Dummy variable multiple regression analysis of matched samples

Abstract
This paper presents and discusses the use of dummy variable multiple regression techniques in the analysis of samples drawn from several related or dependent populations ordinarily appropriate for random effects and mixed effects derived from the two factor analysis of variances model with one observation per-cell or treatment combinations. Using the extra sum of squares principle the method develops necessary sums of squares, degrees of freedom and the F-ratios required to test the significance of factor level effects thereby helping to resolve the problem of one observation per treatment combination, encountered in the usual two factor analysis of variance models with one observation per cell. The method also provides estimates of the overall and factor mean effects comparable to those obtained with the two factor analysis of variance methods. In addition, the method also provides estimates of the total or absolute effects as well as the direct and indirect effects of the independent variables or factors on the dependent or criterion variable which is not ordinarily obtainable with the usual analysis of variance techniques. The proposed method compares favorably with the usual Friedman’s two-way analysis of variance test by ranks using some sample data.

Keywords: friedman’s two-way ANOVA, mixed–effects ANOVA, dummy variable, regression, extra sum of square, treatment

Introduction

Dummy variable analysis of variance technique is an alternative approach to the non-parametric Friedman’s two-way analysis of variance test by ranks used to analyze sample data appropriate for use in parametric statistics for two factor random and mixed effects or analysis of variance models with one replication or observation per treatment combinations.1,2

To develop a non-parametric alternative method for the analysis of matched samples that are appropriate for use with two factor random and mixed-effects analysis of variance models with only one observation per cell or treatment combination, we may suppose that a researcher has collected a random sample of size ‘a’ observations randomly drawn from a population ‘A’ of subjects or blocks of subjects exposed to or observed at some ‘c’ time periods, points in space, experimental conditions, tests, or treatments that are either fixed or random and mixed-effects analysis of samples drawn from several related or dependent populations. This paper presents and discusses the use of dummy variable multiple regression for use with a two factor analysis of variance problem, we as usual present factor A, namely subject or block of subjects with ‘a’ levels is represented with a-1 dummy variables of 1s and 0s, while factor B with c levels is represented by c-1 dummy variables of 1s and 0s.

Now to set up a dummy variable multiple regression model for use with a two factor analysis of variance problem, we as usual present each factor or the so called parent independent variable with one dummy variable of 1s and 0s less than the number of its categories or levels.2 Thus factor A, namely subject or block of subjects with ‘a’ levels is represented with a-1 dummy variables of 1s and 0s, while factor B with c levels is represented by c-1 dummy variables of 1s and 0s.

Hence we may let

\[ x_{i,j} = \begin{cases} 1, & \text{if } y_{ij} \text{ is an observation on the } i \text{th subject or block of subjects} \\ 0, & \text{otherwise} \end{cases} \]  

\[ x_{j,a} = \begin{cases} 1, & \text{if } y_{ij} \text{ is an observation on the } j \text{th level of factor B (treatment)} \\ 0, & \text{otherwise} \end{cases} \]  

for \( i = 1, 2, ..., a \) and all \( j = 1, 2, ..., c \)

Also let

\[ x_{j,a} = \begin{cases} 1, & \text{if } y_{ij} \text{ is an observation or response at the } j \text{th level of factor B (treatment) and ith level of factor A(subject or block of subjects)} \\ 0, & \text{otherwise} \end{cases} \]  

for \( j = 1, 2, ..., c \) and all \( i = 1, 2, ..., a \).

Then the resulting dummy variable multiple regression model fitting or regressing the dependent or criterion variable \( y_{ij} \) on the dummy variables representing factors A (subject or block of subjects) and B (treatment) is

\[ y_{ij} = \beta_0 + \beta_{1,i} x_{1,i} + \beta_{2,j} x_{2,j} + \ldots + \beta_{a-1,i} x_{a-1,i} + \beta_{1,B} x_{1,B} + \beta_{2,B} x_{2,B} + \ldots + \beta_{c-1,B} x_{c-1,B} + \epsilon_{ij} \]  

