AN ALTERNATIVE METHOD OF ANALYZING ONE-WAY CLASSIFICATION DATA BY THE USE OF DUMMY VARIABLES*

*Mariano B. de Ramos***

INTRODUCTION

Linear models applied to experimental designs have the form

$$\underline{y} = X\underline{\beta} + \underline{\epsilon} \qquad [1]$$

where $\underline{y}$ is the N x 1 vector of responses, X is the N x (t + 1) design matrix consisting of 1's and 0's, $\underline{\beta}$ is the (t + 1) vector of parameters, and $\underline{\epsilon}$ is the N x 1 vector or random errors. By applying the least squares procedure to [1] in estimating β, the normal equation is

$$(X'X) \quad \underline{\hat{\beta}} = X'\underline{y} \qquad [2]$$

In the case of experimental designs model no unique solution of $\underline{\hat{\beta}}$ in [2] can be found because the matrix X or $X'X$ has rank lower than the columns of X or XX.

To take a specific example, consider a one-way classification design model with treatments and n_i replications per treatment. The linear model for the response y is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, 3, \ldots, t, \quad j = 1, 2, \ldots, n_i \qquad [3]$$

where μ is the effect common to all observations, τ_i is the effect of treatment i defined as the difference between the mean of treat-

---

*Paper presented at the Third National Convention on Statistics, December 13-14, 1982, Philippine International Convention Center (PICC), Metro Manila.

**Associate Professor of Experimental Statistics, University of the Philippines at Los Banos and Senior Research Fellow, International Rice Research Institute.

ment i, μ_i, and μ the mean of all treatment means. In this case, the matrices involved in the model are:

$$y = \begin{pmatrix} y11 \\ y12 \\ \vdots \\ y_{in_1} \\ \vdots \\ y_{t_2} \\ y_{t_2} \\ \vdots \\ y_{tn_t} \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 1 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & & 0 \\ 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 1 \\ 1 & 0 & & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 1 \end{pmatrix} \quad N \times (t+1)$$

$$\beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau \\ \tau_t \end{pmatrix} \quad (t+1) \times 1 \qquad \epsilon = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{t_1} \\ \xi_{t_2} \\ \vdots \\ \xi \\ \epsilon_{tn_t} \end{pmatrix} \quad N \times 1$$

and $N = n_1 + n_2 + \ldots + n_t$.

The resulting normal equation for estimating β is

$$X'X\hat{\underline{\beta}} = X'\underline{y}$$

where

$$X'X = \begin{pmatrix} N & n_1 & n_2 & \ldots & n_t \\ n_1 & n_1 & 0 & \ldots & 0 \\ n_2 & 0 & n_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ n_t & 0 & 0 & & n_t \end{pmatrix} \qquad (t+1) \times (t+1)$$

$$X'\underline{y} = \begin{pmatrix} Y_{\cdot\cdot} \\ Y_{1\cdot} \\ Y_{2\cdot} \\ \cdot \\ \cdot \\ \cdot \\ Y_{t\cdot} \end{pmatrix} \qquad (t+1) \times 1$$

and $Y_{\cdot\cdot} = \sum_{i}^{t} \sum_{j}^{n_i} Y_{ij}$, $\qquad Y_{i\cdot} = \sum_{j}^{n_i} Y_{ij}$.

There are only t independent columns of rows in the matrix $X'X$ and herefore the rank of $X'X$ is t, hence, there is no unique solution to the parameters $\mu, \tau_1, \tau_2, \ldots \tau_t$ in equation [4].

In practice one usually obtains a unique solution to [4] by, for example, imposing the restriction $\sum_{i}^{t} \hat{\tau}_i = 0$. However, this restriction may not be realistic if the model being used is not a fixed effects

model or Model I, but a random effects model or Model II. Of course, one could always find a solution to [4] by the use of generalized-inverse matrix.

