Whiteboard Notepad

File Insert About

# Supervised Learning

## Topic: Simple Linear Regression

*By*

*Prof. Dr. Sourav Saha*

Type here to search

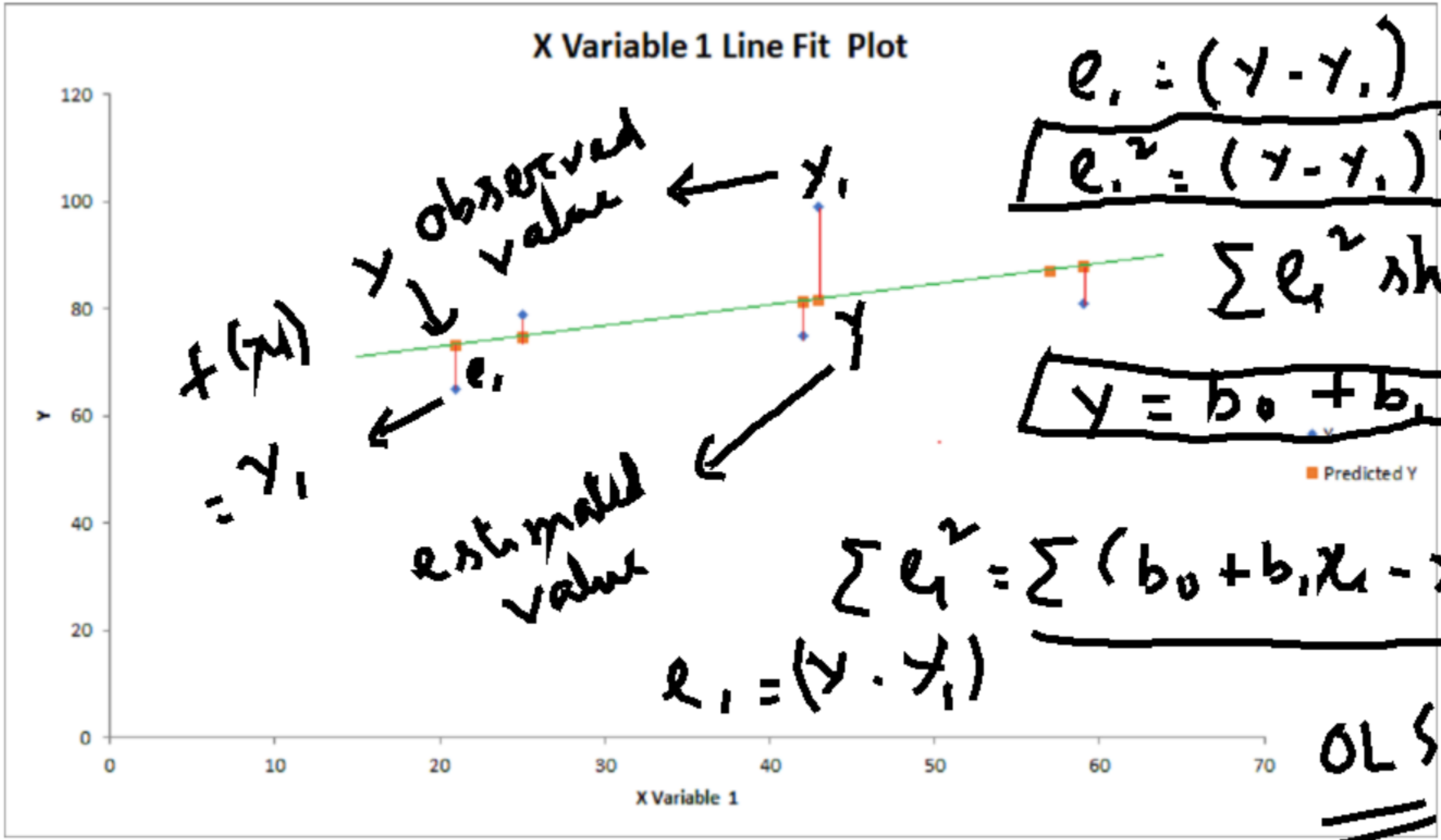
ENG 17:33 30-04-2021

# A simple example of a regression equation to predict the glucose level given the age.

| SUBJECT | AGE X | GLUCOSE LEVEL Y |
|---|---|---|
| 1 | 43 | 99 |
| 2 | 21 | 65 |
| 3 | 25 | 79 |
| 4 | 42 | 75 |
| 5 | 57 | 87 |
| 6 | 59 | 81 |
| 7 | 55 | ? |

