Measures of Association.

1st Oct : M.

Attbs & Variables ⇒

Variable ⇒ A variable is a quantiative character which can be expressed numerically.

eg. height of a person, weight ⇒ Expressed Numerically

ft. 2, 3, 4.5, 6, ----.

Attribute :-

An attb is a qualitative character which cannot be primarily expressible numerically.

eg. sex ; Economic Status < Rich, middle class, poor

level of education < Illiterate, primary, high school.

Measures of Association for attb's (A and B) :-

A: Attacked with Malaria   joint freq^n?

B: Quinine used.

$B' = \beta$ $A' = \alpha$.

Outcome.

Joint freq^n	A (Y/N) Attack with malaria (A)	$A' = \alpha$ Non-attack with malaria (α)	Total
B: Quinine Used (B) (Y/N)	19 (f_{AB})	587 ($f_{\alpha B}$)	606 (f_B)
$B' = \beta$: No Quinine used (β)	193 ($f_{A\beta}$)	2741 ($f_{\alpha\beta}$)	2934 (f_β)
~~Total~~ Total.	212 (f_A)	3328 (f_α)	3540 (n)

marginal freq

attacked with malaria.

Marginal frequencies.

Total people who used quinine.

f: frequency.

<u>Two attbs</u> :- Quinine Used (B) (Y/N)
Attacked with Malaria (A) (Y/N)

Total people in observation or sample

Observations ⇒

$f_B = f_{AB} + f_{\alpha B}$ (people who used Quinine)

$B' = \beta$ $f_\beta = f_{A\beta} + f_{\alpha\beta}$ (people who did not use Quinine)

$A' = \alpha$ $f_A = f_{AB} + f_{A\beta}$ (people attacked with malaria)

$f_\alpha = f_{\alpha B} + f_{\alpha\beta}$ (people Not-attacked with malaria)

$n = f_B + f_\beta = f_A + f_\alpha$

Total people in observation

Quest^n ⇒

If $\frac{f_{AB}}{f_B}$ =

are sam

Quest^n ⇒ Whether attbs A and B are related i.e. whether there's dependency between them.

Used quinine & attacked with malaria

i.e. if we use Quinine (B) then are people attacked with malaria (A) are less or there is no relation.

$$\frac{f_{AB}}{f_B} = \frac{f_{A\beta}}{f_\beta}$$

people who used quinine.

$\frac{f_{AB}}{f_B}$ = attacked with maleria even though they used quinine.

$$\frac{f_{A\beta}}{f_\beta} = \frac{\text{didn't use Q \& attacked}}{\text{people who didn't use quinine}} =$$ attacked with maleria & they don't use quinine.

If $\frac{f_{AB}}{f_B} = \frac{f_{A\beta}}{f_\beta}$ are same

Whether you used Q or not, the no. of people attacked with M is same.

i.e. No affect of Q on Malaria.

i.e. The attbs A and B are unrelated ~~OR~~ independent.

$$\therefore \frac{f_{AB}}{f_B} = \frac{f_{AB} + f_{A\beta}}{f_B + f_\beta} = \frac{f_{A\beta}}{f_\beta}$$ // all three are equivalent

$$= \frac{f_A}{n}$$

$$\frac{f_{AB}}{f_B} = \frac{f_A}{n}$$

$$f_{AB} - \frac{f_A \cdot f_B}{n} = 0$$

$19 - \frac{212 \times 606}{}$

$\delta_{AB} = 0$ when

$$\delta_{AB} = f_{AB} - \frac{f_A \cdot f_B}{n} = 0$$

$\delta_{AB} =$ Joint − marginal $= 0$

Then, Two attbs A & B are independent if $\delta_{AB} = 0$

1st Oct

Opp of independent = Associative.

A & B are called Associative if $\delta_{AB} \neq 0$

$$\delta_{AB} = 19 - \frac{212 \times 606}{3504} = -17.29 \neq 0$$

∴ A & B are associative, not independent
(they are negatively associative)

defn:- A & B are said to be positively associative if $\delta_{AB} > 0$

defn:- A & B are negatively associative if $\delta_{AB} < 0$

* Two attb's A & B are said to be complete positively associated if

$$f_{A\beta} = 0 \quad \text{and / or} \quad f_{\alpha B} = 0$$

i.e. atleast one of $f_{A\beta}$ & $f_{\alpha B}$ is zero.

If $f_{A\beta}$ and $f_{\alpha B}$ are zero then.

	A	α	
B	f_{AB}	0	f_B
β	0	$f_{\alpha\beta}$	f_β
	f_A	f_α.	

then $f_A = f_B = f_{AB}$. ← whoever has att B they all have attb A

$f_\alpha = f_\beta = f_{\alpha\beta}$.

If one of them is zero
↳ ($f_{A\beta}$ or $f_{\alpha B}$)

	A	α
B		*
β	*	

② defⁿ: A & B are complete negatively associated if

$f_{AB} = 0$ and/or $f_{\alpha\beta} = 0$

	A	α	
B	0	$f_{\alpha B}$	f_B
β	$f_{A\beta}$	0	f_β
	f_A	f_α.	

atleast one of off diagon elem = 0

all off-diagon elem = 0

$f_A = f_\beta = f_{A\beta}$.

$f_B = f_{B\alpha} = f_{\alpha B}$

particular cases.

③ defⁿ: A & B are Absolute positively Associated

if $f_{A\beta} = 0$ and $f_{\alpha B} = 0$

④ defⁿ: A & B are Absolute Negatively associated if

if $f_{AB} = 0$ and $f_{\alpha\beta} = 0$

Attbs A and B

positive Asso.
$\delta_{AB} > 0$

Complete +ve Asso

atleast one of off diagonal elem $= 0$

$f_{A\beta} = 0$ and/or $f_{\alpha B} = 0$

Absolute +ve Associatn

all off-diagonal elem $= 0$

$f_{A\beta} = 0$ and $f_{\alpha B} = 0$

Independence
$\delta_{AB} = 0$

Negative Asso.
$\delta_{AB} < 0$

Complete -ve Asso

atleast one of diagonal elem $= 0$

$f_{AB} = 0$ and/or $f_{\alpha\beta} = 0$

Absolute Negative Asso

~~Both~~ all diagonal elem $= 0$.

$f_{AB} = 0$ and $f_{\alpha\beta} = 0$

* Interpreting freq^n distribution in terms of probability distribution / frequency.

If you divide each entity by 'n'.

