Online-Class 18-3-2021

Probability, Statistics and Reliability (MAT3003)

SLOT: B21 + B22 + B23

MODULE - 3

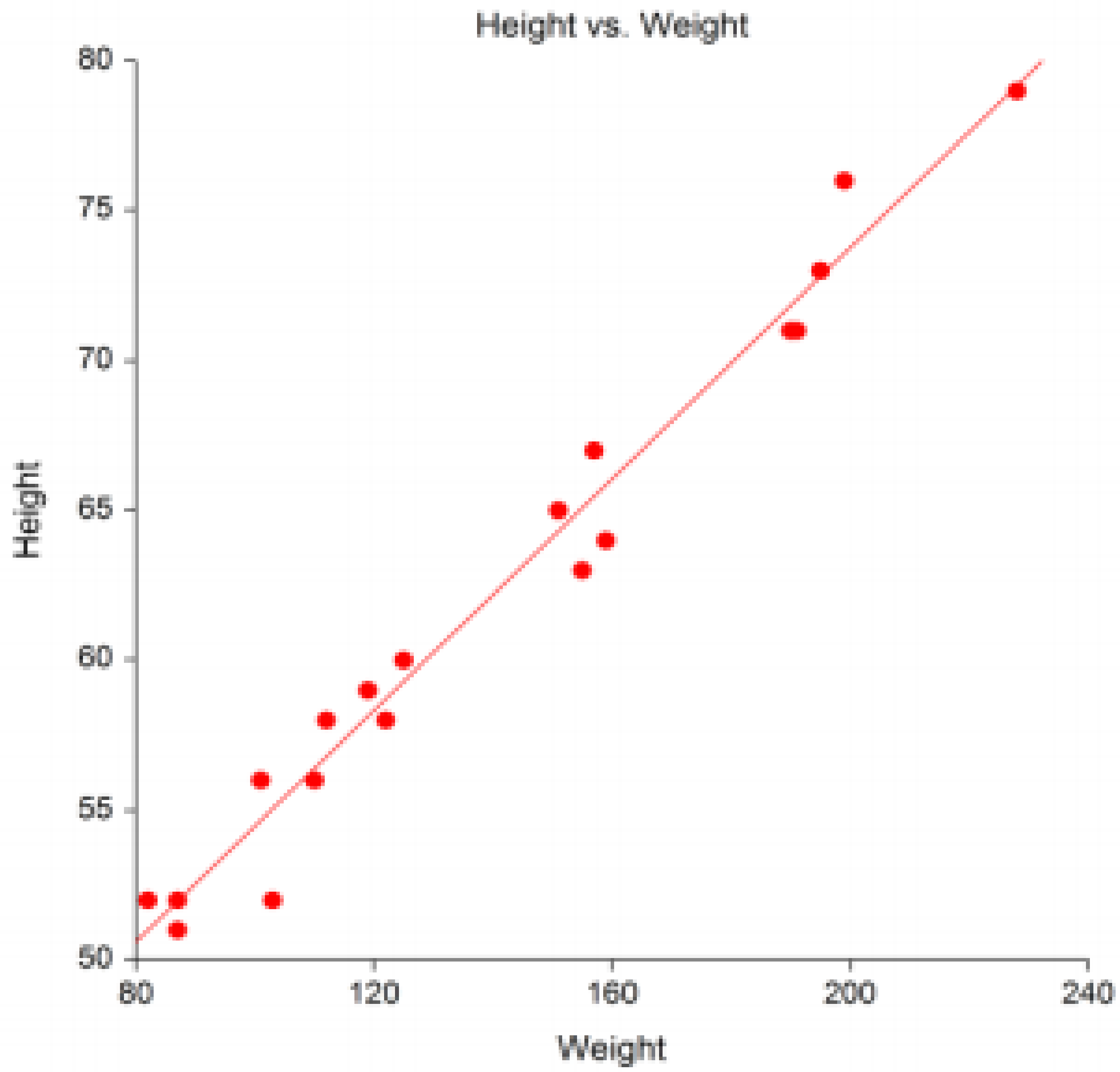
Topic: Regression Analysis

Reference: Probability and Statistics for Engineers & Scientists
by R.E.Walpole, Pearson (2012).

Contents

What is Regression?