For \( i = 1, 2, ..., n = a \cdot c \) sample observations where \( y_{ij} \) is the \( i \)th response or observation on the criterion or dependent variable; \( x_{ij} \) are dummy variables of 1s and 0s representing levels of factors A and B; \( \beta_{ij} \) are partial regression coefficients and \( \epsilon_{ij} \) are error terms, with \( \mathbb{E}(\epsilon_j) = 0 \) for \( i = 1, 2, ..., n = a \cdot c \).
observation per row by column, that is factor A (subject or block of subjects) by factor B (treatment) combination; for one to be able to have an estimate for the error term of squares for the regression model, and hence be able to test desired hypotheses, it is necessary to assume that there are no factors A by B interactions or that such interactions have been removed by an appropriate data transformation. Also note that an advantage of the present method over the extended median 

\[ E(y_i) = \beta_0 + \beta_{1,A} x_{1,i,A} + \beta_{2,A} x_{2,i,A} + \ldots + \beta_{a-1,A} x_{a-1,i,A} + \beta_{1,B} x_{1,i,B} + \beta_{2,B} x_{2,i,B} + \ldots + \beta_{c-1,B} x_{c-1,i,B} \] (4)

To find the expected or mean effect of any of the factors or parent independent variables, we set all the dummy variables representing that factor equal to 1 and all the other dummy variables found in equation 4 equal to 0. Thus for example the expected or mean effect or value of factor A (subject or block of subjects) on the dependent variable is obtained by setting \( x_{1,i,A} = 1 \) and \( x_{j,i,A} = 0 \) in equation 4 for \( i = 1, 2, \ldots, a-1 \).

Similarly the expected or mean value of factor B (treatment) is obtained by setting \( x_{j,i,B} = 1 \) and \( x_{j,i,B} = 0 \) in equation 4 for \( i = 1, 2, \ldots, c-1 \). thereby obtaining

\[ E(y_{i,j}) = \beta_0 + \sum_{i=1}^{c-1} \beta_{i,B} + \sum_{j=1}^{a-1} \beta_{l,B} \] (5)

Now the dummy variable multiple regression model of equation 3 can equivalently be expressed in matrix form as

\[ y = X\beta + e \] (6)

Where \( y \) is an nx1 column vector of observations or scores on the dependent or criterion variables; \( X \) is an nxr design matrix of 'r' dummy variables of 1s and 0s; \( \beta \) is an nx1 column vector of partial regression coefficients; and \( e \) is nx1 column vector of error terms, with \( E(e) = 0 \) where 'n' = number of observations and 'n'= number of factors a + b = number of dummy variables of 1s and 0s included in the regression model.

Similarly the expected value of \( y \) is from equation 4.

\[ E(y) = X\beta \] (7)

Application of the usual methods of least squares to either equation 3 or 6 yields an unbiased estimate of the regression parameter \( \beta \) as

\[ \hat{\beta} = \hat{b} = (XX)^{-1} X'y \] (8)

Where \( (XX)^{-1} \) is the inverse matrix of the non-singular variance-covariance matrix \( XX \). A usually of research interest is the regression model of either equation 3 or 6 fits, or equivalently that the independent variables or factors have no effects on the dependent or criterion variable, meaning that the partial regression coefficient is equal to zero stated symbolically that we have the null hypothesis.

\[ H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0 \] (9)

As in equation 3 this null hypothesis is tested using the usual F-test presented in an analysis of variance Table where the total sum of squares is calculated in the usual way as

\[ SST = y'y - n\bar{y}^2 \] (10)

With \( n=1-a.c-1 \) degrees of freedom where \( \bar{y} \) is the mean value of the dependent variables.

Similarly the treatment sum of squares in analysis of variance parlance which is the same as the regression sum of squares in regression models is calculated as

\[ SSTreatment = SSR = b'X'y - n\bar{y}^2 \] (11)

With \( (a-1)+(c-1) = a+c-2 \) degrees of freedom. The error sum of squares SSE indicates the difference between the total sum of squares, SST and the sum of squares regression SSR; thus,

\[ SSE = SST - SSR = y'y - b'X'y \] (12)

With \( (a.c-1)-(a-1)+(c-1)=(a-1)(c-1) \) degrees of freedom.

