

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 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 are 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

Volume 3 Issue 5 - 2016

 Okeh UM,¹ Oyeka ICA²
¹Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Nigeria

²Department of Statistics, Nnamdi Azikiwe University, Nigeria

Correspondence: Okeh UM, Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Abakaliki Nigeria, Email uzomaokay@gmail.com

Received: April 03, 2016 | **Published:** May 23, 2016

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 randomly drawn from population B experimental conditions, points in time, tests or experiments comprising numerical measurements.

The proposed method

Let y_{ij} be the i^{th} observation drawn from population A, that is the observation on the i^{th} subject or block of subjects exposed to or observed at the j^{th} level of factor B that is j^{th} treatment or time period for $i=1,2,\dots,a$; $j=1,2,\dots,c$.

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

$$x_{i,A} = \begin{cases} 1, \text{if } y_{ij} \text{ is an observation on the } i\text{th subject or block of} \\ \text{subjects and } j\text{th level of factor B (treatment)} \\ 0, \text{otherwise} \end{cases} \quad (1)$$

for $i = 1, 2, \dots, a-1$; and all $j = 1, 2, \dots, c$

Also let

$$x_{j,B} = \begin{cases} 1, \text{if } y_{ij} \text{ is an observation or response at the } j\text{th level} \\ \text{of factor B (treatment) and } i\text{th level of factor} \\ A(\text{subject, or block of subjects}) \\ 0, \text{otherwise} \end{cases} \quad (2)$$

for $j = 1, 2, \dots, c-1$; and all $i = 1, 2, \dots, a$.

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

$$y_i = \beta_0 + \beta_{1,A} x_{11,A} + \beta_{2,A} x_{12,A} + \dots + \beta_{a-1,A} x_{1a-1,A} + \beta_{1,B} x_{11,B} + \beta_{2,B} x_{12,B} + \dots + \beta_{c-1,B} x_{1c-1,B} + e_i \quad (3)$$

For $l=1,2,\dots,n=a.c$ sample observations where y_l is the l^{th} response or observation on the criterion or dependent variable; x_{ls}

are dummy variables of 1s and 0s representing levels of factors A and B; β_{ls} are partial regression coefficients and e_{ls} are error terms, with $E(e_i)=0$, for $l=1,2,\dots,n=a.c$. Note that since there are only one

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 sum 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

test for dependent or matched samples and also over the Friedmans two –way analysis of variance test by ranks is that the problem of tied observations within subjects or blocks of subjects does not arise, and hence unlike in the other two non-parametric methods under reference there is no need to find ways to adjust for or break ties between scores within blocks of subjects.³ The expected or mean value of the criterion variable is from equation 3.

$$E(y_l) = \beta_0 + \beta_{1;A}x_{1l;A} + \beta_{2;A}x_{2l;A} + \dots + \beta_{a-1;A}x_{(a-1)l;A} + \beta_{1;B}x_{1l;B} + \beta_{2;B}x_{2l;B} + \dots + \beta_{c-1;B}x_{(c-1)l;B} \quad (4)$$

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 $x_{l;A}=1$ and $x_{j;B}=0$ in equation 4 for $l=1,2,\dots,a-1; j=1,2,\dots,c-1$.

Similarly the expected or mean value of factor B (treatment) is obtained by setting $x_{l;B}=1$ and $x_{j;A}=0$ in equation 4 for $l=1,2,\dots,c-1; j=1,2,\dots,a-1$ thereby obtaining

$$E(y_{l;A}) = \beta_0 + \sum_{l=1}^{a-1} \beta_{l;A} \text{ and } E(y_{l;B}) = \beta_0 + \sum_{l=1}^{c-1} \beta_{l;B} \quad (5)$$

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

$$\underline{y} = X\underline{\beta} + \underline{e} \quad (6)$$

Where \underline{y} is an $n \times 1$ column vector of observations or scores on the dependent or criterion variables; X is an $n \times r$ design matrix of 'r' dummy variables of 1s and 0s; $\underline{\beta}$ is an $r \times 1$ column vector of partial regression coefficients; and \underline{e} is on $n \times 1$ column vector of error terms, with $E(\underline{e}) = \underline{0}$ where 'n'=a.c observations and 'n'=(a-1)+(c-1)=a+c-2 dummy variables of 1s and 0s included in the regression model.

