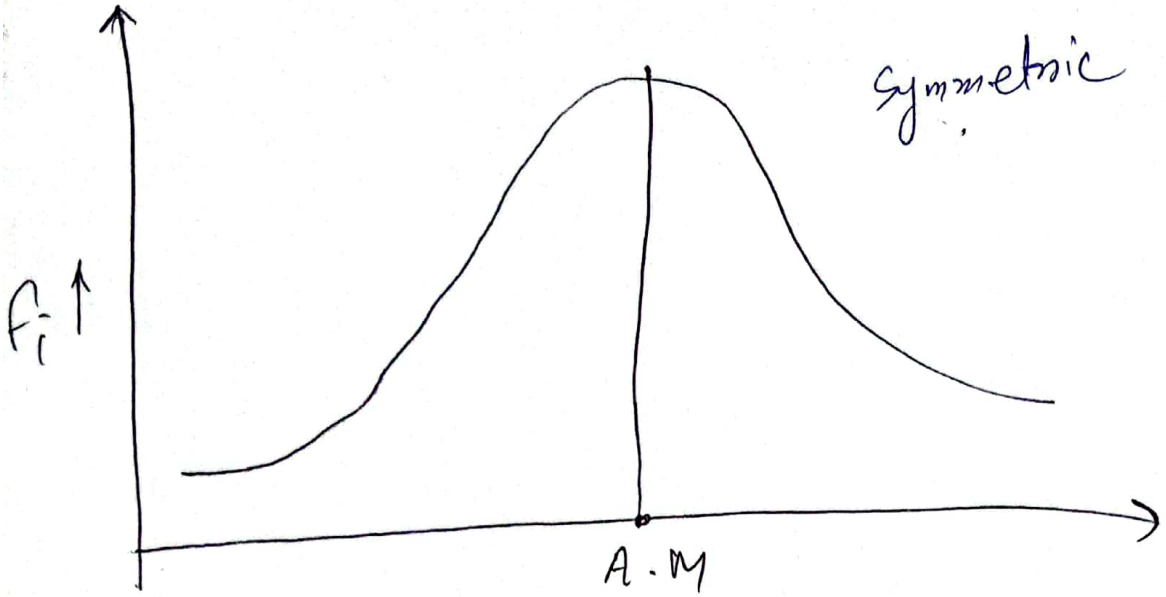
$= 3.08125$ lit/hr

Here, the class with the maximum number of frequencies is 3-4, which is 8. Thus, 3-4 is our modal class

∴ The mode is, $M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$

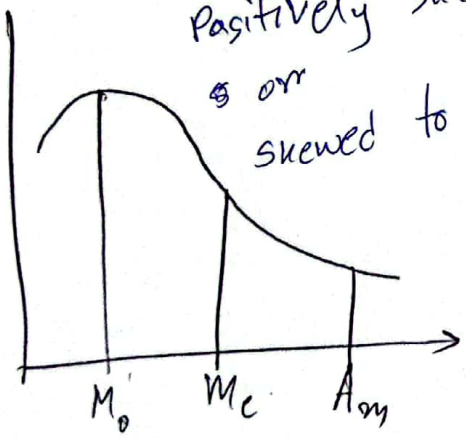
$= 3 + \frac{8-4}{(8-4)+(8-7)} \times 1 = 3 + 0.8$

$= 3.8$ lit/hr.

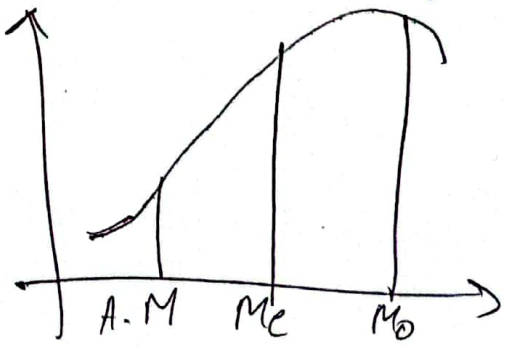


Skewed ⟶

Positively skewed or skewed to the right



Negatively skewed or skewed to the left.



C-8(10-8)