- $Definition$: A regression model is a mathematical equation that describes the relationship between two or more variables.
- When the random variables X and Y are $linearly$ correlated, the points plotted on the scatter diagram, corresponding to n pairs of observed values of X and Y, will have a tendency to cluster round a straight line. This straight is called the regression line.
- The regression line can be taken as the best fitting straight line for the observed pairs of values of X and Y in the least square sense.



Reference: Probability and Statistics for Engineers & Scientists
by R.E.Walpole, Pearson (2012).

Equation of the Regression Line of Y on X

- The regression line of Y on X is the best-fitting straight line for the observed pairs of values $(x_1,\ y_1)$, $(x_2,\ y_2)$,..., $(x_n,\ y_n)$, based on the assumption that x is the independent variable and y is the dependent variable.

- Hence, let the equation of the regression line of Y on X be assumed as

$$y = ax + b.$$

Equation of the Regression Line of X on Y

- Equation of the regression line of X on Y be assumed as

$$x = ay + b.$$

Types of Regression Model:

- Two Types
- Simple Regression Model: includes only two variables – one dependent and one independent.

- Multiple Regression: includes more than two variables – one dependent and two or more independent variables.

Simple vs. Multiple Linear Regression

- **Simple Linear Regression – one independent variable.**

$$y = b_0 + b_1x_1$$

- **Multiple Linear Regression – multiple independent variables.**

$$y = b_0 + b_1x_1 + b_2x_2 \dots + b_nx_n$$

2nd independent variable and weight (coefficient)

nth independent variable and weight (coefficient)

Reference: Text Book: 1.Probability and Statistics for Engineers & Scientists by R.E.Walpole, Pearson (2012).

Simple Regression

Model-1: Straight Line Fit (or Linear Regression)

Fitting a straight line by the method of least squares.

Model-2: Quadratic or Second Degree or Parabolic Regression (Non-linear Regression)

Fitting a second degree curve by the method of least squares.

Normal Equations by the Method of Least Square

Let the Equation of the straight line be

$$y = a + bx \qquad (1)$$

Applying $\sum$ on both sides of $y = a + bx$, the First Normal Equation is given by:

$$\sum y = a.n + b\sum x \qquad (2)$$

Applying $\sum$ on both sides after multiplying by x both sides of y = ax + b,

the Second Normal Equation is given by:

$$\sum xy = a\sum x + b\sum x^2 \qquad (3)$$

Reference: Text Book: 1.Probability and Statistics for Engineers & Scientists by R.E.Walpole, Pearson (2012).

Model-1: Fitting a straight line by the method of least squares

A straight line can be fitted to the given data by the method of least squares. The equation of a straight line or least square line is $Y = a + bX$, where a and b are constants or unknowns.

To compute the values of these constants we need as many equations as the number of constants in the equation. These equations are called normal equations. In a straight line there are two constants a and b so we require two normal equations.

Normal Equation for '*a*' $\quad \sum Y = na + b\sum X$

Normal Equation for '*b*' $\quad \sum XY = a\sum X + b\sum X^2$

The direct formula of finding a and b is written as

$$b = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}, \qquad a = \bar{Y} - b\bar{X}$$

Finally, substituting values of a and b in straight line equation,

$$\boldsymbol{Y = (\ldots) + (\ldots)X}$$

which is the required equation.

Note: $\bar{X} = \frac{\sum X}{n}$, and $\bar{Y} = \frac{\sum Y}{n}$ are the average values of X and Y data.

Question 1 (Fitting a straight line)

Fit a straight line by using the method of least squares, for the following data:

x	5	3	2	4	9
y	24	15	12	18	55

Solution

Let the equation of straight line be:

$$y = a + bx \qquad (1)$$

where a and b are constants.

