

Statistical Correlation

(Part 1)



Basics Concepts
Types of Correlation
Methods

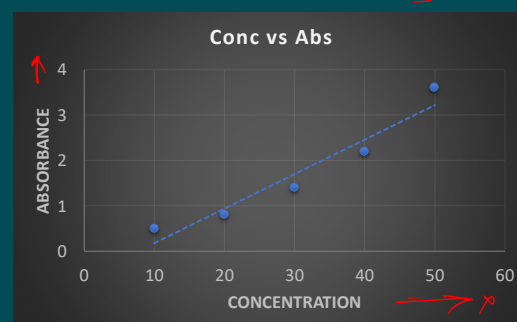
Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

Correlation



- 🔗 **Correlation**- Relationship between two variable
- 🔗 Correlation is a statistical measure that expresses the degree of relationship between two variables
- 🔗 Useful statistical measure in bivariate or multivariate data
- 🔗 If two quantities vary in such way that movement in one are accompanied by movement in the other, these quantities are correlated

SN	Concentration (ug/ml)	Absorbance
1	10	0.5
2	20	0.8
3	30	1.4
4	40	2.3
5	50	3.6



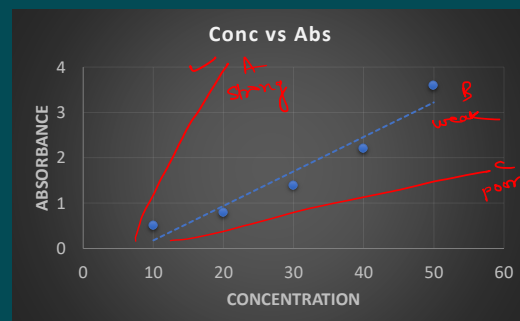
Correlation



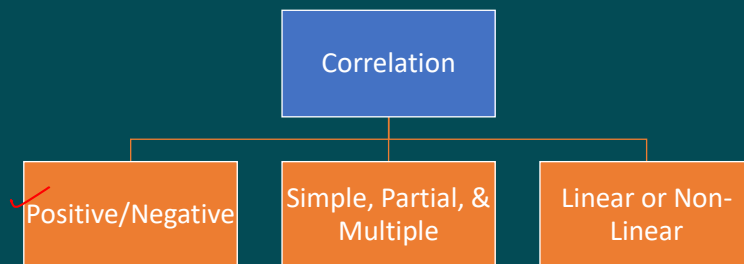
Correlation- Relationship between two variable

The measure of correlation called the **correlation coefficient** or **correlation index**, which is summarized in one figure the **direction** and **degree** of correlation

SN	Concentration (ug/ml)	Absorbance
1	10	0.5
2	20	0.8
3	30	1.4
4	40	2.3
5	50	3.6



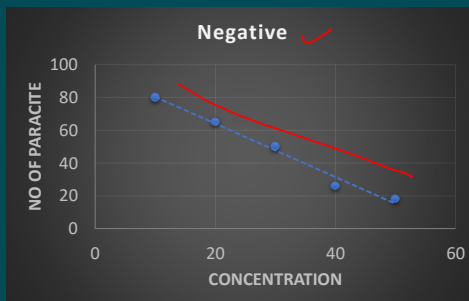
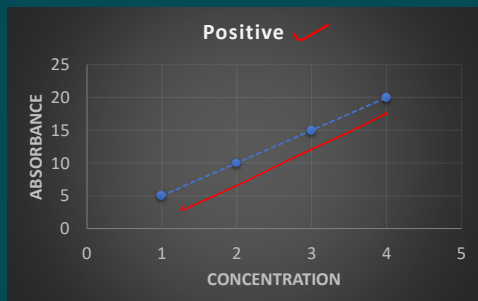
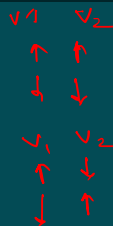
Correlation



Correlation



- Positive**- If one variable increase/decrease then 2nd variable is also increase/decrease
- Negative**- If one variable increase/decrease then 2nd variable is decrease/increase



