Global Journal of Science Frontier Research: F
Mathematics and Decision Sciences
Volume 15 Issue 7 Version 1.0 Year 2015
Type : Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

One Factor Analysis of Variance and Dummy Variable Regression Models

By Okeh UM & Oyeka ICA
Nnamdi Azikiwe University, Nigeria

Abstract- This paper proposes and presents a method that would enable the use of dummy variable regression techniques for the analysis of sample data appropriate for analysis with the traditional one factor analysis of variance techniques with one, equal and unequal replications per treatment combination. The proposed method, applying the extra sum of squares principle develops F ratio-test statistics for testing the significance of factor effects in analysis of variance models. The method also shows how using the extra sum of squares principle builds more parsimonious explanatory models for dependent or criterion variables of interest. In addition, unlike the traditional approach with analysis of variance models, the proposed method easily enables the simultaneous estimation of total or absolute and the so-called direct and indirect effects of independent or explanatory variables on the dependent or criterion variables. The proposed methods are illustrated with some sample data and shown to yield essentially the same results as would the one factor analysis of variance techniques when the later methods are equally applicable.

Keywords: dummy variable regression, analysis of variance, degrees of freedom, treatment, regression coefficient.

GJSFR-F Classification : *MSC 2010: 62J05*

ONEFACTORANALYSISOFVARIANCEANDDUMMYVARIABLEREGRESSIONMODELS

Strictly as per the compliance and regulations of :

One Factor Analysis of Variance and Dummy Variable Regression Models

Okeh UM [α] & Oyeka ICA [σ]

Abstract- This paper proposes and presents a method that would enable the use of dummy variable regression techniques for the analysis of sample data appropriate for analysis with the traditional one factor analysis of variance techniques with one, equal and unequal replications per treatment combination. The proposed method, applying the extra sum of squares principle develops F ratio-test statistics for testing the significance of factor effects in analysis of variance models. The method also shows how using the extra sum of squares principle builds more parsimonious explanatory models for dependent or criterion variables of interest. In addition, unlike the traditional approach with analysis of variance models, the proposed method easily enables the simultaneous estimation of total or absolute and the so-called direct and indirect effects of independent or explanatory variables on the dependent or criterion variables. The proposed methods are illustrated with some sample data and shown to yield essentially the same results as would the one factor analysis of variance techniques when the later methods are equally applicable.

Keywords: dummy variable regression, analysis of variance, degrees of freedom, treatment, regression coefficient.

I. Introduction

Analysis of variance and regression analysis whether single-factor or multi-factor, sometimes both in theory and applications have often been treated and presented as rather different concepts by various authors. In fact only limited attempts seem to have been made to present analysis of variance as a regression problem (Draper and Smith, 1966; Neter and Wasserman, 1974).

Nonetheless analysis of variance and regression analysis are actually similar concepts, especially when analysis of variance is presented from the perspective of dummy variable regression models. This is the focus of the present paper, which attempts to develop a method to use dummy variable regression models and apply the "extra sum of squares principle" in the analysis of one-factor analysis of variance models with unequal replications per treatment combination as a regression problem.

II. Proposed Method

Let $y_i = y_{ij}$ be the response, score, or observation for the ith subject in a random sample of size n_j randomly drawn from population Y_j or administered test or treatment T_j ,for i=1,2,..., n_j; $j = 1, 2, ..., k$ populations or treatments. It is for the moment assumed

Author α: Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Abakaliki Nigeria. e-mail: uzomaokey@ymail.com
Author σ: Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.

that population Y_j or treatment T_j are each continuous measurement on the ratio scale.

Then the usual one-way analysis of variance model reflecting the dependence of subjects' responses or scores as a function of the different effect of treatments may be expressed in the form

$$y_{ij} = \mu + \alpha_j + e_{ij} \tag{1}$$

Notes

Where μ is the grand or overall mean, α_j is the differential effect of treatment T_j and e_{ij} are independent error terms with $E(e_{ij}) = 0,\ for\ i = 1,2,...,n_j; and\ j = 1,2,...,k.$ The treatment differential effect α_j are also subject to the constraint

$$\sum_{j=1}^{k} \alpha_j = 0 \tag{2}$$

The expected value of y_{ij} of equation 1 which is also the mean value or mean score specific to subject, administered test or treatment T_j is

$$E(y_{ij}) = \mu_j = \mu + \alpha_j \tag{3}$$

For j=1,2,…,k.