These results are summarized in an analysis of variance Table (Table 1)

The null hypotheses \( H_0 \) of Equation 13 is tested using the F-ratio of Table 1. The null hypothesis is rejected at the if the calculated F-ratio is greater than the tabulated or critical F-ratio at a specified \( \alpha \) -level of significance, otherwise the null hypothesis \( H_0 \) is accepted.

If the model fits, that if not all the elements of \( \beta \) are equal to zero, that is if the null hypothesis \( H_0 \) of equation 9 is rejected, then one may proceed to test further hypothesis concerning factor level effects, that is one may proceed to test the null hypothesis that factors A (subject or block of subjects) and B (treatment) separately have no effects on the dependent or criterion variable. In other words, the null hypotheses

\[ H_0 : \beta_A = 0 \quad \text{versus} \quad H_1 : \beta_A \neq 0 \] (13, 14)

\[ H_0 : \beta_B = 0 \quad \text{versus} \quad H_1 : \beta_B \neq 0 \]

Where \( \beta_A \) and \( \beta_B \) are respectively the (a-1) and (c-1) vectors of partial regression coefficients or effects of factor A (subject or block of subjects) and B (treatment) on the criterion or dependent variable. However a null hypothesis that is usually of greater interest here is that of equation 14, that is that treatments, points in time or space of tests or experiments do not have differential effects on subjects.
Table 1 Two factor analysis of variance Table for the full model of Equation 6

| Source of variation | Sum of squares | Degrees of freedom | Mean sum of squares | F-ratio |
|---------------------|---------------|--------------------|--------------------|--------|
| Regression(treatment) | $SSR = \sum b_j' X_jy - n\bar{y}^2$ | $a + c - 2$ | $MSR = \frac{SSR}{a + c - 2}$ | $MSR$ |
| Error | $SSE = \sum (y_j - b_j'X_j)^2$ | $(a - 1)(c - 1)$ | $MSE = \frac{SSE}{(a - 1)(c - 1)}$ | $MSE$ |
| Total | $SST = \sum (y_i - \bar{y})^2$ | $(a + c - 1)$ | | |

Now to obtain appropriate test statistics for use in testing these null hypotheses we apply the extra sum of squares principle to partition the treatment or regression sum of squares $SSR$ into its two component parts namely, the sum of squares due to factor A (subject or block of subjects), SSA and the sum of squares due to factor B (treatment), SSB, to enable the calculation of the appropriate F-ratios.

Now the nxr matrix $X$ for the full model of equation 6 can be partitioned into its two component sub-matrices namely $X_A$, an nx(a-1) design matrix of a-1 dummy variables of 1s and 0s representing the included a-1 levels of factor A (subject or block of subjects) and $X_B$, an nx(c-1) matrix of the c-1 dummy variables of 1s and 0s representing the included c-1 levels of factor B (treatment). The partial regression coefficient $\hat{b}$, estimated being an nx1 column vector of regression effects of equation 8 can also be partitioned into the corresponding partial regression coefficients estimated such as, $\hat{b}_A$, which is an (a-1)x1 column vector of partial regression coefficients or effects of factor A and $\hat{b}_B$, which is a (c-1)x1 column vector of the effects of factor B on the dependent variable. Hence the treatment sum of squares SST, that is the sum of squares regression SSR of equation 11 can be equivalently expressed as

$$SSTreatment = SSR = \sum b_j'X_jy - n\bar{y}^2 = (Xb)'b'y - n\bar{y}^2;$$

equivalently as

$$SSR = \left(\begin{array}{c}X_A \\ X_B \end{array}\right)'(b_A b_B)' \left(\begin{array}{c}X_A \\ X_B \end{array}\right) - n\bar{y}^2 = (b_A'X_A y + b_B'X_B y) - n\bar{y}^2$$

or equivalently

$$SSR = \sum b_j'X_jy - n\bar{y}^2 = (b_j'X_jy - n\bar{y}^2) + n\bar{y}^2$$

Which when interpreted is the same as the statement

$$SSTreatment = SSR = SSA + SSB + SS(F = \hat{\mu})$$

Where $SSR$ is the sum of squares of regression for the full model with $r = a + c - 2$ degrees of freedom; $SSA$ is the sum of squares due to factor A (subject or block of subject); with a-1 degrees of freedom; $SSB$ is the sum of squares due to factor B (treatment) with c-1 degrees of freedom; and $SS(F = \hat{\mu})$ 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 equations 6 with $X$, and the reduced regression models of $X_A$ and $X_B$ again separately on the criterion or dependent variable $y$.

