Jour. Ind. Soc. Ag. Statistics
44 (1), 1992 : 37–44

The Use of Dummy Variable Approach in Testing Regressions

T.R. Shanmugam
Tamil Nadu Agricultural University,
Coimbatore - 641 003
(Received : September, 1989)

Summary

It is purposed to show how dummy variables were essentially a data classifying device in that they divided a sample into two size groups viz. small and large based on farm size. There were differences in the quantitative variables in the two groups. The differences were reflected in the change in intercepts with common slope co-efficients. Hence, a pooled function of the two size groups with dammy variable would be more appropriate to explain the phenomena of adoption of new technology in agriculture.

Key Words: Cobb- Douglas function, Regression model, Test for homogeneity,

Introduction

In regression analysis, it frequently happens that the dependent variable is influenced, not only by variables which can be readily quantified on some well defined scale but also by variables which are qualitative in nature. Dummy variables can be used as proxies in regression models to indicate the presence or absence of an attribute as constructed by Koutsoyiannis [3] and Lovell [4]. The present study is done using dummy variables to determine if there is any break-through in production function relationship between two groups of farmers viz., small and large with regard to adoption of a new technology.

2. *Materials and Methods*

The data for the present study were taken from the survey conducted in Periyar district during 1989 to study the economics of hybrid (Varalaxmi) cotton seed production. The size of the sample was fixed at 80 and selected at random. All the eighty farmers were post stratified into small (up to one hectare) and large farmers (more than one hectare). There were 38 small farms (Group I) and 42 large farms (Group II) with the average size of 0.96 and 2.71 hectares respectively.

Regression Function analysis had been made use to study and evaluate the important independent variables influenced the percentage of area under hybrid cotton seed to the gross area sown in both groups. Two Cobb-Douglas type of regression functions were run separately for the two groups and two pooled functions with and without dummy variables were also fitted for all the eighty farmers selected . Totally four regression functions were estimated by OLS tecnique.

2.1 *Size Group Functions*

$$Y = A X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} e^{\mu_i}$$

which may be expressed alternatively as

$$\ln Y = \ln A + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \mu_i$$

where ln = natural log to the base e.

Y = Percentage of area under hybrid cotton seed crop to the gross area sown.

X_1 = Size of the operated holding in hectares

X_2 = Irrigation intensity of the farm expressed in percent.

X_3 = Farm income in rupees.

μ_i = Random error term.

β_i = (i = 1 to 3) Parameters to be estimated

2.2 *Pooled Function Without Dummy Variables*

The model was same as the group function presented above.

2.3 *Pooled Function with Dummy Variables*

$$Y = A X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} D^{\beta_4} e^{\mu_i}$$

where

D	is	dummy variable, and
	=	1 for group I and
	=	2.718 for group II,

When log transformed, the values one and 2.718 became zero and one respectively. The other variables and parameters were same as the group function presented above.

2.4 *Econometric Tests*

To find out whether one aggregate function would be sufficient or group-wise functions would be essential to explain the phenomena of factors influencing the extent of adoption of seed production, the following ecnometric tests discussed by Johnston [2] suggested by Thanodaran, Bhide and Heady [5], and was also followed by Bisaliah [1] were conducted.

2.5 *Test For Group Effect*

A separate function of the Cobb-Douglas type was fitted for each of the the two groups, whose error sum of squares were S_1 and S_2 with n_1 and n_2 degrees of freedom, respectively. An aggregaate function was also fitted for the poled data of both the groups whose error sum of square was e and k_1 being degrees freedom. An F test was performed to determine whether the functional relationships were the same for the two groups.

$$F_1 = \frac{\dfrac{e_1 - S_1 - S_2}{k_1 - n_1 - n_2}}{\dfrac{S_1 + S_2}{n_1 + n_2}}$$

2.6 *Test For Difference In Intercepts*

Besides the two group-wise functions and a pooled function as specified above, another function was also fitted to the pooled function with a dummy variable with values '1' for first group and 2.718 for the second group. Assuming the error sum of squares for this function to be z_1 and m_1 degree of freedom, and F test could be performed to determine the significance of the differences in the intercepts.

$$F = \frac{\dfrac{e_1 - z_1}{k_1 - m_1}}{\dfrac{z_1}{m_1}}$$

However, this test was conditional on a common slope, so the test for differences in slope was performed first before deciding for differences in intercepts.

2.7 *Test for Homogeneity of Slopes*

Based on the error sum of squares mentioned for the above two tests, a test for homogeneity of slopes was performed using an F test.

$$F = \frac{\dfrac{z_1 - S_1 - S_2}{m_1 - n_1 - n_2}}{\dfrac{S_1 + S_2}{n_1 + n_2}}$$

Based on the results of these three tests, the appropriate function relevant for the present study was selected and analysed.

3. *Results and Discussions*

It might be evident from the Table I that all the estimated functions had shown high coefficients of determination (R^2)

The 'F' test was performed to find out whether one aggregate function would be sufficient to explain the phenomena of factors influencing the extent of adoption of hybrid cotton seed production or group-wise functions were essential. The test for group effect was conducted and the computed 'F' value was 1.8637 as against the tabular value of 2.50 at one per cent level of significance (Table II Section-A). Since the 'F' test indicated that there were no significant differences between the two groups of farms in respect of their production relationship, it was inferred that only one aggregate function for the pooled group was sufficient to explain the above phenomena.

