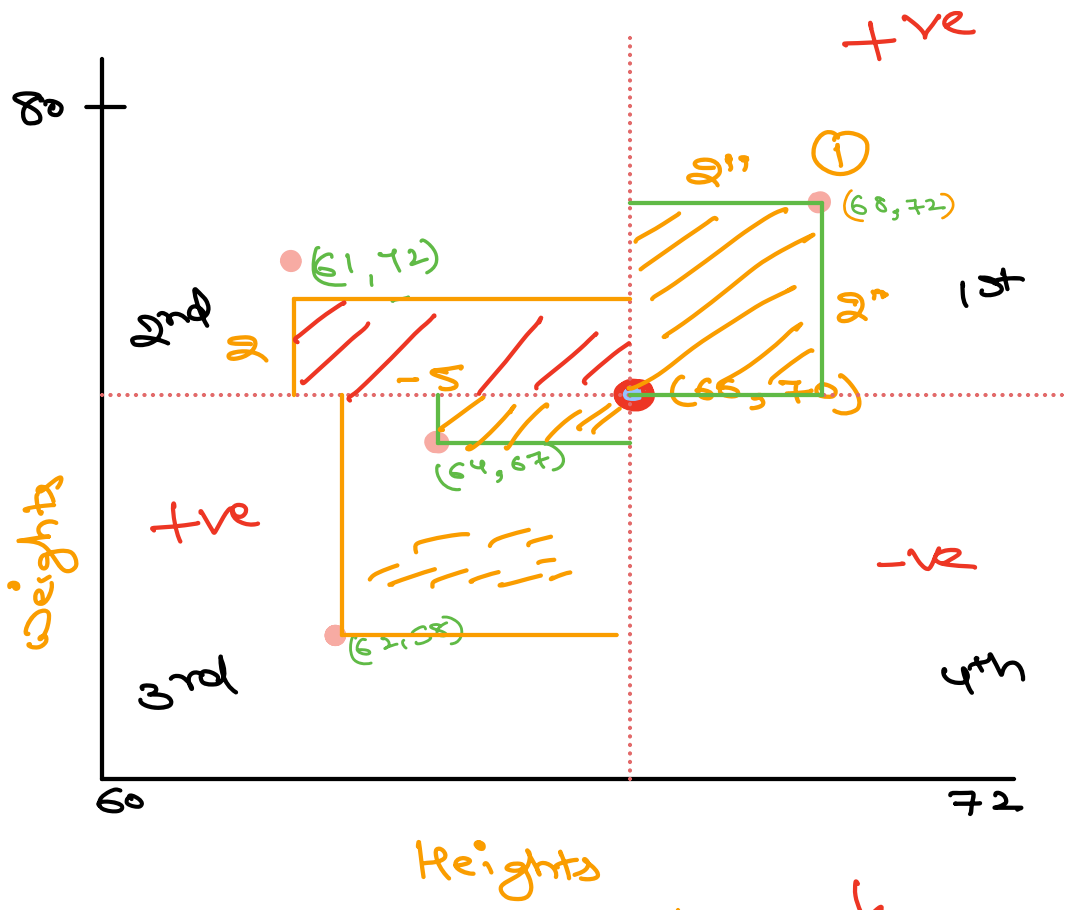
Co-variance

Height (inches)	Weight (kg)
68	72
62	58
64	67
61	72
70	79
66	61
61	68
65	64
71	80
72	79
$\bar{h} = 66$	$\bar{w} = 70$



