No class tomorrow

Next class on Friday

We have a class on Saturday

If you support will have class on Sunday

Saturday and Sunday

Vscode

Github

Will combine batch with morning batch people

Data science == 4 to 5months

1. What is statistics
2. Types of statistics
3. Population vs sample
4. Data types
5. Levels of data
6. Parameter vs statistics
7. Frequency table
8. Frequency distribution table
9. Bar plot
10. Pie chart
11. Histogram
12. Distribution plot
13. Central tendency
14. Mean-median-mode
15. Mean vs median
16. Outlier
17. Types of distribution
18. Neg skew-pos skew-no skew
19. Range- mean deviation – absolute mean deviaton
20. Variance
21. Standard deviation
22. How to find outlier
23. Percentile -Quartile
24. Box plot
25. Univariate-bivariate-multivariate
26. Scatter plot

Will create an assignment

Statistics write up assignment

PART-1

Write up should finish ======= > ppt

Variance:

Is always deal about a single variable : x

Age

30 (30-mean)^2

31 31-mean

32

33

34

Mean= 30+31+32+33+34/5 =

Age income

30 5ok

31 55k

How age and income varying together

Covariance:

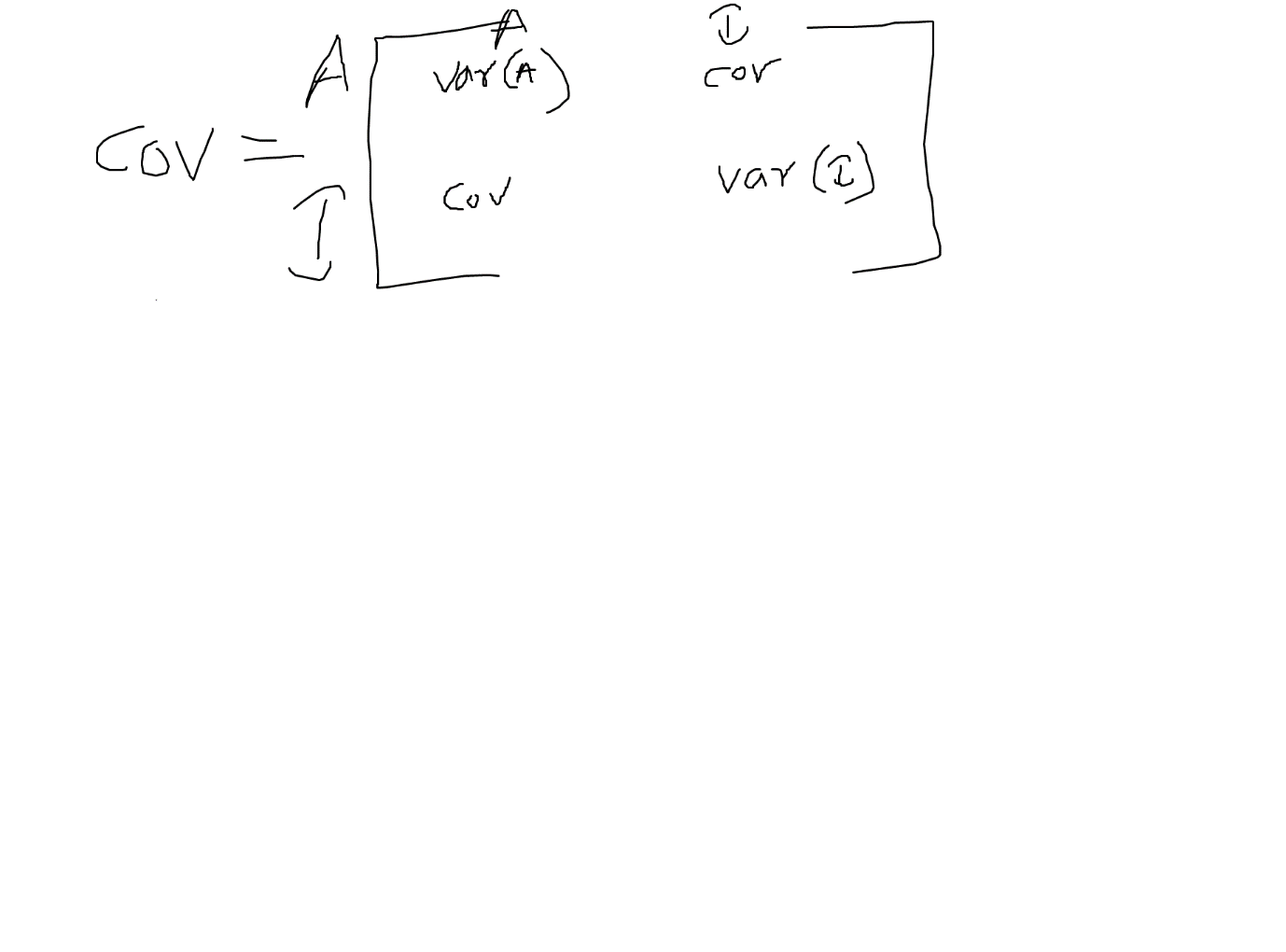
Co ========== > two

Age vs income

Age vs age

Income vs Age

Income vs Income



Hi sir,  
variance and covariance is only applicable for numeric datatype columns right.  
all maths all measurements will applicable only for numerical columns only