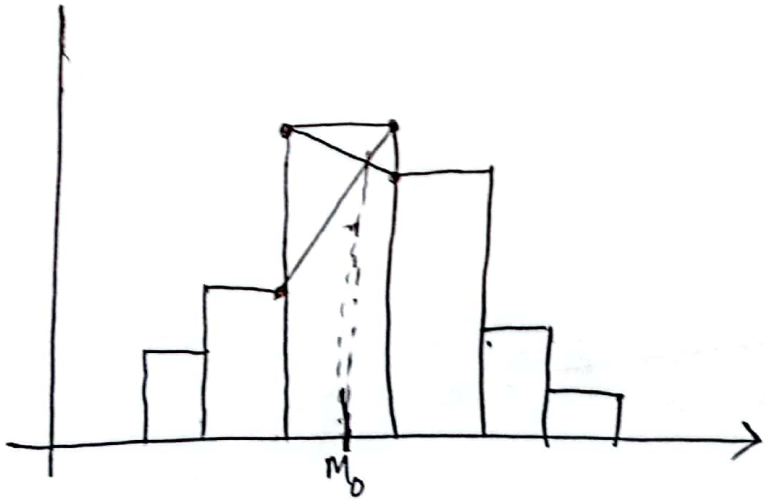
29/05/24

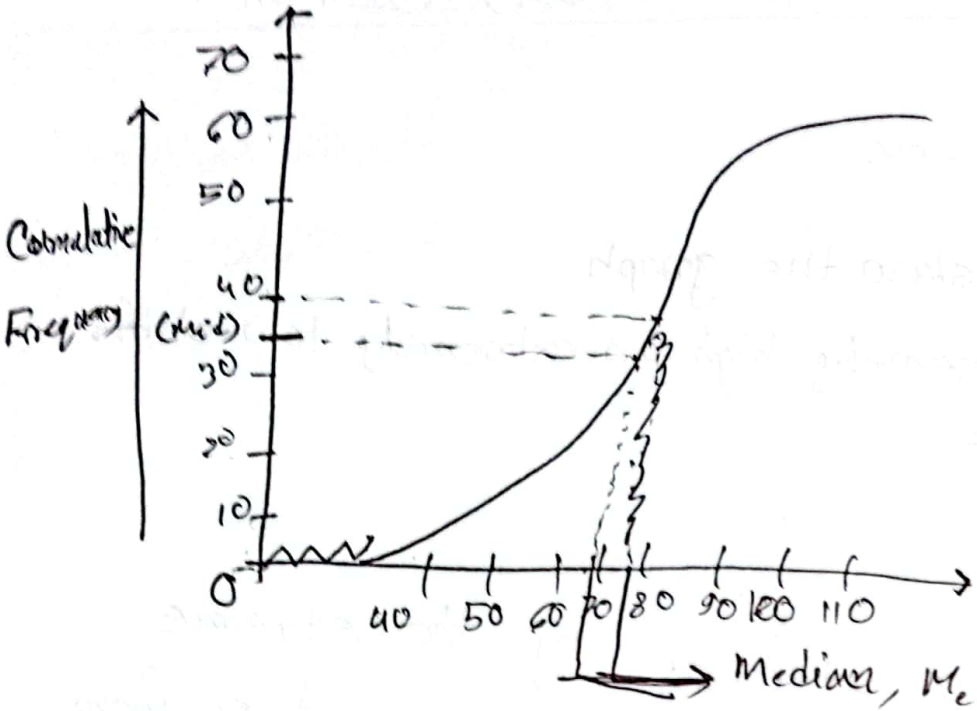
Relative Positions of the Mean, Median and Mode

Reason of Skewness

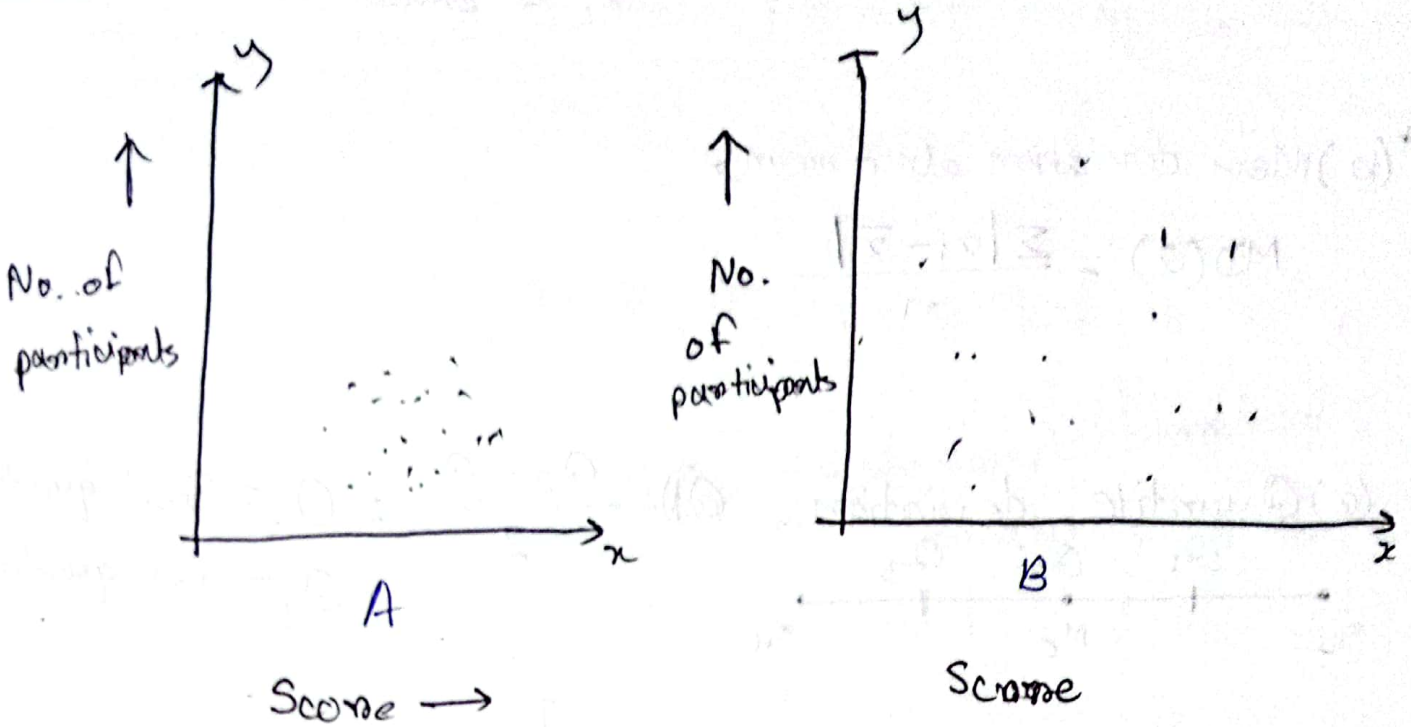
* Extreme values skew the graph
* Values that are extremely high or extremely low shifts the value towards it.

