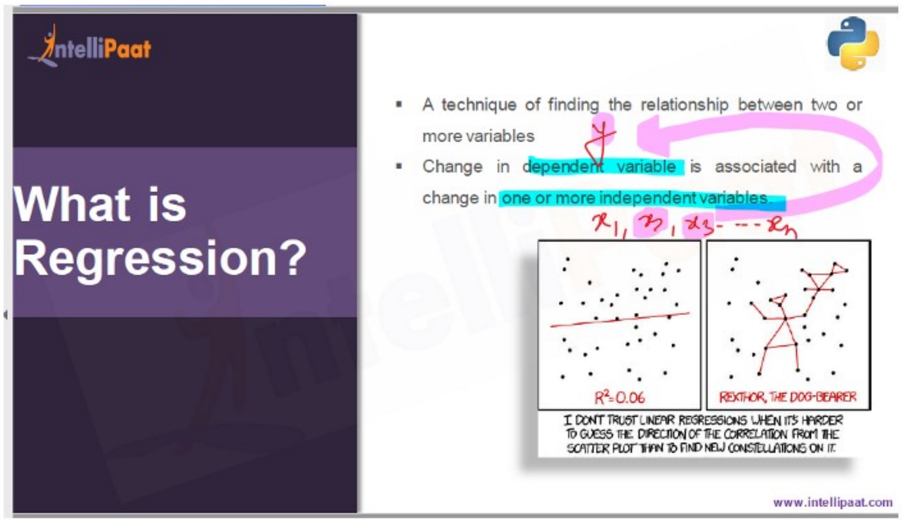
Linear Regression

15 December 2023 07:52

→ model is linear → linear relationship between Y & x_i's

↳ continuous / ~~categorical~~

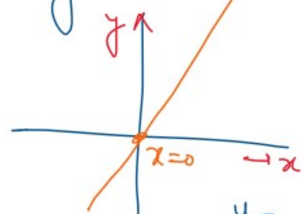
Supervised Learning Algorithm



a function in 'x'

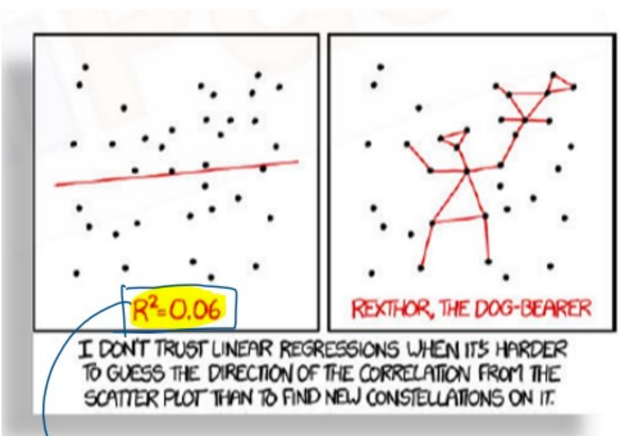
$y = 2x$

$y = mx$



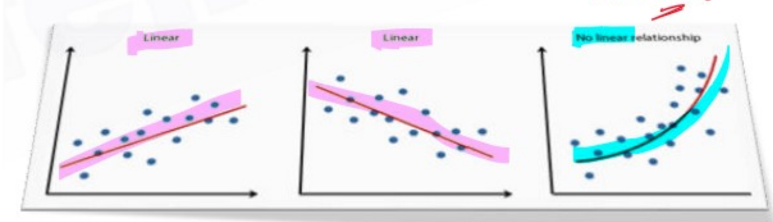
$y = 2x$

y	x
0	0
−2	−1
2	+1



6%

Non-linear



① Intuitive Approach

Function ↓ Limits → Continuity → Differentiability

Calculus

Maxi... & Minima (Global optima...)

① Intuitive Approach
↓
② Mathematical Approach
↓
③ Python Hands-on

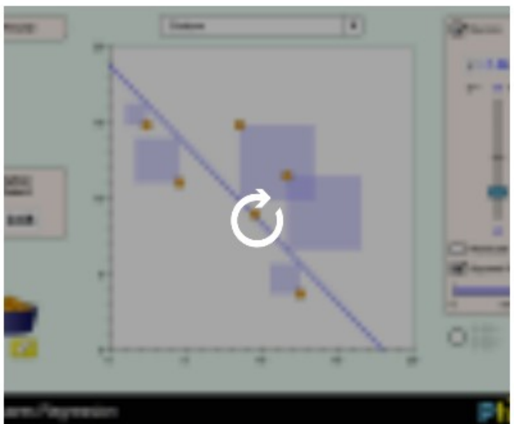
① → Calculus → [Limits → Continuity → Differentiability
Maxima & Minima (Global Optimization)]