	A	α	
B	$\frac{f_{AB}}{n}$	$\frac{f_{\alpha B}}{n}$	$\frac{f_B}{n}$
β	$\frac{f_{A\beta}}{n}$	$\frac{f_{\alpha\beta}}{n}$	$\frac{f_\beta}{n}$
	$\frac{f_A}{n}$	$\frac{f_\alpha}{n}$	1

	A	α	
B	P_{AB}	$P_{\alpha B}$	P_B
β	$P_{A\beta}$	$P_{\alpha\beta}$	P_β
	P_A	P_α	1

Psbl characteristic of Good measure of Associat^n :-

1) It should not depend on the total frequency (n) ie. how much data is there, should not be part of measure ①
2) Measure should give zero value for independent at b/s ②
3) ——"—— +ve value for +ve Associat^n ③
4) ——"—— -ve value for -ve Associat^n ④

β β

① Yule's Q measure :- for association betn A & B.

$$Q_{AB} = \frac{f_{AB} \cdot f_{\alpha\beta} - f_{A\beta} f_{\alpha B}}{f_{AB} \cdot f_{\alpha\beta} + f_{A\beta} f_{\alpha B}}$$ ① depending on relative frequencies

$$= \frac{n \cdot \delta_{AB}}{f_{AB} f_{\alpha\beta} - f_{A\beta} f_{\alpha B}}$$ ② depending on

* Observatn ⇒

① $Q_{AB} = 0$ iff A & B are independent ($\because \delta_{AB} = 0$)

② $Q_{AB} = 1$ iff $f_{A\beta} \cdot f_{\alpha B} = 0$

$f_{A\beta} = 0$ and/or $f_{\alpha B} = 0$.

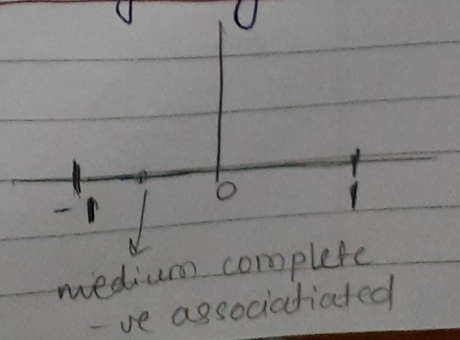
i.e. A and B are complete positively associated

③ $Q_{AB} = -1$ iff $f_{AB} \cdot f_{\alpha\beta} = 0$

$f_{AB} = 0$ and/or $f_{\alpha\beta} = 0$

i.e. A and B are complete negatively associated.

④ $-1 \leq Q_{AB} \leq 1$



classmate

Date

Page

② Yule's Y measure :-

$$Y_{AB} = \frac{\sqrt{f_{AB} \cdot f_{\alpha\beta}} - \sqrt{f_{A\beta} \cdot f_{\alpha B}}}{\sqrt{f_{AB} \cdot f_{\alpha\beta}} + \sqrt{f_{A\beta} \cdot f_{\alpha B}}}$$

① $Y_{AB} = 0$ iff A & B are independent ($\delta_{AB} = 0$)

② $Y_{AB} = 1$ iff A & B are complete positively As

③ $Y_{AB} = -1$ iff A & B are complete -vely Ass

④ $-1 \leq Y_{AB} \leq 1$

③

$$Y_{AB} = \frac{1 - \sqrt{1 - Q^2_{AB}}}{Q_{AB}}$$

Relⁿ betⁿ Yule's Y measure & Q measure

$$Q_{AB} = \frac{2 \cdot Y_{AB}}{1 + Y^2_{AB}}$$

③ Yule's V measure :- (for Absolute association)

$$V_{AB} = \frac{f_{AB} \cdot f_{\alpha\beta} - f_{A\beta} \cdot f_{\alpha B}}{\sqrt{f_A \cdot f_B \cdot f_\alpha \cdot f_\beta}}$$

① off diagonal zero $V_{AB} = 1$ iff $f_{A\beta} = 0$ and $f_{\alpha B} = 0$ ~~$f_{\alpha\beta} = 0$~~

ie when A and B are absolute positively Asso.

② diagonal elem zero $V_{AB} = -1$ iff $f_{AB} = 0$ and $f_{\alpha\beta} = 0$.

i.e. when A & B are absolute negatively Asso

③ $V_{AB} = 0$ iff A and B are independent.
($\delta_{AB} = 0$)

④

Ex.1) From table of malaria :

	A	α	
B	19	587	606
β	193	2741	2934 (f_β)
	212 (f_A)	3328 (f_α)	3540

① $$Q_{AB} = \frac{f_{AB} \cdot f_{\alpha\beta} - f_{A\beta} \cdot f_{\alpha B}}{f_{AB} \cdot f_{\alpha\beta} + f_{A\beta} \cdot f_{\alpha B}}$$

$$= \frac{19 \times 2741 - 193 \times 587}{19 \times 2741 + 193 \times 587}$$

$$Q_{AB} = -0.3701$$

② $$Y_{AB} = \frac{\sqrt{f_{AB} \cdot f_{\alpha\beta}} - \sqrt{f_{A\beta} \, f_{\alpha B}}}{\sqrt{f_{AB} \cdot f_{\alpha\beta}} + \sqrt{f_{A\beta} \cdot f_{\alpha B}}}$$

$$= \frac{\sqrt{19 \times 2741} - \sqrt{193 \times 587}}{\sqrt{19 \times 2741} + \sqrt{193 \times 587}}$$

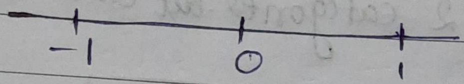
$$= -0.19189.$$

③ $$V_{AB} = \frac{f_{AB} \cdot f_{\alpha\beta} - f_{A\beta} \cdot f_{\alpha B}}{\sqrt{f_A \cdot f_B \cdot f_\alpha \cdot f_\beta}}$$

$$= \frac{19 \times 2741 - 193 \times 587}{\sqrt{212 \times 606 \times 3328 \times 2934}}$$

$$= -0.05465$$

All three measures are coming negative.



[for ϕ_{AB} and Y_{AB}, as it is negative so it is complete negative Association.

for V_{AB} it is Absolute Negative Association]

Conclusion ⇒ But it is not strictly -1, All measures are showing that there is slight negative Association betn Malaria & Quinine.

It shows that there is slight evidence that if Quinine is used more & more then attack with ~~material~~ malaria will be less.

Here we had Two attbs, $\xrightarrow{\&}$ Two categories.

∴ As you ↑s use of Quinine then Malaria attack ↓s.

Manifold two-way (K x m) classification

more than 2 categories but two attb's A & B

$\boxed{\text{Sub I}}$ attb 1

$\boxed{\text{Sub II}}$ → attb 2

K x m manifold ⇒

	VGood	Good	Avg	Below Avg	Poor
VGood	20	30	27	5	2
Good	15	25	50	12	3
Avg	10	17	27	15	9
Below Avg	5	8	11	28	18
Poor	2	3	7	19	32

* Two attributes & they can have K different categories.

$A \leftarrow A_1, A_2, \ldots A_K$

$B \leftarrow B_1, B_2, \ldots B_m$

(K x m)

$\boxed{A}_K$ $\boxed{B}_m$

i ↓ j →

	A_1	A_2	----	A_K	marginal freq^{n}s
B_1	f_{11}	f_{12}	----	f_{1K}	f_{10}
B_2	f_{21}	f_{22}	----	f_{2K}	f_{20}
⋮					⋮
B_m	f_{m1}	f_{m2}	----	f_{mK}	f_{m0}
Marginal freq^{n}s →	f_{01}	f_{02}	----	f_{0K} →	n

first index fixed — f_{i0}

2nd index fixed — f_{0j}

n — Total frequency.

