DATE: / /20
PAGE NO.

Statistic :-

1. Infer Inferential
2. Descriptive

1. Inferential Statistic :-

It deal with taking sample an analysis sample making judgments or claims about population

2. Descriptive Statistic :-

It refer to getting data and talking about it

Measure variable

Quantitative data

Data that is measure in number. it deal with the number that make sense to perform Arithmetic calculations with

quantitative variable
- Height
- weight
- midterm scorse

Categorical data

Refer to the value that place "Thing" into Different group or category

categorical variable
- Hair color
- Type of cat
- Letter Groad

categorical variable

categorical and ordinal

Logical ordering to the value of a categorical variable

Ex :- Letter grade

A+, A, B+, B, C+, C

categorical and Nominal

NO Logical ordering to the value of a categoric variable

Ex :- Hair color

Black, Blonde, Brown

DATE: / /20
PAGE NO.

Quantitative variable

Discrete

Refer to the variable that can only be measured in certain numbers

Ex:- No. of pets

0, 1, 2, 13, ~~2.7~~ ✗

continous

Refer to variable that can takes on Any numerical value

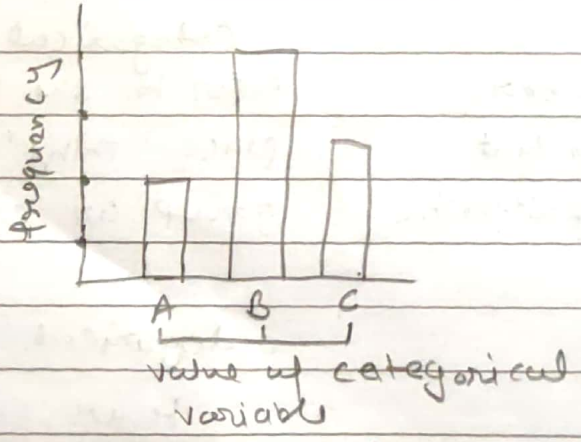
