Regression Analysis (Part 1)

Basics Concepts
Least Square Method
Linear Regression
$Y = a + bX$ & $X = a + bY$

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

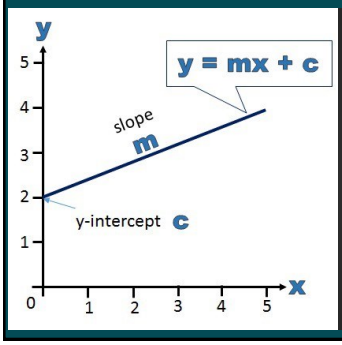
Regression

- **Correlation analysis** established the **r**elationship between two or more variables
- Now with the help of regression analysis we estimate or predict the value of one variable given the value of the another

SN	Over	Runs
1	10	60
2	20	130
3	30	220
4	40	280
5	50	

Regression

- **Regression analysis** established the **average** relationship between two or more variables and helps to estimation or prediction
- Y = mx ± c
- m- Slope
- C – Y-intercept



Regression

X-Axis (Overs)

- Independent/Explanatory/Predictor/RegratorVariable
- Used to prediction the variable of interest

Y-Axis (Runs)

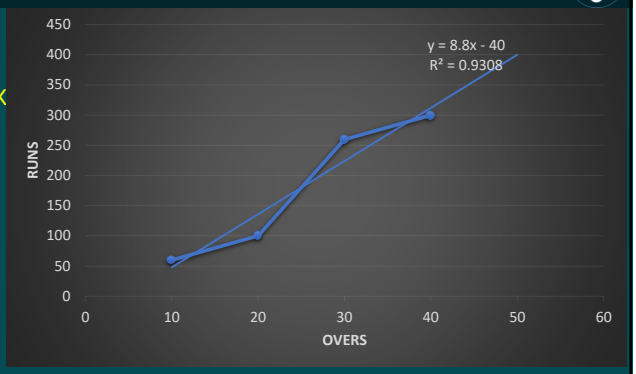
- dependent or Explained Variable
- It is predicted by Explanatory Variable

Analysis

- Simple Linear Regression Analysis
- Regression equation of Y on X- Y = a + bx
- a- constant (y-intercept)
- b- Slop of the regression line, indicate the changes in Y variable for a unit changes in X variable

Regression

- Simple Linear Regression Analysis Method-
- Regression equation of Y on X- Y = a + bx
- **Least Squares method**
- $\sum(y - y_c)^2 = 0$
- For determination a and b
- $\Sigma Y = Na + b\sum x$
- $\sum XY = a\Sigma x + b\Sigma x^2$



Regression

- Suppose Y is independent and X is dependent

Y-Axis (Runs)

- Independent or Explanatory Variable
- Used to prediction the variable of interest

X-Axis (Over)

- dependent or Explained Variable
- It is predicted by Explanatory Variable

Analysis

- Simple Linear Regression Analysis
- Regression equation of X on Y- X = a + By
- $\Sigma X = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

SN	Over	Runs
1	10	60
2	20	130
3	30	220
4	40	280
5		320

Regression

- Simple Linear Regression Analysis Method-
- Regression equation of X on Y- =X = a + bY
- $\Sigma X = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

S N	Over (X)	Runs (Y)	XY	X^2	Y^2
1	10	60	600	100	3600
2	20	130	2600	400	16900
3	30	220	6600	900	48400
4	40	280	11200	1600	78400
	100	690	21000	3000	147300

(100 = 4a + b690) x 172.5 ------1

21000 = a690 + b147300 ------2

17250 = 690a + b119025

21000 = a690 + b147300

-3750 = -28275b

b = 3750/28275 = 0.13

100 = 4a + 0.13x690

100 = 4a + 89.7

a = 10.3/4 = 2.57

- $\Sigma X = Na + b\sum Y$

100 = 4a + b690

- $\sum XY = a\Sigma Y + b\Sigma Y^2$

2100 = a690 + b147300

- X = a + bY
- X = 2.57 + 0.13Y
- X = 2.57 + 0.13 x 320
- X = 2.57 + 41.6
- X = 44.17