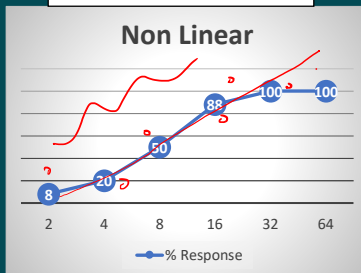
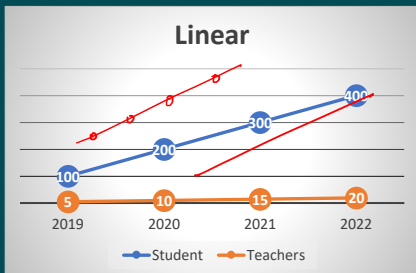
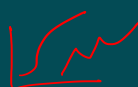
Correlation



- Linear**- Change in variables in constant ratio
- Non-Linear**- Doesn't Change in variables in constant ratio

Year	Age	Student	Teachers
2019	1Y	100	5
2020	2Y	200	10
2021	3Y	300	15
2022	4Y	400	20

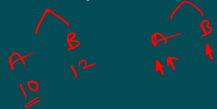
Conc.	% Response
2	8
4	20
8	50
16	88
32	100
64	100



Correlation



- 1 **Simple**- Study of two variables, e.g., Abs. Vs Conc
- 2 **Partial**- Study of three variables, e.g., Rice production depends on- soil, rainfall, fertilizer
- 3 **Multiple**- Study of multiple variables- e.g., Marks depends on- Study, Knowledge, Questions, Health, evaluator etc



Correlation



Methods of Studying Correlation

✓ Karl Pearson's coefficient of Correlation

Spearman's Rank correlation Coefficient

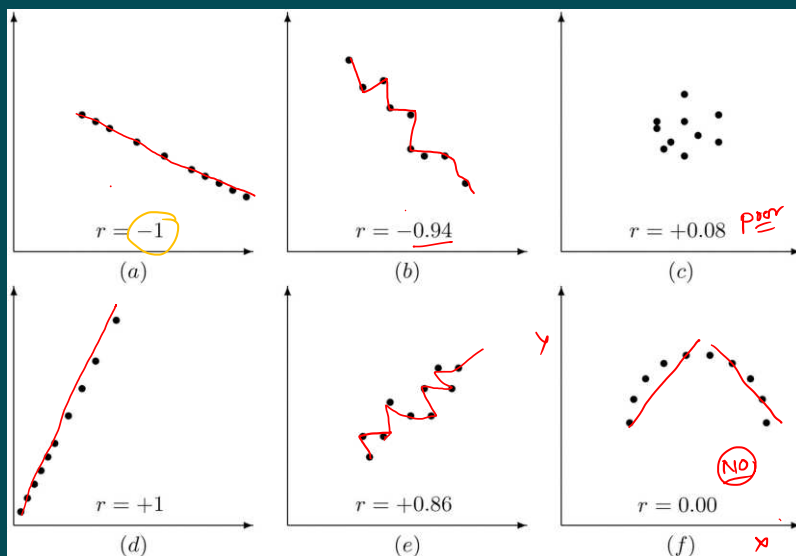
Others

- ✓ Scattered Diagram
- ✓ Graphic
- ✓ Concurrent Deviation
- ✓ Method of Least Squares

$r =$ degree
direction

$r = -1$ to $+1$
 -1 ← 0 → $+1$
 strong -ve +ve strong
 No

Correlation



Statistical Correlation (Part 2)



Karl Pearson's coefficient of correlation

Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

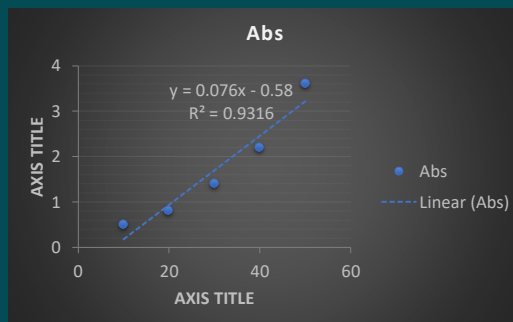
Correlation



Correlation- Relationship between two or more variables

Correlation is a statistical measure that expresses the direction and degree of relationship between variables