① $(68-66) * (72-70) = 4$

② $(61-66) * (72-70) = -10$
$-5 \quad * \quad 2$

③ $(64-66) * (67-70) = 6$
$-2 \quad * \quad -3$

④ $(69-66) * (58-70) = 48$
$-4 \quad * \quad -12$

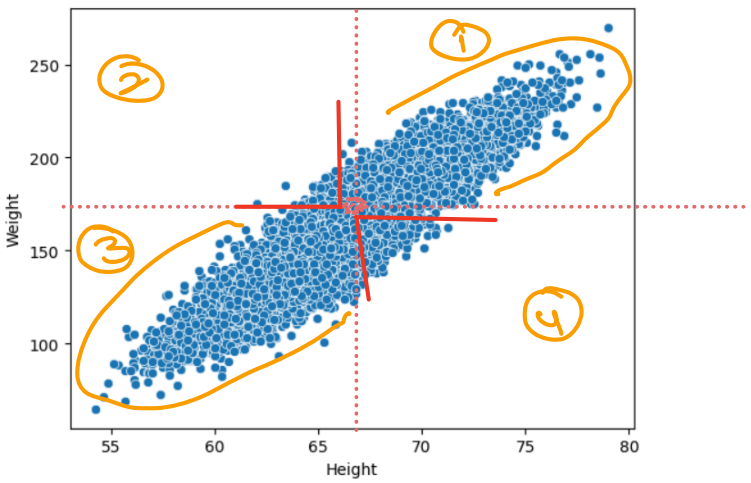
$$\frac{4 + -10 + 6 + 48}{4}$$

x_i, y_i

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad \text{(Covariance)}$$

$$Cov(x,y) = E(x - E(x)) * (y - E(y))$$

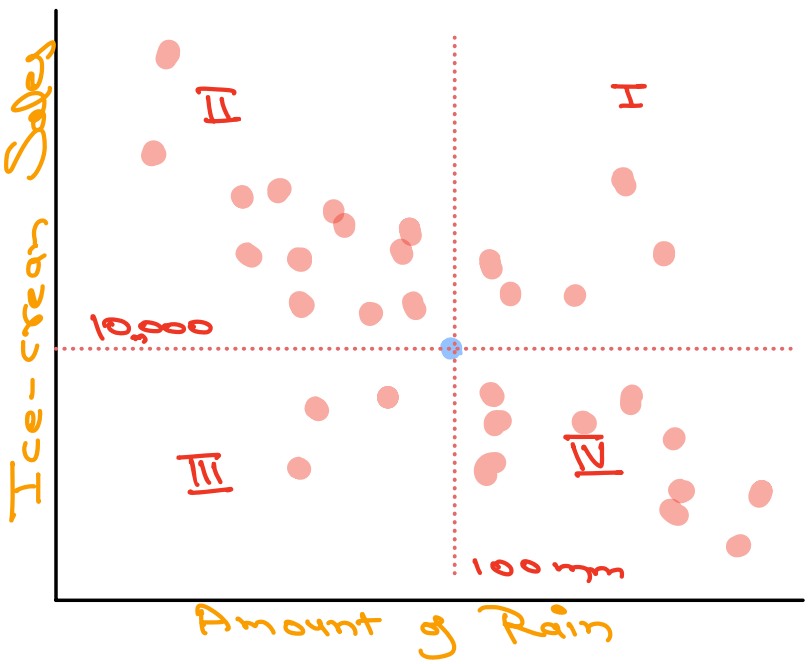
- E → Expected Value
- X and Y will be any two variables