Regression

- Simple Linear Regression Analysis Method-
- Regression equation of Y on X- Y = a + bx
- For determination a and b
- $\Sigma y = Na + b\sum x$
- $\sum xy = a\Sigma x + b\Sigma x^2$

S N	Conc. (X)	Abs (Y)	XY	X^2	Y^2
1	2	1	2	4	1
2	4	2	8	16	4
3	6	3	18	36	9
4	8	4	32	64	16
5	10	5	50	100	25
	30	15	110	220	55

15 = 5a + b30 ------1

110 = a30 + b220 ------2

(15 = 5a + b30) x6

110 = a30 + b220

90 = 30a + 180b

110 = 30a + 220b

-20 = -40b

b = 20/40 = 0.5

15 = 5a + 0.5 x 30

15 = 5a + 15

a = 0/5 = 0

- Y = a + bX
- Y = 0 + 0.5X
- Y = 0.5 X 12
- Y = 6

Regression Analysis (Part 2)

Multiple Regression

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

Regression

- Regression analysis helps to estimate or predict the value of one variable given the value of the another
- Prediction the value of one dependent variable by available multiple independent variable.

SN	Y (Marks)	X (Study hour)
1	2	3
2	4	4
3	6	6
4	8	7
5	10	9

Y on X, $Y = a + bx$

- $\Sigma Y = Na + b\sum x$
- $\sum XY = a\Sigma x + b\Sigma x^2$

SN	Y (Marks)	X (Study hour)	X2 (Class)
1	2	3	2
2	4	4	3
3	6	6	4
4	8	7	5
5	10	9	6

Y on X_1 & X_2
$Y = a + b_1x_1 + b_2X_2$

Regression

- The model should be relevant and reliable
- The model should be linear, and variables must have normal distribution
- The purpose of the constant "a" is denote the dependent variable value in case when the values of independent variable turn to zero

SN	Y (Marks)	X (Study hour)	X2 (Class)
1	2	3	2
2	4	4	3
3	6	6	4
4	8	7	5
5	10	9	6

Y on X_1 & X_2 $Y = a + b_1x_1 + b_2X_2$

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

Regression

SN	Y (Marks)	X1 (Study hour)	X2 (Class)	YX1	YX2	X1X2	Y^2	$X1^2$	$X2^2$
1	2	3	2	6	4	6	4	9	4
2	4	4	3	16	12	12	16	16	9
3	6	6	4	36	24	24	36	36	16
4	8	7	5	56	40	35	64	49	25
5	10	9	6	90	60	54	100	81	36
	30	**29**	**20**	**204**	**140**	**131**		**191**	**90**

Y on X_1 & X_2 $Y = a + b_1x_1 + b_2X_2$

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

Regression

SN	Y (Marks)	X1 (Study hour)	X2 (Class)	YX1	YX2	X1X2	Y^2	$X1^2$	$X2^2$
1	2	3	2	6	4	6	4	9	4
2	4	4	3	16	12	12	16	16	9
3	6	6	4	36	24	24	36	36	16
4	8	7	5	56	40	35	64	49	25
5	10	9	6	90	60	54	100	81	36
	30	**29**	**20**	**204**	**140**	**131**		**191**	**90**

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

30 = 5a + 29b1 + 20b2 ----------1

204 = 29a + 191b1 + 131b2 ----------2

140 = 20a + 131b1 + 90b2 ----------3

Solve the equation 1 & 2

(30 = 5a + 29b1 + 20b2) x 5.8

204 = 29a + 191b1 + 131b2

174 = 29a + 168.2 b1 + 116 b2

204 = 29a + 191b1 + 131b2

-30 = -21.8 b1 -15 b2

30 = 21.8 b1 + 15 b2----------4

Regression

SN	Y (Marks)	X1 (Study hour)	X2 (Class)	YX1	YX2	X1X2	Y^2	$X1^2$	$X2^2$
1	2	3	2	6	4	6	4	9	4
2	4	4	3	16	12	12	16	16	9
3	6	6	4	36	24	24	36	36	16
4	8	7	5	56	40	35	64	49	25
5	10	9	6	90	60	54	100	81	36
	30	**29**	**20**	**204**	**140**	**131**		**191**	**90**