* Arithmetic mean is affected by the extreme value in a data set while the median is free from it (- Explain with example)




70
60
50
40
30
20
10
0
Cumulative
Frequency
(cm)
40
50
60
70
80
90
100
110
Median, M_e

Measures of Dispersion:



A is more consistent/reliable/steady etc than B

*Two types of measures:

i) Absolute measures
- (a) Range (R)
- (b) Mean deviation about mean (MD, $\bar{x}$)
- (c) Quartile deviation (QD)
- (d) Standard " (S)

ii) Relative "
- (a) Coefficient of range (C.R.)
- (b) " " mean deviation about mean (CMD, $\bar{x}$)
- (c) Coefficient of Quartile deviation (CQD)
- (d) Coefficient of standard deviation

or, Coefficient of Variation (C.V.)

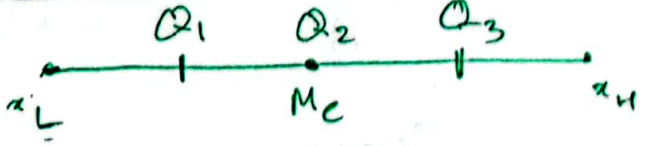
For ungrouped:

(i)(a) Range $= R = x_H - x_L$; x_H = highest value
x_L = lowest value

**(b) Mean deviation about mean,

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

(c) Quantile deviation, (ii) $= \frac{Q_3 - Q_1}{2}$; Q_3 = 3rd quadrant
Q_1 = 1st quadrant



***(d) Standard Deviation, $S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$; $\bar{x} = \frac{\Sigma x}{n}$

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

↳ preferred

For grouped:

(i)(a) No need

*(b) $MD(\bar{x}) = \frac{\Sigma f_i |x_i - \bar{x}|}{N}$; $N = \Sigma f_i$
$\bar{x} = \frac{\Sigma f_i x_i}{N}$

(c) No need

**(d) $S = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{\Sigma f_i x_i^2 - \frac{(\Sigma f_i x_i)^2}{N}}{N-1}}$

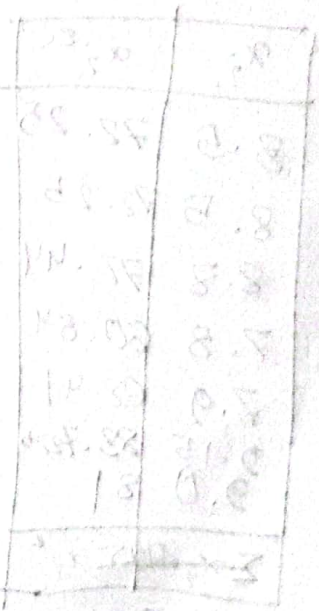
↳ preferred

2 (i)(a) $C.R. = \frac{x_H - x_L}{x_H + x_L} \times 100$

(b) $C.MD(\bar{x}) = \frac{MD(\bar{x})}{\bar{x}} \times 100$

(c) $CQD = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

(d) $C.V = \frac{S}{\bar{x}} \times 100$; S = sample standard deviation



#Example: Suppose, the value from strength test on some samples of yarns from different spinning machines are given (in lbs.)

from 1st machine:

8.5, 8.5, 8.8, 7.8, 7.9, 9.15, 9.0

from 2nd machine:

8.8, 9.2, 8.4, 8.4, 9.2, 9.10, 8.25

Decide which machine looks more reliable regarding their performance variations.

Soln: The calculation table for standard deviation is given below (for first machine)

x_i	x_i^2
8.5	72.25
8.5	72.25
8.8	77.44
7.8	60.84
7.9	62.41
9.15	83.7225
9.0	81
Σx_i	Σx_i^2

$\Sigma x_i = 51.15$ $\Sigma x_i^2 = 509.9125$

∴ the standard deviation for the first spinning machine is, $S = \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}} = 4.7636$

That is, the strengths of yarns are varying by 4.763 lbs from the value of the overall outputs.

Now, the coefficient of variation is,

$$C.V = \frac{s}{\bar{x}} \times 100$$

$$= 65\%$$

$$= 65\%$$

That is, the percentage rate of variation of strengths of yarns is 12% for the 1st machine.

