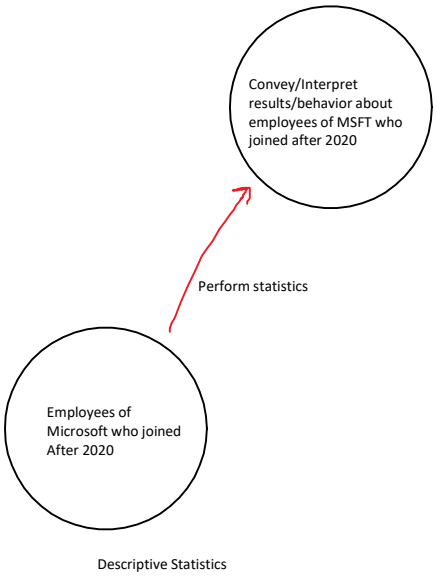
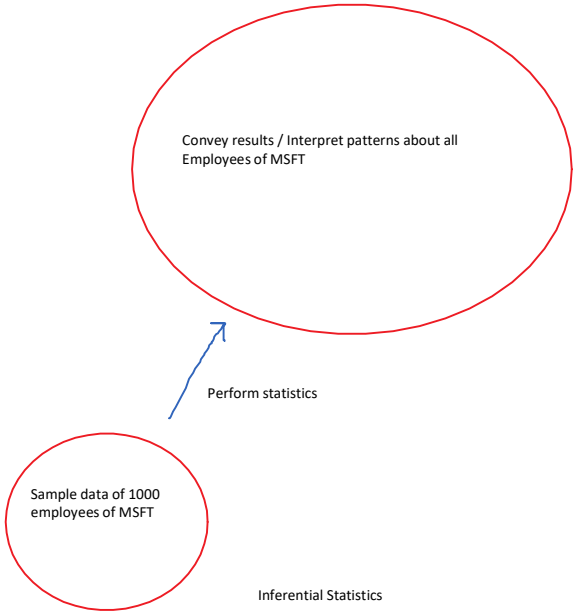
Statistics