SN	Concentration (ug/ml)	Absorbance
1	10	0.5
2	20	0.8
3	30	1.4
4	40	2.3
5	50	3.6



Correlation



Methods of Studying Correlation

✓ Karl Pearson's coefficient of Correlation

Spearman's Rank correlation Coefficient

Others

- ✓ Scattered Diagram
- ✓ Graphic
- ✓ Concurrent Deviation
- ✓ Method of Least Squares

Karl Pearson's coefficient of Correlation



- Used to measure the strength of relationship between two variables - degree
- direction
- Denoted by "r". It is the Pearson's Product
- $r = -1$ to $+1$
- Applied when deviation is taken from actual mean value not to the assume men

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

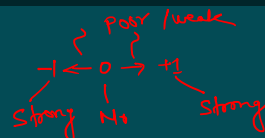
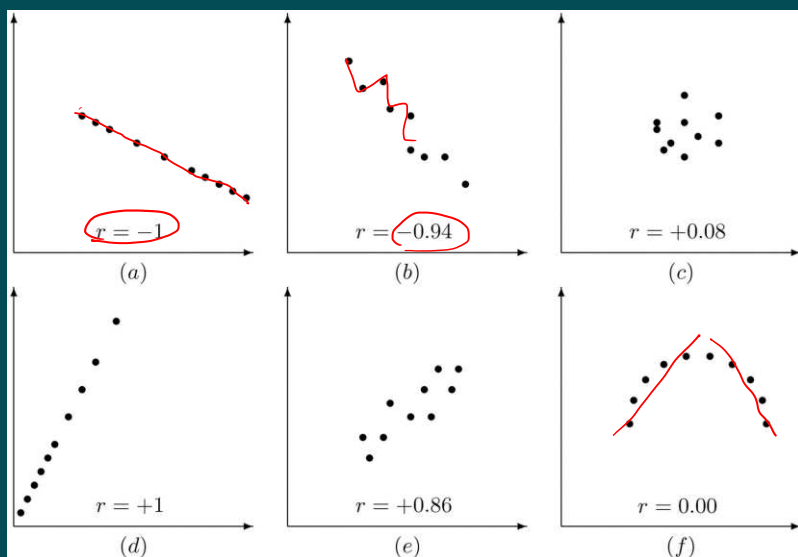
$$x = \underline{x - \bar{x}}$$

$$y = \underline{Y - \bar{Y}}$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \cdot \sqrt{\sum y^2}}$$

r	Inference
+1	perfect positive correlation
0	No correlation
-1	perfect negative correlation

Correlation



Karl Pearson's coefficient of Correlation



Calculate the r value from given data

Student	A	B	C	D	E
Pathology	5	6	7	9	8
Pharmacology	4	7	9	8	7

Student	X	x(x-7)	x ²	Y	y(y-7)	y ²	xy
A	5	-2	4	4	-3	9	6
B	6	-1	1	7	0	0	0
C	7	0	0	9	2	4	0
D	9	2	4	8	1	1	2
E	8	1	1	7	0	0	0
Sum	35		10	35		14	8

$$\bar{x} = \frac{\sum x}{n} = \frac{35}{5} = 7$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$x = x - \bar{x}$
 $y = y - \bar{y}$

$$r = \frac{8}{\sqrt{10 \times 14}}$$

$\sum x^2 = 10$
 $\sum y^2 = 14$
 $\sum xy = 8$

$$r = \frac{8}{\sqrt{140}}$$

$$r = \frac{8}{11.8}$$

$r = 0.67$

Karl Pearson's coefficient of Correlation



Calculate the r value from given data

Patient	A	B	C	D	E
Blood Sugar (X)	120	140	200	240	300
BP (Y)	80	100	110	130	180

Patient	X	x(x-200)	x ²	Y	y(y-120)	y ²	xy
A	120	-80	6400	80	-40	1600	3200
B	140	-60	3600	100	-20	400	1200
C	200	0	0	110	-10	100	0
D	240	40	1600	130	10	100	400
E	300	100	10000	180	60	3600	6000
Sum	1000		21600	600		5800	10800

