Plant Archives Vol. 19 No. 2, 2019 pp. 3685-3688 **e-ISSN:2581-6063 (online), ISSN:0972-5210**

APPLICATION OF SIMPLE LINEAR REGRESSION ANALYSIS PROCEDURE FOR POTATO YIELD

V. Sekhar[1], V. Srinivasa Rao[2], K. Umakrishna[1] and T. Thomson[3]

[1]Department of Statistics, College of Horticulture, Venkataramannagudem -534 101, Dr. YSR Horticultural University, West Godavari (District), Andhra Pradesh, India.

[2]Department of Statistics, Agricultural College, Bapatla - 522 101, Acharya NG Ranga Agricultural University, Guntur (District), Andhra Pradesh, India.

[3]College of Horticulture, Venkataramannagudem - 534 101, Dr. YSR Horticultural University, West Godavari (District), Andhra Pradesh, India.

E-mail : sekhar30mscstat@gmail.com, vs_raoin@yahoo.co.in, umakrishnastat@gmail.com, thomsonmonu@gmail.com

Abstract

This paper was attempted to estimating the potato yield through fitting of simple linear regression model. The data on potato yield (in yield/plot) was collected at COH, Venkataramannagudem. This article explains us when we conduct simple linear regression and what is the advantage of simple linear regression.

Key words : Simple Linear Regression, R^2, Multiple R.

Introduction

The term regression was introduced by Sir Francis Galton. Simple linear regression is a statistical method that allows us to review and study relationships between two continuous (quantitative) variables: One variable, denoted X, is independent variable. The other variable denoted Y is dependent variable.

Regression is also applicable for following

- Farmer income and expenditure on crops
- Yield of a crop and quantity of fertilizer applied

Materials and Methods

The statistical model of simple linear regression is $Y = a + bX + \varepsilon$

Where,

Y is dependent variable

X is independent variable

a is intercept, b is slope or regression coefficient

a and b are also called as parameters

ε is error term (ε = observed value – predicted value)

R-square is a statistical measure of how near the data are to the fitted regression line.

Since R^2 is a proportion, it is always a number between 0 and 1.

If $R^2 = 1$, all of the data points fall perfectly on the regression line.

If $R^2 = 0$, the estimated regression line is perfectly horizontal.

$$R^2 = \frac{Regression\ SS}{Total\ SS}$$

Analysis procedure by manual

To analyze data prepare are given in table 2.

Fitting of regression equation $Y = a + bX + \varepsilon$

$$b = \frac{Cov(X, Y)}{Var(X)} = \frac{\sum X_i Y_i - \frac{\Sigma X_i \Sigma Y_i}{n}}{\left(\sum X_i^2 - \frac{(\Sigma X_i)^2}{n}\right)}$$

$$= \frac{121215 - \frac{(598)(2008)}{10}}{\left(35922 - \frac{(598)^2}{10}\right)} = \frac{121215 - \frac{1200784}{10}}{\left(35922 - \frac{357604}{10}\right)}$$

$$= \frac{121215 - 120078.4}{(35922 - 35760.4)}$$

$$= \frac{1137}{162} = \mathbf{7.0334}$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{2008}{10} = 200.8$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{598}{10} = 59.8$$

$$a = \overline{Y} - b\overline{X}$$

$$a = 200.8 - (7.0334)59.8$$

$$= 200.8 - 420.5973 = \mathbf{-219.7973}$$

Test of Significance of Regression coefficient (b)

$\boldsymbol{H_0}$: The population Regression coefficient $\beta = 0$

Statistic

$$t = \frac{|b|}{SE(b)} = \frac{|b|}{\sqrt{\frac{Var(Y) - b^2 Var(x)}{(n-2)Var(x)}}}$$

$$= \frac{|7.0334|}{\sqrt{\frac{87.4960 - (7.0334)^2 162}{(10-2)162}}}$$

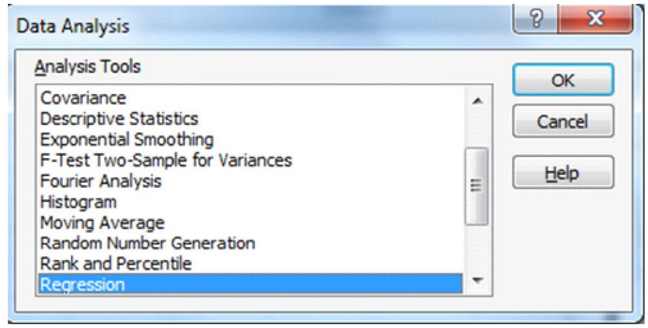
$$= \frac{|7.0334|}{\sqrt{\frac{87.4960 - 49.4687x162}{(8)162}}} = \frac{|7.0334|}{\sqrt{\frac{87.4960 - 8013.9319}{1296}}}$$