② → Linear Algebra

③ → (Probability and Random Variables
⊕
Probability distribution Normal distribution)

④ → statistics

[Least-Squares Regression]



mathematical equation

$Y = 0.83X + 0.75$

$Y = \hat{\beta}_1 x + \hat{\beta}_0$

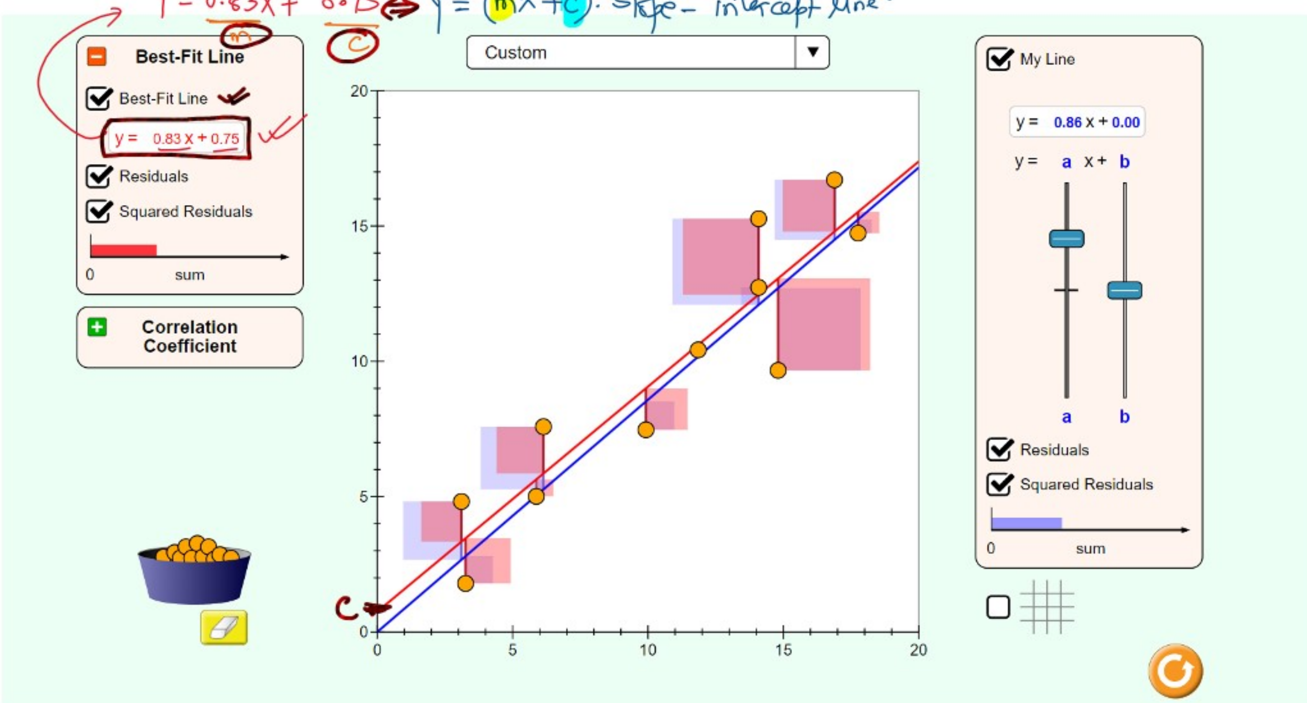
slope

intercept

$Y = 0.83X + 0.75 \Leftrightarrow Y = (mx + c)$: slope - intercept line.