[Do for 2nd machine]

The calculation table for standard deviation is given below (for second machine):

x_{i2}	x_{i2}^2
8.8	77.44
9.2	84.64
8.4	70.56
8.4	70.56
9.2	84.64
9.10	82.81
8.25	68.0625
$\Sigma x_i = 61.35$	$\Sigma x_i^2 = 538.7425$

∴ The standard deviation for the second machine is,

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

$$= \sqrt{\frac{(588.7125) - \frac{(61.35)^2}{7}}{6}}$$

$$= 0.413$$

$$C.V_2 = \frac{S}{\bar{x}} \times 100\ \%$$

$$= \frac{0.413}{8.764} \times 100\ \%$$

$$= 4.7125\ \%$$

$\therefore C.V_2 < C.V_1$

∴ ~~Machine~~ 2nd Machine looks more reliable.

C-9(W-8) *Gupta — Ch-5 (144) 30/5/24

#Slide Example (148) → avoid combined deviation (Online)

(33, 37)

	65-70	70-75	75-85	85-95	95-100
Machine I	1	4	8	6	2
Machine II	2	6	8	4	1

Calculation table for machine - 1

Class interval (what it says on question)	f_i (Frequency)	x_i	$f_i x_i$	$f_i x_i^2$
65-70	1	67.5	67.5	4556.25
70-75	4	72.5	290	21025
75-85	8	80	640	51200
85-95	6	90	540	
	N = 21		$\Sigma f_i x_i$ = 1732.5	$\Sigma f_i x_i^2$ = 144393.75

$$\therefore S = \sqrt{\frac{\Sigma f_i x_i^2 - \frac{(\Sigma f_i x_i)^2}{N}}{N-1}} = 8.5513 \text{ percent spot removal}$$

which means, the amount of variation in spot removal perc is 8.5513 unit

$$\therefore \bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{1732.5}{21} = 82.5$$

$$\therefore CV = \frac{S}{\bar{x}} \times 100\% = 10.36\%$$

That is, the percentage rate of variation in the spot removals of machine-I is 10.36%

[Do point 2]

Simple Linear Regression and Correlation

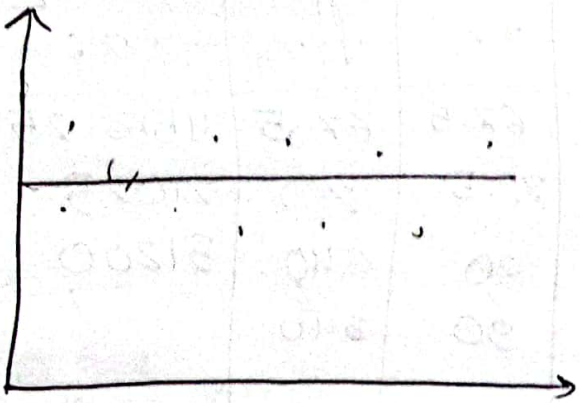
Correlation: relation between two or more variables

a) Simple Correlation : Correlation between only two variables

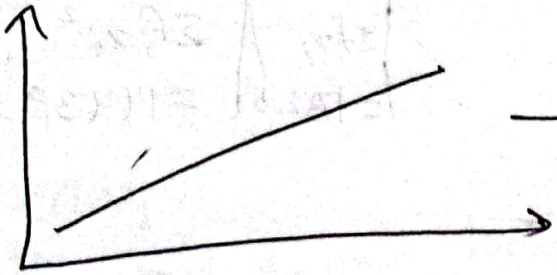
b) Multiple Correlation : Correlation " more then " "

c) Partial : Ignore what you dont want

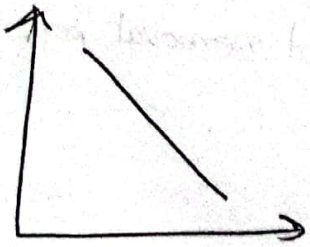
* Graphs with dots are called Scatter plots



