Simple Linear Regression Model

A simple linear regression explains the relationship between an independent and dependant value, which consists of a straight line. A scatter plot is created to asses the direction of the line. The slope can be upwards (positive) or downwards (negative). We begin by finding the equation.

Prediction Line Equation:

$\hat{Y}_i = \beta_0 + \beta_1 X_1$

Where:

β_0 = Y Intercept for the population

- Represents the mean value of Y when X is zero

β_1 = slope for the population

- Represents the expected change in Y per unit change in X

X_1 = independent variable for observation *i*

Y_i = dependant variable for observation *i*

Example:

Square Feet (thousands): X	Sales (millions of dollars): Y
1.7	3.7
1.6	3.9
2.8	6.7
5.6	9.5
1.3	3.4
2.2	5.6
1.3	3.7
1.1	2.7
3.2	5.5
1.5	2.9
5.2	10.7
4.6	7.6
5.8	11.8
3.0	4.1

Calculator Instructions:

CALC → SET:

2Var XList: List 1 (whichever list contains your X values)

2Var Freq: List 2 (whichever list contains your Y values)

EXE

REG → X → "aX + b" or "a + bX"

For aX+b: a is the slope and b is the y-intercept

For a+bX: a is the y-intercept and b is the slope

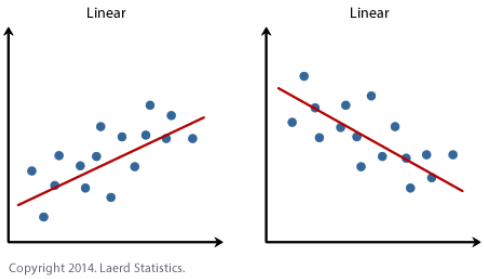
Answer: *"aX+b"* → $\hat{y} = 1.6699x + 0.9645$

"a+bX" → $\hat{y} = 0.9645 + 1.6699x$

TRSM Academic Success Centre | "Train to Learn Effectively"
Visit us at TRS 2-168 | T: (416) 979-5000 x2435 | E: trsm.academicsuccess@ryerson.ca
W: www.ryerson.ca/tedrogersschool/success

$\beta_1 > 0$ → positive relationship between x and y so (x ↑, y↑)
$\beta_1 < 0$ → negative relationship between x and y so (x ↑, y↓)

Interpretation: When the x is zero, the y-intercept is 0.9645. Since $\beta_1 > 0$, it indicates that there is positive relationship between x and y so as x goes up by 1, y goes up by 1.6699.



Coefficient of Correlation and Determination:

- **Coefficient of correlation = r**
  - It is used to measure the strength of association between two variables
- **Coefficient of determination =** r^2
  - the % of variation in the dependent (y) variable is explained by its relationship to the independent (x) variable in the regression model

Testing Linear Regression:

- We are now testing to see whether there is evidence of a linear relationship between x and y

  Ho: $B_1 = 0$
  Ha: $B_1 \neq 0$

Calculator Instructions:

Test: TEST → T → REG → ≠ → XLIST: List 1 and YLIST: List 2 → EXE
Critical Value: DIST → T → InvT → "Variable" → Area: α or α/2

Example: Centre Tail Test, with a confidence level of 95%
Area: 1–0.95/2 = 0.025
df = n – 2: (n=14 → df = 12)

Analysis:

Reject Ho and conclude that there <u>is</u> a linear relationship between x and y
Do not reject Ho and conclude that there <u>is no</u> linear relationship between x and y

Equations:

$b1 = SSXY / SSX$
$SSXY = (XY\ Total) - [(X\ Total * Y\ Total) / n]$
$SSX = X^2\ Total - [(X\ total)^2 / n]$
$b0 = \bar{y} - b1\ \bar{x}$
$\bar{y} = y\ total / n$
$\bar{x} = x\ total / n$
$r^2 = SSR / SST$
$SST = SSR + SSE$
$SST = Y^2\ Total - [(Y\ Total)^2 / n]$
$SSR = (Y\ Total * b_0) + (XY\ Total * b_1) - [(Y\ Total)^2 / n]$
$SSE = Y^2\ Total - (Y\ Total * b_0) - (XY\ Total * b_1)$
$S_{XY} = \sqrt{}\ SSE / (n-2)$
$e = y - \hat{y}$

TRSM Academic Success Centre | "Train to Learn Effectively"
Visit us at TRS 2-168 | T: (416) 979-5000 x2435 | E: trsm.academicsuccess@ryerson.ca
W: www.ryerson.ca/tedrogersschool/success

Multiple Regression Model

Multiple Regression has more than two intendent (X) values, whereas Linear Regression has just one. For both, there is only one dependent (Y) value.

Equation:

$$\mathbf{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots}$$

Where:

β_0 = Y Intercept
β_1 = slope of Y with variable X_1, holding variable X_2 constant
β_2 = slope of Y with variable X_2, holding variable X_1 constant
E_i = random error in Y for observation i

SPSS Output:

Coefficients

Model	Unstandardized coefficient		~~Standardized coefficient~~	T	Sig.
	B	Std. Error	~~Beta~~		
1 (Constant)	58.157	~~2.658~~		~~21.878~~	~~0.000~~
Horsepower	-0.118	~~0.033~~	~~-0.392~~	~~-3.600~~	~~0.001~~
WEIGHT	-0.0069	~~0.001~~	~~-0.534~~	~~-4.903~~	~~0.000~~

Equation: $\hat{Y} = 58.157 + (-0.118)\, X_1 + (-0.0069)\, X_2$

Interpretation: Holding X_2 constant, X_1 goes up by one, and $\hat{Y}$ goes down by 0.118
Holding X_1 constant, X_2 goes up by one, and $\hat{Y}$ goes down by 0.0069

Coefficient of Determination

Model	R	R Square	Adjusted R Square	Std. Error. Of the Estimate
1	0.866[a]	0.749	0.832	4.1777

Interpretation: 74.9% of variation of Y is explained by X_1 and X_2 and the remaining 25.1 is not explained by either variable

Which model is the optimal (most efficient)?

- The Model with the highest Adjusted R^2 is the most optimal

ANOVA

Sources	df	Sum of sq.	Mean of sq. (var)	F STAT	Sig.
Regression	**k** # of x's	SSR	MSR = $\frac{SSR}{k}$	F = MSR/MSE	P-value
Error	**n – k – 1**	SSE	MSE = $\frac{SSE}{N - k - 1}$		
Total	**N – 1**	SST			

TRSM Academic Success Centre | "Train to Learn Effectively"
Visit us at TRS 2-168 | T: (416) 979-5000 x2435 | E: trsm.academicsuccess@ryerson.ca
W: www.ryerson.ca/tedrogersschool/success