$n = f_{ok} + f_{mo}$

row $f_{io} = \sum_{j=1}^{k} f_{ij}$, for $i = 1, 2, \ldots, m$

column $f_{oj} = \sum_{i=1}^{m} f_{ij}$, for $j = 1, 2, \ldots, k$.

$$\cancel{n = \sum_{i=1}^{m} f_{ip} \pm \sum_{j=1}^{k} f_{ij}} = \sum_{i=1}^{m} \sum_{j=1}^{k} f_{ij}$$

$$\cancel{n = \sum_{j=1}^{m} f_{io} = \sum_{i=1}} \qquad n = \sum_{j=1}^{k} f_{oj} = \sum_{i=1}^{m} f_{io}$$

defn ⇒ Two attbs A and B are said to be Independent if

$$\delta_{ij} = f_{ij} - \frac{f_{io} \times f_{oj}}{n} = 0, \quad \forall \; i = 1, 2, \ldots m \;\; \& \;\; j = 1, 2, \ldots k$$

for all.

for each i & j this should satisfy

$$\frac{f_{1j}}{f_{10}} = \frac{f_{2j}}{f_{20}} = \text{-----} = \frac{f_{mj}}{f_{mo}}$$

for each $j = 1, 2 \ldots k$

it doesn't matter no of people in category B_1 or B_2, the proportion of category A is same then A & B are independent.

for any two category ~~A1k or B4~~. A_i and B_j are independent as then we can say two attbs A and B are independent.

$n = f_{oK} + f_{mo}$

row $f_{io} = \sum_{j=1}^{K} f_{ij}$, for $i = 1, 2, \ldots m$

column $f_{oj} = \sum_{i=1}^{m} f_{ij}$, for $j = 1, 2, \ldots K$.

$$\cancel{n = \sum_{i=1}^{m} f_{ij} \pm \sum_{j=1}^{K} f_{ij}} = \sum_{i=1}^{m} \sum_{j=1}^{K} f_{ij}$$

$$\cancel{n = \sum_{j=1}^{m} f_{io} = \sum_{i=1}^{K}} \qquad n = \sum_{j=1}^{K} f_{oj} \neq \sum_{i=1}^{m} f_{io}$$

def^n ⇒ Two attbs A and B are said to be Independent if

$$\delta_{ij} = f_{ij} - \frac{f_{io} \times f_{oj}}{n} = 0, \quad \forall\ i = 1, 2, \ldots m \ \& \ j = 1, 2, \ldots K$$

for each i & j this should satisfy

$$\frac{f_{1j}}{f_{10}} = \frac{f_{2j}}{f_{20}} = \cdots = \frac{f_{mj}}{f_{mo}}$$

for each $j = 1, 2 \ldots K$

it doesn't matter no of people in category B_1 or B_2, the proportion of category A_i is same then A & B are independent.

for any two category ~~A_K or B_4~~ A_i and B_j are independent as then we can say two attbs A and B are independent.

(i.e. for certain pair)

* If for some $i \neq j$

$\delta_{ij} \neq 0$ then A & B are associated

We can't define +ve/ -ve association here.

(in table)

If all elems except diagonal elems are Zero then we can define the concepts of association like complete +ve/-ve asso OR Absolute +ve/-ve association

$$\delta_{ij} = f_{ij} - \frac{f_{io} \times f_{oj}}{n}$$

Measure of Association:-

$$\chi^2_{AB} = \sum_{i=1}^{m} \sum_{j=1}^{K} \left(\frac{n \cdot \delta_{ij}^2}{f_{io} \times f_{oj}} \right)$$

Karl pearson measure.

$$= n \left[\sum_{i=1}^{m} \sum_{j=1}^{k} \left(\frac{f_{ij}^2}{f_{io} \times f_{oj}} \right) - 1 \right]$$

If $\chi^2_{AB} = 0 \iff$ if $\delta_{ij} = 0 \;\; \forall \; i = 1, 2, \ldots m$ & $j = 1, 2, \ldots K$.

$\iff$ Attributes A & B are independent

drawback χ^2_{AB} measure dependents on freq^n 'n', so as $n \uparrow$s then χ^2_{AB} also takes higher value. & may take ∞ value also.

Karl pearson's coefficient of contingency $\Rightarrow$ Also associat^n can't be described properly. ↖ Solution

$$C_{AB} = \sqrt{\frac{\chi^2_{AB}}{n + \chi^2_{AB}}}$$

Obsv $\Rightarrow$ (1) $C_{AB} = 0 \iff \chi^2_{AB} = 0 \iff$ A & B are independent.

(2) $\chi^2_{AB} < n + \chi^2_{AB} \iff C_{AB} < 1$ +ve measure

$$\Rightarrow \boxed{0 \leq C_{AB} < 1}$$

eg $C_{AB} = 0.2$ low degree of Asso bet^n A & B

$C_{AB} = 0.7$ high ———— " ————

C_{AB} will never attain value 1 even though there is absolute / Perfect Association

↓

all off-diagonal elems $= 0$.

diagonal elems $\neq 0$.

	A_1	A_2	- - - - -	A_K	
B_1	f_{11}	0	0 — 0	0	f_{10}
B_2	0	f_{22}		0	f_{20}
⋮	0	0	- - - -	0	⋮
B_K	0	0	0 0 - -	f_{KK}	f_{K0}
	f_{01}	f_{02}		f_{0K}	n

$$\left.\begin{aligned} f_{11} &= f_{10} = f_{01} \\ f_{22} &= f_{20} = f_{02} \\ f_{KK} &= f_{K0} = f_{0K} \end{aligned}\right\} \quad f_{ij} = \begin{cases} 0, & i \neq j \\ f_{ii}, & i = j \end{cases}$$

So above is full absolute associatⁿ still C_{AB} isn't taking value 1.

classmate

Date

Page

$$\chi^2_{AB} = n\left[\sum_{j=1}^{m}\sum_{i=1}^{k}\left(\frac{f_{ij}^2}{f_{io}\times f_{oj}}\right)^2 - 1\right]$$

$$= n\left[\sum_{S=1}^{k}\left(\frac{f_{ii}^2}{f_{io}\times f_{oi}}\right) - 1\right]$$

$$= n\left[\sum_{S=1}^{k}(1) - 1\right]$$

$$\boxed{\chi^2_{AB} = n(k-1)}$$

for perfect/absolute Association

$$\boxed{C_{AB} = \sqrt{\frac{k-1}{k}}}$$

for $k\times k$ table. C_{AB} is not taking value 1 so this drawback removed by

Carl pearson measure will take

Tschuprow's measure of Asso.

$$T_{AB} = \sqrt{\frac{\chi^2_{AB}}{n\sqrt{(k-1)(m-1)}}}$$ for $k\times m$ table.

obsv ⇒

① $T_{AB} = 0$ iff $\chi^2_{AB} = 0 \Leftrightarrow$ A & B are independent

② for $k\times k$ table, the max^m value of $T_{AB} = 1$ (full Associat^n)

$$T_{AB} = \sqrt{\frac{\cancel{n(k-1)}}{\cancel{n}\sqrt{\cancel{(k-1)}(k-1)}}} = 1$$

Date
Page

$\therefore$ when ~~Abs~~ full Asso bet^n two attbs. A and B
then ~~it~~ $T_{AB} = 1$.