The mean effect μ_j and the treatment effect α_j are estimated using the usual one-way analysis of variance techniques. The statistical significance of the differential effects α_j is also determined using the usual F-ratio with k-1 and n-k degrees of freedom where $n = \sum_{j=1}^{k} n_j$.

We here however obtain alternative methods for estimating these effects and determine their statistical significance using dummy variable regression techniques. To do this we will k-1 dummy variables of 1's 0's to represent the k level of treatments or populations. We have here used k-1 dummy variables of 1s and 0s to represent the k treatments in order to adjust for the constraints imposed on the α_{js} by equation (2) and to ensure that the variance-covariance matrices $X'X$ resulting from the designed metrics X for the regression model is of full column rank and hence non-singular. To obtain the design metrics X for the dummy variable regression model we may let

$$x_{ij} = \begin{cases} 1, if\ \ the\ response\ or\ score\ y_i = y_{ij}\ by\ the\ ith\ subject \\ is\ reported\ or\ observed\ \ for\ test\ or\ treatment\ T_j \\ o, otherwise \end{cases} \tag{4}$$

$$for\ i = 1,2,...,n_j;\ j = 1,2,...,k-1.$$

Then a dummy variable regression model expressing the dependent of the responses or scores y_i of all the $n = \sum_{j=1}^{k} n_j$ subjects on the k-1 dummy variables of 1s and 0s representing k treatments may be expressed as

Global Journal of Science Frontier Research (F) Volume XV Issue VII Version I Year 2015

Notes

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j . x_{ij} + e_i \quad (5)$$

Where y_{is} are response scores by subjects, β_{js} are partial regression coefficients, $x_{ij}s$ are dummy variables of 1s and 0s,and ei's are error terms uncorrelated with x_{ij}, with $E(e_i) = 0, for\ i = 1, 2, ..., n = \sum_{j=1}^{k} n_j$.Now the expected value or mean value of y_i from equation (5).

$$E(y_i) = \beta_0 + \sum_{j=1}^{k-1} \beta_j x_{ij} . \quad (6)$$

In particular, note that the expected value or mean response or mean score by subjects at treatment T_j is obtained from equation (6) by setting $x_{ij} = 1$ and $x_{il} = 0, for\ all\ l \neq j; j = 1, 2, ..., k-1$ yielding

$$E(y_i)u_j = \beta_0 + \beta_j \quad (7)$$

Note from equations (3) and (7) that since $\beta_0 = \mu$ we will have that

$$\beta_j = \alpha_j \quad (8)$$

$$for\ \ j = 1, 2, .., k-1.$$

Equation can equivalently be expressed in its matrics form as

$$\underline{y} = X . \underline{\beta} + \underline{e} \quad (9)$$

Where $\underline{y}$ is an nx1 column vector of response scores, X is an nxk design metrics of 1s and 0s; $\underline{\beta}$ is a kx1 column vector of partial regression coefficient; and $\underline{e}$ is an nx1 column vector of error terms uncorrelated with X, $E(\underline{e}) = 0$.

Application of the usual least squares techniques to either equations (5) or (9) yields unbiased estimates of the partial regression coefficients, β as

$$\hat{\underline{\beta}} = \underline{b} = (X'X)^{-1} X'\underline{y} \quad (10)$$

Where $(X'X)^{-1}$ is the non-singular matrices inverse of the variance-covariance matrix $X'X$. This result will yield the fitted dummy variable regression model

$$\hat{\underline{y}} = X . \underline{b} \quad (11)$$

A null hypothesis that is usually of interest is that the regression models of either equations (5) or (9) kicks that is that not all the partial regression coefficients are zero. In other words the null hypothesis

$$H_0 : \underline{\beta} = \underline{0} \text{ versus } H_1 : \underline{\beta} \neq \underline{0} \tag{12}$$