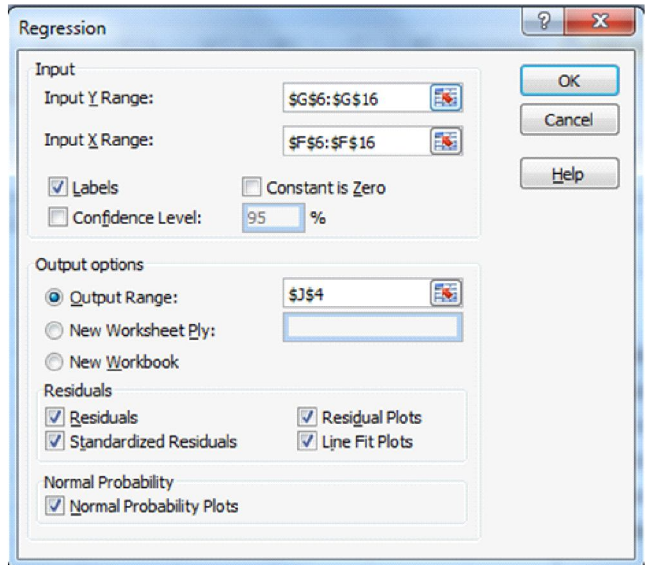
$$= \frac{|7.0334|}{\sqrt{\frac{735.6681}{1296}}} = \frac{|7.0334|}{\sqrt{0.5676}} = \frac{|7.0334|}{0.7534}$$

$$= 9.3355^{**}$$

t-tabulated value at 5% level of significance with (n-2) = (10-2) = 8 d.f. is 2.3060

t-tabulated value at 1% level of significance with 8 d.f. is 3.3554





Analysis procedure by MS-Excel

To perform regression analysis by using the Data Analysis add-in, do the following :

1. Choose Data Analysis command button on the Data tab.
2. When Excel displays the Data Analysis dialog box, select the Regression tool from the Analysis Tools list and then click OK (fig. 1).
3. Give Input Y Range, Input Y Range, select Labels, select output range, give output range (single cell), choose required check boxes and then click OK (fig. 2).
4. The Excel gives results in fig. 3.

Results and Discussion

The predicted or estimated regression equation is $\boldsymbol{Y = -219.7973 + (7.0334)X}$ because of a = -219.7973 and b = 7.0334

t-calculated value 9.3355 is greater than t-tabulated value 2.3060 and also greater than 3.3554. So, we reject our Null Hypothesis. *i.e.* The population Regression coefficient $\beta \neq 0$ *i.e.* significant.