m

c

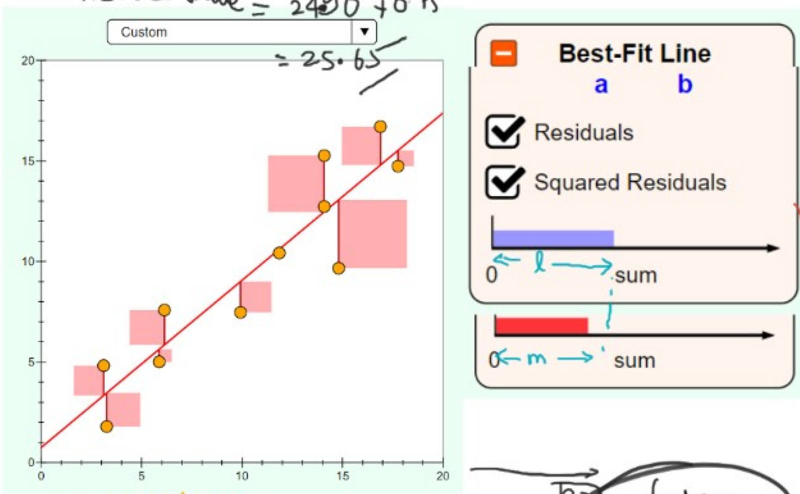


$y = 0.83x + 0.75$

$y_{\text{Predicted value}} = 0.83 \times 30 + 0.75$ at $x = 30$

$= 24.90 + 0.75$

$= 25.65$



l: intuitive best fit line

m: LR model best fit line

a) $l > m$

b) $l < m$

c) $l = m$

d) Not sure

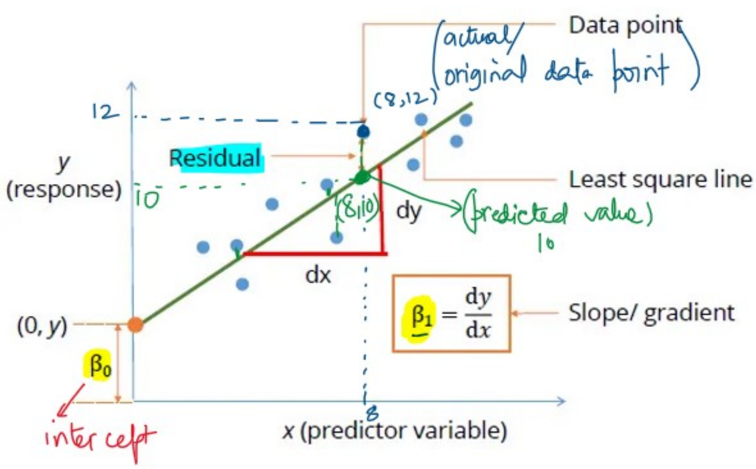
30 → future

● : data points

— : Linear Regression Line

• data point

— (Linear Regression / Best fit / least square line)



(Predicted − actual)

(10 − 12)

= −2

$y = c + mx + u$

Residual

$$\boxed{y = \beta_0 + \beta_1 x + u}$$

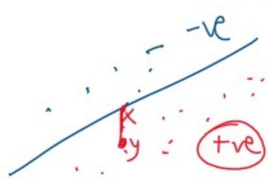
Actual value

Predicted value

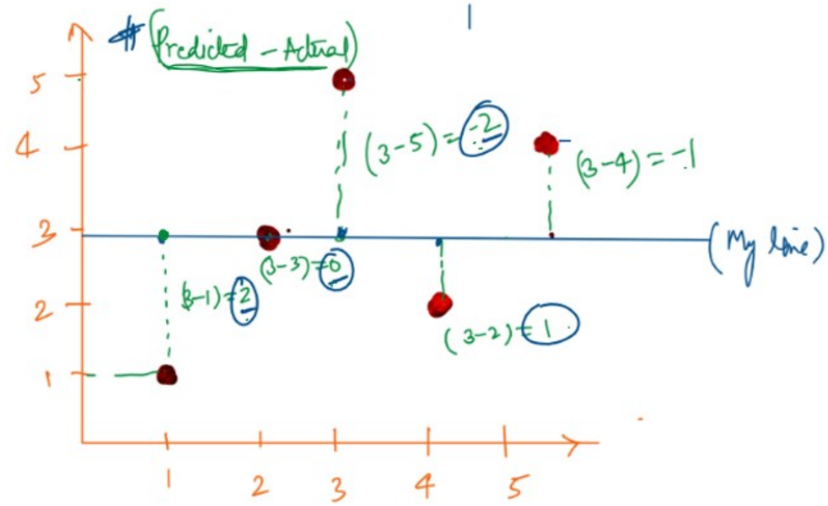
slope/ gradient

$\hat{\beta_0}$, $\hat{\beta_1}$: derive it as well.