This null hypothesis is tested using the usual F-ratio with k-1 and n-k degrees of freedom presented in the form of an analysis or variance table (Table 1)

Table 1 : Analysis of Variance Table for full Regression Model of equation (9)

Source of Variation	Sum of squares	Degree of freedom	Mean sum of squares	F-ratio
Regression	$SSR = \underline{b}'X'\underline{y} - n.\bar{y}^2$	k-1	$MSR = \frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \underline{y}'.\underline{y} - b'.X'.\underline{y}$	n-k	$MSE = \frac{SSE}{n-k}$	
Total	$SST = \underline{y}'.\underline{y} - n.\bar{y}^2$	n-1		

If the null hypothesis fits, that is if not all the regression coefficient in $\underline{\beta}$ are equal to zero, then one may proceed to test further null hypothesis involving the β_{js} in a regression model, or equivalently α_{js} in the corresponding one-way analysis of variance order. In other words, one may proceed to test other null hypothesis concerning the β_{js}; *for some* $j = 1, 2, ..., k-1$.

These tests are usually conducted using the traditional students' t test statistic with n-k degrees of freedom. We will here however propose and present an alternative and perhaps more generalized method for the same purpose based on the so called extra sum of squares principle (Drapa and Smith,1996;Netter and Wassemann,1974;Oyeka and Okeh, 2014;Boyle,1974).

Now suppose the null hypothesis H_0 of equation 12 is rejected based on the full model of equation 9;that is suppose not all the β_{js} *in* $\underline{\beta}$ are equal to zero, then the regression coefficient $\underline{\beta}$ and hence its estimated value $\hat{\underline{\beta}}$ and the corresponding design matrix X can be partitioned into at most k-1 mutually exclusive portions or groups. However, for the present presentation and the application of the extra sum of squares principle, we will here partition the design matrix X into two mutually exclusive groups or sub-matrices, namely group A with sub-matrix X_A with 'a' dummy variables of 1s and 0s and group B with sub-matrix X_B with 'b' dummy variables of 1s and 0s such that $a = (k-1)$, *where* $a, b \geq 1$. The treatments in groups A and B may for some reasons be similar within but dissimilar between themselves. The regression vector $\underline{\beta}$ and its sample estimate $\underline{b}$ are similarly partitioned into $\underline{b}_A$ *and* $\underline{b}_B$ with a and b rows respectively.

In this situation the treatment sum of squares SST in analysis parlance which is also the regression sum of squares in regression models can similarly be partitioned as

$$SST = SSR = \underline{b}'.X'.\underline{y} - n.\bar{y}^2 = (X.b)'.\underline{y} - n.\bar{y}^2$$

Notes

or equivalently as

$$SSR = \left(X_A - X_A \begin{pmatrix} \hat{b}_A \\ \hat{b}_B \end{pmatrix} \right)' \underline{y} - n\,\bar{y}^2 = \left(\underline{b}'_A X'_{A.\underline{y}} + \underline{b}'_B X'_{B.\underline{y}} \right) - n\,\bar{y}^2 \qquad (13)$$

or

equivalently

$$SSR = \underline{b}' X' \underline{y} - n\,\bar{y}^2 = \left(\underline{b}'_A X'_{A.\underline{y}} - n\,\bar{y}^2 \right) + \left(\underline{b}'_B X'_{B.\underline{y}} - n\,\bar{y}^2 \right) + n\,\bar{y}^2 \qquad (14)$$

Which when interpreted is the same as the statement

$$SST = SSR = SSA + SSB + SS(\bar{y} = \hat{\mu}) \qquad (15)$$

Where SSR is the sum of square regression for the full model of equation 9 with design matrix X and with k-1 degrees of freedom; SSA is the sum of squares regression for group A for the reduced model with design matrix X_A with 'a' degrees of freedom; SSB is the sum of squares regression for the reduced model for group B with design matrix X_B with degrees of freedom b=(k-1)-a; and $SS(\bar{y} = \hat{\mu})$ is an additive correction factor due to mean effect.