→ observed value

$y = f(x)$




$e_1 = (y - y_1)$

$\boxed{e_1^2 = (y - y_1)^2}$

$y_1$ ← y observed value

$f(x)$

$e_1$

$= y_1$

y

estimated value

$\sum e_1^2$ should be min

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

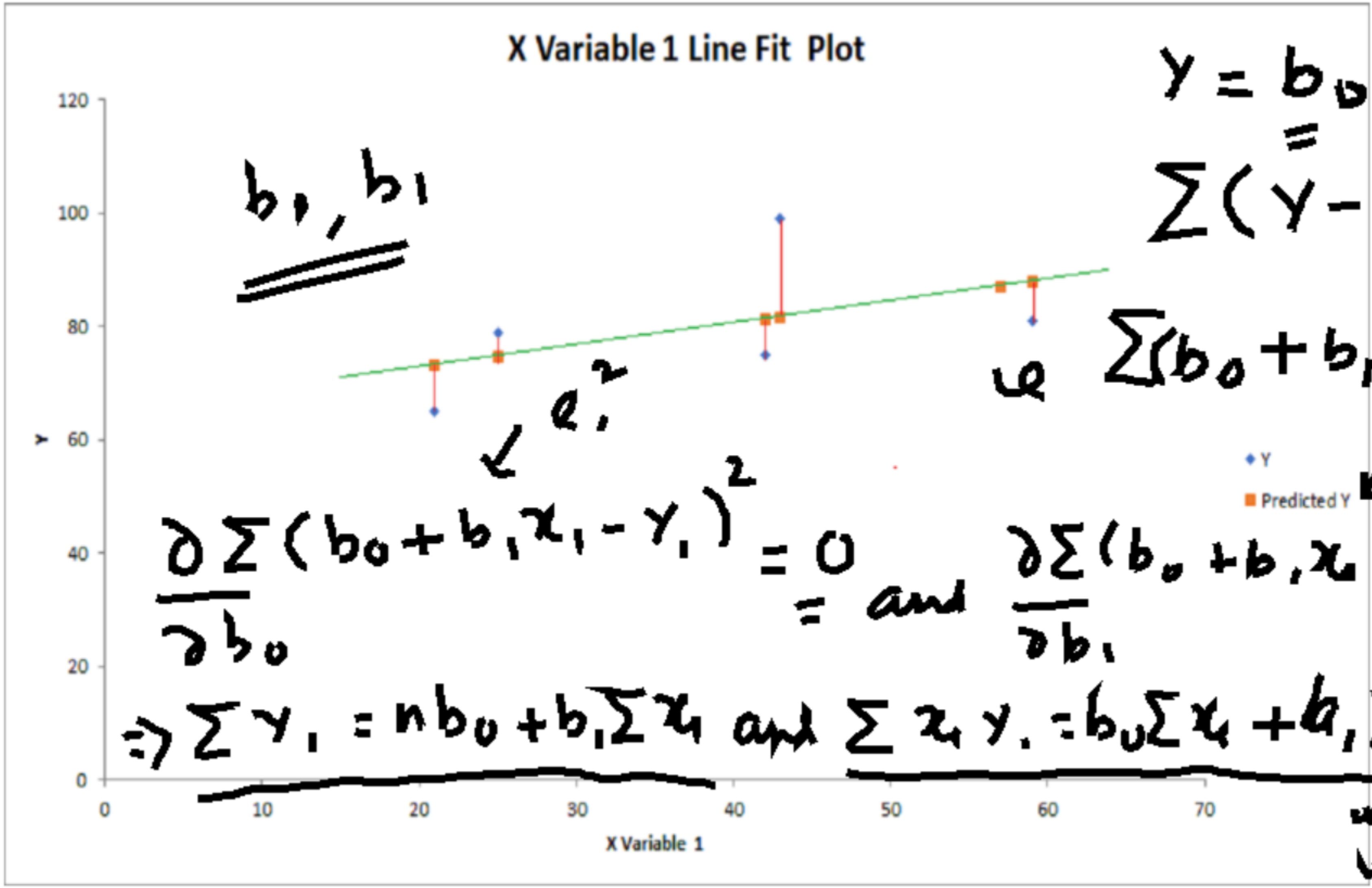
$\sum e_1^2 = \underline{\sum (b_0 + b_1 x_1 - y_1)^2}$ should be min