However the differences in the phenomena of factors influencing the extent of adoption of new technology might be reflected in the functions either with a change in its intercepts or slopes. The F test was conducted to find out whether the slopes of the two group-wise functions were the same. The computed 'F' was 0.2608 as against the tabular value of 2.74 at one per cent level of significance (Table II, Section B). Hence, it was inferred that the slopes of the two group-wise production functions estimated were not significantly different. Thus the above test indicated that one aggregate function for both the groups with common slopes would be enough to explain the phenomena. However, the differences in the phenomena might be reflected in the function with a change in its intercept. As discussed earlier, the F test was conducted to find out whether the intercepts of two groups were the same. The computed F value was 7.2928 as against the tabular value of 6.99 at one per cent level of significance (Table II, Section C). Hence, it was inferred that intercepts of the two group-wise functions estimated were significantly different. Thus the above three tests indicated that one aggregate function with common slope and with a change in its intercept would be essential to explain the phenomena of factors influencing the extent of adoption of hybrid cotton seed production. Hence, a pooled function with dummy variable was selected to explain the phenomena.

3.1 *Selected Function*

$$\ln Y = -1.4265 + 0.2075 \ln X_1 + 0.8257 \ln X_2 + 0.3971 \ln X_3 + 0.1492 \ln D$$

Using dummy variables the above function could be rewritten group-wise as follows.

3.2 *Size Group I*

$$\ln Y = -1.4265 + 0.2075 \ln X_1 + 0.8257 \ln X_2 + 0.3971 \ln X_3$$

3.3 *Size Group II*

$$\ln Y = -1.2773 + 0.2075 \ln X_1 + 0.8257 \ln X_2 + 0.3971 \ln X_3$$

Where, the variables, standard errors, coefficient of multiple determination (R^2) and F statistic were the same as for the aggregate production function with dummy, presented in the Table 1.

The R^2 of the estimated function was 0.7213. This indicated that nearly 72 per cent of the variation in the percentage of area under hybrid cotton seed crop to the gross area sown was explained by the variables included in the anlysis. The variable, size of the operational holding in hectares (X_1) was found to have a positive but non-significant impact on percentage of area under hybrid cotton seed crop to the gross area sown. It indicated that this new technology was scale (size) neutral. Intensity of irrigation expressed in percentage (X_2) was found to have a positive and significant impact on percentage of area under hybrid cotton seed crop to the gross area sown. As expected, irrigation was the most important factor that influenced the farmers to adopt a new technology. The third variable, farm income in rupees (X_3) was also found to have a positive and significant impact on percentage of area under hybrid cotton seed crop to the gross area sown. It denoted the investment capacity of the farmer which was needed for the seed production since it was a capital intensive enterprise.

The advantages of the dummy variable technique can be readily seen from this analysis. We need to run only a single regression with dummy because the individual regressions can be deduced from it in the manner indicated. Finally, pooling increases the degrees of freedom and it will improve the relative precision of the estimated parameters.

REFERENCES

[1] Bisaliah, S., 1982. Technological Change and Functional Distribution Effects in Indian Agriculture. An Economic Analysis, *Artha Vijnana*, **24(1)**, 5-7.

[2] Johnston, J., 1972. *Econometric Methods*, McGraw-Hill, Kogakusha Ltd., Tokyo.

[3] Koutsoyiannis, A. 1987. *Theory of Econometrics*, Macmillan Education Limited, London.

[4] Lovell, M.C., 1963. Seasonal Adjustment of Economic Time Series, *Journal of the American Statistical Association*, **58**, 993-1010.

[5] Thamodaran, R., Shashanka Bide and Earl O. Heady, 1982. An Economic Analysis of Water Management Systems in Southern Tamil Nadu – Production Function and Programming Approach., *Indian Journal of Agricultural Economics*, **37(1)**, 45.

Table I. Estimated Regressions

Variables	Size Group I	Size Group II	Pooled	
			Without Dummy	With Dummy
Intercept	-1.1324^{S} (0.2669)	-1.2759^{S} (0.2743)	-1.3856^{S} (0.2567)	-1.4265^{S} (0.2493)
X_1	0.1563^{NS} (0.1812)	0.1749^{NS} (0.1943)	0.1952^{NS} (0.1897)	0.2075^{NS} (0.1763)
X_2	0.3452^{S} (0.3794)	0.2792^{S} (0.3564)	0.9643^{S} (0.3468)	0.8257^{S} (0.2164)
X_3	0.2947^{NS} (0.2736)	0.2849^{NS} (0.2571)	0.3214^{S} (0.0467)	0.3971^{S} (0.0628)
D				0.1492^{S} (0.0251)
R^2	0.6471	0.6538	0.7195	0.7213
N	38	42	80	80
F	20.7815^{S}	23.9211^{S}	64.9816^{S}	48.5266^{S}

Figures in parenthesis indicate the standard errors

NS Non significant at one per cent level

S Significant at one per cent level

Table II. Testing the Functions Using Dummy Variables

S. No.	Group	Error sum of Squares	Degrees of freedom	Calculated 'F' Value	Table 'F' Value at one percent level
		A. Test for Group Effect			
1.	Group I	0.3164	34		
2.	Group II	0.3278	38	1.7637[NS]	2.50
3.	Pooled without dummy		76		
		B. Test of Homogeneity of Slopes			
1.	Group I	0.3164	34		
2.	Group II	0.3278	38		
3.	Pooled with dummy	0.6479	75	0.2608[NS]	2.74
		C. Tets for Difference in intercepts			
1.	Pooled without dummy	0.7109	76		
2.	Pooled with dummy	0.6479	75	7.2928[NS]	6.99

NS Non significant at one per cent level of significance

S Significant at one percent level