By the method of least squares, we write the Normal equations for Equation (1), as follows:

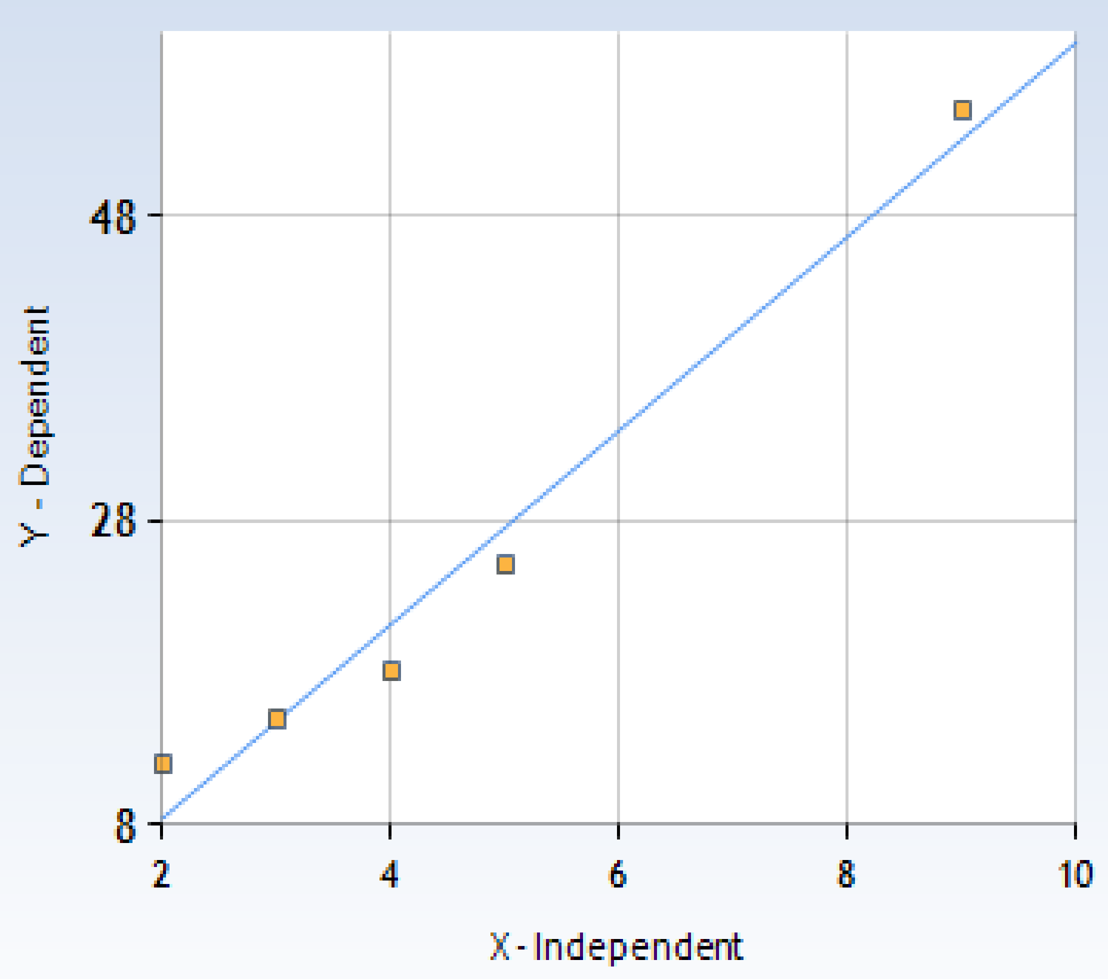
$$\sum y = an + b\sum x \qquad (2)$$

$$\sum xy = a\sum x + b\sum x^2 \qquad (3)$$

where n is the number of data points.

Now, we determine the values of $\sum x$, $\sum y$, $\sum xy$ and $\sum x^2$ with the help of given data, by preparing a Table as follows:

x	y	xy	x^2
5	24	120	25
3	15	45	9
2	12	24	4
4	18	72	16
9	55	495	81
$\sum x$=23	$\sum y$=124	$\sum xy$=756	$\sum x^2$=135



From the table, we have

$$\sum x = 23\,,\ \sum y = 124\,, \sum xy = 756 \quad \text{and} \quad \sum x^2 = 135$$

Substituting these values in Normal Equations (2) & (3), and then solving these Equations for a and b.

$$\mathrm{a} = -4.30\ldots\ldots$$

$$\mathrm{b} = +6.3698$$

Substituting values of a and b in Equation (1),

$$\mathbf{y = -4.30 + 6.3698}\boldsymbol{x}$$

which is the required equation of straight line.

