Linear Regression.

How to calculate linear regression using least square method ?

Independent Variable	Dependent Variable
x	y
1	2
2	4
3	5
4	4
5	5
Mean 3	4

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	2	-2	-2	4	4
2	4	-1	0	1	0
3	5	0	1	0	0
4	4	1	0	1	0
5	5	2	1	4	2
				10	6

$b_1 = \frac{6}{10} = 0.6$

$\hat{y} = b_0 + b_1 x$ [b_0 is calculated using the mean Coordinate (3, 4)]

$4 = b_0 + 0.6 \times 3$

$b_0 = 2.2$

$\hat{y} = 2.2 + 0.6x$

How to predict for a new x value:

if x is given as 10, predicted y value is

$\hat{y} = 2.2 + 0.6 \times 10$

$\boxed{\hat{y} = 8.2}$

How to calculate R^2 using Regression Analysis?

x	y	$y-\bar{y}$	$(y-\bar{y})^2$	$\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$
1	2	-2	4	2.8	-1.2	1.44
2	4	0	0	3.4	-0.6	0.36
3	5	1	1	4	0	0
4	4	0	0	4.6	0.6	0.36
5	5	1	1	5.2	1.2	1.44

mean = 4   6   3.6

$$R^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{(y-\bar{y})^2} = \frac{3.6}{6} = 0.6$$

R^2 is closer to 1 means the distance between actual and predicted value is less/closer

R^2 is closer to 0 means the distance between actual and predicted value is more/greater.

Therefore its prefered to have R^2 closer to 1.

Example: 2 $\boxed{\text{Linear Regression}}$ $\boxed{\hat{y} = b_0 + b_1 x}$

(y-intercept) $b_0 = \dfrac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$

(Coefficient of x) $b_1 = \dfrac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$

	Age x	Glucose level y	xy	x^2
	43	99	4257	1849
	21	65	1365	441
	25	79	1975	625
	42	75	3150	1764
	57	87	4959	3249
	59	81	4779	3481
Sum:	247	486	20485	11409

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{486 \times 11409 - 247 \times 20485}{6 \times 11409 - (247)^2}$$

$$b_0 = \frac{484897 9}{7445} = 65.14$$

$$b_1 = \frac{6 \times 20485 - 247 \times 486}{6 \times 11409 - (247)^2} = \frac{2868}{7445} = 0.385$$

For given x value Predict y:

$x = 55$, Regression line Equation: $\hat{y} = 65.14 + (0.385)x$

$\hat{y} = 65.14 + 0.385 \times 55$

$= 86.327$

Predicted Glucose for age 55 is 86.3