Simple Linear Regression by hand step by step

Overview
We are going to predict the Y values depends upon the given X values using simple linear regression model. Using the formula we are going to find Linear Equation and calculate following values using step by step.

Formula’s required for Simple Linear Regression single values

- Linear Equation $Y = W_0 + W_1 X$
- Slope of Regression Line $W_1 = r * \frac{S_y}{S_x}$
- Y Intercept of Regression Line $W_0 = \bar{Y} - W_1 * \bar{X}$
- R Pearson Correlation Coefficient r = $\frac{\sum((X-\bar{X})*(Y-\bar{Y}))}{\sqrt{\sum(X-\bar{X})^2 * \sum(Y-\bar{Y})^2}}$
- Standard Deviation of Y $S_y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}}$
- Standard Deviation of X $S_x = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$
- X Mean $\bar{X} = \frac{sum(X)}{n}$
- Y Mean $\bar{Y} = \frac{sum(Y)}{n}$

Given Data

X	Y
110	179
113	172
119	200
119	216
120	185
120	188
123	200
124	180
125	240
127	220
133	245
137	265
140	255

Step 1: Find Necessary Values required for S_x, S_Y and r

Required Formula’s

- $\bar{X} = \frac{sum(X)}{n}$
- $\bar{Y} = \frac{sum(Y)}{n}$

- n value = 13
- n-1 value = 1
- X Mean $\bar{X}$= $\frac{sum(X)}{n}$ = $\frac{1610}{13}$ =123.85
- Y Mean $\bar{Y}$= $\frac{sum(Y)}{n}$ = $\frac{2745}{13}$ = 211.15
- $(X - \bar{X}), (X - \bar{X})^2$, $(Y - \bar{Y}), (Y - \bar{Y})^2$, $(X * \bar{X}) * (Y * \bar{Y})$

X	**Y**	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X * \bar{X}) * (Y * \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
110	179	-13.85	-32.15	445.21	191.72	1033.87
113	172	-10.85	-39.15	424.67	117.64	1533.02
119	200	-4.85	-11.15	54.05	23.49	124.41
119	216	-4.85	4.85	-23.49	23.49	23.49
120	185	-3.85	-26.15	100.59	14.79	684.02
120	188	-3.85	-23.15	89.05	14.79	536.10
123	200	-0.85	-11.15	9.44	0.72	124.41
124	180	0.15	-31.15	-4.79	0.02	970.56
125	240	1.15	28.85	33.28	1.33	832.10
127	220	3.15	8.85	27.90	9.95	78.25
133	245	9.15	33.85	309.82	83.79	1145.56
137	265	13.15	53.85	708.28	173.02	2899.41
140	255	16.15	43.85	708.28	260.95	1922.49
123	**211.15**			**2882.31**	**915.69**	**11907.69**

- $\sum((X - \bar{X}) * (Y - \bar{Y})) = 2882.31$
- $\sum(X - \bar{X})^2 = 915.69$
- $\sum(Y - \bar{Y})^2 = 11907.69$

Step 2: Find standard Deviation of X, Y and Slope of Regression Line

- Standard Deviation of Y $S_y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}}$
  - $\sum(Y - \bar{Y})^2 = 11907.69, n - 1 = 12$

- $S_y = \sqrt{\frac{11907.69}{12}}$
- $S_y = 9.094$

- Standard Deviation of X $S_x = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$
  - $\sum(X - \bar{X})^2 = 915.69, n - 1 = 12$
  - $S_x = \sqrt{\frac{915.69}{12}}$
  - $S_x = 2.522$

- Pearson correlation coefficient r = $\frac{\sum((X-\bar{X})*(Y-\bar{Y}))}{\sqrt{\sum(X*\bar{X})^2 * \sum(Y*\bar{Y})^2}}$
  - $\sum((X - \bar{X}) * (Y - \bar{Y})) = 2882.31$
  - $\sum(Y - \bar{Y})^2 = 11907.69$
  - $\sum(X - \bar{X})^2 = 915.69$
  - r = $\frac{2882.31}{\sqrt{915.69 * 11907.69}}$
  - r = 0.87

Step 3 Find the Slope of Regression Line W_1 = $r * \frac{S_y}{S_x}$

- Correlation Coefficient r = 0.87
- Standard Deviation of X $S_x = 2.522$
- Standard Deviation of Y $S_y = 9.094$
- Slope of Regression Line $W_1 = r * \frac{S_y}{S_x}$
- $W_1 = 0.87 * \frac{9.094}{2.522}$
- $\boldsymbol{Slope\ of Regression\ Line\ W_1 = 3.15}$

Step 4 Intercept of Regression Line W_0= $\bar{Y}$ - $W_1 * \bar{X}$

- X Mean $\bar{X}$= 123.85
- Y Mean $\bar{Y}$= 211.15
- $Slope\ of Regression\ Line\ W_1 = 3.15$
- Intercept of Regression Line W_0 = 211.15 -3.15 * 123.85
- W_0 = 211.15 -3.15 * 123.85
- ***Intercept of Regression Line W_0 = -178.98***

Step 5 Find Linear Equation $\boldsymbol{Y = W_0 + W_1 X}$

- W_0 = -13.744
- $Slope\ of Regression\ Line\ W_1 = 3.15$
- *Intercept of Regression Line* W_0 = -178.98
- Linear Equation $Y = W_0 + W_1 X$
- **Linear Equation $\boldsymbol{Y = -178.98 + 3.15 * X}$**