Ex. 17. Classif^n of 830 prof workers.

		Activity Status (A)			
		A_1	A_2	A_3	
Occupation	B_1	169 f_{11}	21 f_{12}	140 f_{13}	330 (f_{10})
	B_2	83 f_{21}	25 f_{22}	68 f_{23}	176 (f_{20})
	B_3	286 f_{31}	10 f_{32}	28 f_{33}	324 (f_{30})
	Marginal freq^n	538 (f_{01})	56 (f_{02})	236 (f_{03})	830 (n)

We are int whether there is any Association bet^n occupation & Activity Status.

$$\chi^2_{AB} = n\left[\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{f_{ij}^2}{f_{i0}\times f_{0j}} - 1\right]$$

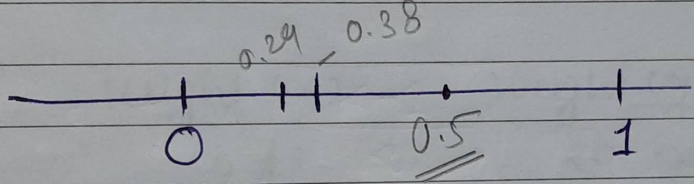
$$= 830\left[\frac{169^2}{330\times 538} + \frac{(21)^2}{330\times 56} + \cdots\right.$$

$$\left.+ \frac{28^2}{236\times 324} - 1\right] = 140.18f$$

now we can't say doA from χ^2_{AB} so calc. CAB

$$C_{AB} = \sqrt{\frac{\chi^2_{AB}}{n + \chi^2_{AB}}} = \sqrt{\frac{140.187}{830 + 140.187}} = 0.380$$

$$T_{AB} = \sqrt{\frac{\chi^2_{AB}}{n\sqrt{2 \times 2}}} = \sqrt{\frac{140.187}{830 \times 2}} = 0.290.$$



Both the measures indicate that there is ~~Very slight Assoc.~~ moderate degree of Associatn betn A & B
i.e., Not complete dependency but dependent to some extent.

Next (Measure of Asso betn Variables)

(ii) Abs/Perfect Associatⁿ.

eg

subject 2		subject 1 A1	A2	A3	
VGood	B1	69	0	0	69
Avg	B2	0	56	0	56
Poor	B3	0	0	33	33
		69	56	33	

if yes → degree of Association

Tell whether Sub 1 & Sub 2 are Associated

Now if student

Measures of Association betⁿ Variables :-

Variable is quantiative chara & can be measured numerically.
Attb's cant be measured numerically.

eg Height, weight, income, marks, etc.

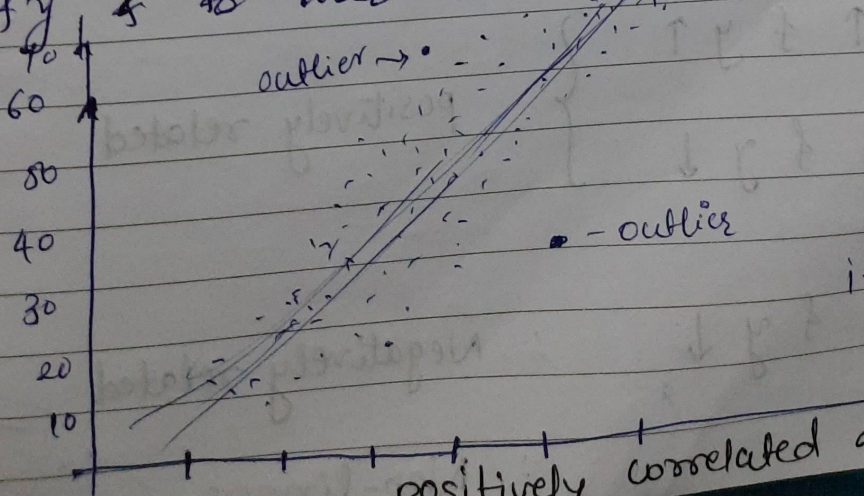
If height is less can we say weight is less
If incomes of two countrys are related, etc.

we check for evidence / relation betⁿ diff variables.

eg.

Height (x)	Weight (y)
3 ft	20 kg.
4	35
3.5	32
4.5	45
5	55
6	72
6	73
5.5	65

So we have to see whether there is relⁿ betⁿ x & y, & ~~is there~~ and what is that relⁿ



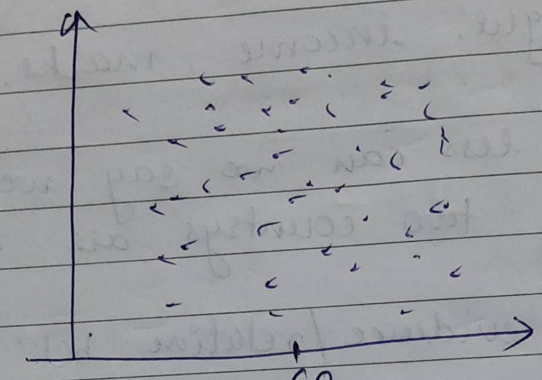
So we can say there is linear relⁿ i.e. positively correlated.

positively correlated data

eg 2) Maths English Is there any relⁿ?

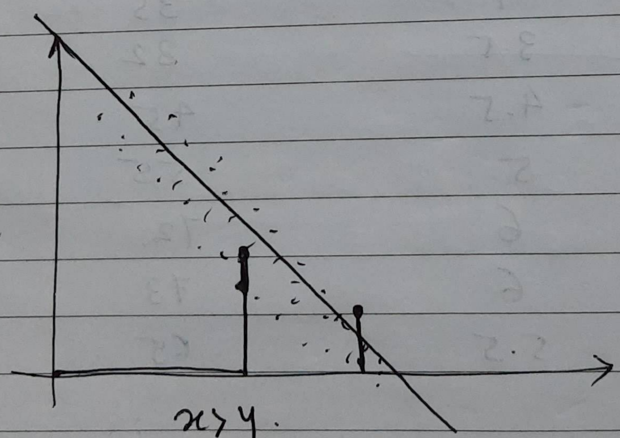
→ No cz if somebody is Good in math does it mean he is Good in English

∴ No relation betⁿ them



Scatter plot

* Maths Stats → can have linear relationship.