Now if the full model of equation 6 fits, that is if the null hypothesis of equation 9 is rejected, then the additional null hypotheses of equations 13 and 14 may be tested using the extra sum of squares principle [4,5]. If we denote the sums of squares due to the full model of equation 6 and the reduced models due to the fitting of the criterion variables $y^*$ 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 of each reduced model and the error sum of squares due to a given factor is calculated as

$$ESS = SS(F) - SS(R)$$

(18)

With degrees of freedom obtained as the difference between the degrees of freedom of $SS(F)$ and $SS(R)$; that is as $Edf = df(F) - df(R)$. Thus the extra sums of squares for factors A (subject or block of subjects) and B (treatment) are obtained as follows respectively

$$ESSA = SSR - SSA; ESSB = SSR - SSB$$

(19)

With $(a - 1) + (c - 1) - (a - 1) = b - 1$ degrees of freedom and $(a - 1) + (c - 1) - (a - 1) - 1 = a - 1$ degrees of freedom.

Note that since each of the reduced models and the full model have the same total sum of squares SST, the extra sum of squares may alternatively be obtained as the difference between the error sum of squares of each reduced model and the error sum of squares of the full model. In other words, the extra sum of squares is equivalently calculated as

$$ESS = SS(F) - SS(R) = SST - SS(F) - SS(R) = SSE(R)$$

(20)

With degrees of freedom similarly obtained. Thus the extra sum of squares due to factors A (subject or block of subjects) and B (treatment) are alternatively obtained as follows respectively.

$$ESSA = SSA - SSE$$

(21)

With c-1 and a-1 degrees of freedom. Where SSR and SSE are respectively the regression sum of squares and the error sum of squares for the full model and SSEA and SSEB are respectively the
error sums of squares for the reduced models for factors A and B. The
null hypotheses of equations 13 and 14 are tested using the F-ratios

\[ F_A = \frac{MESA}{MSE} \]  

(22)

With a-1 and (a-1)(c-1) degrees of freedom where

\[ MESA = \frac{ESSA}{c-1} \]  

(23)

Is the mean extra sum of squares due to factor A (subject or block of
subjects) and

\[ F_B = \frac{MESB}{MSE} \]  

(24)

With a-1 and (a-1)(c-1) degrees of freedom where

\[ MESB = \frac{ESSB}{a-1} \]  

(25)

Is the mean extra sum of squares due to factor B (treatment). These
results are summarized in Table 2a which for ease of presentation also
includes the sum of squares and other values of Table 1 for the full
models.

If the various F–ratios and in particular the F-ratios based on
the extra sums of squares of Table 2b indicate that the independent
variables or factor levels have differential effects on the response,
dependent, or criterion variable, that is if the null hypotheses of either
equation 13 or 14 or both are rejected, then one may proceed further
to estimate desired factor level effects and test hypotheses concerning
them.

### Table 2a

Table showing two factor Analysis of Variance for Sums of Squares for the full model and due to reduced models and other statistics