$y = 2 + 0.5x$



+ve residues cancel -ve residues

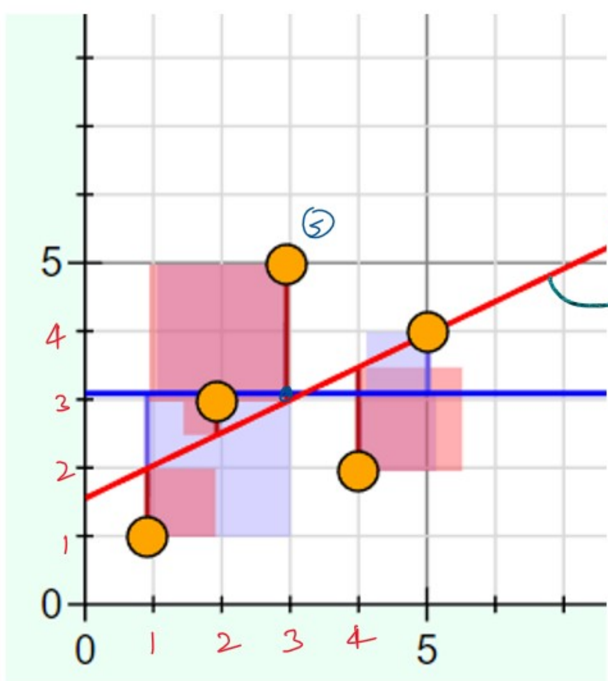


① sum of residues $= \cancel{2} + 0 \cancel{-2} \cancel{+1} \cancel{-1} = \underline{\underline{0}}$.

(zero sum of residues)

⇓

(zero error in the model)



Best fit line → $\boxed{7.8}$ ← (10) (my line)

ordinary least square

↓

(minimum)

Observation: Residues nullify each other and may not be truly representing the residues.

OLS Method: ordinary least squares

OLS Method: ordinary least squares

sum of squared residues

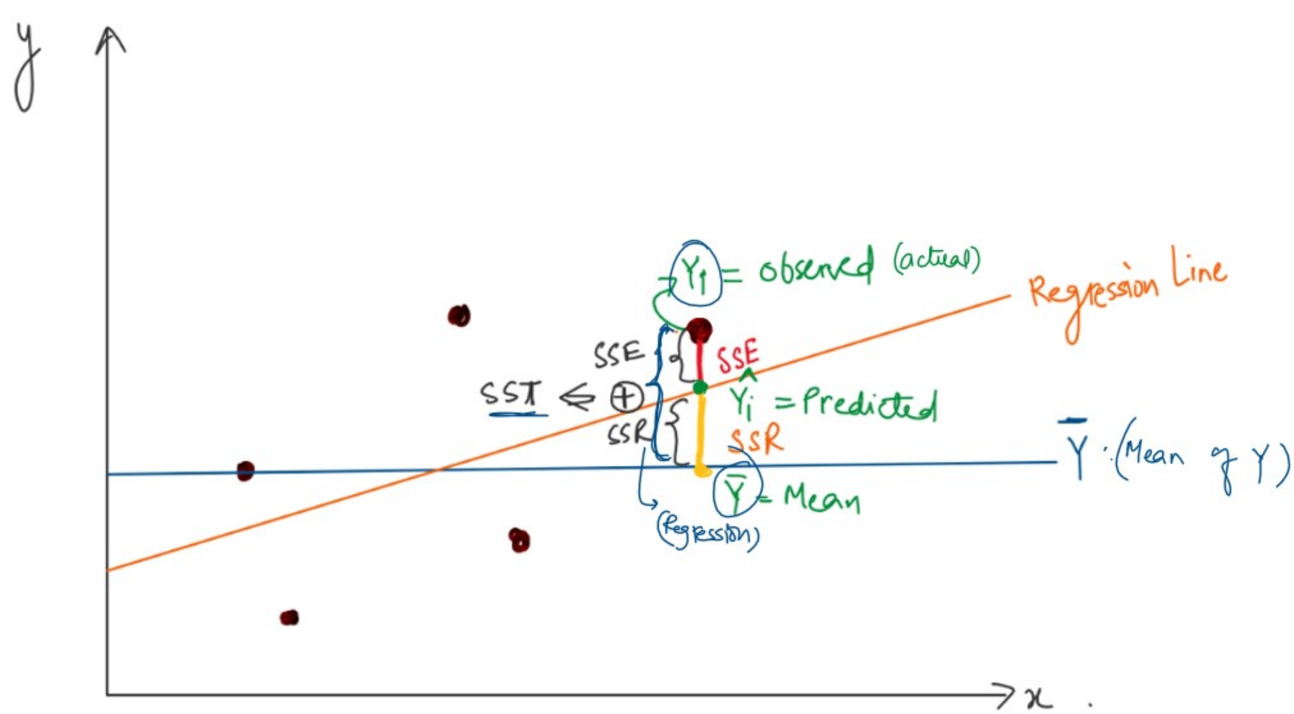
sum of residues $= \cancel{2} + 0 \cancel{-2} \cancel{+1} \cancel{-1} = 0$.

sum of squared residues $= 2^2 + 0^2 + (-2)^2 + 1^2 + (-1)^2$

$= 4 + 0 + 4 + 1 + 1$

$= 10$.