Dummy variables are predefined set of independent variables corresponding to a set of categorical or nominal factor. These variables have values 1's and 0's indicating the presence or absence of the factor level, respectively. The role of dummy variables in multiple regression analysis has been demonstrated to be particularly useful in the analysis of data in the fields of social sciences and economics.

In one-way classification design the treatments or groups are usually level of nominal variables like types of variety, types of insecticide, kinds of cultural management practices, etc. In many cases biological experiments would involve a set of treatments in which one is a control or standard. But even if there is no control in a set of treatments, one may designate one of the treatments as the reference treatment from which the rest of the treatments will be compared with. In this situation the use of dummy-variable regression model will be in order.

The main objective of this paper is to demonstrate the use of dummy variables in analyzing one-way classification experiments in which one of the treatments is a control. Specifically, the objectives are: (i) to show the definitions of the dummy variables as used in one-way classification design, (ii) to demonstrate the estimation and hypothesis testing procedures, (iii) and to relate the results of analysis using dummy variables approach with that of the conventional analysis of variance.

THE DATA AND REGRESSION MODEL

The usual format of a one-way classification data with one treatment as control is shown in the following tables:

TABLE 1. TABULAR FORMAT OF A ONE–WAY CLASSIFICATION DATA WITH ONE TREATMENT AS CONTROL

Treatment	*Replication* 1	2	3	..	*j*	...	n_i	*No. of Reps.*	*Totals*	*Means*
1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1n_1}	n_1	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{12}	y_{22}	y_{23}	...	y_{2j}	...	y_{2n_2}	n_2	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3n_3}	n_3	$y_{3\cdot}$	$\bar{y}_{3\cdot}$
.	.	.		.		.		.	.	.
.	.	.		.		.	.	.	.	.
.	.	.		.		.	.	.	.	.
i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{in_i}	n_i	$y_{i\cdot}$	$\bar{y}_{i\cdot}$
.	.	.		.		.	.	.	.	.
.	.	.		.		.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
t (Control)	y_{t1}	y_{t2}	y_{t3}	...	y_{tj}	...	y_{tn_t}	n_t	$y_{t\cdot}$	$\bar{y}_{t\cdot}$
Totals	N				$y_{\cdot\cdot}$					

For each observation Y_{ij} we shall define a set of t – 1 dummy variables denoted by X and these are defined as follows:

X_{1ij} = 1 if i = 1 and 0 otherwise,

X_{2ij} = 1 if i = 2 and 0 otherwise,

$X_{(t-1)ij}$ = 1 if i = t – 1 and 0 otherwise.

Also $X_{1ij}, X_{2ij}, \ldots, X_{(t-1)ij}$ = will all be 0 when $i = t$.

This relationship between the dummy variables and treatment is shown in Table 2.

TABLE 2. RELATIONSHIP BETWEEN THE TREATMENTS AND DUMMY VARIABLES

Treatment	*Dummy variable*						
	X_1	X_2	X_3	...	X_i	...	X_{t-1}
1	1	0	0	...	0	...	0
2	0	1	0	...	0	...	0
3	0	0	1	...	0	...	0
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
i	0	0	0		*1*		*0*
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
$t-1$	0	0	0		0	...	1
t	0	0	0		0	...	0

Following the definition given above, the multiple regerssion model has the form

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \ \ldots \ + \beta_{t-1} X_{t-1\,ij} + \epsilon_{ij} \qquad [5]$$
$$i = 1, 2, 3, \ \ldots, \ t, \quad j = 1, 2, \ \ldots, \ n_i$$

where β_0 is the intercept, the β_i.s are partial regression coefficients, and ϵ_{ij} are the random error terms.