These sums of squares namely SSR,SSA and SSB are obtained by separately fitting the full model of equation 9 with design matrix X, and the reduced regression model with design matrices X_A and X_B respectively again separately on the criterion or dependent variable $\underline{y}$.

Now if the full model of equation 9 fits, that is if the null hypothesis H_0 of equation 12 is rejected, then an additional null hypothesis such as the null hypothesis that treatment in group A on the average have different effects on subjects and also that treatments in group B on the average have different effects on subjects may be tested. These null hypothesis may be tested using the extra sum of squares principle (Drapa and Smith,1966;Netter and Wasserman,1974;Oyeka et al,2013).

Now if we denote the sums of squares due to the full model of equation 9 and the reduced models due to the fitting of the criterion variable $\underline{y}$ to any of the reduced design matrices X_A and X_B by SS (F) and SS(R) respectively, then following the extra sum of squares principle the extra sum of squares due to a given factor is calculated as

$$ESS = SS(F) - SS(R) \qquad (16)$$

With degrees of freedom obtained as the difference between the degrees of freedom of SS(F) and SS(R) ,that is as

$$df(ESS) = df(F) - df(R) \qquad (17)$$

Thus the extra sums of squares regression due to group A and group B are obtained as respectively.

$$ESSA = SSR - SSA;\ ESSB = SSR - SSB \qquad (18)$$

Notes

With (k-1)-a=b degrees of freedom and (k-1)-b=a degrees of freedom respectively.

The corresponding extra sum of squares error for factors A and B are obtained from equation 16 using

$$ESSE = SSE(F) - SSE(R) \tag{19}$$

That is as

$$ESSEA = SSE - SSEA; ESSEB = SSE - SSEB \tag{20}$$

The corresponding degrees of freedom are obtained using equation 17 as

$$\left.\begin{array}{l} df(ESSEA) = \left((n-1)-a\right)-(n-k) = k-1-a = b; \\ df(ESSEB) = \left((n-1)-b\right)-(n-k) = k-1-b = a \end{array}\right\} \tag{21}$$

Table 2 : Analysis of variance table for multiple comparisons in dummy variable Regression models

Sources of variation	Sum of squares(SS)	df	MSE	F-ratio	ESS	df	EMSS	F-ratio
Full model								
Regression	$SSR = \underline{b}'X'.\underline{y} - n.\overline{y}^2$	$k-1$	$MSR = \frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$	ESSR = SSR	k-1		
Error	$SSE = \underline{y}'.\underline{y} - \underline{b}'X'.\underline{y}$	$n-k$	$MSE = \frac{SSR}{n-k}$		ESSE = SSE	n-k		
Reduced model								
Group A Regression	$SSA = \underline{b}'_A X'_A.\underline{y} - n.\overline{y}^2$	a	$MSR = \frac{SSA}{a}$	$F_A = \frac{MSA}{MSEA}$	ESSA = SSR-SSA	k-1-a=b		
Error	$SSEA = \underline{y}'.\underline{y} - \underline{b}'_A X'_A.\underline{y}$	$n-1-a$	$MSEA = \frac{SSEA}{n-1-a}$		ESSEA=SSEA-SSE=ESSA	k-n-a=b		
Group B Regression	$SSB = \underline{b}'_B X'_B.\underline{y} - n.\overline{y}^2$	b	$MSB = \frac{SSB}{b}$	$F_B = \frac{MSB}{MSEB}$	ESSB = SSR-SSB	k-1-b=a		
Error	$SSEB = \underline{y}'.\underline{y} - \underline{b}'_B X'_B.\underline{y}$	$n-1-b$	$MSEB = \frac{SSR}{n-1-b}$		ESSEB=SSEB-SSE=ESSB	n-1-b		
Total	$SST = \underline{y}'.\underline{y} - n.\overline{y}^2$	$n-1$						

The null hypothesis H_0 of equation 12 is tested using the usual F-ratio for the full model with k-1 and n-k degrees of freedom. If this null hypothesis is rejected, then one may apply the extra sum of squares principle to obtain the results of Table 2.The null hypothesis that treatments in group A have different effects on subject may be tested using the $F_A^* - ratio$ with b and n-k degrees of freedom while the null hypothesis that treatments in group B have different effects on subjects may be tested using the $F_B^* - ratio$ with a and n-k degrees of freedom.