Ex:- weight

105 185 170 683

Charts

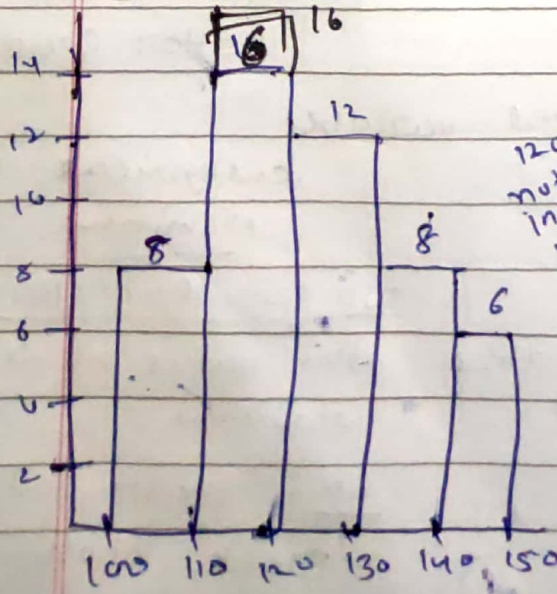
Categorical data :-

for categorical data we use **bar chart** or **pie chart**



Quantitative data :-

Stemplot, Histogram, Time Plot



120 is not included in this interval

weight	frequence
100 - 110	8
110 - 120	16
120 - 130	12
130 - 140	8
140 - 150	6

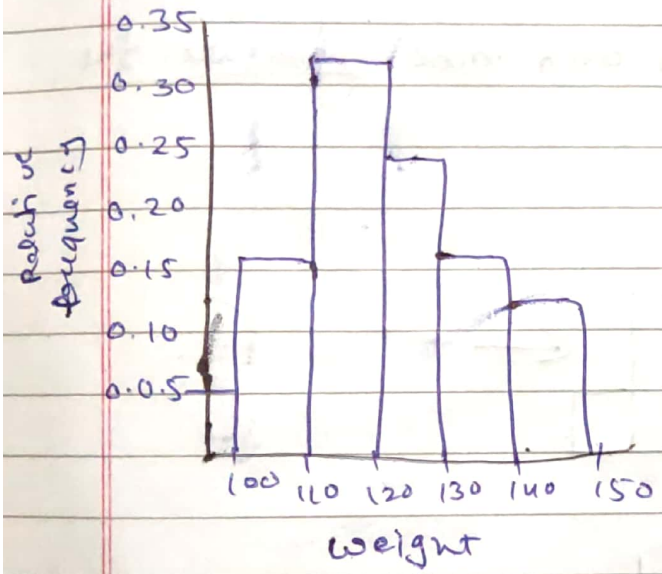
frequen distribution

DATE: / /20
PAGE NO.

Relative frequency Distributions

weight	~~Relative~~ frequency		Relative frequency
100 - 110	8	8 ÷ 50	0.16
110 - 120	16	16 ÷ 50	0.32
120 - 130	12	12 ÷ 50	0.24
130 - 140	8	8 ÷ 50	0.16
140 - 150	6	6 ÷ 50	0.12
	50		1

Relative frequency Histogram



STEMPLOTS

STEMP | LEAF

Refer to the all of the other number exlept the last number

Refer to the very last number

117

11 | 7

, 69

6 | 9

30 32 33

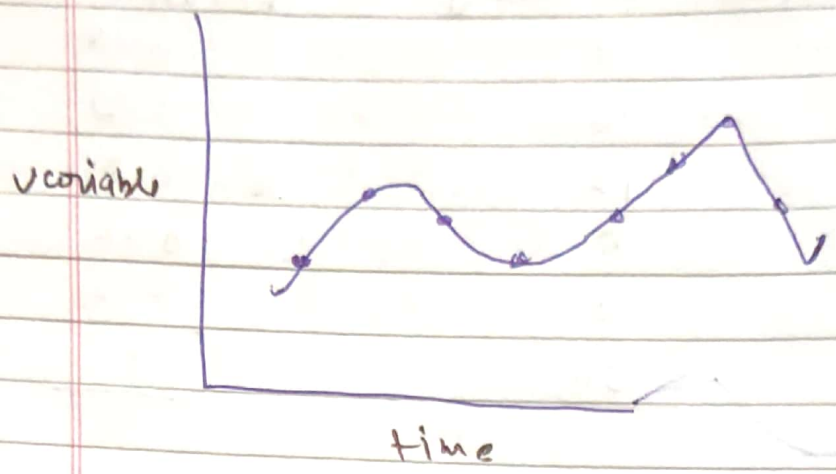
3 | 0 1 2 5 3

31 35

DATE : / /20
PAGE NO.

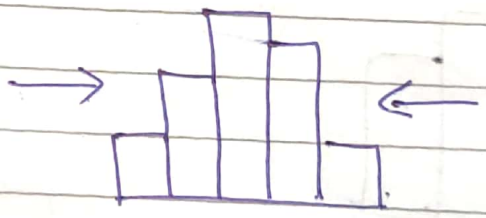
TIMEPLOT

Show how a variable changes over time



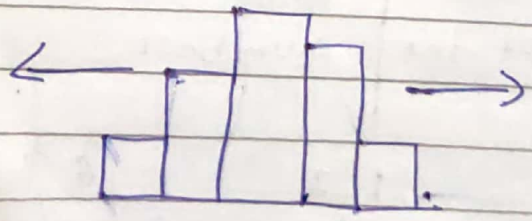
Note :-

mean, median and mode measure the centeral tendency



Note :-

Range and SD measure the spread of data



DATE: / /20
PAGE NO.