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

30 = 5a + 29b1 + 20b2 ----------1

204 = 29a + 191b1 + 131b2 ----------2

140 = 20a + 131b1 + 90b2 ----------3

Solve the equation 1 & 3

(30 = 5a + 29b1 + 20b2) x 4

140 = 20a + 131b1 + 90b2

120 = 20a + 116 b1 + 80 b2

140 = 20a + 131b1 + 90b2

-20 = -15 b1 -10 b2

20 = 15 b1 + 10 b2----------5

Regression

SN	Y (Marks)	X1 (Study hour)	X2 (Class)	YX1	YX2	X1X2	Y^2	$X1^2$	$X2^2$
1	2	3	2	6	4	6	4	9	4
2	4	4	3	16	12	12	16	16	9
3	6	6	4	36	24	24	36	36	16
4	8	7	5	56	40	35	64	49	25
5	10	9	6	90	60	54	100	81	36
	30	**29**	**20**	**204**	**140**	**131**		**191**	**90**

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

30 = 5a + 29b1 + 20b2 ----------1

204 = 29a + 191b1 + 131b2 ----------2

140 = 20a + 131b1 + 90b2 ----------3

Solve the equation 4 & 5

(30 = 21.8 b1 + 15b2) x 2

(20 = 15 b1 + 10 b2) x 3

60 = 43.6b1 + 30b2

60 = 45b1 + 30b2

----

0 = -1.4b1 the b1 = 0

From eq 5

20 = 15 b1 + 10 b2

20 = 0 + 10b2 then b2 =2

Regression

SN	Y (Marks)	X1 (Study hour)	X2 (Class)	YX1	YX2	X1X2	Y^2	$X1^2$	$X2^2$
1	2	3	2	6	4	6	4	9	4
2	4	4	3	16	12	12	16	16	9
3	6	6	4	36	24	24	36	36	16
4	8	7	5	56	40	35	64	49	25
5	10	9	6	90	60	54	100	81	36
	30	**29**	**20**	**204**	**140**	**131**		**191**	**90**

1. $\Sigma Y = Na + b_1 \sum X_1 + b_2 \sum X_2$
2. $\Sigma YX_1 = a\Sigma X_1 + b_1 \sum X^2{}_1 + b_2 \sum X_1 X_2$
3. $\Sigma YX_2 = a\Sigma X_2 + b_1 \sum X_1 X_2 + b_2 \sum X^2{}_2$

30 = 5a + 29b1 + 20b2 ----------1

204 = 29a + 191b1 + 131b2 ----------2

140 = 20a + 131b1 + 90b2 ----------3

b1 = 0

B2 = 2

From eq 1

30 = 5a + 29 x 0 + 20 x 2

30 = 5a +40

-10/5 = a

a = -2

So finally $Y = a + b_1x_1 + b_2X_2$

$Y = -2 + 2X_2$

$Y = 2X_2 - 2$

Regression Analysis (Part 3)

Standard Error of Regression

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

Standard Error of Regression

- Estimate the deviation from actual value of variables (X or Y)

SN	X	Y	X^2	Y^2	XY
1	2	3			
2	4	4			
3	5	8			
4	7	9			
5	8	10			
	26	**34**			

Y on X, Y = a + bx

- $\Sigma Y = Na + b\sum X$
- $\sum XY = a\Sigma X + b\Sigma X^2$

X on Y, X = a + bY

- $\Sigma X = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

Standard Error of Regression

- Estimate the deviation from actual value of variable

SN	X	Y	X^2	Y^2	XY
1	2	3	4	9	6
2	4	4	16	16	16
3	5	8	25	64	40
4	7	9	49	81	63
5	8	10	64	100	80
	26	**34**	**158**	**270**	**205**