$$\bar{x} = 200$$

$$\sum x^2 = 21600$$

$$\sum y = 600$$

$$\sum y^2 = 5800$$

$$\sum xy = 10800$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$x = x - \bar{x}$
 $y = y - \bar{y}$

$$r = \frac{10800}{\sqrt{21600 \times 5800}}$$

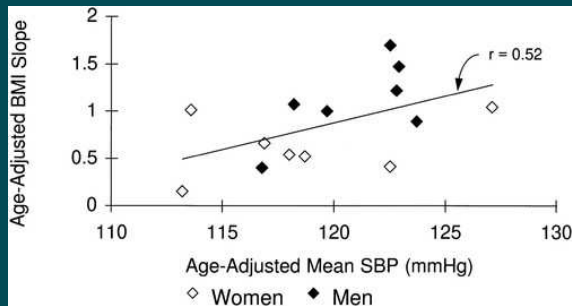
$$r = \frac{10800}{\sqrt{125280000}}$$

$$r = \frac{10800}{11192.8}$$

$r = 0.96$

positive strong

Karl Pearson's coefficient of Correlation



<https://www.ahajournals.org/doi/full/10.1161/01.hyp.30.6.1511>

Statistical Correlation (Part 3)



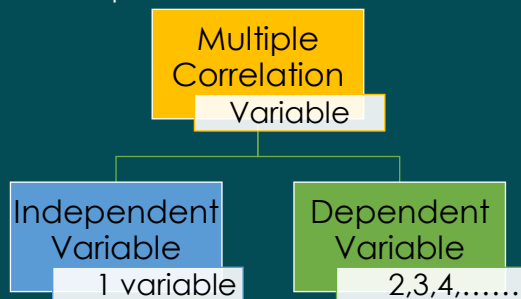
Multiple correlation

Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

Multiple Correlation



- 📌 **Correlation**- Relationship between two or more variables
- 📌 Correlation is a statistical measure that expresses the direction and degree of relationship between variables



- 📌 **Multiple Correlation Coefficient** establishes/indicates the Relationship between one variable with multiple other variables

Multiple Correlation



- 📌 **Multiple Correlation Coefficient** establishes/indicates the Relationship between one variable with multiple other variables
- 📌 Denoted by 'R'

S N	Students	BMI (A)	BP (B)	Blood Sugar (C)
1	Sham	15	80	120
2	Ram	20	100	140
3	Raju	25	110	200
4	Ramu	30	130	240
5	Shetu	35	180	300

Karl Pearson's coefficient of Correlation

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \quad \begin{array}{l} x = x - \bar{x} \\ y = Y - \bar{Y} \end{array}$$

$$r = \frac{\Sigma xy}{N \cdot \sigma_x \sigma_y}$$

Multiple Correlation



Multiple Correlation Coefficient establishes/indicates the Relationship between one variable with multiple other variables

Denoted by 'R'

The multiple correlation coefficient is denoted as $R_{A.BC.....}$

A (independent variable) which correlate with dependent Variable (B,C,D...k)

S N	Students	BMI (A)	BP (B)	Blood Sugar (C)
1	Sham	15	80	120
2	Ram	20	100	140
3	Raju	25	110	200
4	Ramu	30	130	240
5	Shetu	35	180	300

$$R_{A.BC} = \sqrt{\frac{\gamma_{ab}^2 + \gamma_{ac}^2 - 2(\gamma_{ab} \cdot \gamma_{ac} \cdot \gamma_{bc})}{1 - \gamma_{bc}^2}}$$

