Sub :

Time : Date : / /

CampusX (Statistics and Probability)

I would note the concept that I didn't find in PW-skills part

Relative frequence : It is the percentage value of the frequency value of a categorical data set.

Example →

Type	Freq.	Relative Freq.
Beach	60	0.3
City	40	0.2
Adventure	30	0.15
Nature	35	0.175
others	15	0.075

population = 200

$= \frac{60 \times 100}{200}$

$= 30\% = 0.3$

Contingency table : When for visualization, you change ~~a column to row~~ row values to column values.

Survived	Pclass
0	1
1	2
	3

→

	Pclass		
Survived	1	2	3
0	42	40	31
1	49	62	12

This table can be created for categorical-categorical data.

Scatter-plot : for numerical – numerical data

Sub : _____ Page : _____ Time : _____ Date : / /

Percentile : Represent the percentage of observations in a dataset thats fall below a particular value. For example, the 75th percentile is the value below which 75% of the observations in the dataset fall.

How to calculate percentile?

Formula : $PL = \frac{P}{100}(N+1)$

PL : the desired percentile value location

N : The total number of observations in the dataset

P : The percentile rank (expressed as a percentage)

Example 1 : Find the 75th percentile score from the below data

78, 82, 84, 88, 91, 93, 94, 96, 98, 99

Step 1 : Sort the data → 78, 82, 84, 88, 91, 93, 94, 96, 98, 99
1 2 3 4 5 6 7 8 9 10

$PL = \frac{75}{100} \times (10+1) = \frac{3}{4} \times 11 = 8.25$

75th percentile value is coming between 8 and 9 position values.

The actual value is $96 + 0.25(98-96)$

75th percentile = 96.5

Sub :

Day

Time : Date : / /

Example 2 :

Data in sorted order → $\underbrace{78, 82, 84}_{x=3}, \underbrace{88}_{y=1}, 91, 93, 94, 96, 98, 99$

Find the percentile of 88 →

Formula = $\frac{x + 0.5y}{n}$

x = number of values below the given value

y = Number of values equal to the given value

n = total number of values in the dataset.

∴ Percentile Rank = $\frac{3 + 0.5 \times 1}{10}$ = 3.5/10 = 0.35 = 35% = 35th percentile

Box : You can see the video when you need (session 39 - Campusx)

→ Box Plot visualisation

→ How it is calculated ?

→ What the plot describes ?

Scatter plot : To show numerical - numerical relationship, we can draw scatter plot.

Sub :

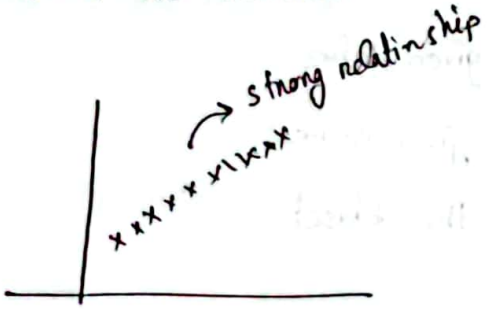
Day

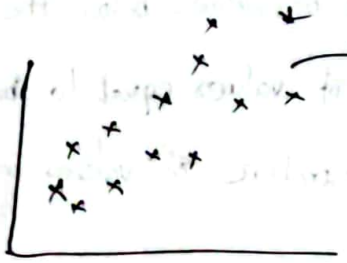
Time : Date : / /

Covariance: It is noted properly in the pw-skills part.

Disadvantage of covariance:

It does not tell us about the strength of the relationship between two variables, since the magnitude of covariance is affected by the scale of the variables.





[Can't find proper relation between two columns]

Correlation:

This technique is noted down on pw-skills part.