Using matrix notations, the model is written as

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon} \qquad [6]$$

where

$$\underline{Y} = \begin{pmatrix} y11 \\ y12 \\ \vdots \\ y1n_1 \\ y21 \\ y22 \\ \vdots \\ y2n_2 \\ \vdots \\ yt_1 \\ yt_2 \\ \vdots \\ ytn_i \end{pmatrix} \quad N \times 1$$

$$X = \begin{pmatrix} \underline{j}1 & \underline{j}1 & \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{j}2 & \underline{0} & \underline{j}2 & \underline{0} & \cdots & \underline{0} \\ \vdots & & & & & \\ \underline{j}_{t-1} & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{j}_{t-1} \\ \underline{j}_t & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} \end{pmatrix} \quad N \times t$$

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{t-1} \end{pmatrix} \quad t \times 1 \qquad \underline{j}_l = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad n_l \times 1 \qquad \underline{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad n_i \times 1$$

$$\underline{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \cdot \\ \cdot \\ \epsilon_{1n_1} \\ \cdot \\ \cdot \\ \epsilon_{t1} \\ \epsilon_{t2} \\ \cdot \\ \cdot \\ \epsilon_{tn_t} \end{pmatrix} \qquad N \times 1$$

ESTIMATION

Consider equation [6] $\underline{Y} = X\underline{\beta} + \underline{e}$. Let us assume that the random vector $\underline{e}$ has the following properties:

$$\mathrm{E}(\underline{\epsilon}) = \underline{0} \qquad \mathrm{E}(\underline{\epsilon}\,\underline{\epsilon}') = \sigma^2 I_N$$

Minimizing the residual sum of squares $\underline{\epsilon}'\underline{\epsilon}$ w. r. t. $\hat{\beta}$ yields the normal equation

$$X'X\,\underline{\hat{\beta}} = X'\underline{y} \qquad [7]$$

where

$$X'X = \begin{pmatrix} N & n_1 & n_2 & \dots & n_{t-1} \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & \dots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ n_{t-1} & 0 & 0 & & n_{t-1} \end{pmatrix} \quad t \times t \qquad X'\underline{y} = \begin{pmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \\ \cdot \\ \cdot \\ y_{t-1} \end{pmatrix} \quad t \times 1$$

The matrix $X'X$ is of rank t (full rank and therefore the unique solution to equation [7] is

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

$$(X'X)^{-1} = \begin{pmatrix} \frac{1}{n_t} & -\frac{1}{n_t} & -\frac{1}{n_t} & \cdots & -\frac{1}{n_t} \\ -\frac{1}{n_t} & \frac{1}{n_1}+\frac{1}{n_t} & \frac{1}{n_t} & \cdots & \frac{1}{n_t} \\ -\frac{1}{n_t} & \frac{1}{n_t} & \frac{1}{n_2}+\frac{1}{n_t} & & \frac{1}{n_t} \\ \vdots & \vdots & \vdots & & \\ -\frac{1}{n_t} & \frac{1}{n_t} & \frac{1}{n_t} & \cdots & \frac{1}{n_{t-1}}+\frac{1}{n_t} \end{pmatrix}_{t \times t}$$

and

$$\hat{\beta} = \begin{pmatrix} \bar{y}_{t\cdot} \\ \bar{y}_{1\cdot} - \bar{y}_{t\cdot} \\ \bar{y}_{2\cdot} - \bar{y}_{t\cdot} \\ \vdots \\ \bar{y}_{t-1} - \bar{y}_{t\cdot} \end{pmatrix}_{t \times 1} \qquad [9]$$

Therefore, the estimate of the intercept β_0 and partial regression coefficient β_i are simple functions of the mean of control ($\bar{y}_{t\cdot}$) and means of treatments ($\bar{y}_{i\cdot}$). Thus

$\hat{\beta}_0 = \bar{y}_{\cdot\cdot}$,

$\hat{\beta}_0 = \bar{y}_{t\cdot}$, the mean of control group,

$\hat{\beta}_1 = \bar{y}_{1\cdot} - \bar{y}_{t\cdot}$, the mean of treatment 1 minus the mean of control,

$\hat{\beta}_2 = \bar{y}_{2\cdot} - \bar{y}_{t\cdot}$, mean of treatment 2 minus the mean of control

$\vdots$

$\hat{\beta}_{t-1} = \bar{y}_{t-1} - \bar{y}_{t\cdot}$, the mean of treatment t−1 minus the mean of control.