SSE, SSR and SST



sum of squares error (residues) (SSE)

- it is the sum of the squared difference between the observed (actual) value and

- it is the sum of the squared difference between the observed (actual) value (Y_i) and predicted value ($\hat{Y}_i$)

- it shows the unexplained variance by regression.

$$\boxed{SSE = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}$$

sum of squares regression (SSR)

- it is the sum of the squared differences between predicted value ($\hat{Y}_i$) and the mean of the dependent variable ($\bar{Y}$)

- it is a measure that describes how well our line fits the data.

$$\boxed{\underset{\text{(regression)}}{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - \underset{\text{Mean}}{\bar{Y}})^2}$$

sum of squares total (SST)

- it is the squared differences between observed dependent variable and its mean.

$$\boxed{SST / TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

- it is a measure of total variability of the dataset.

Mathematically,

Mathematically,

SST = SSR + SSE

$$\Rightarrow \sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2 + \sum_{i=1}^{n}(\hat{Y}_i-Y_i)^2$$

Total variability of the dataset (SST)

Variability explained by the regression line (SSR)

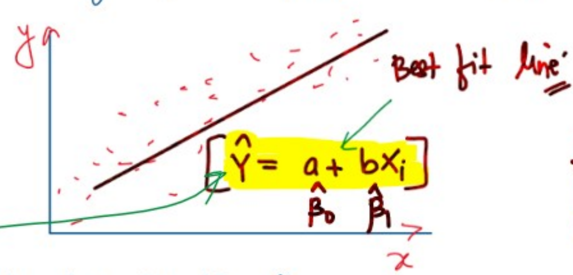
unexplained variability (SSE)

100% = 75% + 25%

Accuracy

Derivation of linear regression equations

Given a set of 'n' points (x_i, y_i) on a scatter plot.



$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
$$\hat{Y} = a + bx_i$$

find the best fit line; $\hat{Y} = a + bx_i$

such that the sum of squared errors in Y : $\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$ is minimized

actual (Observed) (Predicted)

$$Q = \sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = \sum_{i=1}^{n}\left[Y_i-(a+bX_i)\right]^2$$

$$Q = \sum_{i=1}^{n}(Y_i - a - bX_i)^{\boxed{2}}$$

Using maxima/minima concept, let us partially differentiate

Using maxima/minima concept, let us partially differentiate with 'a' and 'b' respectively.

$$z = x^2 y$$

$$\frac{\partial z}{\partial x} = 2xy \quad \Big| \quad \frac{\partial z}{\partial y} = x^2$$

-('y' constant) ('x' constant)

$$\boxed{\frac{\partial Q}{\partial a} = 0}$$

Differentiation

$y = x^2$, $\frac{dy}{dx} = 2x$

$y = x^3$, $\frac{dy}{dx} = 3x^2$

$y = c$, $\frac{dy}{dx} = 0$

$y = x^n$, $\frac{dy}{dx} = n x^{n-1}$

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2 \left(\boxed{Y_i - a - bX_i}\right)^1 \times \underbrace{(0 - 1 - 0)}_{\text{(chain rule)}}$$

$$\frac{d}{\partial a}(Y_i - a - \underline{b}\,\underline{x_i}) = (0 - 1 - 0)$$

const.