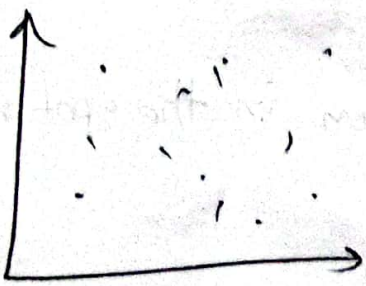
~~No~~ zero correlation



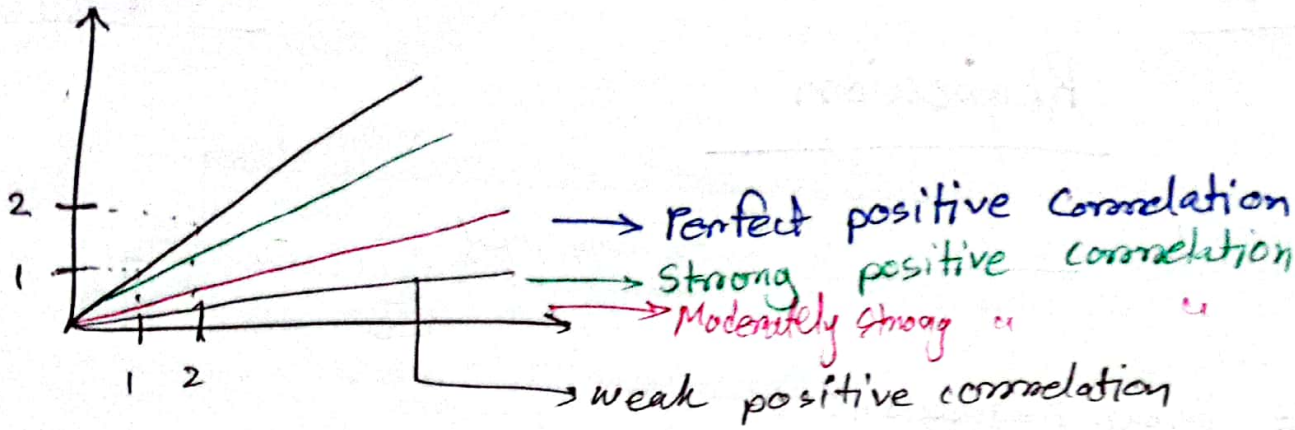
positive correlation

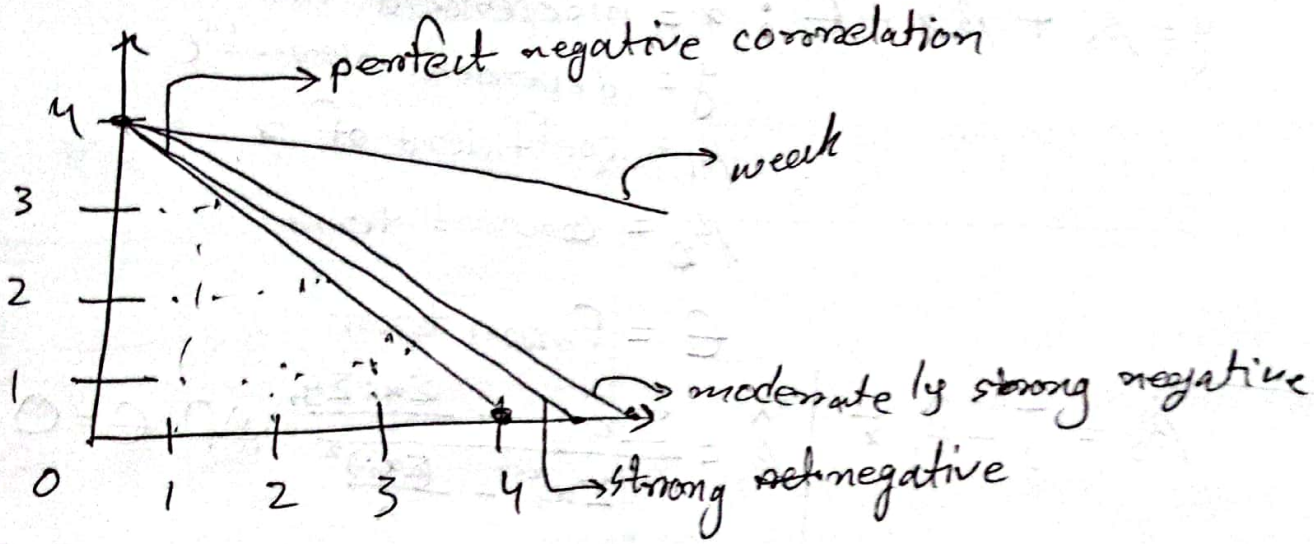


Negative correlation



No significant evidence of correlation





Pearson's Correlation Coefficient,

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

→ Covariance (x, y)
↳ determines tve or -ve

$-1 \leq r \leq 1$;

1 → perfect
0.7 - <1 → strong positive
0.4 <0.7 → moderately strong
>0 - <0.4 → weak

C-10(W-8)

31/05/24

Regression

* If indicates _____ is working well

Simple linear regression:

$y = \beta_0 + \beta_1 x + \epsilon$; x = independent variable
y = dependent variable
β_1 = coefficient of x
β_0 = Constant term
ϵ = Error term

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1 = \dfrac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}}$ ∴ $\sum \epsilon = 0$

*Error - will happen naturally

#Regression:

Let,

X = [data] (unit)

Y = [data] (unit)

The simple linear regression equation is, → The formula from last page

The calculation table of regression model/parameter is given below:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
$\sum x_i =$	$\sum y_i =$	$\sum x_i^2 =$	$\sum y_i^2 =$	$\sum x_i y_i =$

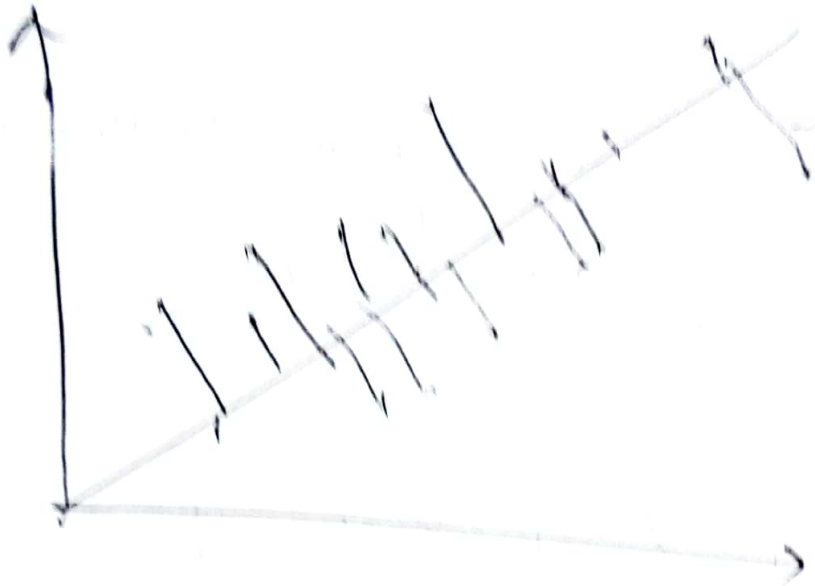
Thus, the estimates of the regression parameters are,