These null hypothesis are each rejected at the α level of significance if

$$F_A^* \geq F_{1-\alpha;b,n-k;} F_B^* \geq F_{1-\alpha;a,n-k} \tag{22}$$

Otherwise the null hypothesis is accepted.

Notes

Notes

Note that if the null hypothesis to be tested is that the regression effect β_j in regression models or α_j in analysis of variance parlance is not different from zero, for some j=1,2,….,k-1,then design matrix X may be partitioned in such a way that X_A, say has only dummy variable of 1s and 0s so that only dummy variable of 1s and 0s so that a=1,and X_β has b=k-1-a=k-2,dummy variable of 1s and 0s.In this case the null hypothesis that a given regression coefficient is equal to zero may be tested and rejected if the F_A^* of equation 22 is satisfied with a-1 and b=k-2.Similarly,to test the null hypothesis

$$H_0:\beta_j-\beta_l=0 \; versus \; H_0:\beta_j-\beta_l\neq 0 \quad (23)$$
$$for\ some\ j,l=1,2,...,k-1;\ j\neq l$$

We may partition the design matrix X into X_A,say and X_B an nx2 and (k-3) design matrices of 1s and 0s respectively.

The null hypothesis H_0 of equation 23 is rejected at the α level of significance if the calculated $F^*-ratios$ satisfy the equation

$$F_A^*\geq F_{1-\alpha;k-3,n-k;}F_B^*\geq F_{1-\alpha;2,n-k} \quad (24)$$

Otherwise the null hypothesis is accepted.

An additional advantage of using dummy variable regression models in either two-way or one-way analysis of variance problems is that the method also more easily enables the simultaneous estimation of factor level or treatment effects separately of several factors on a specified dependent or criterion variable through the effects of their representative dummy variables of 1s and 0s. Specifically,the method enables the estimation of the direct effects of a given factor or independent variable on a dependent or criterion. This is a waited sum of the partial regression coefficients or effects of the set of dummy variables of 1s and 0s representing that independent variable(Wright,1973).That is the direct effect of an independent variable, referred to here as population or treatment $X_i\ or\ T$ on a dependent or criterion variable Y is the weighted sum of the partial regression coefficients β_j or effects of the set of dummy variables x_{ij} of 1s and 0s representing that independent variable. The weights θ_j are the simple regression coefficients of each representative dummy variable x_{ij} regressing on the specified independent variable or treatment X_j represented by numerical codes.

Thus suppose we here without loss of generality represent populations or treatments X_j with the numerical code $c+j$; for j=1,2,…,k-1, where c is any real number then a simple regression equation expressing the dependence of the dummy variable x_{ij} of 1s and 0s on its parent variable X_j may be written as

$$x_{ij}=\theta_0+\theta_j X_j+e_i \quad (25)$$

Where θ_j a simple regression coefficient and ei is are error terms uncorrelated with X_j with $E(e_i)=0,\ for\ j=1,2,...,k-1; l=1,2,...,n.$

Now taking the partial derivative of the expected value of equation (25) with respect to X_j, we have

$$\frac{dE(x_{ij})}{dX_j} = \theta_j \tag{26}$$

Now taking the partial derivative of equation (6),the expected value of y_i with respect to X_j,and substituting equation (26) we obtain

$$\frac{dE(y_i)}{dX_{ij}} = \sum_{j=1}^{k-1} \beta_j . \frac{dE(x_{ij})}{dX_j} = \sum_{j=1}^{k-1} \theta_j . \beta_j = \beta_{dir} \tag{27}$$

Notes

Where β_{dir} is the so-called direct effect of the treatment X and T on the dependent or criterion variable Y when these treatments are represented by set of dummy variable of 1s and 0s.