Multiple Correlation



S N	Students	BMI (A)	a (A-Mean)	a ²	BP (B)	b (B-Mean)	b ²	Blood Sugar (C)	c (C-Mean)	c ²	ab	ac	bc
1	Sham	15	-10	100	80	-40	1600	120	-80	6400	400	800	3200
2	Ram	20	-5	25	100	-20	400	140	-60	3600	100	300	1200
3	Raju	25	0	0	110	-10	100	200	0	0	0	0	0
4	Ramu	30	5	25	130	10	100	240	40	1600	50	200	400
5	Shetu	35	10	100	180	60	3600	300	100	10000	600	100	6000
	Sum	125		250	600		5800	1000		21600	1150	1400	10800
	mean	25			120			200					

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$

$$R_{A.BC} = \sqrt{\frac{\gamma_{ab}^2 + \gamma_{ac}^2 - 2(\gamma_{ab} \cdot \gamma_{ac} \cdot \gamma_{bc})}{1 - \gamma_{bc}^2}}$$

Multiple Correlation



S N	Students	BMI (A)	a (A- Mean)	a ²	BP (B)	b (B- Mean)	b ²	Blood Sugar (C)	c (C- Mean)	c ²	ab	ac	bc
1	Sham	15	-10	100	80	-40	1600	120	-80	6400	400	800	3200
2	Ram	20	-5	25	100	-20	400	140	-60	3600	100	300	1200
3	Raju	25	0	0	110	-10	100	200	0	0	0	0	0
4	Ramu	30	5	25	130	10	100	240	40	1600	50	200	400
5	Shetu	35	10	100	180	60	3600	300	100	10000	600	100	6000
	Sum	125		250	600		5800	1000		21600	1150	1400	10800
	mean	25			120			200					

$$r = \frac{\sum ab}{\sqrt{\sum a^2 \times \sum b^2}}$$

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

$$\sigma_a = \sqrt{\frac{\sum a^2}{N}} = \sqrt{\frac{250}{5}} = 7.07$$

$$r_{AB} = \frac{\sum ab}{N \cdot \sigma_a \cdot \sigma_b} = \frac{1150}{5 \cdot 7.07 \cdot 34} = \frac{1150}{1190} = 0.96$$

$$r_{AB}^2 = 0.92 \quad | \quad r = 0.96$$

$$r = \frac{\sum ab}{N \cdot \sigma_a \cdot \sigma_b}$$

$$\sigma_b = \sqrt{\frac{\sum b^2}{N}} = \sqrt{\frac{5800}{5}} = 34.05$$

$$\sigma_c = \sqrt{\frac{\sum c^2}{N}} = \sqrt{\frac{21600}{5}} = 65.72$$

Multiple Correlation



S N	Students	BMI (A)	a (A- Mean)	a ²	BP (B)	b (B- Mean)	b ²	Blood Sugar (C)	c (C- Mean)	c ²	ab	ac	bc
1	Sham	15	-10	100	80	-40	1600	120	-80	6400	400	800	3200
2	Ram	20	-5	25	100	-20	400	140	-60	3600	100	300	1200
3	Raju	25	0	0	110	-10	100	200	0	0	0	0	0
4	Ramu	30	5	25	130	10	100	240	40	1600	50	200	400
5	Shetu	35	10	100	180	60	3600	300	100	10000	600	100	6000
	Sum	125		250	600		5800	1000		21600	1150	1400	10800
	mean	25			120			200					

$$r = \frac{\sum ab}{\sqrt{\sum a^2 \times \sum b^2}}$$

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

$$\sigma_a = \sqrt{\frac{\sum a^2}{N}} = \sqrt{\frac{250}{5}} = 7.07$$

$$r_{AC} = \frac{\sum ac}{N \cdot \sigma_a \cdot \sigma_c} = \frac{1400}{5 \cdot 7.07 \cdot 65.7} = \frac{1400}{2299.5} = 0.6$$