$Cov(x, y) \rightarrow +V_2$

Ex-2 Rain vs Ice-cream Sales

$Cov(x, y) \rightarrow -V_2$



① $S_1 = 0.8$ $S_2 = -0.6$

V_1 → 2000 V_2 → -1000

$V_1 > V_2$

② V_1 → 1000 V_2 → -2000

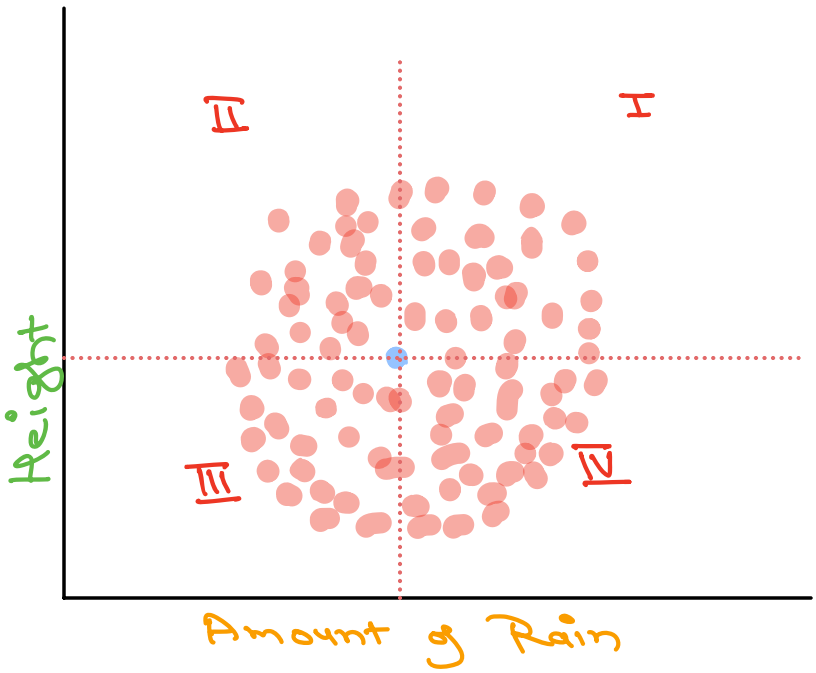
$V_1 > V_2$

S_1	S_2
0.7	-0.75

Which one is strongly related?

We cannot say just with covariance

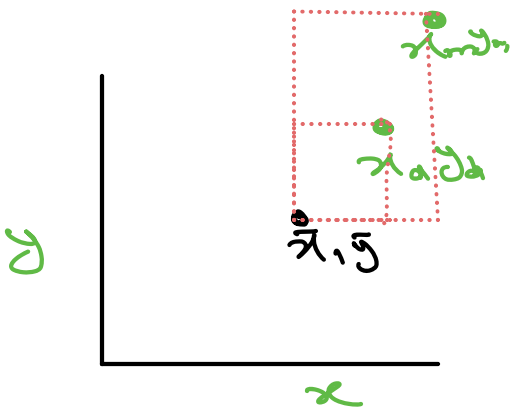
Ex - 3



Cov (x, y) ≤ 0

(No relation b/w x, y)

Pearson Correlation



Scenario

- Let's say researcher 1 measures H and W in inches and Kg
- While researcher 2 measures H and W in cm and pounds

1 inch = 2.54 cm
1 kg = 2.2 pounds

Questions

Does this Unit of measurement impact Variance?

$Cov_{r_1} < Cov_{r_2}$

r_1 r_2

β_1 β_2

0.8 0.4

Does the relationship actually changes across two scenarios?

No

Pearson Correlation

$$\rho = \frac{\text{Covariance}(X, Y)}{\sigma_x * \sigma_y}$$

For H and W:

$$\text{i.e } \rho = \frac{1}{n} \sum_{i=1}^{n} (h_i - \bar{h})(w_i - \bar{w})$$

Range $(-1, 1)$

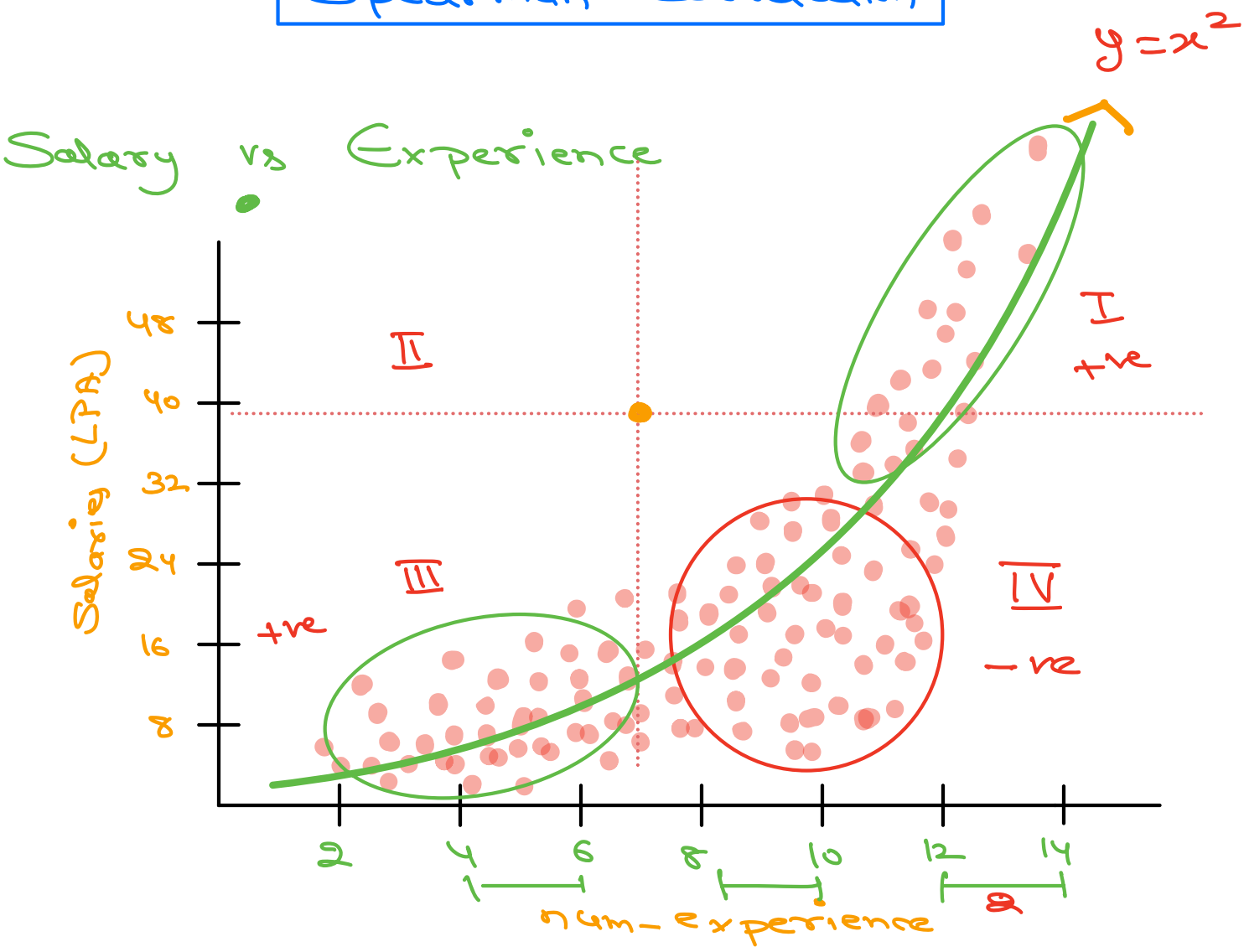
Intuition

Cor ≈ 1 → Strong positive relationship

Cor ≈ -1 → Strong Negative relationship

Cor ≈ 0 → Weak relationship

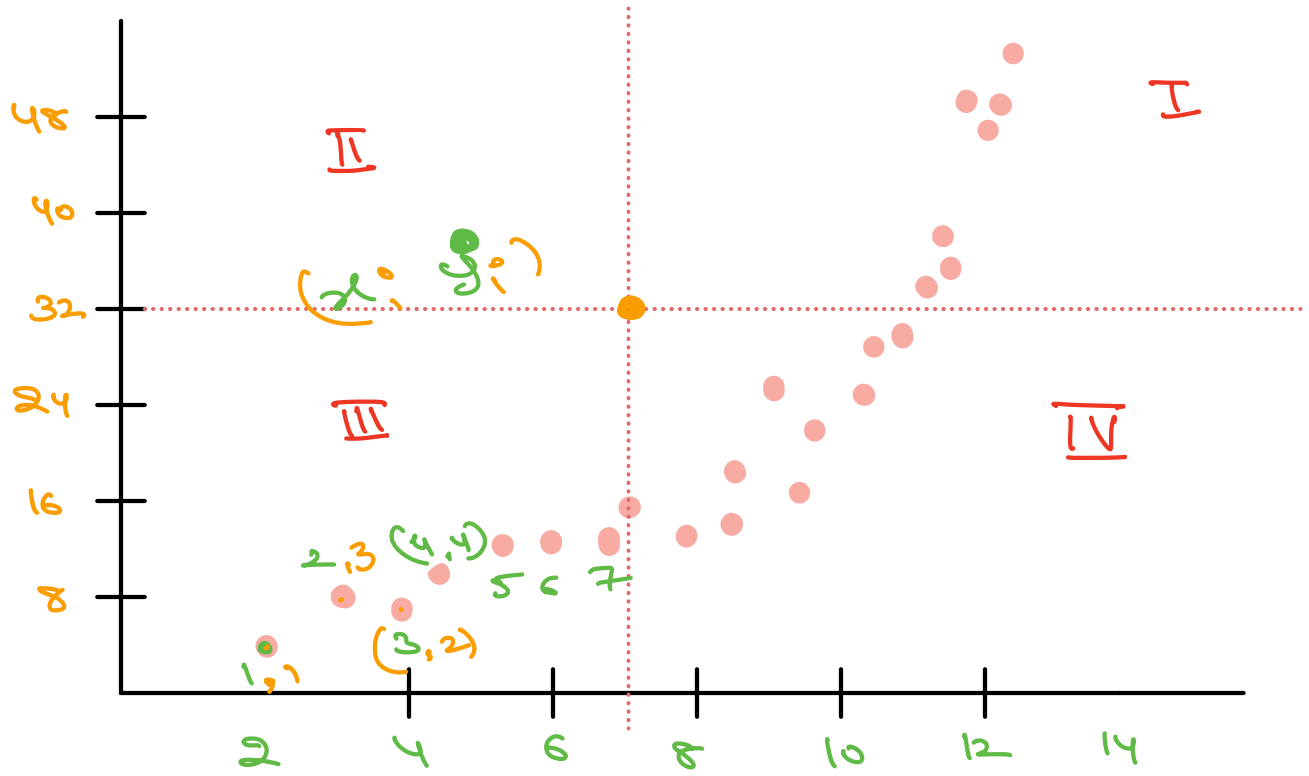
Spearman Correlation



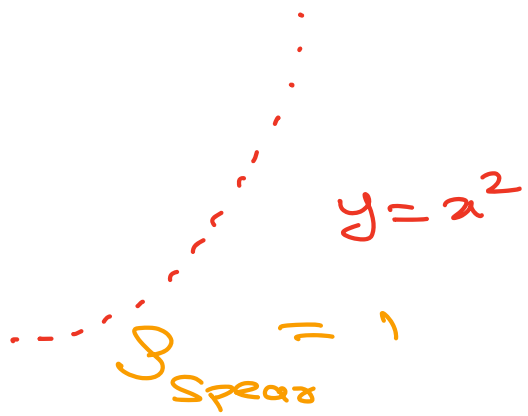
$\rho \approx 0$

Pearson correlation fails to Capture Non-Linear Relationship

Non-linear Relationships Spearmann correlation works better



Rank X	Rank Y	d_i
1	1	0
2	3	-1
3	2	1
4	4	0
...	...	