measure of center :-

sample Data = 139, 140, 154, 154, 154, 155, 180, 192, 192

mode = 154 , median = 154

mean = 165.6

measure of Spread :-

$$\text{Range} = \max - \min$$

Sample S.D

Rang = 192 − 139 = 43

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

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
10	−5.4	29.16
12	3.4	11.56
16	0.6	0.36
19	3.6	12.96
20	4.6	21.16
		75.6

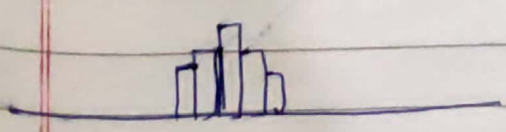
$\bar{x} = 15.4$

$$S.D = \sqrt{\frac{75.2}{4}}$$

$$= \sqrt{18.8}$$

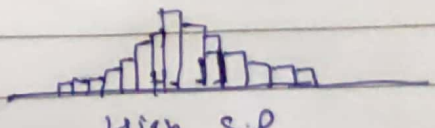
$$S.D = 4.336$$

Note :- S.D tell us how cause the value in a Data Set are to the mean



Small S.D

Note :- Small amount of variability of Dataset



High S.D

Note :- higut amount of variability of data Set

DATE: / /20
PAGE NO.

variance

sample variance ↓

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

\# what is the mean weight and ~~weight~~ S.D ?

weight of Student →	105	156	145	172	100

$\bar{x} = 135.6$ S.D = 31.75

Suppose that during the winter, each person wear an extra 5 pounds of cloths, Under these conditions, what is the new mean and S.D ?

$\bar{x}_{new} = 135.6 + 5 = \boxed{140.6}$

$S.D_{new} = 31.75$ → bcz spread of data is same for old and new data

Transforming Data

measure of ~~spread~~ center (mean, median, mode)

Note :
Affected by
(+) (−) (*) (%)

measure of spread (range, S.D)

Note :
Affected by
(*) (%)

DATE: / /20
PAGE NO.

Suppose that in order to stay Hydrated, these Student drink 2.5 ml of water for every pound of they weight Plus 750 ml of water every day. what is the mean and S.D for the amount of water tomorrow every day?

Weight = 105 156 145 172 100

$\bar{x} = 135.6$

$S.D = 31.75$

$\bar{x}_{new} = (135.6) \times 3.5 + 750$

$\boxed{\bar{x}_{new} = 1089}$

$S.D_{new} = (31.75) + 3.5$

$\boxed{S.D_{new} = 79.38}$

measure of center

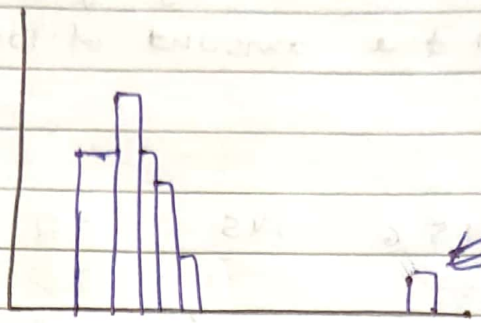
$$\boxed{center_{new} = (center_{old}) * x + b}$$

measure of spread

$$\boxed{spread_{new} = (spread_{old}) * x}$$

DATE: / /20
PAGE NO.

The effect of outlier on spread and center



outlier

9 x 1/6 21 20 22 25 21 (90) 23

(3) 51 56 69 73 81 80

outlier

Note :-

outlier can effect measure of center and spread.

Ex :- temperature of winnipeg on July

year	Temp
2015	26°
2014	15°
2013	20.5°
2012	31°
2011	-350°
2010	31°
2009	30.5°

outlier

$\bar{x} =$ [-28°] ← not possible

↑ mean is effected by outlier

- Range is also effected by outlier
- as we know mean is effected by outlier & as a result S.D also effecd by outlier

DATE: / /20
PAGE NO.