The corresponding sample estimate of β_{dir} is obtained in terms of the sample estimates of β_j namely b_j as

$$\hat{\beta}_{dir} = b_{dir} = \sum_{j=1}^{k-1} \theta_j b_j \tag{28}$$

III. Illustrative Example

In a study of the effect of granulated starch on the disintegration time of tablets a random sample of four local sources of starch was chosen and five measurements of the disintegration time of a tablet were made for each source yielding the following results (in seconds) (Table 3)

Table 3 : Disintegration time (in seconds) of starch Tablets by source of starch

Maize	Cassava	Yam	Cocoyam
29	35	100	116
120	145	120	180
114	122	245	90
75	70	240	310
	55	180	
	75	246	

To use the proposed method to analyze the data, we would first use three dummy variables of 1s and 0s to represent the four different sources of starch by applying equation 4 to the data of Table 3 to obtain the design matrix X of Table 4.

Table 4 : Design Matrix X for the sample Data of table 3

S/N	y_i	X_{i0}	X_{i1}	X_{i2}	X_{i3}
1	29	1	1	0	0
2	120	1	1	0	0
3	114	1	1	0	0
4	75	1	1	0	0
5	35	1	0	1	0
6	145	1	0	1	0
7	122	1	0	1	0
8	70	1	0	1	0
9	55	1	0	1	0
10	75	1	0	1	0
11	100	1	0	0	1

12	120	1	0	0	1
13	245	1	0	0	1
14	240	1	0	0	1
15	180	1	0	0	1
16	246	1	0	0	1
17	116	1	0	0	0
18	180	1	0	0	0
19	90	1	0	0	0
20	310	1	0	0	0

Notes

Fitting the full model of Eqn 9 using the design matrix X of table 4, we obtain the fitted regression equation

$$\hat{y}_i = 33.322 + 20.242x_{i1} - 11.525x_{i2} + 8.122x_{i3}\ (Pvalue = 0.0000) \quad (29)$$

A P-value of 0.0000 clearly shows that the model fits.

The expected result of measurements of the disintegration time of a tablet for the first treatment is obtained by setting $x_{i1} = 1$, and all other $x_{is} = 0$ in equation (29) giving

$$\hat{y}_i = 33.322 + 20.242 = 53.564$$

The estimated response measurement result of the disintegration time of a tablet for the second treatment is estimated by setting

$x_{i2} = 1$ and all other $x_{i's} = 0$ in Equation (29) yielding

$\hat{y}_i = 33.322 - 11.525 = 21.797$

The estimated response measurement result of the disintegration time of a tablet for the third treatment is similarly estimated by setting

$x_{i3} = 1$ and all other $x_{i's} = 0$ in Equation (29) yielding

$\hat{y}_i = 33.322 + 8.122 = 41.444$

The corresponding analysis of variance table for the full model is presented in Table 5.

Table 5 : Anova Table for the Full Model of Equation (29)

Source of Variation	Sum of Squares (SS)	Degrees of freedom (Df)	Mean Sum of Squares (MS)	F-Ratio	P-Value
Regression (treatment)	3648.412	3	1216.137	9.859	0.0000
Error	1973.529	16	123.346		
Total	5621.941	19			

Having fitted the full model which is here seen to fit, we now proceed to fit the dependent variably y separately on each of the sub matrices x_{i1}, x_{i2} *and* x_{i3} each with two dummy variables of Is and 0s to obtain the corresponding sum of squares due to each of these factors. The sums of squares due to factor A and B are calculated following Equation (13). The results are summarized in a one factor analysis of variance Table with extra sums of squares (Table 6).