29 April 2022 14:19

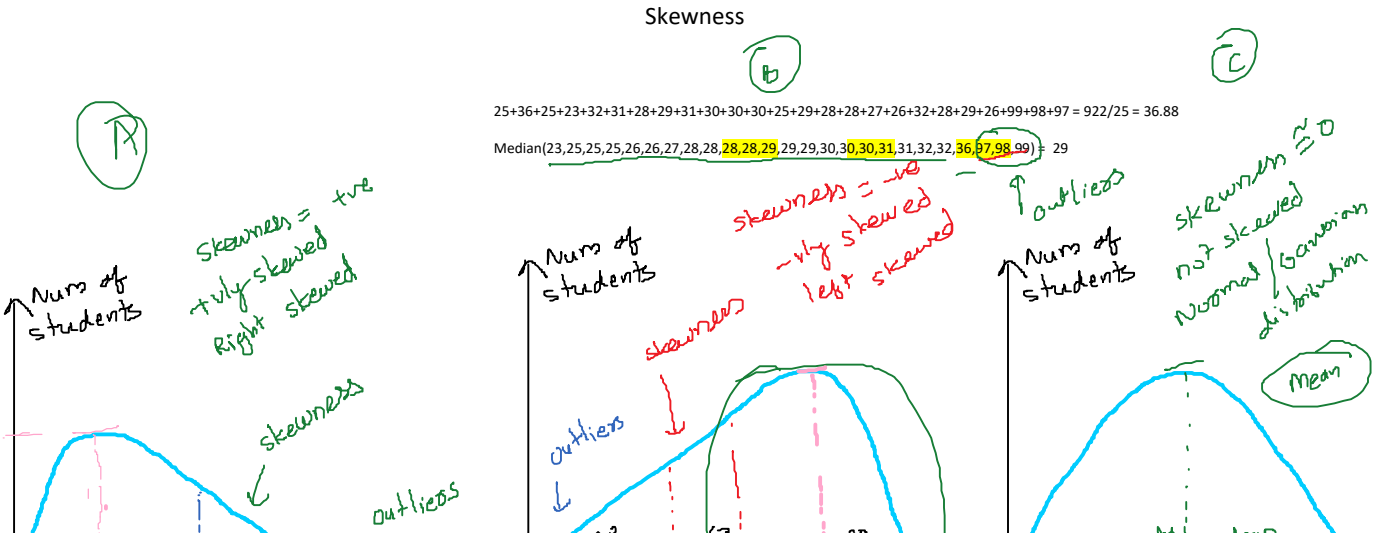


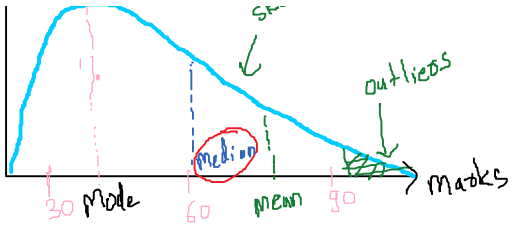


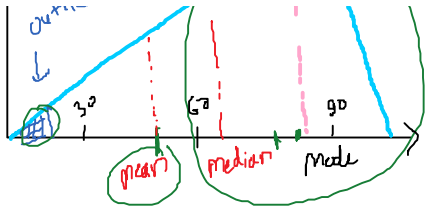
Type of variable

- Categoric
  - Nominal - Mode
  - Ordinal - Median
- Numeric
  - Skewness - skewed - median
  - Not skewed - mean

Skewness







Measure of Spread

Variance

Delhi

19
22
26
32
38
44
39
42
37
34
38
32
36
28
27
22

$\bar{x} = 34$

Mumbai

21
22
21.5
22.2
21.8
21.03
22.6
22.4
21.3
22.3
22.4

$\bar{x} = 22$

$$\text{Variance} = \sigma^2 = \frac{\sum (x - \bar{x})^2}{N}$$

°C

Machine	Temp	Hum	Vib	Damaged
m1				Y
m2				Y
m3				N
...				...
m1000				N

$\sigma^2 = 0$

$\bar{x} = 22\ °C$

$\sigma^2 = 9\ °C^2$

standard deviation = $\sqrt{\text{variance}}$

$\sigma = 3\ °C$

	A	B	C
$\bar{x}$	$1000	$1000	$1000
σ	$240	$320	$75

Coefficient of variance

$$CV = \frac{\sigma}{\mu}$$ this is always < 1

Range = measure of spread

Range = max − min

25, 28, 32, - - - - , 96, 98, 99

Range = max − min

= 99 − 25 = 74

Percentage v/s Percentile

25, 28, 32, 36, 38, 45, 49, 55, 59, 63, 65, 65, 69, 75, 72, 79, 85, 86, 88, 92, 95

25th percentile Q1

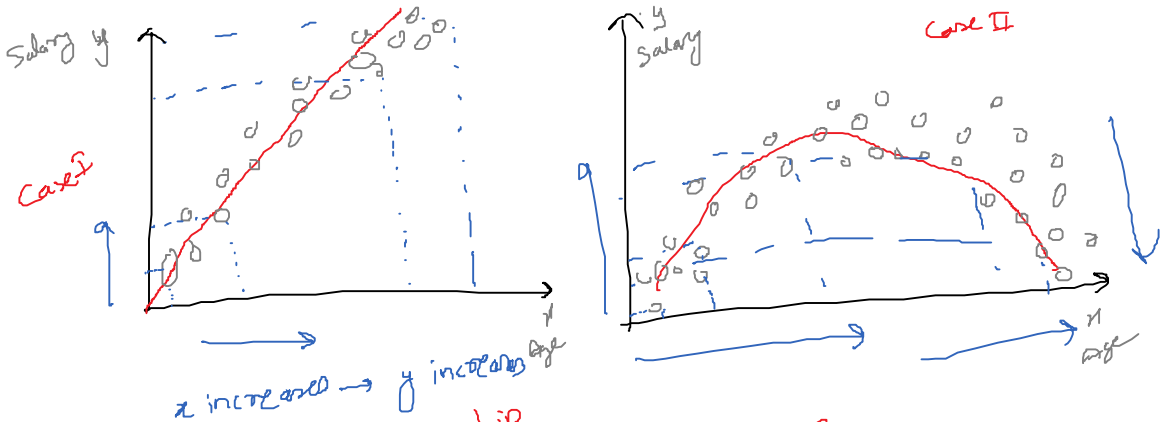
50th percentile Q2

75th percentile Q3

1st student, percentage = (25*100)/ 100 = 25
Last student = (95*100)/100 = 95

Percentile = Number of values below *100/ total number of values
1st student = 0*100 / 21 = 0
Last student = 20*100 / 21 = 95.2381