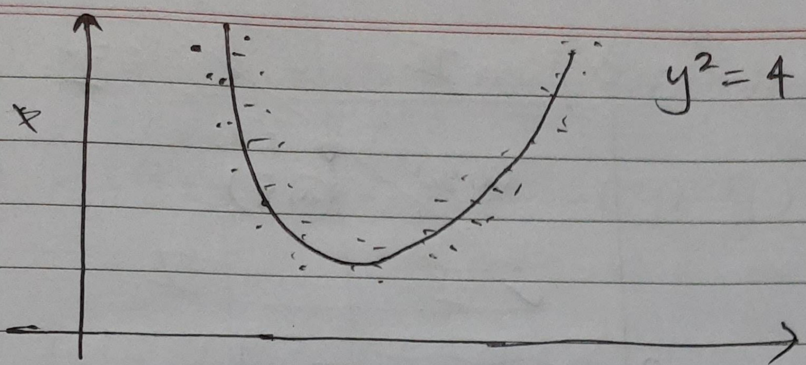
Negatively related data

(1) x ↑ & y ↑ } positively related
(2) x ↓ & y ↓ }

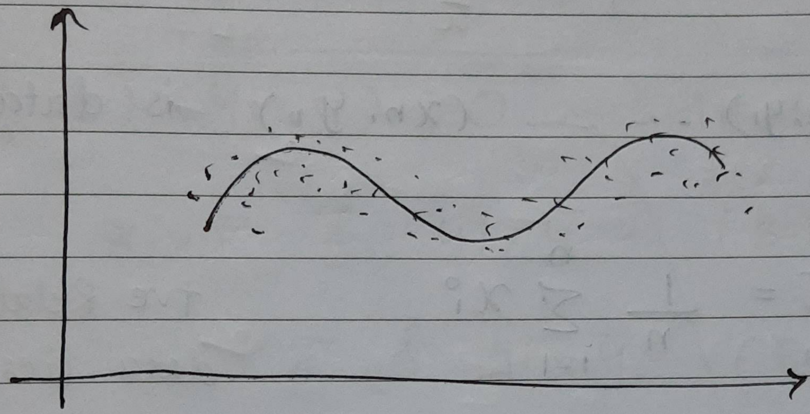
(3) x ↑ & y ↓ : Negatively related

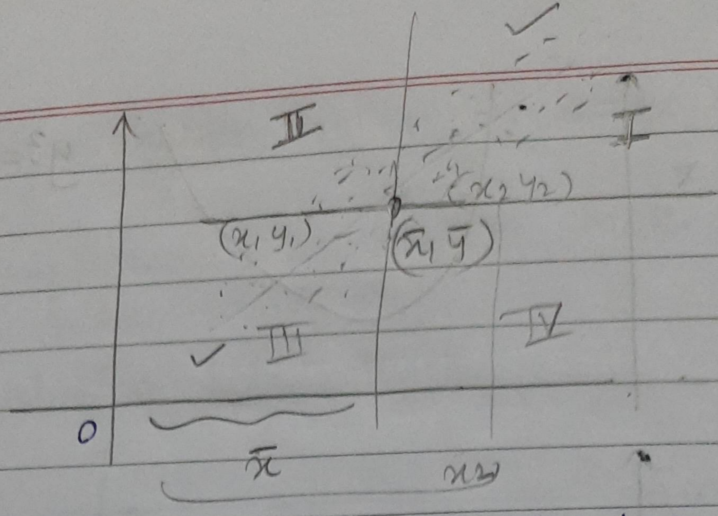
} Linear Relationship

(4) : Non-linear (Not in syll)



$y^2 = 4$





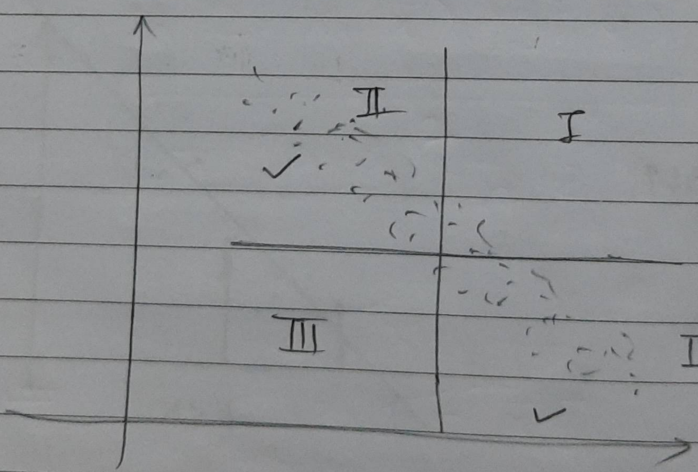
(x_1, y_1) --- --- (x_n, y_n) is dataset.

Centre of x: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

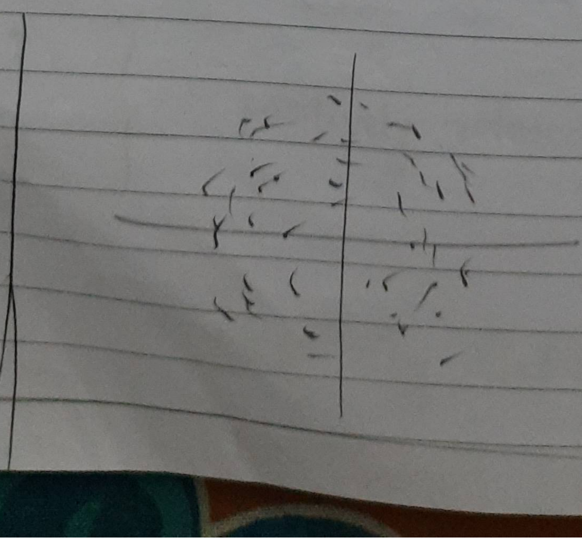
+ve Relationship data lies in → I, II

center of y: $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

−ve relationship ⇒ II, IV



(Uncorrelated)
No Relation
data distributed throught all quadrants



$(x_1-\bar{x}), (x_2-\bar{x}) \dots (x_n-\bar{x})$

$(y_1-\bar{y}), (y_2-\bar{y}) \dots (y_n-\bar{y})$

for ith data pt

If $(x_i-\bar{x})(y_i-\bar{y}) > 0$ ← will be +ve

$\Rightarrow$ $x_i > \bar{x}$ & $y_i > \bar{y}$ (Both +ve)

or $x_i < \bar{x}$ & $y_i < \bar{y}$ (Both -ve)

then

* Accumulated distance

Avg Total dist of x_i, y_i from center $(\bar{x}, \bar{y})$

$$\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y}) = \text{Total dist of pts } x_i\text{'s } \& \text{ } y\text{'s from centers } \bar{x} \text{ and } \bar{y}$$

$$\frac{1}{n}\sum (x_i-\bar{x})(y_i-\bar{y}) \Leftarrow cov(x,y)$$

avg distance

covariance

how far my pts x_i and y_i are from centre $\bar{x}$ and $\bar{y}$

Date
Page

① Product correlation coefficient :-

$$r_{xy} = \frac{cov(x,y)}{\sqrt{var(x)\, var(y)}}$$

$$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$var(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$r_{x,y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$\boxed{r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2\right)}}}$$