HYPOTHESIS TESTING

By assuming further that the random errors ϵ_{ij} in [5] are normally distributed, one can proceed to testing some relevant hypotheses. In this case, the analysis of variance has the following components:

$$SS_{\text{Total (uncorrected)}} = \underline{y}'\underline{y} = \sum_{i}^{t} \sum_{j}^{n_i} y_{ij}^2 \qquad \text{with N d.f.,}$$

$$SS_{\text{reg}} = \underline{\hat{\beta}}\,'X'\underline{y} = \sum_{i}^{t} y_{i.}^2/n_i \quad \text{with t d. f.,}$$

$$SS_{\text{resid}} = \underline{y}'\underline{y} - \underline{\hat{\beta}}\,'X'\underline{y} = \sum_{i}^{t} \sum_{j}^{n_i} y_{ij}^2 - \frac{\sum_i y_i^2}{\bar{n}_i} \quad \text{with N–t d. f.}$$

The sum of squares due to regression, SSreg, is partitioned into two parts, namely,

$$SS\text{due to } \hat{\beta} = y_{\cdot\cdot}^2/N \text{ with 1 d. f.,}$$

and SSdue to $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{t-1} = \sum_{i}^{t} y_{i.}^2/n_i$ with t–1 d. f.

As in the usual model of full rank, the unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\underline{y}'\underline{y} - \underline{\hat{\beta}}\,'X'\underline{y}}{N-t} = \frac{SS_{\text{resid.}}}{N-t} = MS_{\text{resid.}}$$

Finally, the analysis of variance table is given in Table 3.

TABLE 3. ANOV TABLE FOR ONE-WAY CLASSIFICATION USING DUMMY VARIABLES.

SV	*DF*	*SS*	*MS*
Regression	t	$\underline{\hat{\beta}}'X'\underline{y}$	
Due to $\hat{\beta}_0$	1	$y_{..}^2/N$	
Due to $\beta_1, \beta_2, \ldots, \hat{\beta}_{t-1}$	$t-1$	$\underline{\hat{\beta}}'X'\underline{y} - y_{..}^2/N$	MS_{Reg}
Residual	$N-t$	$\underline{y}'\underline{y} - \underline{\hat{\beta}}'X'\underline{y}$	MS_{resid}
Total	N	$\underline{y}'\underline{y}$	

The following null hypotheses can be tested:

a. All the effects of treatments over the control are zero or

$H_0: \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_{t-1} = 0$

Test statistic: $F = \dfrac{MS_{reg}}{MS_{resid}} \sim F_{(t-1,\, N-t)}$

b. The effect of treatment i over the control is equal to the effect of treatment i' over the control or

$H_0: \beta_i = \beta_{i'}$ for all $i \neq i'$.

Test statistics: $t = \dfrac{\hat{\beta}_i - \hat{\beta}_{i'}}{\sqrt{MS_{resid}\left(\dfrac{1}{n_i} + \dfrac{1}{n_{i'}}\right)}} \sim t_{(N-t)}$

c. The effect of any treatment i over the control is zero or

H_0: $\beta_i = 0$ for all $i = 1, 2, \ldots, t-1$.

Test statistic: $$t = \frac{\hat{\beta}_i}{\sqrt{MS_{resid}\left(\frac{1}{n_i} + \frac{1}{n_t}\right)}} \sim t_{(N-t)}.$$

In general, one may wish to test any linear hypothesis concerning the β_i's such as $H_0 : \underline{a}'\underline{\beta} = \delta^*$.
The linear unbiased estimates of $a\beta$ is $a\hat{\beta}$ with variance estimate

$$\begin{aligned} \text{Var}\ (\underline{\hat{a}'\hat{\beta}}) &= \underline{a}'\text{var}\,(\hat{\beta})\,\underline{a} \\ &= \hat{\sigma}^2\underline{a}'(X'X)^{-1}\underline{a} \end{aligned}$$

Therefore the statistic

$$t = \frac{\underline{a}'\hat{\beta} - \delta^*}{\sqrt{\hat{\text{Var}}\ (\underline{\hat{a}'\hat{\beta}})}} \qquad \tilde{} t_{(N-t)}$$

Example:

To illustrate the method of analyzing one-way classification design with one of the treatments as control using dummy variables let us consider the data on rice yield given by Gomez and Gomez (1976). The original data which were in kg/ha were converted into tons/ha to reduce the number of figures in the computations.