Student with marks 65 = 10*100/21 = 47.619



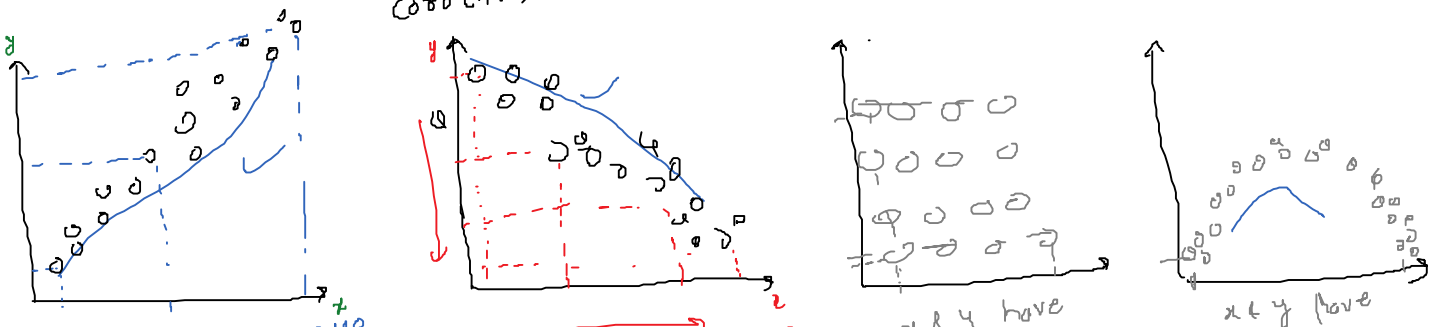
* linear relationship

non linear relationship

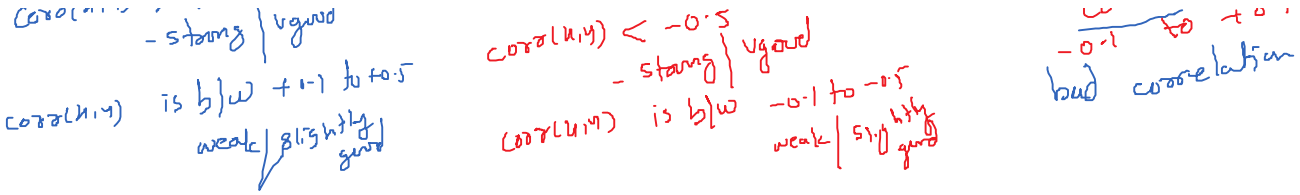
correlation - measure of linear relationship

$$Corr(x,y) = \frac{\frac{1}{n}\sum (x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y}$$

corr(x,y) = −1 to +1



corr(x,y) > +0.5 - strong ⎫ good
corr(x,y) < -0.5 - strong ⎭ good



A drug production company introduces a drug and claims that this drug can cure the disease in 15 days.
Your research company has to test the claim for this drug, after taking multiple samples on different patients below are number of days it took for them to get cured.
Days = [12,14,12,13,16,18,17,19,18,15,14,15,13,12,15,15,14,16,14,12,12]

We need to test whether the claim by the company is correct or not, the business also says that out of tests they have done, 95% of the patients were cured with in 15 days.

Pop-mean = 15
CI = 95%. Alpha = 1- CI = 1-0.95 = 0.05

$$\sigma = \frac{s}{\sqrt{n}}$$

n = sample size, n = 20

We need to test whether the population mean (15) is similar to the sample mean of tests done on 20 patients.

Z test - when we know population standard deviation
- Used to compare population mean with sample mean

T test - when we do not know population standard deviation
- Used to compare population mean with sample mean

$$z = \frac{x - \mu}{\sigma}$$

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