In MS-Excel, by using data analysis option, we will

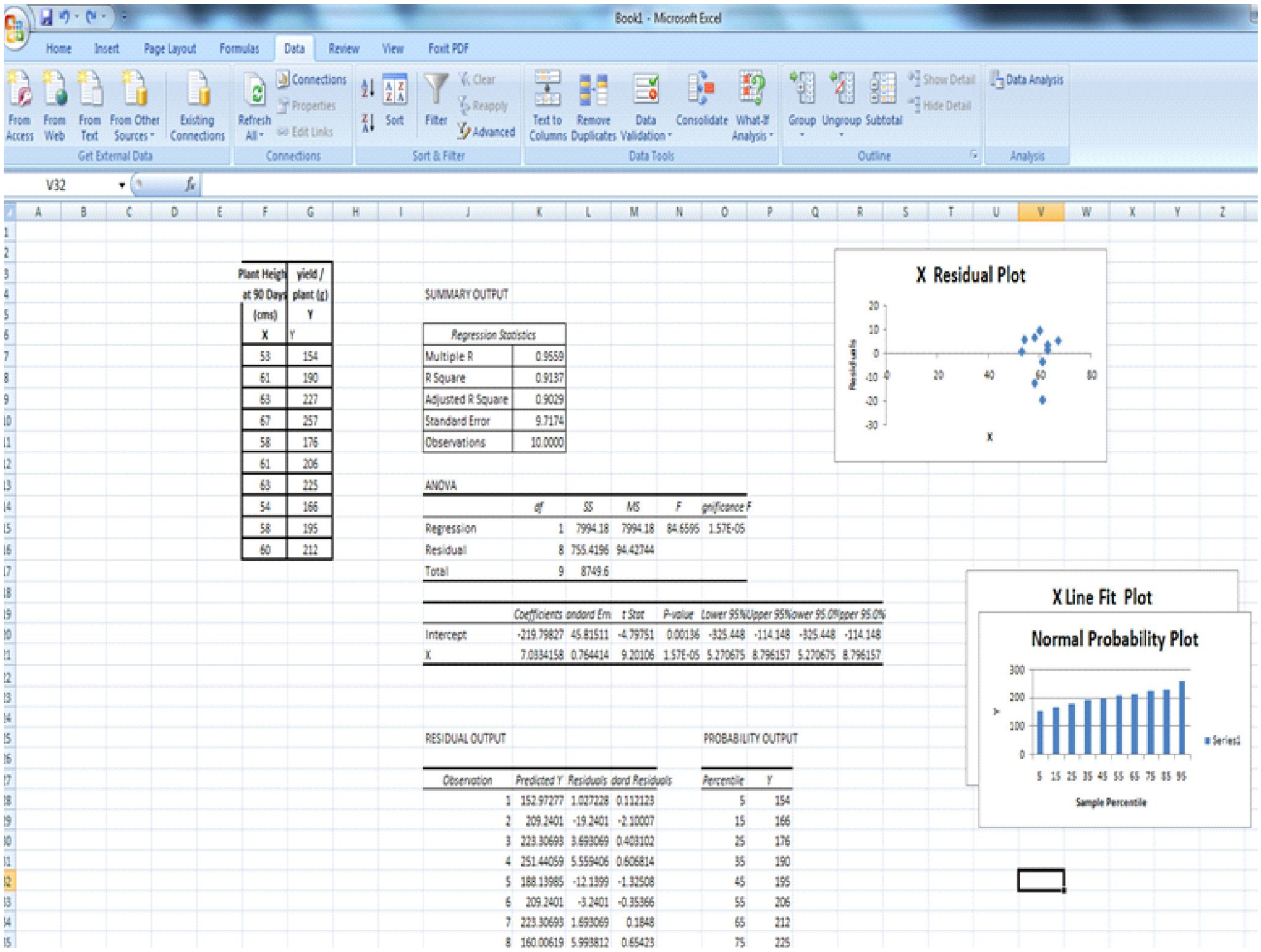


Table 1 : Collected potato (Plant height and plant yield) data pertaining to 10 potato plants.

Plant Height at 90 Days (cms) **X**	53	61	63	67	58	61	63	54	58	60
Yield /plant (g) **Y**	154	190	227	257	176	206	225	166	195	212

Table 2 :

X	Y	X^2	Y^2	XY
53	154	2809	23716	8162
61	190	3721	36100	11590
63	227	3969	51529	14301
67	257	4489	66049	17219
58	176	3364	30976	10208
61	206	3721	42436	12566
63	225	3969	50625	14175
54	166	2916	27556	8964
58	195	3364	38025	11310
60	212	3600	44944	12720
$\Sigma X =$ **598**	$\Sigma Y =$ **2008**	$\Sigma X^2 =$ **35922**	$\Sigma Y^2 =$ **411956**	$\Sigma XY =$ **121215**

Table 3 : ANOVA.

	df	SS	MS	F	Significance F
Regression	1	7994.18	7994.18	84.6595	1.57484E-05
Residual	8	755.4196	94.42744		
Total	9	8749.6			

In above table Significance F is 1.57484E-05 < 0.05, so, Regression coefficient is significant.

Table 4 : Regression Statistics.

Multiple R	0.9559
R Square	0.9137
Adjusted R Square	0.9029
Standard Error	9.7174
Observations	10

get results are given in table 3.

- In table 4, Multiple R is 0.9559 *i.e* the correlation between original yield and predicted yield values is 0.9559.
- In table 4, R square is 0.9137, theoretically that will come from below formula :

$$R^2 = \frac{Regression\ SS}{Total\ SS} = \frac{7994.18}{8749.6} = 0.9137$$

- R^2 = 0.9137 *i.e.* 91.37% is high, that lies between 85% and 100%, indicates there is high linearity between plant yield and plant height at 90 days.
- Our equation is simple linear regression, So, there is no need to consider Adjusted R Square

Conclusion

Suppose, if we consider potato plant height is 59 cms and then what is the value of yield?

Plant height (X) = 59 cms

Predicted simple linear regression

$$\begin{aligned} Y &= -219.7973 + (7.0334)X \\ &= -219.7973 + (7.0334)59 \\ &= -219.7973 + 414.9706 \\ &= 195.173 \end{aligned}$$

So, yield value is 195.173

In simple linear regression, we can easily estimate the value of dependent variable value (195.173) by using known value of independent variable value (59).

Regression expresses the relationship in the form of an equation. In this article, we have used simple example and excel to illustrate simple linear regression analysis and encourage the readers to analyze their data by these manual and excel procedures.

References

Agarwal, B. L. (2009). *Programmed Statistics (Questions-Answers)*. 2nd Edition, New Age International Publishing Limited, New Delhi.

Cini Varghese, MS-Excel:Statistical Procedures, IASRI, Library Avenue, New Delhi-110012.

Horticulture Statistics Division (2018). Department of Agriculture, Cooperation and farmers welfare. *Monthly report potato*. Government of India, New Delhi, January 2018.

Nageswara Rao, G. (2007). *Statistics for Agricultural Sciences*. BS Publications, Hyderabad (Second Edition).

Rangaswamy, R. (2010). *A text Book of Agricultural Statistics*. New Age International Publishing Limited, Hyderabad.

Srinivasa Rao, V. and R. Srinivasulu (2013). *Practical manual of STCA-101 (for class use only)*. Agricultural College, Bapatla.