$e_1 = (y - y_1)$

OLS

Whiteboard Notepad

File Insert About




$b_0 , b_1$

$Y = b_0 + b_1 x$

$\sum (Y - Y_1)^2$ is min

ie $\sum (b_0 + b_1 x_1 - Y_1)^2$ min

$e_1^2$

$e_1^2$

$\frac{\partial \sum (b_0 + b_1 x_1 - Y_1)^2}{\partial b_0} = 0$ and $\frac{\partial \sum (b_0 + b_1 x_1 - Y_1)^2}{\partial b_1} = 0$

$\Rightarrow \sum Y_1 = n b_0 + b_1 \sum x_1$ and $\sum x_1 Y_1 = b_0 \sum x_1 + b_1 \sum x_1^2$

$b_0 = ?$

$b_1 = ?$

Type here to search

ENG 17:33 30-04-2021

# A simple example of a regression equation to predict the glucose level given the age.

$$b_0 = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad \text{(i)}$$

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad \text{(ii)}$$

$y = b_0 + b_1 x$

$\sum x_i y_i$

$\sum x_i$

$\sum y_i$

$\sum x_i^2$

Whiteboard Notepad

File Insert About

# A simple example of a regression equation to predict the glucose level given the age.

**Step 1:** *Make a chart of your data, filling in the columns in the same way as you would fill in the chart if you were finding the Pearson's Correlation Coefficient*

| SUBJECT | AGE X | GLUCOSE LEVEL Y | XY | $X^2$ | $Y^2$ |
|---|---|---|---|---|---|
| 1 | 43 | 99 | 4257 | 1849 | 9801 |
| 2 | 21 | 65 | 1365 | 441 | 4225 |
| 3 | 25 | 79 | 1975 | 625 | 6241 |
| 4 | 42 | 75 | 3150 | 1764 | 5625 |
| 5 | 57 | 87 | 4959 | 3249 | 7569 |
| 6 | 59 | 81 | 4779 | 3481 | 6561 |
| Σ | **247** | **486** | **20485** | **11409** | **40022** |

Type here to search

ENG 17:34 30-04-2021

# A simple example of a regression equation to predict the glucose level given the age.

**Find $b_1$:**

$$b_1 = \frac{n\,(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b_1 = \frac{6\,(20485) - (247)(486)}{6(11409) - (247)^2}$$