what will be the scenario for categorical columns.===== > ML models

will convert cat ====== numerical columns

case status case status

accepted 1

denied 0

generative ai ======== >

good what oppostite word ========== bad

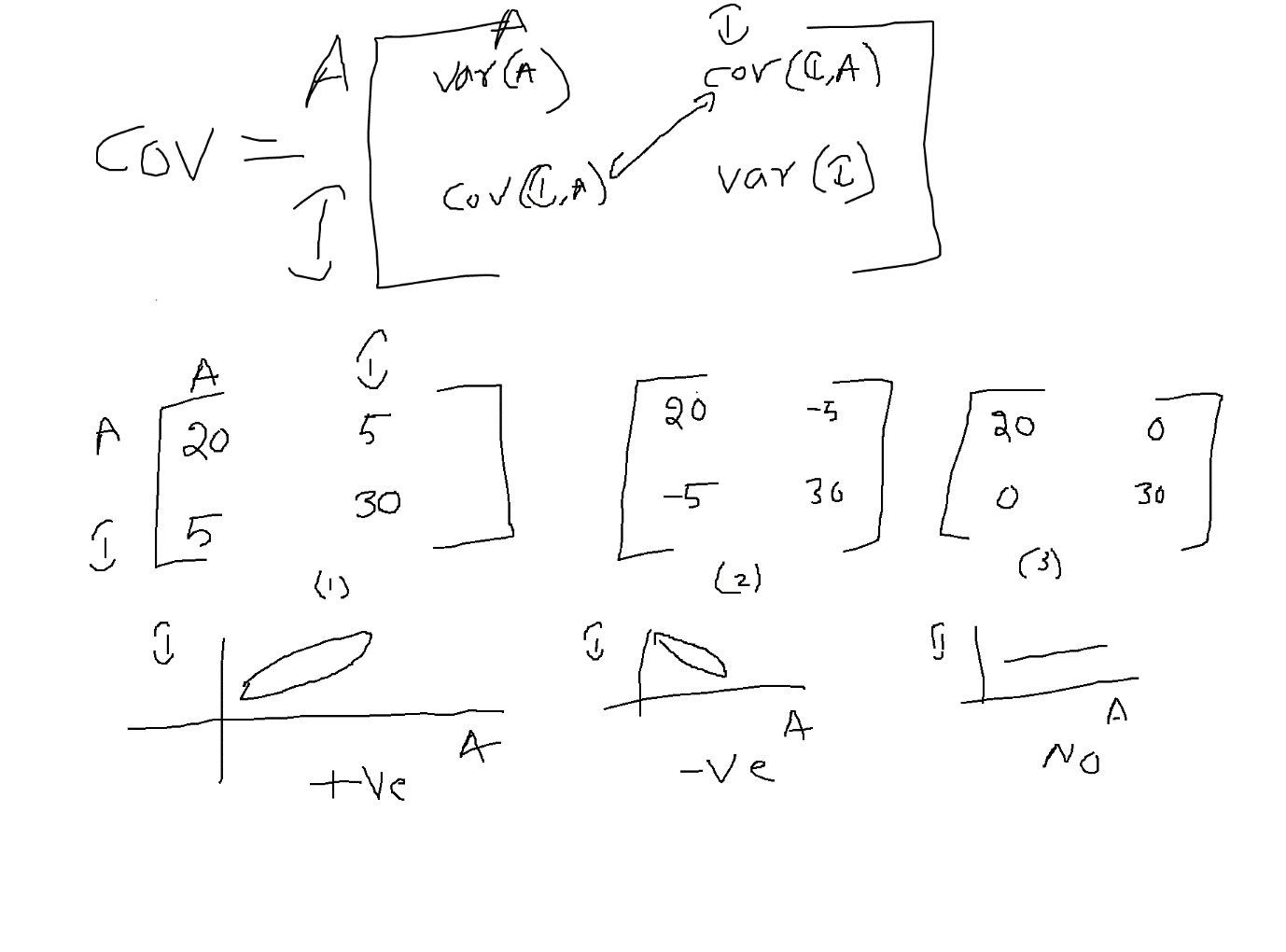
NLP

how does covariance help in analysis

age increase income increase

age incases---------

no not that why sir... just want to understand.... in realtime situations how does it help



Covariance will say just , a relation is there between two variables or not

But it never provides how much percentage two variables are related each other

A and B both are good friends

Can you mention by using a number

70%

80%

b.tech ====== >

he is my friend ====== pos ========== 90%

2 days ========= 85%

-90%

0

Positive 0 to 1 ======== 0 to 100%

Negative -1 to 0 ======== -100% to 0

No relation 0

-1 =======neg========= 0 =========pos==========1

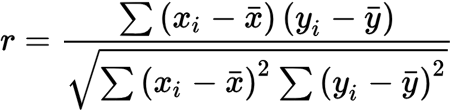
Correlation:

Pearson correlation is denoted with r

Pearson correlation coefficient

Covariance: it has relation or not

Correlation: how much the relation



can we apply variance for univariant as well in above formulae.(don’t confuse)

Variance is one measurement

Mean is a measurement

75% is a measurements

Mean-median-mode-variance-std-75%-25%-50% ===== Age column

| no\_of\_employees |
| --- |
| 14513 |
| 2412 |
| 44444 |
| 98 |
| 1082 |
| 2339 |
| 4985 |

Age Income

30 50k

35 60k

40 70k

Covariance value:

| Age(X) | Income(y) |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 30 | 50k | 35-30=5 | 60-50=10 | 25 | 100 |
| 35 | 60k | 35-0=0 | 60-60=0 | 0 | 0 |
| 40 | 70k | 35-40=-5 | 60-70=-10 | 25 | 100 |
| Mean=35 | Mean= 60k |  |  | Root(50)= | Root(200) |

Sir actually i did not understand in matrix why a12 position value is equals to a21

X\*y = y\*x