Table 6 : One factor Analysis of Variance Table with Extra Sums of Squares for the Sample Data of Table 3

Source of Variation	Sum of Squares (SS)	Degrees of freedom (Df)	Mean of sum of squares MS	F-Ratio	Extra Sum of Squares (ESS)	Degrees of freedom (Df)	Extra mean sum of squares (EMS)	F- Ratio	Critical F value P- value
Full Model									
Regression	3648.412	3	1216.137	9.859	3648.412	3	1216.137	9.859	2.030*
Error	1973.529	16	123.346		1973.529	16	123.346		
Group A									
Regression	1834.132	4	458.533	2.609	1814.28	6	302.38	6.852	5.86*
Error	2635.451	15	175.697		661.922	15	44.128		
Group B									
Regression	952.112	6	158.69	0.704	2696.3	4	674.075	9.148	4.00*
Error	2931.421	13	225.49		957.892	13	73.684		
Total									

*Note: * indicates statistical significance at the 5 percent level*

These analyses indicate that the hypothesized model fits, that is that not all the factor level effects are zero. However group A and B have significant effects on the criterion variable Y.

Finally to estimate the direct effect or partial regression coefficient of group A and group B say, represented by the dummy variables x_{i1}, x_{i2} and x_{i3}, we first estimate the simple regression coefficient resulting when theses dummy variables are each regressed on A using Equation 26, yielding.

$$\alpha_{i1} = 2.13; \alpha_{i2} = 0.34; \alpha_{i3} = -0.23$$

Using these results with Equation (29) in (28), we obtain an estimate of the direct effect of A and B on 'y' as

$$b_{A\,and\,B}dir = (20.242 \times 2.13) + (-11.525 \times 0.34) + (8.122 \times -0.23) = 37.3289$$

The estimated simple regression coefficient or effect of A and B on y is $b_{A\ and\ B}$ =37.3289.

Hence the estimated indirect effect of A on'y' is

$$b_{A\,and\,B}\,ind = 37.3289 - 24.832 = 12.4969$$

IV. Summary and Conclusion

We have in this paper proposed and developed a method that enabled the use of dummy variable multiple regression techniques for the analysis of data appropriate for use with two factor analysis of variance models with unequal observations per treatment combination and with interactions. The proposed model and method employed the extra sum of squares principle to develop appropriate test statistics of F ratios to test for the significance of factor and interaction effects.

The method which was illustrated with some sample data was shown to yield essentially the same results as would the traditional two factor analysis of variance model with unequal observations per cell and interaction. However the proposed method is more generalized in its use than the traditional method since it can easily be

Notes

used in the analysis of two-factor models with one observation, equal, and unequal observations per cell as a rather unified analysis of variance problem.

Furthermore unlike the traditional analysis of variance models the proposed method is able to enable one using the extra sum of squares principle, to determine the relative contributions of independent variables or some combinations of these variables in explaining variations in a given dependent variable and hence build a more parsimonious explanatory model for any variable of interest. In addition, the method enables the simultaneous estimation of the total or absolute, direct and indirect effects of a given independent variable on a dependent variable, which provide additional useful information.

Notes

References Références Referencias

1. Boyle, Richard P (1974) Path Analysis and Ordinal Data. In Blalock, H M (ed) Causal Model in the Social Sciences Aldine Publishing Company Chicago 1974.
2. Draper, N. R and Smith, H. (1966). Applied Regression Analysis: John Wiley & sons, Inc., New York.
3. Neter, J. and Wasserman, W. (1974). Applied Linear Statistical Models. Richard D. Irwin Inc, ISBN 0256014981, ISSN- 101-423-199. New York.
4. Oyeka, I. C. A, Uzuke C. U, Obiora-ilouno H. O and Mmaduakor C (2013): Ties Adjusted Two way Analysis of Variance tests with unequal observations per cell. Science Journal of Mathematics & Statistics (SJMS), ISSN:2276-6324.
5. Oyeka ICA & Okeh UM(2014). Two Factor Analysis of Variance and Dummy Variable Multiple Regression Models. Global Journal of Science Frontier Research: F Mathematics and Decision Sciences Volume 14 Issue 6 Version 1.0 Year 2014.
6. Wright, Sewall (1934): The Methods of Path Coefficients. Annals of Mathematical Statistics: Vol 5.

This page is intentionally left blank