$$b_1 = \frac{2868}{7445} = 0.385335$$

# A simple example of a regression equation to predict the glucose level given the age.

- **Step 3:** *Insert the values into the equation.*

$$y' = b_0 + b_1 x$$

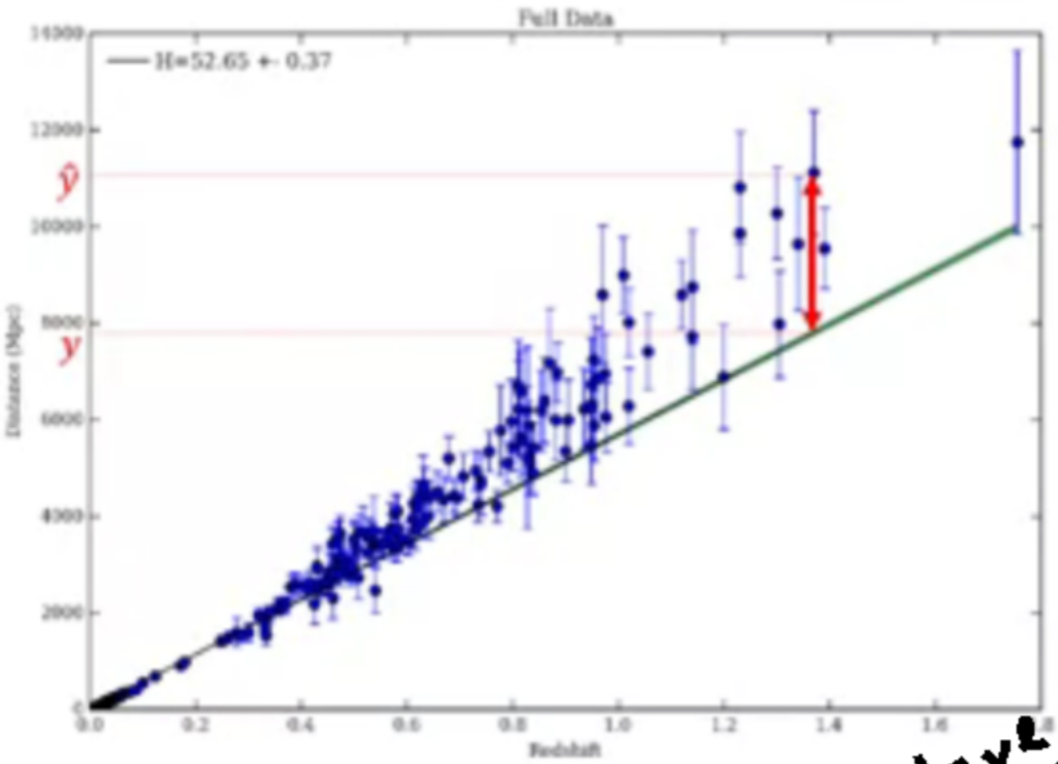
$$\boxed{y' = 65.14 + 0.385225x}$$

- **Step 4:** *Prediction – the value of y for the given value of x = 55*

$$y' = 65.14 + (0.385225 * 55)$$
$$y' = 86.327$$

**Hence, the glucose level for the given age 55 is 86.327**

# What is an error of the model? Evaluation?




$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

obs
pred. (ŷ) value

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

$$RAE = \frac{\sum_{j=1}^{n}|y_j - \hat{y}_j|}{\sum_{j=1}^{n}|y_j - \bar{y}|}$$

obs

$$\boxed{RSE} = \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$$

Relative Square Error

$$\boxed{R^2 = 1 - RSE}$$

higher value indicates better accuracy

Whiteboard Notepad

File Insert About

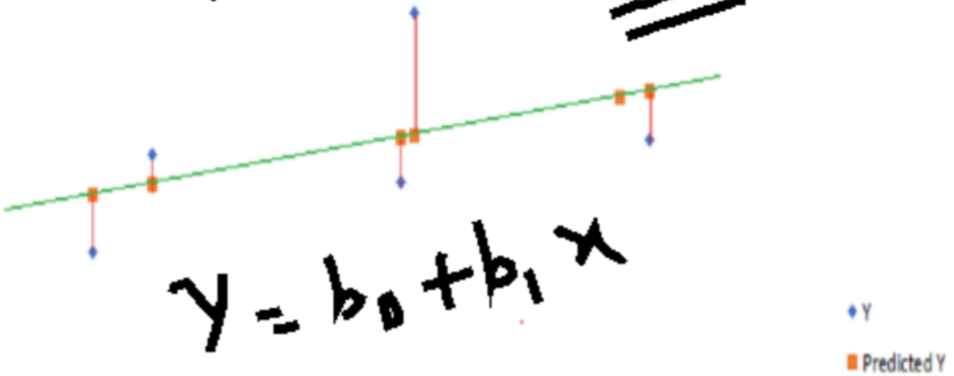
| A ($x$) | B ($y$) | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 43 | 99 | 81.70625 | 17.2937542 | 299.0739342 | 324 | | |
| 21 | 65 | 73.2313 | -8.231296172 | 67.75423667 | 256 | | |
| 25 | 79 | 74.7722 | 4.227803895 | 17.87432578 | 4 | | |
| 42 | 75 | 81.32102 | -6.321020819 | 39.9553042 | 36 | | |
| 57 | 87 | 87.0994 | -0.099395567 | 0.009879479 | 36 | | |
| 59 | 81 | 87.86985 | -6.869845534 | 47.19477766 | 0 | | |
| | 81 | | | 471.862458 | 656 | | |
| | | | | | | 0.71930253 | 0.280697 |

$\bar{y}$  $y$  $y - y_1$  $(y-y_1)^2 = e_1^2$  $(y-\bar{y})^2$

$$\text{RSE} = \frac{\sum (y-y_1)^2}{\sum (y-\bar{y})^2}$$

$R^2$: R-squared $= 1 - \text{RSE}$

$\boxed{28\%}$




$$y = b_0 + b_1 x$$

$> 80\% \rightarrow$ good

$< 80\% \rightarrow$ bad

Type here to search

ENG 17:35 30-04-2021

# THANK YOU