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = [\ ]$$

$$\beta_0 = [\ ]$$

∴ The regression model is,

$$y = \ldots + \ldots x$$



→ Amount of scatter is standard deviation
↓ But its not S.D.
Standard Error Estimate

Standard Error of the Estimate:

The measure of the dispersion or scatter of the observed values (sample observations) around the estimated regression line.

$$S_{yx} = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}} \overset{**}{=} \sqrt{\frac{\Sigma y_i^2 - \hat{\beta}_0 \Sigma y_i - \beta_1 \Sigma x_i y_i}{n-2}}$$

↳ preferred

$$= \sqrt{MSE} = \sqrt{\text{Mean Square Error}}$$

Coefficient of Determination

The amount of variation in the response variable (y) that can be explained by the variation of independent variable (x).

$$R^2 = \frac{RSS}{TSS} \overset{**}{=} 1 - \frac{ESS}{TSS};$$

RSS = Regression sum square = $\Sigma(\hat{y}_i - \bar{y})^2$

ESS = Error sum of square = $\Sigma(y_i - \hat{y}_i)^2$

Here, $TSS = RSS + ESS$ $[0 \le R^2 \le 1]$ ↳ preferred

where, TSS = Total sum of square = $\Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$

C-13(W-10)

26/06/24

*Ask P

Probability

* Randomness = uncertain outcome
* Variables used in probability are called Random Variables
* Sample Space : set of all possible outcomes. → S
* Event : Subset of sample space ~~for from~~ that contains required outcomes.

* Discrete : Individual
  Continuous : Between
* Mutually Exclusive : Two or more events with no common event
  ↳ $A \cap B = \emptyset$

Probability :

(Old) Let, a random experiment has 'n' outcomes
A be = any specific characteristic
n_A = number of outcomes having characteristic 'A'

$$\therefore P(A) = \frac{n_A}{n}$$

Axioms of Probability:

Let,

a random experiment has 'n' outcomes:

$$A_1, A_2, A_3 \ldots A_n$$

(i) $\Sigma P(A_i) = 1$

(ii) $0 \leq P(A_i) \leq 1$ { for any A_i }

(iii) For any two events A_1 and A_2, if $A_1 \cap A_2 \neq \emptyset$, then,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Rules of Solving Problems:

(i) Complementary Rule: $P(A) = 1 - P(A')$

(ii) Additional Rule: For any two events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B) \text{ ; [mutually exclusive]}$$

(iii) Conditional Rule: If A depends on B then,

A|B ≡ A given B → depends on; domain

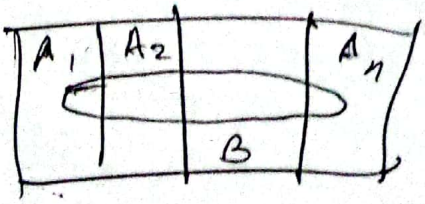
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ ; } [P(B) \neq 0]$$

(iv) Independence Rule: Any two events A and B are said to be independent of each other if,

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

(v) Bayes' Rule: $P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\Sigma P(A_i) P(B|A_i)}$

→ total sum



Suppose, in a controlled room, 3 operators are giving commands for completing a mission. Their previous records say that:

The chances of failures of the three operators are 9%, 12% and 7.5%.

The given mission will be considered completed if the 3 ops work together and each operator is independent of the other two ops.

Find the probability that a newly given mission will be completed successfully by them.

Let,

A = first operator works successfully

B = second " " "

C = third " "

Given that,

their failure rates in terms of probability are:

$P(A') = 0.09$, $P(B') = 0.12$, $P(C') = 0.075$

∴ Their success rates = $P(A) \cdot P(B) \cdot P(C)$

$= \{1 - P(A')\} \cdot \{1 - P(B')\} \cdot \{1 - P(C')\}$; []

$= 1 - 0.09$ $= 1 - 0.12$ $= 1 - 0.075$

$= 0.91$ $= 0.88$ $= 0.925$

∴ The probability that the given mission will be completed successfully is,

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

$$= (0.91)(0.88)(0.925); \text{ [ ]}$$

$$= 0.74074$$

$$= 74.074\% \quad \text{(Ans.)}$$

+ 25/09/22

$\cup$ { either or, neither nor, minimum of, at least

$\cap$ { and, all, both, including all, all inclusive, nothing will be left alone etc.

Conditional Rule { such that, given that, subjected to the, condition that, restricted to, constrained to etc

C-21 (00-14) 25/09/24

Related data

Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	940

(Ch-2, Ex-2.19)

Suppose Si wafers are to selected. In total 582 are of low level contamination, 112 wafers are high but center. There are 84314 edge ones.