Y on X, Y = a + bx

- $\Sigma Y = Na + b\sum X$
- $\sum XY = a\Sigma X + b\Sigma X^2$

X on Y, X = a + bY

- $\Sigma Y = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

Y on X, Y = a + bx

(34 = 5a + 26b) x 5.2----------1

205 = 26a + 158b----------2

From eq 1 & 2

176.8 = 26a + 135.2b

205 = 26a + 158b

-28.2 = -22.8b

b = 28.2/22.8 = **1.2**

From eq 1

34 = 5a + 26b

34 = 5a + 31.2

2.8/5 = a = **0.56**

Y = 0.56 + 1.2X

Standard Error of Regression

- Estimate the deviation from actual value of variable

SN	X	Y	Yc	Y-Yc	$(Y-Yc)^2$
1	2	3	2.96	0.04	0.0016
2	4	4	5.36	-1.36	1.84
3	5	8	6.56	1.44	2.07
4	7	9	8.96	0.04	0.0016
5	8	10	10.16	-0.16	0.02
	26	**34**			**3.39**

Y = 0.56 + 1.2X

$$Syx = \sqrt{\frac{(Y-Yc)^2}{N}} \qquad Sxy = \sqrt{\frac{(X-Xc)^2}{N}}$$

$$Syx = \sqrt{\frac{3.93}{5}}$$

$$Syx = \sqrt{0.78}$$

Syx = 0.88

Y on X, Y = a + bx

- $\Sigma Y = Na + b\sum X$
- $\sum XY = a\Sigma X + b\Sigma X^2$

X on Y, X = a + bY

- $\Sigma Y = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

Standard Error of Regression

- Estimate the deviation from actual value of variable

SN	X	Y	X^2	Y^2	XY
1	2	3	4	9	6
2	4	4	16	16	16
3	5	8	25	64	40
4	7	9	49	81	63
5	8	10	64	100	80
	26	**34**	**158**	**270**	**205**

Y on X, Y = a + bx

- $\Sigma Y = Na + b\sum X$
- $\sum XY = a\Sigma X + b\Sigma X^2$

X on Y, X = a + bY

- $\Sigma X = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

X on Y, X = a + bY

(26 = 5a + 34b) x 6.8----------1

205 = 34a + 270b----------2

From eq 1 & 2

176.8 = 34a + 231b

205 = 34a + 270b

-28.2 = -39b

b = 28.2/39 = **0.72**

From eq 1

26 = 5a + 34b

26 = 5a + 24.5

1.5/5 = a = **0.3**

X = 0.3 + 0.72X

Standard Error of Regression

- Estimate the deviation from actual value of variable

SN	X	Y	Xc	X-Xc	$(X-Xc)^2$
1	2	3	2.46	-0.46	0.21
2	4	4	3.2	0.8	0.64
3	5	8	6	-1	1
4	7	9	6.8	0.2	0.04
5	8	10	7.5	0.5	0.25
	26	**34**			**2.14**

X = 0.3 + 0.72Y

$$Syx = \sqrt{\frac{\Sigma(Y-Yc)^2}{N}} \quad Sxy = \sqrt{\frac{\Sigma(X-Xc)^2}{N}}$$

$$Sxy = \sqrt{\frac{2.14}{5}}$$

$$Sxy = \sqrt{0.43}$$

Sxy = 0.65

Y on X, Y = a + bx

- $\Sigma Y = Na + b\sum X$
- $\sum XY = a\Sigma X + b\Sigma X^2$

X on Y, X = a + bY

- $\Sigma Y = Na + b\sum Y$
- $\sum XY = a\Sigma Y + b\Sigma Y^2$

Thanks for Watching

Subscribe my YouTube Channel