TABLE 4. GRAIN YIELD OF RICE USING DIFFERENT FOLIAR AND GRANULAR INSECTICIDES FOR THE CONTROL OF BROWN PLANTHOPPERS AND STEM BORERS
(Design: CRD with n_i = reps).

Treatment	*Replication* 1	2	3	4	*Totals*	*Means*
1. Dol-Mix (1 kg)	2.5	2.1	2.1	1.8	8.5	2.125
2 Dol-Mix (2 kg)	3.4	2.6	2.2	2.5	10.7	2.675
3 DDT + δ – BHC	2.5	2.5	2.8	2.4	10.2	2.550
4 Azodrin	2.4	2.4	1.6	2.1	8.5	2.125
5 Dimeeron-Boom	2.0	1.7	1.6	1.9	7.2	1.80
6 Dimeeron-Knap	1.8	1.7	1.9	1.3	6.7	1.675
7 Control	1.4	1.5	1.3	1.1	5.3	1.325

$N = 28 \qquad y.. = 57.1$

From these data, the following statistics were computed using [7] and [8]:

(1)
$$X'X = \begin{pmatrix} 28 & 4 & 4 & 4 & 4 & 4 & 4 \\ 4 & 4 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 4 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 4 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 4 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 4 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix} \quad 7 \times 7$$

(2)
$$(X'X)^{-1} = \tfrac{1}{4}\begin{pmatrix} 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 2 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 2 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 & 2 & 1 \\ -1 & 1 & 1 & 1 & 1 & 1 & 2 \end{pmatrix} \quad 7 \times 7$$

(3)
$$X'\underline{y} = \begin{pmatrix} y. \\ y1. \\ y2. \\ y3. \\ y4. \\ y5. \\ y6. \end{pmatrix}_{7 \times 1} = \begin{pmatrix} 57.1 \\ 8.5 \\ 10.7 \\ 10.2 \\ 8.5 \\ 7.2 \\ 6.7 \end{pmatrix}_{7 \times 1}$$

$$\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \end{pmatrix} = \begin{pmatrix} \bar{y}_1 . - y_c . \\ \bar{y}_2 . - y_c . \\ \bar{y}_3 . - y_c . \\ \bar{y}_4 . - y_c . \\ \bar{y}_5 . - y_c . \\ \bar{y}_6 . - y_c . \end{pmatrix} = \begin{pmatrix} 1.325 \\ .80 \\ 1.35 \\ 1.225 \\ .80 \\ .475 \\ .350 \end{pmatrix}_{7 \times 1}$$

(5) Sums of Squares:

a. $SS\text{Total (uncorrected)} = \underline{y}'\underline{y}$
$= (2.5)^2 + (3.4)^2 + \ldots + (1.1)^2$

$= 123.91$

b. $SS_{\text{reg}} = \underline{\hat{\beta}}' X'\underline{y}$
$= (57.1)(1.325) + (8.5)('80) + \ldots + (5.3)(.350)$
$= 121.96$

b1) Due to $\hat{\beta}_0 = y.^2_.\,/N$
$= \dfrac{(57.1)^2}{28}$
$= 116.44$

b2) Due to $\hat{\beta}_1, \hat{\beta}_2, \ldots \hat{\beta}_6/\hat{\beta}_0 = \underline{\hat{\beta}}' X'\underline{y} - y.^2_.\,/N$
$= 121.96 - 116.44$
$= 5.52$

c. $SS_{\text{resid}} = SS_{\text{Total}} - SS_{\text{reg}}$

$= 123.91 - 121.96$

$= 1.95$

(6) ANOV Table

SV	*DF*	*SS*	*MS*
Regression	(7)	121.96	
Due to $\hat{\beta}_o$	1	116.44	
Due to $\hat{\beta}_1, \hat{\beta}_2 \ldots \hat{\beta}_6 / \hat{\beta}_0$	6	5.52	$MS_{reg} = 0.92$
Residual	21	1.95	$MS_{resid} = .09286$
Total	28	123.91	