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^{n} (Y_i - a - bX_i)$$

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/v/partial-derivatives-introduction

$$\frac{\partial Q}{\partial a} = -2 \left[\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} a - b \sum_{i=1}^{n} X_i \right] = 0$$

$\sum_{i=1}^{n} a \rightarrow (na)$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\boxed{\Rightarrow \sum_{i=1}^{n} X_i = n\bar{X}}$$

$$\frac{\partial Q}{\partial a} = -2n\bar{Y} + 2na + 2bn\bar{X} = 0$$

Dividing by $\underline{2n}$

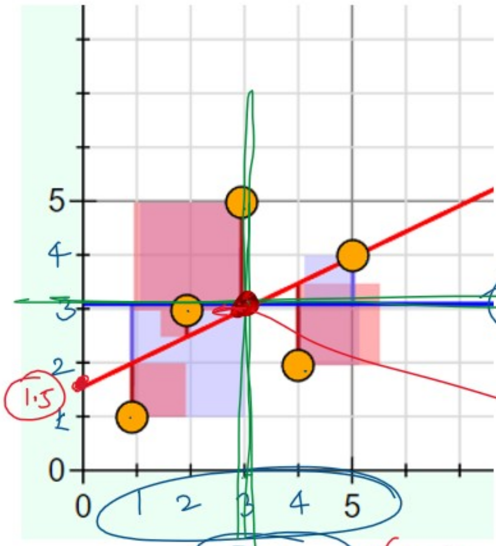
$$\Rightarrow -\bar{Y} + a + b\bar{X} = 0$$

$$\boxed{\Rightarrow a = \bar{Y} - b\bar{X}}$$

$a \downarrow \hat{\beta}_0$, $b \downarrow \hat{\beta}_1 = ??$

$= a + b$

'a' is constant (Y-intercept) is such that the line must go through the mean of 'X' and 'Y'



$\bar{Y} = 3$: (center of Y)

(this data point- center of data cloud.)

$\bar{X} = 3$ → (center of 'X')

$\boxed{a = 1.5}$

https://www.youtube.com/watch?v=4b4MUYve_U8 : **Andrew Ng - DO NOT WATCH NOW -- try it after our course is over**

Second condition for minimizing 'Q' is:

$$\boxed{\frac{\partial Q}{\partial b} = 0}$$

$$Q = \sum_{i=1}^{n} (Y_i - a - bx_i)^2$$

$$\frac{\partial Q}{\partial b} = \sum_{i=1}^{n} 2\,(Y_i - a - bx_i) \times (0 - 0 - x_i)$$

$$\frac{\partial Q}{\partial b} = (-2) \sum_{i=1}^{n} (Y_i - a - bx_i)(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[Y_i x_i - a x_i - b x_i^2 \right] = 0$$

Let us substitute 'a' from $a = \bar{Y} - b\bar{X}$

$$\Rightarrow \sum_{i=1}^{n} \left[Y_i X_i - (\bar{Y} - b\bar{X}) x_i - b x_i^2 \right] = 0$$

$$\Rightarrow \sum^{n} \left[Y_i x_i \quad \bar{Y} x \ldots \right.$$

$$\Rightarrow \sum_{i=1}^{n}\left[\underbrace{Y_i x_i - \bar{Y}x_i} + \underbrace{b\bar{x}x_i - bx_i^2}\right] = 0$$

$$\sum_{i=1}^{n} b = b\sum_{i=1}^{n} 1$$

$b + b + \dots$ n times | $b(1+1+\dots+n)$

(bn) | $b(n)$

bn

$$\Rightarrow \sum_{i=1}^{n}\left(Y_i x_i - \bar{Y}x_i\right) + b\sum_{i=1}^{n}\left(\bar{x}x_i - x_i^2\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(Y_i x_i - \bar{Y}x_i\right) - b\sum_{i=1}^{n}\left(x_i^2 - \bar{x}x_i\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(Y_i x_i - \bar{Y}x_i\right) = b\sum_{i=1}^{n}\left(x_i^2 - \bar{x}x_i\right)$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n}\left[Y_i x_i - \bar{Y}x_i\right]}{\sum_{i=1}^{n}\left[x_i^2 - \bar{x}x_i\right]}$$

sum product$(x_i\ y_i)$

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

$$\Rightarrow b = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{Y}\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i}$$

$(n\bar{x})$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{Y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}$$

Intuitively by using two expressions:

$$\sum_{i=1}^{n}\left(\bar{x}^2 - x_i\bar{x}\right) = 0$$

$$= n\bar{x}^2 - \sum x_i\bar{x}$$

$$= n\bar{x}^2 - \bar{x}\cdot\sum_{i=1}^{n} x_i \rightarrow n\bar{x}$$

$$= \cancel{n\bar{x}^2} - \cancel{n\bar{x}^2}$$

$\bar{x} = k$

$\bar{x}^2 = k^2$

$\sum_{i=1}^{n} k^2 = nk^2$

$n\bar{x}^2$

$\sum_{i=1}^{n} x_i \rightarrow n\bar{x}$

$$= n\bar{x}^2 - \bar{x}\cdot\left(\sum_{i=1}^{n} x_i\right) \rightarrow (n\bar{x})$$