| Source of variation | Sum of squares (SS) | Degrees of freedom(DF) | Mean sum of squares(MS) | F-ratio |
|---------------------|---------------------|------------------------|------------------------|---------|
| Full model          |                     |                        |                        |         |
| Regression          | \( b'y \bar{X}'y - \bar{y}^2 \) | a+c-2                  | \( MSR = \frac{SSR}{a+c-2} \) | \( F = \frac{MSR}{SSR} \) |
| Error               | \( SSE = y'\bar{y} - b'y \bar{X}'y \) | (a-1)(c-1)             | \( MSE = \frac{SSE}{(a-1)(c-1)} \) |         |
| Factor A (Subjects on block of subjects) | | | | |
| Regression          | \( b'y \bar{X}'_A y - n \bar{y}^2 \) | a-1                    | \( MSA = \frac{SSA}{a-1} \) |         |
| Error               | \( SSEA = y'\bar{y} - b'y \bar{X}'_A y \) | a(c-1)                 | \( MSEA = \frac{MSA}{a(c-1)} \) |         |
| Factor B (Treatment) | | | | |
| Regression          | \( b'y \bar{X}'_B y - n \bar{y}^2 \) | c-1                    | \( MSB = \frac{SSB}{c-1} \) | \( F = \frac{MSB}{MSEB} \) |
| Error               | \( SSEB = y'\bar{y} - b'y \bar{X}'_B y \) | c(a-1)                 | \( MSEP = \frac{MSEB}{c(a-1)} \) |         |
| Total               | \( y'\bar{y} - n \bar{y}^2 \) | a(c-1)                 |                        |         |

### Table 2b

Two-factor Analysis of Variance Table for the Extra sums of Squares due to reduced models and other statistics (Continuation)

| Extra sum of squares | Degrees of freedom(DF) | Extra mean sum of squares (EMSA) | F-ratio |
|----------------------|------------------------|---------------------------------|---------|
| ESR=SSR              | a+c-2                  | \( EMSR = \frac{SSR}{a+c-2} \) | \( F = \frac{MSR}{MSR} \) |
| ESER=SSB             | (a-1)(c-1)             | \( EMSB = \frac{SSB}{(a-1)(c-1)} \) |         |

Factor A

| ESSA=SSR-SSA         | c-1                    | \( EMSA = \frac{ESSA}{c-1} \) | \( F = \frac{EMS}{EMS} \) |
| ESSEA-SEESSA=ESEA    | c-1                    | \( EMS = \frac{ESSEA}{c-1} \) |         |

Citation: Okeh UM, Oyeka ICA. Dummy variable multiple regression analysis of matched samples. Biom Biostat Int J. 2016;3(5):158–165.
DOI: 10.15406/bbij.2016.03.00077
Dummy variable multiple regression analysis of matched samples

Table Continued

| Table Continued |
|-----------------|
| **Extra sum of squares (ESS=SS(F)-SS(R))** | **Degrees of freedom (DF)** | **Extra mean sum of squares (EMSA)** | **F-ratio** |
| **Factor B** | | | |
| ESSB=SSR-SSB | a-1 | EMSB=ESSB/a-1 | F_b=EMSB/MSE |
| ESSEB=SSEB-SSE=ESSB | a-1 | EMSEB=ESSB/a-1 | |

In fact an additional advantage of using dummy variable regression models in two factor or multiple factor analysis of variance type problems is that the method also more easily enables the estimation of factor level effects separately of several factors on a specified dependent or criterion variable. For example it enables the estimation of the total or absolute effect, the partial regression coefficient or the so called direct effect of a given independent variable here referred to as the parent independent variable on the dependent variable through the effect of its representative dummy variables as well as the indirect effect of that parent independent variable through the mediation of other independent variables in the model. The total or absolute effect of a parent independent variable on a dependent variable is estimated as the simple regression coefficient of that independent variable represented by codes assigned to its various categories when regressed on the dependent variable. The direct effect of a parent independent variable on a dependent variable is the weighted sum of the partial regression coefficients or effects of the dummy variables representing that parent independent variable on the dependent variable, where the weights are the simple regression coefficients of each representative dummy variable regressing on the specified parent independent variable represented by codes. The indirect effect of a given parent independent variable on a dependent variable is then simply the difference between its total and direct effects.