(7) Test of Hypotheses

$H_0: \beta_1 = \beta_2 = \ldots = \beta_6 = 0$

$$F. = \frac{MS_{\text{reg}}}{MS_{\text{resid}}} = \frac{0.92}{0.09286} = 9.9^{**} \quad (P < .01)$$

Conclusion: The effects of treatments over the control are not all zero.

b. $H_0: \beta_i = \beta_{i'}$, e.g. $H_0: \beta_1 = \beta_2$

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{(.09286)(1/4 + 1/4)}} = \frac{.80 - 1.35}{0.2155} = -2.55^{**} \quad (P < .01)$$

Conclusion: The effect of treatment 1 over the control is not the same as the effect of treatment 2 over the control.

c. $H_0: \beta_i = 0$, e.g., $H_0: \beta_6 = 0$

$$t = \frac{\hat{\beta}_6}{\sqrt{(0.09286)(1/4 + 1/4)}} = \frac{0.350}{.2155} = 1.62^{NS} \quad (P > .05)$$

Conclusion: The effect of treatment 6 over the control is zero.

d. Consider $a' = (0, 1, 1, 1, -1, -1, -1)$,
$H_0: a'\hat{\beta} = 0$

$$t = \frac{a'\hat{\beta}}{\sqrt{\text{var}\,(a'\hat{\beta})}} = \frac{1.75}{\sqrt{(.09286)\frac{6}{4}}}$$

$$= \frac{1.75}{.3732}$$

$$= 4.689^{**} \quad (P < .01)$$

Conclusion: The effects of treatments 1, 2, and 3 over the control are not equal to the effects of treatments 4, 5 and 6 over the control.

SUMMARY AND CONCLUSION

If one uses the ordinary one-way analysis of variance linear model, the resulting matrix $X'X$ is not of full rank and therefore the parameter vector $\underline{\beta}$ cannot be estimated uniquely. By using dummy variables the resulting matrix $X'X$ is of full rank and yields a unique solution for $\underline{\beta}$.

By using dummy variables, the resulting estimates of the regression parameters are functions of the mean of control and the means of treated. In fact,the estimate of the intercept is the mean of control and the estimate of partial regression coefficients are deviations of effects of treatment means over the control mean.

The sum of squares due to the intercept is equivalent to the correction factor in the ordinary analysis of variance, whereas the sum of squares due to the partial regression coefficients is the same as the treatment sum of squares in the ordinary ANOV.

Any linear comparison among the partial regression coefficients in the dummy variable regression model is equivalent to the linear comparison among the treatment means by the ordinary analysis of variance.

The technique is particularly useful when one can not assume that the effects τ_i in the ordinary ANOV sum up to zero thus encountering the problem of singularity of the matrix $X'X$.

REFERENCES

1. Draper, N. and H. Smith (1966). Applied Regression Analysis, John Wiley and Sons, Inc., New York.
2. Gomez, K.A. and A. A. Gomez (1976). Statistical Procedures for Agricultural Research with emphasis on Rice, the International Rice Research Institute, Los Baños, Laguna.
3. Graybill, F.A. (1961). An Introduction fo Linear Statistical Models, Vol. 1 MacGraw-Hill Book Co., New York.
4. Johnston, J. (1960). Econometric Method, McGraw -Hill Book Co., New York.
5. Nie, N.H. et al (1970). Statistical Packages for Social Sciences, 2nd edition, Mc.Graw-Hill Book Co., New York.