Five Number Summary

Min	1st Quartile	Median	3rd Quartile	Max
10	25	33	36	59

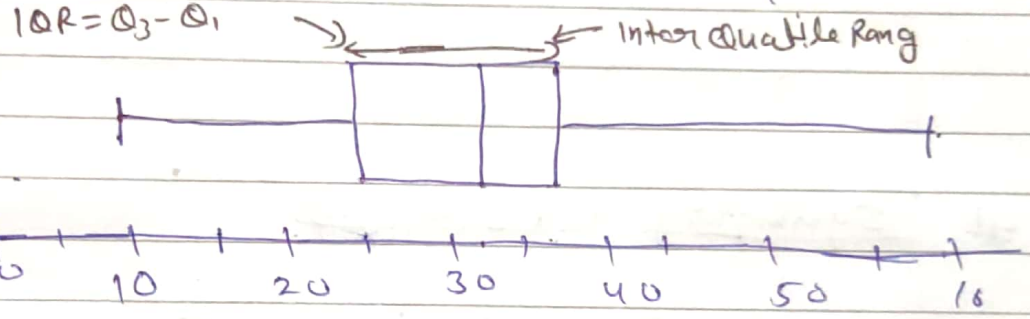
Data set

10 11 12 [25] 25 27 31 [33] 34 34 35 [36] 43 50 59

7 7

⇒ Boxplot :-

Gives us a visual representation of the 5 No. summary



⇒ Modified Boxplot :-

$$IQR = Q_3 - Q_1$$

A data value is considered to be an outlier if

$$\text{Data value} < Q_1 - 1.5(IQR)$$

or

$$\text{Data value} > Q_3 + 1.5(IQR)$$

$IQR = Q_3 - Q_1 = 36 - 25 = 11$

Data value is an outlier if it is

less than : $Q_1 - 1.5(IQR) = 8.5$

greater than : $Q_3 + 1.5(IQR) = 52.5$ outlier

Data Set

10 11 12 25 25 27 31 33 34 34 35 36 43 50 (59)

DATE: / /20
PAGE NO.

Percentile :-

Describle the % of data values that fall at or below another data value

Q Suppose we have 15 student in class and a test is conducted of 50 marks. mark obtain by student is