Now the direct effect or partial regression coefficient of a given parent independent variable on a dependent variable is obtained by taking the partial derivative of the expected value of the corresponding regression model with respect to that parent independent variable. For example the direct effect of the parent independent variable ‘A’ say on the dependent variable Y is obtained from equation 5 as

$$\beta = \frac{dE(Y)}{dA} = \frac{\sum_{i=1}^{a} \beta_{i,A} dE(x_{i,A})}{dA} + \sum_{i=1}^{a} \beta_{i,Z} \frac{dE(x_{i,Z})}{dA}$$

or

$$\beta = \frac{\sum_{i=1}^{a} \beta_{i,A} dE(x_{i,A})}{dA}$$

or

$$\frac{\sum_{i=1}^{a} \beta_{i,A} dE(x_{i,A})}{dA} + \sin \sum_{i=1}^{a} \beta_{i,Z} \frac{dE(x_{i,Z})}{dA} = 0$$

For all other independent variable ‘Z’ in the model different from ‘A’,

The weight \( \frac{dE(x_{i,Z})}{dA} \) is estimated by fitting a simple regression line of dummy variable, \( x_{i,A} \) regressing on its parent independent variable, A represented by codes and taking the derivative of its expected value with respect to ‘A’. Thus, if the expected value of the dummy variable \( x_{i,A} \) regressing on its parent independent variable ‘A’ is expressed as \( E(x_{i,A}) = c_{0} + \alpha_{i,A} \cdot A \)

Then the derivative of this expected value with respect to A is

$$\frac{dE(x_{i,A})}{dA} = \alpha_{i,A}$$

(27)

Hence using Equation 27 in Equation 26 gives the direct effect of the parent independent variable A on the dependent variable Y as

$$\beta_{A} \frac{dA}{dA} = \sum_{i=1}^{a} \alpha_{i,A} \frac{dA}{dA}$$

(28)

Whose sample estimate is from Equation 8

$$\hat{\beta}_{A} = b_{A} \frac{dA}{dA} = \sum_{i=1}^{a} c_{i,A} b_{i,A}$$

(29)

The total or absolute effect of ‘A’ on ‘Y’ is estimated as the simple regression coefficient or effect of the parent independent variable ‘A’ represented by codes on the dependent variable ‘Y’ as

$$\hat{\beta}_{A} = b_{A}$$

(30)

Where \( b_{A} \) is the estimated simple regression coefficient or effect of ‘A’ on ‘Y’. The indirect effect of ‘A’ on ‘Y’ is then estimated as the difference between \( \frac{dA}{dA} \) and \( b_{A} \frac{dA}{dA} \), that is as

$$\hat{\beta}_{A} \frac{dA}{dA} = b_{A} \frac{dA}{dA} = b_{A} - b_{A} \frac{dA}{dA}$$

(31)

The total, direct and indirect effects of factor B are similarly estimated.

**Illustrative example 1**

The body weights of a random sample of 10 Broilers here termed “subject or block of subjects” regarded as factor ‘A’ with ten levels and types of weighing machine here termed “treatment” regarded as factor ‘B’ with five levels are shown below.

To set up a dummy variable regression model of body weight (y) regressing on “subject or block of subjects” here termed factor ‘A’ with ten levels and types of weighing machine, here termed “treatments” treated as factor ‘B’ with five levels, we as usual represent factor ‘A’ with nine dummy variables of 1s and 0s and factor ‘B’ with four dummy variables of 1s and 0s, using Equation 1.

**Citation:** Okeh UM, Oyeka ICA. Dummy variable multiple regression analysis of matched samples. Biom Biostat Int J. 2016;3(5):158–165.

DOI: 10.15406/bbij.2016.03.00077
The resulting design matrix ‘X’ for the full model is presented in Table 3 where \( x_{1,i} \) represents level 1 or broiler No.1; \( x_{1,B} \) represents levels 9 or broiler No.9 and so on. Similarly \( x_{1,B} \) represents weighing machine No.1 or treatment 1, \( x_{2,B} \) represents weighing machine No.2 or treatment 2 and so on, until \( x_{4,B} \) represents weighing machine No.4 or treatment 4.