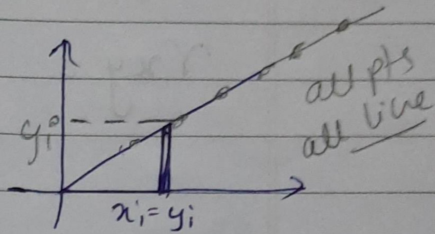
$$\frac{(x_i - \bar{x})^2}{n^2} \quad (x_i - \bar{x}^2)$$

① $-1 \leq r_{xy} \leq 1$

② $r_{xy} = 0$ iff $cov(x, y) = 0$ iff x and y uncorrelated

③ $r_{xy} = 1$ if $x_i = y_i$.

x & y are +tvely correlated



④ $r_{xy} = -1$ iff $x_i = -y_i$

x & y are negatively correlated

ex.1)

X (height of father)	~~67~~ 65	66	67	68	69	70	72	~~67~~
Y (height of son)	~~71~~ 67	68	65	68	72	72	69	~~71~~

$$\bar{x} = \frac{1}{8} \sum_{i=1}^{8} x_i = \frac{1}{8}(65 + 66 + \cdots + 67) = 68.$$

$$\bar{y} = \frac{1}{8} \sum_{i=1}^{8} y_i = \frac{1}{8}(67 + 68 + \cdots + 71) = 69.$$

$$\sum_{i=1}^{n} x_i^2 = (65^2 + 66^2 + \cdots + 67^2) = 37028$$

$$\sum_{i=1}^{n} y_i^2 = (67^2 + 68^2 + \cdots + 71^2) = 38132$$

$$\sum_{i=1}^{n} x_i y_i = (65 \times 67 + 66 \times 68 + \cdots + 67 \times 71)$$

$$= 37560$$

$$r_{xy} = \frac{\frac{1}{8} \times 37560 - 68 \times 69}{\sqrt{\left(\frac{1}{8} \times 37028 - 68^2\right)\left(\frac{1}{8} \times 38132 - 69^2\right)}}$$

$$= 0.603$$

There is moderate degree of +ve Association bet^n x and y.

Two chara-

Case I 1)

② Rank Correlation Coefficient :-

sometimes variable might not be measurable or don't have instrument to measure or it may be costly to measure. But if we can order dataset then

eg Preference is there But not necessarily the
or Ranking its numerical values we know

* We can rank or order

* Sometimes we have only ranking than the actual / raw data eg rank 1 marks (85-95)

or Grade B (75 – 85)

So now based on ranking we have to figure out dependency

* **Spearman's ρ measure** :-

Two chara- { A : u_1 u_2 ... u_n — eg rankings given By 2 judges.
B : v_1 v_2 ... v_n

Case I 1) No tie case ⇒ Two candidates cannot have same ranking.

product Correl^n coeff

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i v_i - \bar{u}\cdot\bar{v}}{\sqrt{\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} u_i^2 - \bar{u}^2\right)}_{Su^2}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} v_i^2 - \bar{v}^2\right)}_{Sv^2}}}$$

$\because$ No tie case

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}(1+2+\dots+n)$$

$$= \frac{1}{\cancel{n}}\frac{\cancel{n}(n+1)}{2} = \frac{n+1}{2}$$

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n}(1+2+\dots+n) = \frac{n+1}{2}$$

$$Su^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2 - \bar{u}^2$$

$$= \frac{1}{n}\left(1^2+2^2+\cdots+n^2\right) - \left(\frac{n(n+1)}{2}\right)^2$$

$$= \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n^2-1}{12}$$

Spear Rank correl coe

Let $d_i = u_i - v_i \quad , i = 1, 2, \ldots n$

①

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - v_i)^2$$

②

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(u_i-\bar{u}) - (v_i-\bar{v})\right]^2$$

③

$$= \cancel{\frac{1}{n}\sum_{i=1}^{n}}\Big[$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(u_i-\bar{u})^2 + \sum_{i=1}^{n}(v_i-\bar{v})^2 - 2\sum_{i=1}^{n}(u_i-u)(v_i-v)\right]$$

$$= Su^2 + Sv^2 - 2\,\mathrm{cov}(u,v)$$

$$r = \frac{cov(u,v)}{S_u \cdot S_v}$$

variance.

$$\frac{1}{n} \sum_{i=1}^{n} d_i^2 = \frac{n^2-1}{12} + \frac{n^2-1}{12} - 2r \cdot S_u \cdot S_v$$

$$= \frac{n^2-1}{12} + \frac{n^2-1}{12} - 2r \frac{n^2-1}{12}$$

Spearmans Rank correlation coefficient

$$\rho_{u,v} = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \sum_{i=1}^{n} (u_i - v_i)^2}{n(n^2-1)}$$

① $-1 \leq \rho_{u,v} \leq 1$

② $\rho_{uv} = 1, \quad \sum_{i=1}^{n} d_i^2 = 0 \Leftrightarrow u_i = v_i$ Complete Agreement

③ $\rho_{u,v} = -1$ iff $v_i = (n+1) - u_i$ complete disagreement

$u_1 \quad u_2 \quad - - \quad u_n$

$1 \quad 2 \quad - - \quad n$

$v_1 = n+1-1 = n$

$v_2 = n+1-2 = n-1$

$v_3 = n+1-3 = n-3$

$\vdots$

$v_n = \cancel{n}+1-\cancel{n} = 1.$

complete disagreement

Both A & B's ranking are complete opposite.

Ex.1)

Candidats	Ranks by judge I (u)	Ranks by judge-2 (v)	$d_i = u_i - v_i$	d_i^2
A	3	6	-3	9
B	8	4	4	16
C	5	7	-3	9
D	4	5	-1	1
E	7	10	-3	9
F	10	3	7	49
G	1	2	-1	1
H	2	1	1	1
I	6	9	-3	9
J	9	8	-1	1

$\sum d_i^2 = 100$

Spearman's ρ is

$$\rho_R = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 100}{10 \times 99}$$

$$\rho_R^{(u,v)} = 0.394$$

Tie Situation ⇒

Ex.1) 10 students

(x) ST	(y) AST	Rank(x)	Rank(y)
68	62	4	5
64	58	6	7
75	68	2.5	3.5
50	45	9	10
64	81	6	1
80	60	1	6
75	68	2.5	3.5
40	48	10	9
55	50	8	8
64	70	6	2

Tie Situation

$\frac{2+3}{2} = 2.5$ rank for 75 marks.

for 64 3 students: $\frac{5+6+7}{3} = \frac{18}{3} = 6$.

$\frac{3+4}{2} = \frac{7}{2} = 3.5$ for 68.