15 20 25 31 34 34 35 35 36 38 36 38 38 39 43

31 → 25th percentile

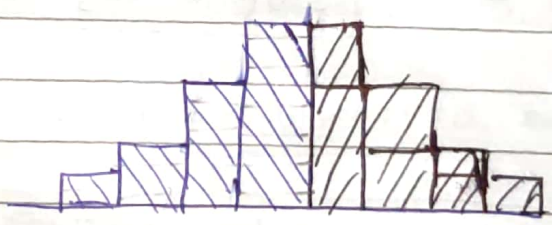
35 → 50th percentile

38 → 75th percentile

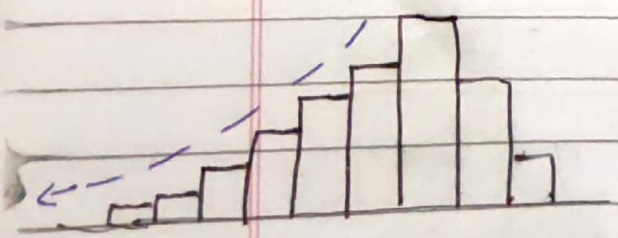
this student beat 75% of the class

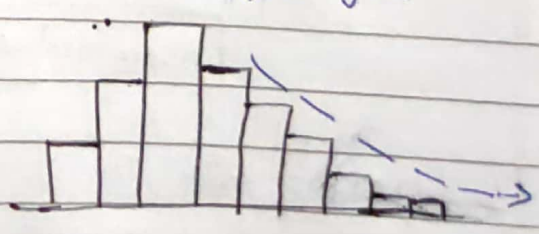
Symmetry and Skewness

Distribution : Symmetric



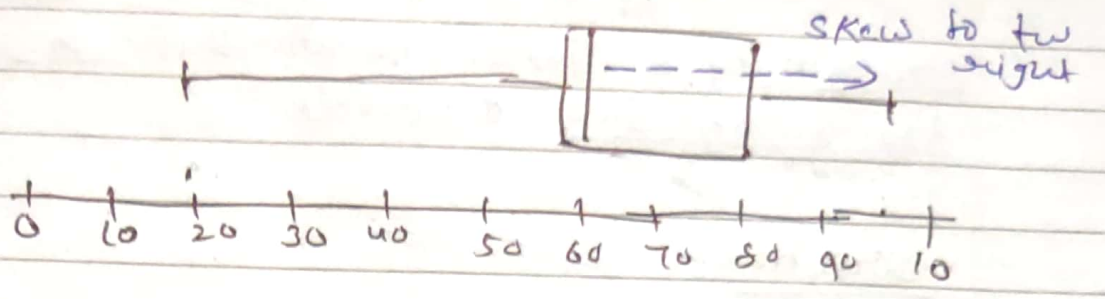
Skewness Refer to Asymmetric



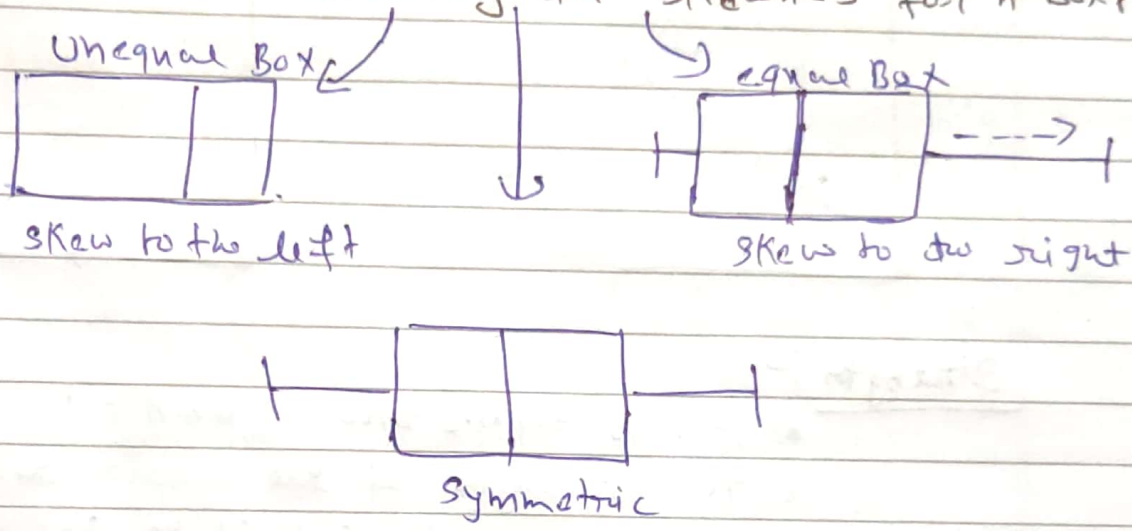


DATE: / /20
PAGE NO.

Boxplot :- 21, 50, 51, 52, 60, 60, 61, 61, 62, 63, 63, 70, 70
71, 71, 80, 80, 80, 90, 93



Strategies for determining the Skewness for A Boxplot



Note :-

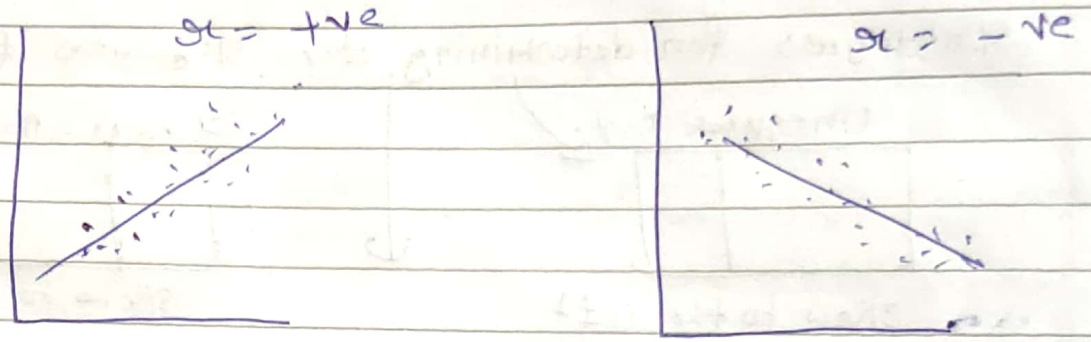
1. if Distribution is Skewed to left then mean is less than median
2. if distribution is Skewed to right then median is greater than median.