Using the design matrix X of Table 3 for the full model of Equation 6 we obtain the fitted regression Equation expressing the dependent of broiler body weight on, that is as a function of broiler (subject) treated as factor A and type of weighing machine (treatment) treated as factor B, both represented by dummy variables of 1s and 0s, as

\[
\hat{y}_i = 2.302 - 0.593x_{1,i} + 3.175x_{2,i} + 0.212x_{3,i} - 2.023x_{4,i} - 1.491x_{7,B} + 0.352x_{6,A}
\]

\[
-1.219x_{2,i} + 0.123x_{9,A} - 2.185x_{4,i} - 0.094x_{8,A} - 0.235x_{12,B} + 2.329x_{3,B} - 0.029x_{14,B}
\]

Now to estimate the total or absolute effect of type of weighing machine (treatment), ‘B; or body weight \( y \) of broilers, we regress \( i=y \) on ‘B’ represented by codes to obtain \( \hat{\beta}_B = 0.054 \). The weights \( \alpha_{i,B} \) to be applied to Equation 6 to determine the direct effect are obtained as explained above by taking the derivative with respect to ‘B’ of the expected value of the simple regression equation expressing the dependence of the dummy variable \( x_{ij} \) of 1s and 0s on its parent variable ‘B’ represented by codes yielding

\[
\alpha_{1,B} = -0.20, \alpha_{2,B} = -0.10, \alpha_{3,B} = 0.00 \text{ and } \alpha_{4,B} = 0.10 .
\]

Using these values in Equation 6, we obtain with Equation 6 the partial or the so called direct effect of type of weighing machine (treatment) ‘B’ on body weight ‘y’ of broilers as

\[
\hat{\beta}_{B \text{dir}} = \hat{\beta}_B \cdot \alpha_{i,B} = 0.0146 \text{.}
\]

Hence the corresponding indirect effect is estimated using Equation 6 as

\[
\hat{\beta}_{B \text{indir}} = \hat{\beta}_B \cdot \alpha_{i,B} = 0.0146 .
\]

The total or absolute, direct and indirect effects of the subjects or block of subjects called factor A are similarly calculated.

It would for comparative purpose be instructive to also analyze the data of example 1 using Friedman two-way analysis of variance test by ranks.

To do this we first rank for each broiler (subject) the body weight as obtained using the five weighing machines (treatment) from the smallest ranked ‘1’ to the largest ranked ‘5’. All tied body weights for each broiler are as usual assigned their mean ranks. The results are presented in Table 4.

Using the ranks shown in Table 4, we calculate the Friedmans test statistic as

\[
\hat{\chi}^2 = \frac{12}{(c-1)} \sum_{i=1}^{c} R_i^2 - 3r(e+1) = \frac{12(1^2+3^2+3^2+27^2+40.5^2+36.5^2)}{(10)(5)(5+1)} - 3(10)(5+1) = 198.38 - 180 = 17.38
\]

Which with \( c-1=5-1=4 \) degrees of freedom is statistically significant \( \chi^2_{0.004,4}=13.277 \), indicating that weighing machines probability differ in the values of body weights of broilers obtained using them. This is the same conclusion that is also reached using the present method.