Spearman's ρ ⇒

A u_1 u_2 u_3 --- u_n u_n

K_1 tie K_2 tie K_s tie — s ties

B v_1 v_2 v_3 --- v_n v_n

K_1' ties K_2' ties - K_t' ties — t ties

No tie ⇒

$$\rho_{A,B} = \frac{\frac{n^2-1}{12} - \frac{T_u+T_v}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2}{\sqrt{\left(\frac{n^2-1}{12} - T_u\right)\left(\frac{n^2-1}{12} - T_v\right)}}$$

where,

$$T_u = \sum_{i=1}^{s} \left(\frac{k_i^3 - k_i}{12n}\right)$$

$$T_v = \sum_{i=1}^{t} \left(\frac{k_i'^3 - k_i'}{12n}\right)$$

$$d_i = u_i - v_i$$

Observation ⇒

1. $\rho_{A,B} = 1$ complete agreement $(u_i = v_i \neq 0)$

2. $\rho_{A,B} = -1$ compl. disagree. $v_i = (n+1) - u_i$

3.

Workers	Sup-I Ranks		Sup II
A	5		5.5
B	6		5.5
C	1		2
D	3	2	2
E	8.5	3	2
F	8.5	8.5	9
G	4	8.5	7
H	7	4	4
I	11	7	8
J	10	11	10.5
K	12	10	12
L		12	10.5

$k' = 2$

$k_1' = 2$

$k_2' = 3$

$k_3' = 2$

$$T_u = \sum_{i=1}^{1} \left(\frac{2^3 - 2}{12 \times 12} \right) = \frac{8-2}{12 \times 12} = 0.0417$$

$$T_u = \sum_{i=1}^{n} \left(\frac{k_i'^3 - k_i'}{12n} \right) = \left(\frac{k_1'^3 - k_1'}{12 \times 12} \right) + \left(\frac{k_2'^3 - k_2'}{12 \times 12} \right) + \left(\frac{k_3'^3 - k_3'}{12 \times 12} \right)$$

$$= \frac{1}{12 \times 12} \left(8 - 2 + 27 - 3 + 8 - 2 \right)$$

$$= \frac{1}{144} \times \quad = 0.25$$

$d_i = u_i - v_i$

$\Sigma d_i^2 =$

& put in formula –

$$\rho_{A,B} = \frac{\frac{n^2-1}{12} - \frac{T_u + T_v}{2} - \frac{1}{2n} \Sigma d_i^2}{\sqrt{\left(\frac{n^2-1}{12} - T_u\right)\left(\frac{n^2-1}{12} - T_v\right)}}$$

$<$ $<$

concordant (x_1, y_1) (x_2, y_2)

$>$ $>$

$<$ $>$

discordant (x_1, y_1) (x_2, y_2)

$>$ $<$

8/10. eve

Kendall's τ rank correlation coefficient $\Rightarrow$

Let (x_1, y_1) - - - - (x_n, y_n) are n-set of pairs of

* Concordant (Agreement) :-

A pair of rankings (x_i, y_i) and (x_j, y_j) is said to be concordant if.

$x_i > x_j$ and $y_i > y_j$

or $x_i < x_j$ and $y_i < y_j$

* Discordant (disagreement)

A pair of ranking (x_i, y_i) and (x_j, y_j) are in discordant if

$x_i > x_j$ and $y_i < y_j$

or $x_i < x_j$ and $y_i > y_j$

eg $(2, 6)$ $(3, 7)$ concordant

$(3, 7)$ $(4, 5)$ discordant

classmate
Date
Page

(Two of them getting same ranking)

If $x_i = x_j$ and/or $y_i = y_j$ then pairs (x_i, x_j) and (y_i, y_j) are neither concordant nor discordant.

judge I, judge II

Case I ⇒ No Tie

I II

n-rankings:
(x_1, y_1)
(x_2, y_2)
⋮
(x_n, y_n)

no. of pairs psbl $= {}^nC_2 = \frac{n!}{2!(n-2)!}$

↳ then no. of concordant & discordant pairs?

Kendall's τ = $\frac{\text{Total score}}{\text{Maximum psbl score}}$

If two pairs are concordant then you can assign score '+1' & for discordant '-1' score

Kendall's τ = $\frac{\text{total score}}{\text{Max}^m \text{ psbl score}} = \frac{P - Q}{\binom{n}{2}}$

P : No of concordant

Q : No of discordant

$$\binom{n}{2} = P + Q \quad \text{(Total pairs)}$$

$$\boxed{\text{Kendall's } \tau = \frac{P-Q}{\binom{n}{2}} = \frac{P-Q}{P+Q}}$$

$$\boxed{\Rightarrow P = \binom{n}{2} - Q}$$

$$\boxed{\Rightarrow Q = \binom{n}{2} - P}$$

Observations ⇒

① $-1 \leq \tau \leq 1$ $\quad \left(-1 \leq \frac{P-Q}{P+Q} \leq 1\right)$

② $\tau = 1$ iff $Q = 0$ i.e. no of discordant pairs is 0. (full agreement) same ranking

③ $\tau = -1$ iff $P = 0$ i.e. no of concordant pairs is zero i.e. full disagreement opp. ranking.

No Tie Case

Ranking given

Ex.

Competitors	Judge A	Judge B	Natural order of A	Ranking of B wrt A	Concordant	discordant
A	5	10	1	5	7	4
B	1	5	2	2	9	1
C	4	1	3	4	7	2
D	2	2	4	1	8	0
E	7	3	5	10	2	5
F	3	4	6	7	4	2
G	6	7	7	3	5	0
H	8	6	8	6	4	0
I	10	8	9	11	1	2
J	9	11	10	8	2	0
K	11	9	11	9	1	0
L	12	12	12	12	0	0

nos which are greater than given no ↓ 5 → 10 is bigger than 5

nos which are less than given number

50 ✓ + 16 ✓ ⇒ 66 ✓

Check how much they agree/disagree.

→ $n = 12$

1) no of pairs psbl $= \binom{12}{2} = \frac{12 \times 11}{2} = 66$ pairs.

2) Check for Concordant & discordant pairs.

$(x_i > x_j)$ and $(y_i > y_j)$

concordant pairs   discordant pairs

$$\tau = \frac{P - Q}{{}^{n}C_2} = \frac{50 - 16}{66}$$

$$\tau = 0.515$$

Kendall's τ ⇒

Case II : Tied Ranks

→ Let there are 's' ties of length $k_1, k_2, \ldots k_s$ in the ranking wrt the first character.

→ Let there be 't' ties of length $k_1', k_2', \ldots k_t'$ in the ranking wrt the 2nd character.

$$A \quad u_1 \quad \underbrace{u_2 \; - \; - \; -}_{k_1} \; - \; \underbrace{- \; - \; -}_{k_2} \; \ldots \; \underbrace{- \; - \; -}_{k_s} \; u_n$$

$$B \quad v_1 \quad \underbrace{v_2 \; - \; - \; -}_{k_1'} \; \underbrace{- \; - \; -}_{k_2'} \; \ldots \; \underbrace{- \; - \; -}_{k_t'} \; v_n$$

Rankings given

∴ You can choose 2 people from $k_1 \ldots k_s$ or $k_1' \ldots k^t$

${}^{k_1}C_2$ or ${}^{k_s}C_2$ ..