Question 2 (For Students)

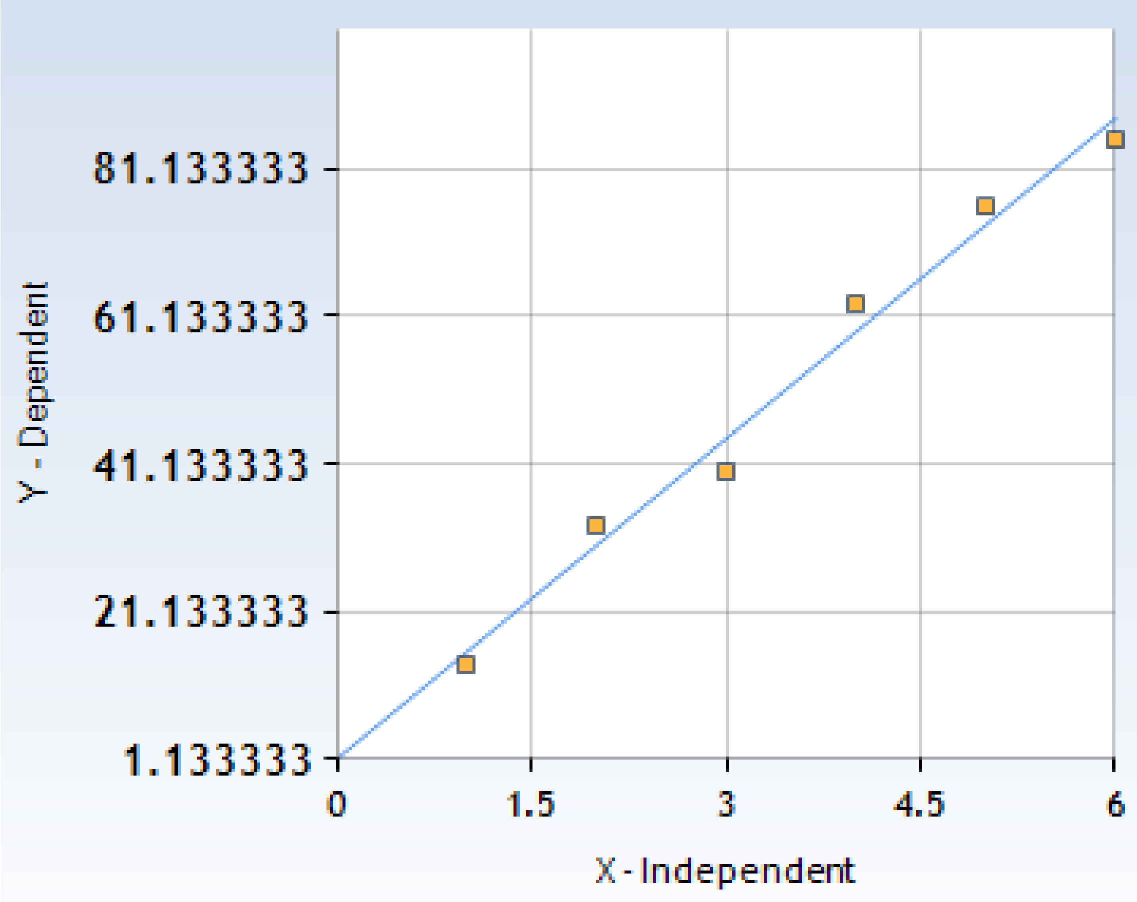
In the following table, X is the tensile force applied to a steel specimen in thousands of pounds, and Y is the resulting elongation in thousands of an inch:

X:	1	2	3	4	5	6
Y:	14	33	40	63	76	85

a) Plot the data to verify that it is reasonable to assume that the regression of Y on X is linear.

b) Find the equation of the least squares line, and use it to predict the elongation when the tensile force is 3.5 thousand pounds.

Solution



$$a = 1.13333, \; b = 14.48571$$

$$y = 1.133333 + 14.48571x \qquad \text{Ans.}$$

Practice Questions

1. Fit a straight line y = a + bx using the following data

x	5	4	3	2	1
y	1	2	3	4	5

2. Fit a straight line y = a + bx using the following data

x	3	5	7	9	11
y	2.3	2.6	2.8	3.2	3.5

3. Fit a straight line to the following data on production.

Year	1996	1997	1998	1999	2000
Production	40	50	62	58	60

THANK YOU