Table 3 Design matrix for the sample data of example 1

| S/no(\( l \)) | Body weight \( y_i \) | \( x_{1,i} \) | \( x_{2,i} \) | \( x_{3,i} \) | \( x_{4,i} \) | \( x_{5,i} \) | \( x_{6,A} \) | \( x_{7,B} \) | \( x_{8,A} \) | \( x_{9,A} \) | \( x_{10,B} \) | \( x_{11,B} \) | \( x_{12,B} \) | \( x_{13,B} \) | \( x_{14,B} \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 2.1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 | 2.1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1.9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 1.7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 7 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 1.8 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 2.1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S/no (I) | Body weight (yi) | $\lambda_{1:4}$ |
|---------|-----------------|----------------|
| 10      | 2               | 1 0 0 0 0 0 0 0|
| 11      | 1.9             | 1 0 0 0 0 0 0 0|
| 12      | 2.2             | 1 0 0 0 0 0 0 0|
| 13      | 1.9             | 1 0 0 0 0 0 0 0|
| 14      | 2.2             | 1 0 0 0 0 0 0 0|
| 15      | 2.2             | 1 0 0 0 0 0 0 0|
| 16      | 1.8             | 1 0 0 0 0 0 0 0|
| 17      | 2.2             | 1 0 0 0 0 0 0 0|
| 18      | 2.1             | 1 0 0 0 0 0 0 0|
| 19      | 2               | 1 0 0 0 0 0 0 0|
| 20      | 2.1             | 1 0 0 0 0 0 0 0|
| 21      | 1.9             | 1 0 0 0 0 0 0 0|
| 22      | 1.8             | 1 0 0 0 0 0 0 0|
| 23      | 1.9             | 1 0 0 0 0 0 0 0|
| 24      | 2.2             | 1 0 0 0 0 0 0 0|
| 25      | 2.1             | 1 0 0 0 0 0 0 0|
| 26      | 1.8             | 1 0 0 0 0 0 0 0|
| 27      | 2               | 1 0 0 0 0 0 0 0|
| 28      | 2.1             | 1 0 0 0 0 0 0 0|
| 29      | 2.1             | 1 0 0 0 0 0 0 0|
| 30      | 2.1             | 1 0 0 0 0 0 0 0|
| 31      | 1.8             | 1 0 0 0 0 0 0 0|
| 32      | 2.1             | 1 0 0 0 0 0 0 0|
| 33      | 1.9             | 1 0 0 0 0 0 0 0|
| 34      | 2.2             | 1 0 0 0 0 0 0 0|
| 35      | 2               | 1 0 0 0 0 0 0 0|
| 36      | 1.7             | 1 0 0 0 0 0 0 0|
| 37      | 2.1             | 1 0 0 0 0 0 0 0|
| 38      | 1.9             | 1 0 0 0 0 0 0 0|
| 39      | 1.9             | 1 0 0 0 0 0 0 0|
| 40      | 2.1             | 1 0 0 0 0 0 0 0|
| 41      | 1.8             | 1 0 0 0 0 0 0 0|
| 42      | 1.9             | 1 0 0 0 0 0 0 0|
| 43      | 2               | 1 0 0 0 0 0 0 0|
| 44      | 2.1             | 1 0 0 0 0 0 0 0|
| 45      | 2.1             | 1 0 0 0 0 0 0 0|
| 46      | 2               | 1 0 0 0 0 0 0 0|
| 47      | 2.1             | 1 0 0 0 0 0 0 0|
| 48      | 2               | 1 0 0 0 0 0 0 0|
| 49      | 2.1             | 1 0 0 0 0 0 0 0|
| 50      | 2.1             | 1 0 0 0 0 0 0 0|

Table Continued...
Summary and conclusion

This paper has proposed the use of dummy variable multiple regression methods for the analysis of several related or dependent samples appropriate for random effects and mixed effects two factor analysis of variance with one observation per cell or treatment combination.

Using the extra sum of squares principle, the method developed necessary sums of squares, degrees of freedom and the F-ratios required in testing for the significance of factor level effects.

The method provided estimates of the overall and factor mean effects comparable to those obtained with the two factor analysis of variance method. In addition the method also provided estimates of the total or absolute effects as well as the direct and indirect effects of the independent variables or factors on the dependent or criterion variable which are not ordinarily obtainable with the usual analysis of variance techniques. The proposed method is illustrated with some sample data and shown to compare favorably with the usual Friedmans two-way analysis of variance test by ranks often used for the same purpose.

Conflict of interest

None.

References

1. Oyeka ICA. Ties Adjusted Two way Analysis of Variance tests with unequal observations per cell. Science Journal of Mathematics & Statistics. 2013;(2013):1–6.
2. Boyle, Richard P. Path Analysis and Ordinal Data. American Journal of Sociology. 1970;75(4):461–480.
3. Oyeka ICA, Afuecheta EO, Ebuh GU, et al. Partitioning the total chi-square for matched Dichotomous Data. International Journal of Mathematics and Computations. 2012;16(3):41–50.
4. Draper NR, Smith H. Applied Regression Analysis. John Wiley & sons, New York, USA. 1966.
5. Neter J, Wasserman W, Christopher N. Applied Linear Statistical Models. New York, USA. 1974.
6. Wright, Sewall. The Methods of Path Coefficients. The Annals of Mathematical Statistics. 1934;5(3):161–215.