$$\sum_{i=1}^{n} x_i \rightarrow n\bar{x}$$

$$= \cancel{n\bar{x}^2} - \cancel{n\bar{x}^2}$$

$$= 0$$

lly, $\sum_{i=1}^{n} (\bar{y}^2 - y_i\bar{y}) = 0$

$$b = \frac{\sum_{i=1}^{n} (x_i y_i - x_i\bar{y}) + \underbrace{\sum_{i=1}^{n} (\bar{x}\bar{y} - y_i\bar{x})}_{\to 0}}{\sum (x_i^2 - x_i\bar{x}) + \underbrace{\sum_{i=1}^{n} (\bar{x}^2 - x_i\bar{x})}_{\to 0}}$$

reverse

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) / n}{\sum_{i=1}^{n} (x_i - \bar{x})^2 / n}$$

*
$$\boxed{b = \frac{cov(x, y)}{var(x)}}$$

$$\boxed{a = \bar{y} - b\bar{x}}$$

Covariance Formula

For Population

$$Cov(x,y) = \frac{\Sigma\,(x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

For Sample

$$Cov(x,y) = \frac{\Sigma\,(x_i - \bar{x}) * (y_i - \bar{y})}{(N-1)}$$

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

R^2: Coefficient of determination

* It is a statistical measure of how well the regression line approximates the actual data.
* It is a measure about the goodness of fit of a model.

SST = SSR + SSE

$$\Rightarrow \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(\hat{y}_i-y_i)^2$$

Total variability of the dataset (SST)

Variability explained by the regression line (SSR)

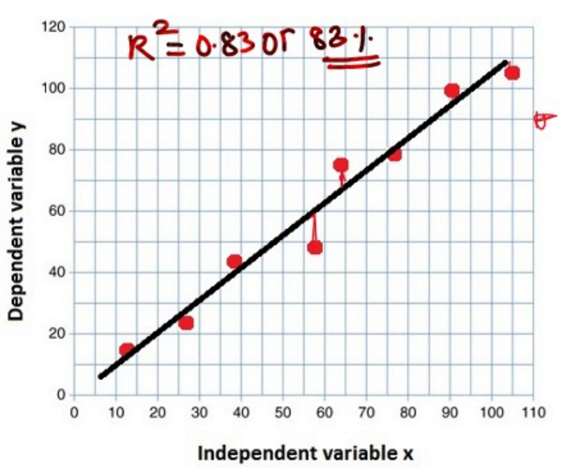
unexplained variability (SSE)

100% = 75% + 25%

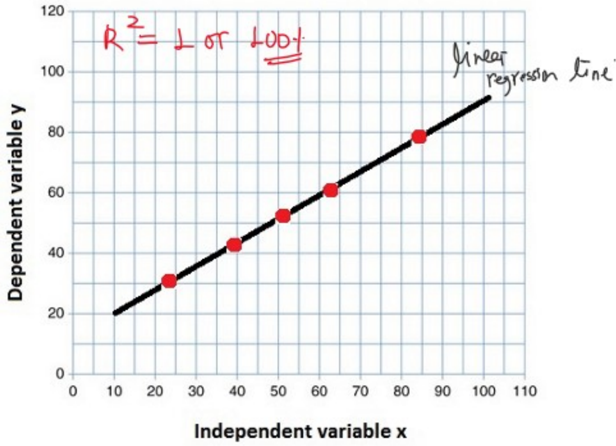
$$R^2 = 1 - \frac{\text{sum squared error (SSE)}}{\text{total sum of squares (SST)}}$$

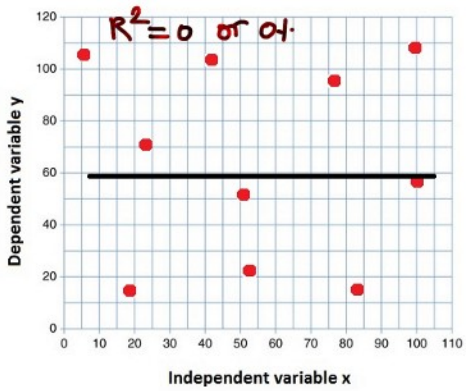
$$R^2 = \frac{\text{sum of squared regression (SSR)}}{\text{total sum of squares (SST)}}$$

Let us say $R^2 = 83\%$ which means that 83% of the variation in 'y' values is accounted for by the x values.