DATE: / /20
PAGE NO.

correlation :-

It tells you about the direction and strength of a linear relationship shared b/w two Quantitative variable.

Direction



Strength :-

- r has values b/w 1 and -1
- The strength of the linear relationship increase as r got close to 1 or -1

$$\text{correlation } (r) = \frac{1}{(n-1)\,S.D_x\,SD_y} \Sigma(x_i - \bar{x})(y_i - \bar{y})$$

Ex:- A teacher wants to Determine the correlation b/w the no. of hours spent studying and test score.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
13	53	0.5	-15	-7.5
15	69	2.5	1	2.5
7	92	-5.5	24	-132
3	10	-9.5	-58	551
10	85	-2.5	17	-42.5
27	99	14.5	31	449.5

$\bar{x} = 12.5 \quad \bar{y} = 68$

821

DATE: / /20
PAGE NO.

$SD_x = 8.28 \quad S.D_y = 32.91$

$$r = \frac{1}{(6-1)(8.28)(32.91)} (821) = 0.602$$

$r = 0.602$

Regression and R-Square

Regression line :- (increase/decrease)

It predict the change in "y" when "x" increase by one unit

$\hat{y} = b_0 + b_1 x$ ← Any value of x

- $\hat{y}$: Predicted value of y
- b_0 : Intercept
- b_1 : Slop

$$b_1 = r \frac{S.D_y}{S.D_x}$$

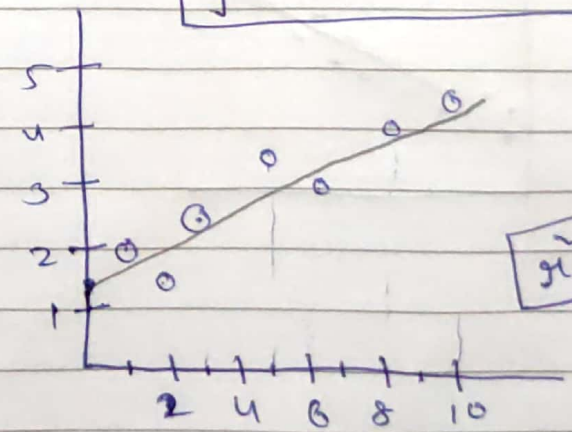
$$b_0 = \bar{y} - b_1 \bar{x}$$

Imp

Q:- Suppose researcher want to predict a student's GPA from the amount of time they study each week

$\hat{y} = 1.45 + 0.311 * x$

study time	y_i
1	2.0
2	1.5
3	2.5
5	3.5
6	3.0
8	4.0
10	4.5



$r = 0.88$

$\bar{x} = 5$ $\quad \bar{y} = 3$

$S.D_x = 3.26$ $\quad S.D_y = 1.08$

$r = 0.94$

$b_1 = 0.311$

$b_0 = 1.45$

DATE: / /20
PAGE NO.

R-Squared :-

	r	r^2
1.	Has value b/w -1 and 1	Has value b/w 0 and 1
2.	Measure the relationship b/w two Quantitative variable with respect to Direction and strength	It's a measure of how close each data point fit to the Regression line. it tell us how well the regression line predict the Actual values

Effect of outlier and Extrapolation on Regression