$$\sigma_{AC} = 0.6 \quad r_{AC}^2 = 0.36$$

$$r = \frac{\sum ab}{N \cdot \sigma_a \cdot \sigma_b}$$

$$\sigma_b = \sqrt{\frac{\sum b^2}{N}} = \sqrt{\frac{5800}{5}} = 34.05$$

$$\sigma_c = \sqrt{\frac{\sum c^2}{N}} = \sqrt{\frac{21600}{5}} = 65.72$$

Multiple Correlation



S N	Students	BMI (A)	a (A- Mean)	a ²	BP (B)	b (B- Mean)	b ²	Blood Sugar (C)	c (C- Mean)	c ²	ab	ac	bc
1	Sham	15	-10	100	80	-40	1600	120	-80	6400	400	800	3200
2	Ram	20	-5	25	100	-20	400	140	-60	3600	100	300	1200
3	Raju	25	0	0	110	-10	100	200	0	0	0	0	0
4	Ramu	30	5	25	130	10	100	240	40	1600	50	200	400
5	Shetu	35	10	100	180	60	3600	300	100	10000	600	100	6000
	Sum	125		250	600		5800	1000		21600	1150	1400	10800
	mean	25			120			200					

$$r = \frac{\sum ab}{\sqrt{\sum a^2 \times \sum b^2}}$$

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

$$\sigma_a = \sqrt{\frac{\sum a^2}{N}} = \sqrt{\frac{250}{5}} = 7.07$$

$$\sigma_b = \sqrt{\frac{\sum b^2}{N}} = \sqrt{\frac{5800}{5}} = 34.05$$

$$\sigma_c = \sqrt{\frac{\sum c^2}{N}} = \sqrt{\frac{21600}{5}} = 65.72$$

$$r_{bc} = \frac{\sum bc}{N \sigma_b \sigma_c} = \frac{10800}{5 \times 34.05 \times 65.72} = \frac{10800}{11169} = 0.96$$

$$r_{bc} = 0.96 \quad r_{bc}^2 = 0.92$$

$$r = \frac{\sum ab}{N \cdot \sigma_a \sigma_b}$$

Multiple Correlation



S N	Students	BMI (A)	a (A- Mean)	a ²	BP (B)	b (B- Mean)	b ²	Blood Sugar (C)	c (C- Mean)	c ²	ab	ac	bc
1	Sham	15	-10	100	80	-40	1600	120	-80	6400	400	800	3200
2	Ram	20	-5	25	100	-20	400	140	-60	3600	100	300	1200
3	Raju	25	0	0	110	-10	100	200	0	0	0	0	0
4	Ramu	30	5	25	130	10	100	240	40	1600	50	200	400
5	Shetu	35	10	100	180	60	3600	300	100	10000	600	100	6000
	Sum	125		250	600		5800	1000		21600	1150	1400	10800
	mean	25			120			200					

$$rab^2 = 0.92 \quad rab = 0.96$$

$$rac^2 = 0.36 \quad rac = 0.6$$

$$rbc^2 = 0.92 \quad rbc = 0.96$$

$$R_{A.BC} = \sqrt{\frac{r_{ab}^2 + r_{ac}^2 - 2(r_{ab} \cdot r_{ac} \cdot r_{bc})}{1 - r_{bc}^2}} = \sqrt{\frac{0.92 + 0.36 - 2 \cdot 0.96 \cdot 0.6 \cdot 0.96}{1 - 0.92}}$$

$$= \sqrt{\frac{1.28 - 1.10}{0.08}} = \sqrt{0.18} = 0.42$$

Multiple Correlation



🔊 Calculate $R_{2,13}$

🔊 $r_{12} = 0.8$ $r_{12}^2 = 0.64$

🔊 $r_{13} = 0.7$ $r_{13}^2 = 0.49$

🔊 $r_{23} = 0.6$ $r_{23}^2 = 0.36$

$$\begin{aligned}
 R_{2,13} &= \sqrt{\frac{\sigma_{21}^2 + \sigma_{23}^2 - 2\sigma_{21}\sigma_{23}r_{12}}{1 - r_{13}^2}} \\
 &= \sqrt{\frac{0.64 + 0.36 - 2 \times 0.8 \times 0.7 \times 0.6}{1 - 0.49}} \\
 &= \sqrt{\frac{1 - 0.67}{0.51}} \\
 &= \sqrt{\frac{0.33}{0.51}} \\
 &= \sqrt{0.64} = 0.8
 \end{aligned}$$



Thanks for Watching



Subscribe my **YouTube**
Channel