17% of total 'y' variance is unexplained variability.





R-squared or coefficient of determination is a measure that gives proportion of variation in target variable (y) explained by the linear regression model.

$$\uparrow\uparrow R^2 = \left(1 - \frac{SSE\downarrow\downarrow}{SST}\right)$$

Given regression line is getting very close to actual data points, sum of squares error (SSE) decrease and hence R^2 increases

$$\downarrow\downarrow R^2 = \left(1 - \frac{SSE\uparrow\uparrow}{SST}\right)$$

Problem with R^2 statistic

R^2 value never decreases no matter the no. of variables we add to our regression

R^2 value never decreases no matter the no of variables we add to our regression model.

R^2 either remains the same or increases with the addition of new independent variables.

$$\hat{Y} = \hat{a} + \hat{b}x$$

$$\boxed{\hat{Y} = \hat{\beta_0} + \hat{\beta_1}x_1} + \hat{\beta_2}x_2 + \hat{\beta_3}x_3 + \cdot - \hat{\beta_n}x_n \quad \text{— MLR}$$

75%.

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1}x_1 \quad \text{— } 0.75 \text{ or } 75\%$$

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1}\hat{x_1} + \hat{\beta_2}\hat{x_2} \quad \text{— } 85\%$$

R^2

'Inflation in accuracy'

Adjusted R^2:

Unlike the std. R^2, which simply tells you the proportion of variance explained by the model, Adjusted R^2 takes into account the no. of predictors (x_i's), independent variables in the model

$$\text{Adjusted } R^2 = \left\{ 1 - \frac{\left[(1-R^2)(n-1)\right]}{(n-K-1)} \right\}$$

where n: represents the no of data points

K: " " no of variables

R^2: std. R^2 value.

$(101-1) = 100$

$$\left[1 - \frac{(1-R^2\uparrow)(n-1)}{(n-K\uparrow-1)}\right]$$

→ denominator

$R^2 = 0.85$

$(1-R^2) = 0.15$

$(1-R^2)(n-1) = 0.15 \times 100 = 15$

$(n-K-1) = (101-1-1) = 99$

$$\text{Adj. } R^2 = \left(1 - \frac{15}{99}\right)$$

1-(15/99)=0.8485 $\simeq 85\%$.

$\boxed{K=10}$

$$\left(\frac{15}{101-10-1}\right) = \left(\frac{15}{90}\right)$$

$$\left(\frac{15}{101-10-1}\right) = \left(\frac{15}{90}\right)$$

$$\text{Adj } R^2 = \left(1 - \frac{15}{90}\right)$$

1 - (15/90)=0.8333

$\boxed{83.33}$ ↓↓

ovf

97.1 %

OLS Regression Results

Dep. Variable:	marks	R-squared:	0.971
Model:	OLS	Adj. R-squared:	0.967
Method:	Least Squares	F-statistic:	265.0
Date:	Sat, 30 May 2020	Prob (F-statistic):	2.04e-07
Time:	23:41:47	Log-Likelihood:	-17.372
No. Observations:	10	AIC:	38.74
Df Residuals:	8	BIC:	39.35
Df Model:	1		
Covariance Type:	nonrobust		

"random independent variable"

97.1 %

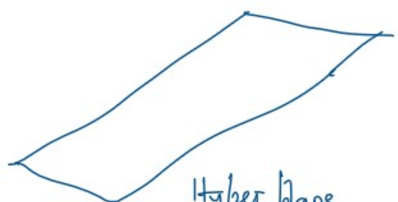
OLS Regression Results

Dep. Variable:	marks	R-squared:	0.971
Model:	OLS	Adj. R-squared:	0.962
Method:	Least Squares	F-statistic:	116.1
Date:	Sat, 30 May 2020	Prob (F-statistic):	4.28e-06
Time:	23:44:34	Log-Likelihood:	-17.364
No. Observations:	10	AIC:	40.73
Df Residuals:	7	BIC:	41.64
Df Model:	2		
Covariance Type:	nonrobust		

96.7 %

$\boxed{96.2\%}$ ↓↓

$$\hat{Y} = 0.25 + 1.5x_1 + 2x_2$$



Hyper plane

{ overfitting vs underfitting } { bias vs variance }

Multicollinearity

- Causes, effects and detection using VIF
  ↳ (variance inflation factor)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

obese weight

BMI

Height

$$x_2 = f(x_1, x_3)$$

$$BMI = f(H, W)$$