Suppose one wafer is selected at Random. Find the probability that

i) If is located in the edge area or it has low level contamination

(ii) If is located in the edge area and it has low level contamination

(iii) If is located in the edge area such that it has low level contamination

(iv) If is located in the edge area but it has low level contamination

~~The infom~~

Sol^n: ~~The in~~ Let,

H = high level contamination

L = low "

C = centen area

E = edge "

∴ The information table is,

Contamination Table	Location: Edge (E)	Centre (C)	Total
High (H)	246	112	358
Low (L)	68	514	582
Total	314	626	940

∴ Now,

the probability that a randomly selected wafer is

(i) located in the edge area or it has low level contamination

$= P(E \cup L) = P(E) + P(L) - P(E \cap L)$

$= \frac{340}{940} + \frac{582}{940} - \frac{68}{940} =$

(ii) Located in the edge and it has low level contamination $= P(E \cap L) = \frac{68}{940} =$

(iii) Located in the edge such it has low level contamination $= P(E|L) = \frac{P(E \cap L)}{P(L)}$

$$= \frac{\frac{68}{940}}{\frac{582}{940}}$$

$$= \frac{68}{582} = \cdots$$

(iv) [Same as (ii)]

* Most something (Red book): 2.20, 2.22, ~~2.9~~

Exercise: 2.78, 2.77, 2.81, 2.90, 2.94

2.7. Bayes Theorem:

A_1	A_2	A_3
35%	40%	25%
.01% error	0.011% error	0.012% error

Suppose there is an error.
Which person has the most probability of doing the error

Soln:

Let,

A_1 = The work is done by the first person
A_2 = " " " " " second "
A_3 = " " " " " third "

and E = The work is error

Given that,

the overall probability of ~~banks~~ of the three persons are

$P(A_1) = 0.35$

$P(A_2) = 0.4$

$P(A_3) = 0.25$

The errors rates of the ~~true~~ three ~~persons~~ are:

$P(E|A_i) = 0.0001$

$P(E(A_2) = 0.00011$

$P(E|A_3) = 0.00012$

~~Hence~~, the probability that a randomly found errors in the work of a given day is done by first person is $P(A_1|E) = \frac{P(A_1) \times P(E|A_1)}{\sum P(A_i) P(E|A_i)}$

$$= \frac{P(A_1) \times P(E|A_1)}{P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + P(A_3)P(E|A_3)}$$

$$= \frac{0.35 \times 0.0001}{(0.35)(0.0001) + (0.4)(0.00011) + (0.25)(0.00012)}$$

2nd Person = $\frac{P(A_2)P(E|A_2)}{\boxed{\quad}}$

$=$

3rd Person $= \frac{P(A_3)\ P(E|A_3)}{\boxed{\quad}}$

Maybe, [greatest value] is most responsible for the errors.

2.38, 2.148,

SP Gupta: Ch-11: 400 → 14, 15, 27, 30

Exercise → 29, 30, 31

Probability Distribution

Binomial Probability Distribution:

summation of Bernoulli trial is Binomial Distribution

$$f(x) \longrightarrow p(x) = \binom{n}{x} p^x q^{n-x} ; \quad x = 0, 1, 2, \ldots, n$$

n = Trial number

x = No. of success

p = Probability of success (from past experience)

q = " " failure

and, $p + q = 1$

Mean, $E(x) = np$

Variance, $V(x) = npq$

C-22(10-14)

26/09/22

Bernoulli Trial:

Trial with only two possible outcomes; expected as failure.

Binomial Distribution:

$$P(x) = f(x) = {}^{n}C_{x}\, p^{x} q^{n-x} \; ; x = 0, 1, 2, \ldots, n$$

where,

n = trial number

x = no. of success → x is whatever you want to find

p = probability of success (from prev. experiences)

q = " " failure

and, $p + q = 1$

Mean $= E(x) = np$

Variance, $V(x) = npq$

$np > npq$

If the chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also assume that the transmission trials are independent. If 4 bits are transmitted, find the probabilities that

(i) Exactly 2 bits will be errors
(ii) At least 2 bits will " "
(iii) At best 1 bits " " "
(iv) More than 1 but not more than 3 bits will be errors.
(v) At best 3 bits will be correct

Soln:

Let, x be the numbers of bits received in error which follows a Binomial distribution

Given that,

trial numbers, $n = 4$

Probability of success, $p = 0.1$

" " failure, $q = 1-p = 0.9$

Thus,

Probability function is,

$$P(x) = \binom{n}{x} p^x q^{n-x} = \binom{4}{x} (0.1)^x (0.9)^{4-x} \; ; [x = 0, \ldots, 4]$$

(i)

∴ The probabilities that,

(i) Exactly 2 bits will be in error

$= P(2) = {}^4C_2 (0.1)^2 (0.9)^{4-2}$

=

(ii) At least 2 bits will be in error

$= P(x \geq 2) = P(2) + P(3) + P(4)$

=

=

(iii) At best 1 bits in error

$= P(x \leq 1) = P(0) + P(1)$

=

=

(iv) More than 1 but less than 3.