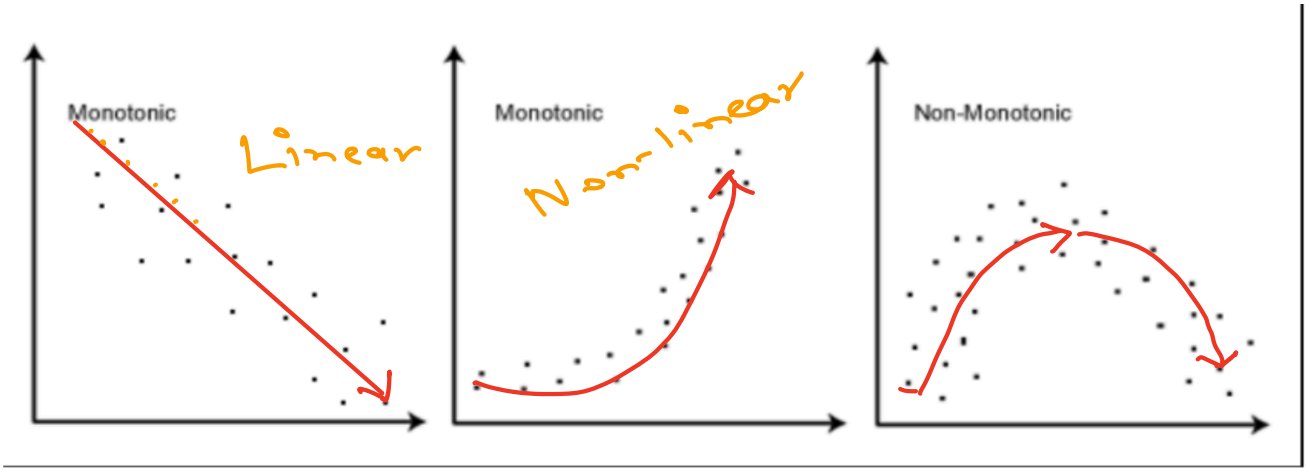
$$\rho_{Spearman} = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Range $(-1, 1)$

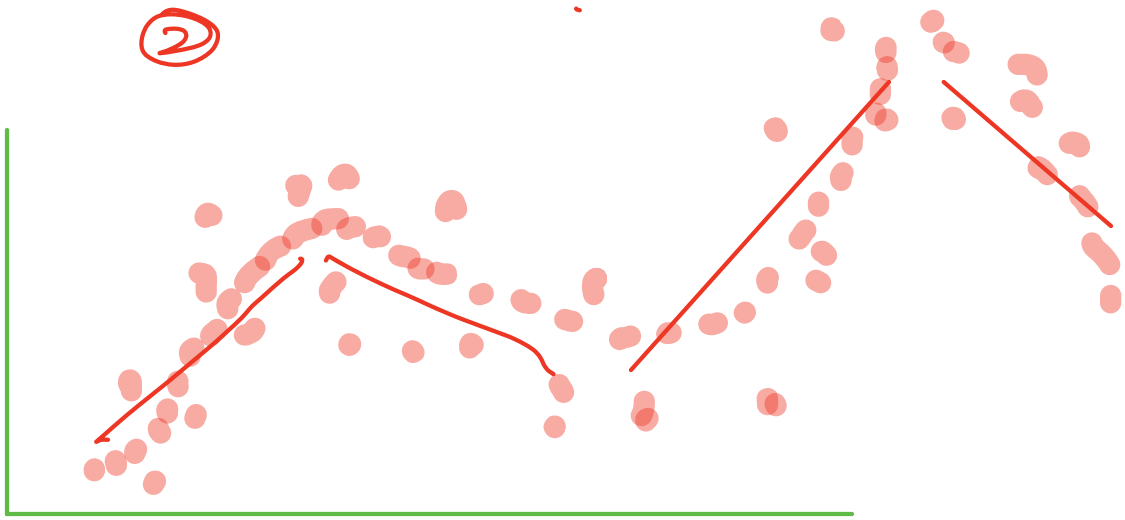
d_i = Difference b/w two ranks of Obs i

n = num_observation

Monotonic vs Non-monotonic



(1) (2)



X Experience (years)	Rank (Experience)	Y Salary ($)	Rank (Salary)	$d_i(RankX_i - RankY_i)$	d_i^2
1	1	1000	1	0	0
3	3	2000	2	1	1
4	4	5000	5	−1	1
5	5	4000	4	1	1
2	2	3000	3	−1	1

$$\sum d_i^2 = 0+1+1+1+1 = 4$$

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 4}{5(5^2-1)} = 1 - \frac{1}{5} = 1 - 0.2 = 0.8$$