

A Study On Multiple Regression Analysis In Era Of Demonetization

Dr. Vinay Pandit
Mathematics and Statistics Department
Lala Lajpatrai College, Mumbai
Email:doctorvinaypandit@gmail.com

Abstract

Multiple Regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable). In the current research paper dependent variable chosen was expenditure of family and independent variables chosen were income of individual and income of family. The research highlights the relationship between these dependent and independent variables during the time of demonetization. Also research paper highlights whether the relation in form of the equation is significant or not by applying t test and ANOVA.

Key words: Simple Regression, Multiple Regression, Demonetization

1.0 Introduction

In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X . The case of one explanatory variable is called **simple linear regression**.

Multiple Regression is an extension of simple linear **regression**. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable)

The main aim of this paper was to find the equation between income and expenditure and extend it to one more variable as income of family and thus find relationship between all during Demonetization

2.0 Research methodology

2.1 Research Objective

To study the significant relation between income and expenditure of a family.

2.2 Scope of Research

Scope of study was confined to service class and business class family. Also the research was restricted to two independent variables.

2.3 Method of data collection

The data collected was primary in nature. The method deployed was interviewing all respondents.

2.4 Sampling technique

A simple random sampling method was deployed for convenience

2.5 Research design

Since the research is statistical in nature the research design deployed was descriptive in nature.

2.6 Limitations of research

The study was restricted to a sample size of 19 families residing in south Mumbai. Also the study was restricted to service class and business class family and was restricted to two independent variables.

3.0 Data Analysis and Findings

3.1 Simple Regression of expenditure on income of individual

Table 3.1.1

<i>Regression Statistics</i>	
Multiple R	0.912139117
R Square	0.831997768
Adjusted R Square	0.822115284
Standard Error	1.618047626
Observations	19

ANOVA Table 3.1.2

	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	1	220.4137246	220.4137246	84.1891321
Residual	17	44.50732802	2.618078119	
Total	18	264.9210526		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.819188059	0.928843367	0.881944242	0.390107099
Income	0.777148832	0.08469859	9.175463591	5.38193E-08

3.2 Simple Regression of expenditure on income of family

Table 3.2.1

<i>Regression Statistics</i>	
Multiple R	0.975255419
R Square	0.951123133
Adjusted R Square	0.948248023
Standard Error	0.872741155
Observations	19

ANOVA table 3.2.2

	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
--	-----------	-----------	-----------	----------	-----------------------

Regression	1	251.9725415	251.972542	330.812799	1.40419E-12
Residual	17	12.94851111	0.76167712		
Total	18	264.9210526			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
Intercept	-2.436797391	0.640636715	-3.8037117	0.00141936	-3.788422712
X Variable 1	0.906461855	0.049837745	18.1882599	1.4042E-12	0.801313405

3.3 Multiple Regression

Table 3.3.1

<i>Regression Statistics</i>	
Multiple R	0.975811042
R Square	0.95220719
Adjusted R Square	0.946233088
Standard Error	0.889568774
Observations	19

ANOVA Table 3.3.2

	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	252.259731	126.129865	159.389193	2.72208E-11
Residual	16	12.66132166	0.7913326		
Total	18	264.9210526			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
Intercept	-2.657994499	0.749141636	-3.5480534	0.0026775	-4.246103822
X Variable 1	-0.086697719	0.143913973	-0.6024274	0.55533221	-0.391781713
X Variable 2	0.995953275	0.156996874	6.34377774	9.75E-06	0.66313477

4.0 Interpretation and conclusion

From the above calculation it is found that coefficient of correlation FOR X1 (Income of individual) and y (Expenditure) is $r = 0.484$ - Weak Positive and that to FOR X2 (Income of family) and y (Expenditure) is $r = 0.7012$ - Strong Positive.

For (Income of individual) and y (Expenditure) the simple regression equation is

$y = 8.63 + 1.18 X_1$ which means that $a = 8.63$ signifies y intercept (when there is no income), and 1.18 signifies regression coefficient.

$b = 1.18$ signifies difference in predicting value of y for each one unit of difference in

And for X_2 (Income of family) and y (Expenditure) is

$y = 8.63 - 1.37 X_2$ which means that $a = 8.63$ signifies y intercept (when there is no income), and 1.37 signifies regression coefficient.

$b = 1.37$ signifies difference in predicting value of y for each one unit of difference in x_2 .

Also we would expect an average expenditure of Rs. -2.65 with no income of individual and Family. However this is only a meaningful interpretation if it is reasonably that there is no income of individual and family and if the dataset actually included values of x_1 and x_2 that were near to 0. If neither of these conditions are true $a = -2.65$ has no meaningful interpretation. $b_1 = -0.08$ Represents the difference in predicted value of expenditure for each one Unit difference in x_1 if x_2 is constant

5.0 Reference

Armstrong, J. Scott (2012). "Illusions in Regression Analysis". International Journal of Forecasting (forthcoming). **28** (3): 689. doi:10.1016/j.ijforecast.2012.02.001.

C.F. Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum*. (1809)

C.F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae*. (1821/1823)

David A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press (2005)

Fisher, R.A. (1922). "The goodness of fit of regression formulae, and the distribution of regression coefficients". *Journal of the Royal Statistical Society*. Blackwell Publishing. **85** (4): 597–612. doi:10.2307/2341124. JSTOR 2341124.

Fotheringham, A. Stewart; Brunson, Chris; Charlton, Martin (2002). *Geographically weighted regression: the analysis of spatially varying relationships* (Reprint ed.). Chichester, England: John Wiley. ISBN 978-0-471-49616-8.

Fotheringham, AS; Wong, DWS (1 January 1991). "The modifiable areal unit problem in multivariate statistical analysis". *Environment and Planning A*. **23** (7): 1025–1044. doi:10.1068/a231025.

Mogull, Robert G. (2004). *Second-Semester Applied Statistics*. Kendall/Hunt Publishing Company. p. 59. ISBN 0-7575-1181-3.

M. H. Kutner, C. J. Nachtsheim, and J. Neter (2004), "Applied Linear Regression Models", 4th ed., McGraw-Hill/Irwin, Boston (p. 25)

N. Ravishankar and D. K. Dey (2002), "A First Course in Linear Model Theory", Chapman and Hall/CRC, Boca Raton (p. 101)

R.Dennis Cook; Sanford Weisberg *Criticism and Influence Analysis in Regression*, Sociological Methodology, Vol. 13. (1982), pp. 313–361

Steel, R.G.D, and Torrie, J. H., *Principles and Procedures of Statistics with Special Reference to the Biological Sciences.*, McGraw Hill, 1960, page 288.

Waegeman, Willem; De Baets, Bernard; Boullart, Luc (2008). "ROC analysis in ordinal regression learning". *Pattern Recognition Letters*. **29**: 1–9. doi:10.1016/j.patrec.2007.07.019

Yule, G. Udny (1897). "On the Theory of Correlation". *Journal of the Royal Statistical Society*. Blackwell Publishing. **60** (4): 812–54. doi:10.2307/2979746. JSTOR 2979746.