$= P(1 < x \leq 3) = P(2) + P(3)$

$=$

(v) Hence,

x is the number of bits received correctly

$\therefore$ Probability of success, $p = 0.9$

$\therefore$ " " failure, $q = 0.1$

$\therefore P(x) = {}^{4}C_{x}(0.9)^{x}(0.1)^{4-x}$; $[x = 0, 1, 2, 3, 4]$

Now,

Probability of at best 3 bits will be received correctly $= P(x \leq 3) = P(0) + P(1) + P(2) + P(4)$

$= 1 - P(4) =$

Mean → 3.18,

Ex → 3.84, 3.87, 3.91

SP Gupta → 421 (Ch-12) → 4, 5, 35

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

Here,

Mean, $E(x) = \mu$; $[-\infty < \mu < \infty]$

Variance, $V(x) = \sigma^2$; $0 < \sigma^2 < \infty$

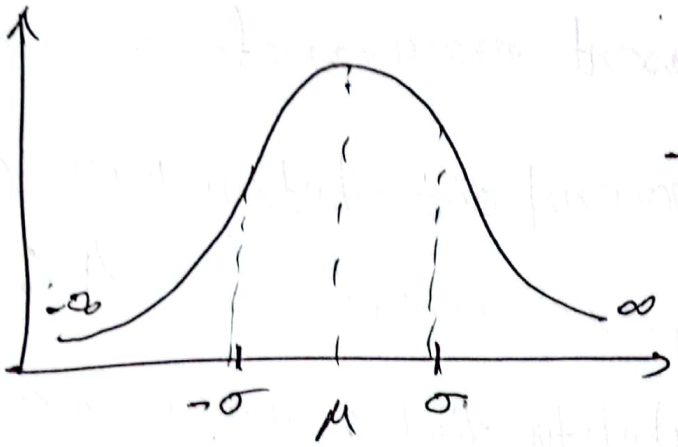
Standard deviation $= \sigma$

Standard Normal Distribution:

When $\mu = 0$ and $\sigma^2 = 1$, then, it is a special case of Normal Distribution, which is called standard normal distribution.

The variable is z where, $z = \frac{x-\mu}{\sigma}$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$



$$\int_{\mu-\sigma}^{\mu+\sigma} f(x)\,dx = 0.6827 \;;\; \mu \pm \sigma$$ covers 68.27% of area

$$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x)\,dx = 0.95445 \;;\; \mu \pm 2\sigma$$ " 95.45% " "

$$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x)\,dx = 0.9972 \;;\; \mu \pm 3\sigma$$ " 99.72% " "

4.10, 4.13, 4.14

Assume that the current measurements in a strip of wire follow a normal distribution with a mean of 10 milliamperes and variance 4 (mA)

(i) what is the probability that a measurement exceeds 13 mA?

(ii) what is the probability that a current measurement is between 9 and 11 mA?

(iii) Determine the value for which the probability that a current measurement is below this value is 0.98

[Use the stat table]

Mont — 4.15, 4.16

Ex → 4.58, 4.61, 4.62, 4.64, 4.65

Soln:

Let,

x be the current measurement which follows normal distribution

Given that,

mean, $\mu = 10$ mA

variance, $\sigma^2 = 4$ mA

$\therefore$ Standard deviation, $\sigma = 2$ mA

The probability that the current measurement for a randomly given strip of wire is

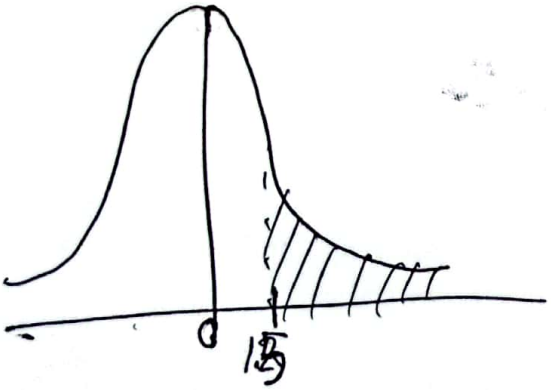
(i) will exceed 13 is,

$$P(x > 13) = P\left(\frac{x-\mu}{\sigma} > \frac{13-\mu}{\sigma}\right)$$

$$= P(z > 1.5) = 1 - P(z < 1.5)$$

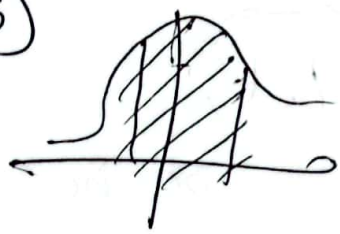
$$= 1 - 0.9332$$

$$= 0.0668$$



(ii) between 9 and 11

$$P(9 < x < 11) = P(-0.5 < z < 0.5)$$



$$= P(z < 0.5) - P(z < -0.5)$$

$$= 0.6915 - 0.3085$$

$$= 0.3830$$

(iii) the value for which the probability that a current is below the value is 0.98

$$z = 2.05$$

$$\text{or, } \frac{x - \mu}{\sigma} = 2.05$$

$$\therefore x = 14.1 \text{ mA}$$