So, for Kendall's τ measure, maximum score will be reduced bcz of the tie situations, as they must be Ruled out.

Q: No. of discordant pairs
P: No of concordant pairs.

Here,
$$\boxed{\tau = \frac{P-Q}{\sqrt{\left[\binom{n}{2}-T_u\right] \times \left[\binom{n}{2}-T_v\right]}}}$$

Subtracting common cases from Total cases

$$T_u = \sum_{i=1}^{s} \frac{1}{2} k_i (k_i - 1) = \binom{k_1}{2} + \binom{k_2}{2} + \dots + \binom{k_s}{2}$$

$$T_v = \sum_{i=1}^{t} \frac{1}{2} k_i' (k_i' - 1) = \binom{k_1'}{2} + \binom{k_2'}{2} + \dots + \binom{k_t}{2}$$

$\binom{n}{2}$ = Total no. of pairs.

2	2
2	3
2	4
8	5
5	6

2, 2, 2: same

total pairs = $^5C_2 = 10$

Here $K = 3$, So choose one among them

(2,2) (2,2) (2,2)

Total pairs = $^5C_2 - {^3C_2} = 10 - 3 = 7$ pairs.

Desc

don't consider tie

ST	AST	Rank(A) natural	Rank(B)	One in natural order	Rank B	C	D
65	84	4.5	1.5	1.5	1.5	7	0
65	56	4.5	4.5	1.5	6.5	1	6
85	84	1.5	1.5	3	3	6	1
75	64	3	3	4.5	4.5	5	0
85	30	1.5	8.5	4.5	4.5	4	0
30	30	9	8.5	6	4.5	4	0
45	56	6	4.5	7	6.5	2	0
30	19	9	10	9	8.5	0	0
41	41	7	6.5	9	10	0	0
30	41	9	6.5	9	6.5	0	0
						29	7

LHS shouldn't match

check match with LHS

match. greater then less than.

$\frac{4+5}{2} = \frac{9}{2} = 4.5$

$\frac{8+9+10}{3} = \frac{27}{3} = 9$

$(x_i > y_i)$

$x_i > x_j$ & $y_i > y_j$

$\frac{4+5}{2} = 4.5$

$x_i < x_j$ & $y_i > y_j$

OR

$x_i > x_j$ & $y_i < y_j$

$P = 29$, $Q = 7$

$k_1 = 2$, $k_2 = 2$, $k_3 = 3$

$T_u = \frac{1}{2}(2\times1 + 2\times1 + 3\times2)$

$\Rightarrow \frac{2+2+6}{2} = \frac{10}{2} = 5$

$k_1' = 2$, $k_2' = 2$, $k_3' = 2$, $k_4' = 2$

$T_v = \frac{1}{2}(2+2+2+2) \Rightarrow \frac{8}{2} = 4$

$\frac{5+6+7}{3}$

Kendall's τ :- common tie situation. 09/10/Eve.

as per given in questn OR Natural way

common tie

	A	B	Rank (A)	Rank (B)	concordant pairs	discord pairs
1	80	40	1.5	11	6	9
2	80	40	1.5	11	0	9
3	75	79	3	1	9	0
4	71	61	4	6	3	3
5	65	70	6	3.5	4	0
6	65	61	6	6	3	1
7	65	74	6	2	5	0
8	45	52	9	8.5	0	2
9	45	40	9	11	0	2
10	45	52	9	8.5	0	2
11	31	61	11.5	6	0	0
12	31	70	11.5	3.5	0	0
			Asc order	Reorder	24	28

avg

$\frac{8+9+10}{3} \Rightarrow \frac{27}{3} = 9$

$\frac{5+6+7}{3} = \frac{18}{3} = 6$

$\frac{11+12}{2} = \frac{23}{3} = 11.5$

$\frac{17}{2} = 8. \quad \frac{17}{2} = 8.5$

$\frac{10+11+12}{3} \Rightarrow \frac{33}{3} = 11$

concordant $x_i > x_j$ & $y_i > y_j$ or $x_i < x_j$ & $y_i < y_j$ } 24 pairs

discordant $x_i > x_j$ & $y_i < y_j$ or $x_i < x_j$ & $y_i > y_j$ } 28 pairs.

$p = 24, \quad q = 28$

Total pairs $\Rightarrow \binom{n}{2} = \binom{12}{2} = \frac{\overset{6}{\cancel{12}} \times 11}{\cancel{2}} = 66.$

length of ties.

$$Tu = \sum_{i=1}^{s} \frac{1}{2} k_i (k_i - 1)$$

$k_1 = 2, \quad k_2 = 3, \quad k_3 = 3, \quad k_4 = 2.$

$$Tu = \frac{1}{2}\left[2\times 1 + 3\times 2 + 3\times 2 + 2\times 1\right] \quad \underline{\underline{k_iC_2}}$$

$= \frac{1}{2}\left[2 + 6 + 6 + 2\right]$

$= \frac{16}{2}$

$\boxed{Tu = 8}$

$$Tv = \sum_{i=1}^{t} \frac{1}{2} k_i'(k_i' - 1)$$

$\overset{11}{k_1' = 3} \quad \overset{3.5}{k_2' = 2} \quad \overset{6}{k_3' = 3} \quad k_4' = 2$

$Tv = \cancel{\frac{1}{2}(2\times1+2\times1)}\ \frac{1}{2}(3\times 2 + 2 + 3\times 2 + 2)$

$= \cancel{3}$

$\Rightarrow \frac{6+2+6+2}{2} \Rightarrow \frac{16}{2} = 8$

K common ties

$$\tau = \frac{24-28}{\sqrt{(66-8)(66-8)}} = \frac{-4}{58} = -0.0689.$$

* $\binom{n}{2} = P + Q \longrightarrow$ when there is No tie

* $\binom{n}{2} = P + Q +$ No. of individual ties (No. of ties)
(no. of ties of A + no. of ties of B).

But for common ties this doesn't hold. This holds for No common ties.

Case III :- when there is common tie

$\binom{n}{2} = P + Q +$ (No. of ties of A + No. of ties of B)
− No. of common ties of both A & B.

K common ties $\Rightarrow \underline{\underline{{}^{K}C_2}}$ common ties

Tu Tv

Add

$\binom{n}{2} = P + Q +$ A's tie + B's tie + common ties

$66 = \underbrace{24 + 28 + 8 + 8}_{68} + \underbrace{({}^2C_2) - (2C_2)}_{2}$ $= 2.$

$66 \overset{\checkmark}{=} 66$

$-4C_2 - 5C_2$ common ties

$\begin{pmatrix} 4 & 1 \\ 4 & 1 \\ 4 & 1 \\ 4 & 1 \end{pmatrix} \Rightarrow (4C_2)$ common tie

$\begin{pmatrix} 6 & 1 \\ 6 & 1 \\ 6 & 1 \\ 6 & 1 \\ 6 & 1 \end{pmatrix} \Rightarrow (5C_2)$

$(P + Q + \text{A's tie} + \text{B's tie}) - 4C_2 - 5C_2$