Similarly the expected value of \underline{y} is from equation 4.

$$E(\underline{y}) = X\underline{\beta} \quad (7)$$

Application of the usual methods of least squares to either equation 3 or 6 yields an unbiased estimate of the regression parameter $\underline{\beta}$ as

$$\hat{\underline{\beta}} = \underline{b} = (X'X)^{-1} X'\underline{y} \quad (8)$$

Where $(X'X)^{-1}$ is the inverse matrix of the non-singular variance-covariance matrix $X'X$. A hypothesis that is usually of research interest is that 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.

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

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

$$SSTotal = \underline{y}'\underline{y} - n.\bar{y}^2 \quad (10)$$

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

$$SSTreatment = SSR = \underline{b}'X'\underline{y} - n.\bar{y}^2 \quad (11)$$

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,

$$SSE = SST - SSR = \underline{y}'\underline{y} - \underline{b}'X'\underline{y} \quad (12)$$

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 α -level of significance, otherwise the null hypothesis H_0 is accepted.

If the model fits, that if not all the elements of $\underline{\beta}$ 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

$$H_0 : \underline{\beta}_A = \underline{0} \text{ versus } H_1 : \underline{\beta}_A \neq \underline{0} \quad \text{and} \quad (13, 14)$$

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

Where $\underline{\beta}_A$ and $\underline{\beta}_B$ 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)	$SSR = \underline{b}'_A X'_A \underline{y} - n \bar{y}^2$	$a+c-2$	$MSR = \frac{SSR}{a+c-2}$	$\frac{MSR}{MSE}$
Error	$SSE = \underline{y}' \underline{y} - \underline{b}'_A X'_A \underline{y}$	$(a-1)(c-1)$	$MSE = \frac{SSE}{(a-1)(c-1)}$	
Total	$SST = \underline{y}' \underline{y} - n \bar{y}^2$	$(a.c)-1$		

Now to obtain appropriate test statistics for use in testing these null hypothesis 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 \underline{b}_A , estimated being an rx1 column vector of regression effects of equation 8 can also be partitioned into the corresponding partial regression coefficients estimated such as, \underline{b}_A , which is an (a-1)x1 column vector of partial regression coefficients or effects of factor A and \underline{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

$$SSTreatment = SSR = \underline{b}'_A X'_A \underline{y} - n \bar{y}^2 = (\underline{X} \underline{b})' \underline{y} - n \bar{y}^2;$$

equivalently as

$$SSR = \left(\begin{pmatrix} X_A & X_B \end{pmatrix} \begin{pmatrix} \underline{b}_A \\ \underline{b}_B \end{pmatrix} \right)' \underline{y} - n \bar{y}^2 = \left(\underline{b}'_A X'_A \underline{y} + \underline{b}'_B X'_B \underline{y} \right) - n \bar{y}^2 \quad (15)$$

or equivalently

$$SSR = \underline{b}'_A X'_A \underline{y} - n \bar{y}^2 = \left(\underline{b}'_A X'_A \underline{y} - n \bar{y}^2 \right) + \left(\underline{b}'_B X'_B \underline{y} - n \bar{y}^2 \right) + n \bar{y}^2 \quad (16)$$

Which when interpreted is the same as the statement

$$SSTreatment = SSR = SSA + SSB + SS(\bar{y} = \hat{\mu}) \quad (17)$$

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 $SS(\bar{y}=\hat{\mu})$ 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 equations 6 with X, and the reduced regression models of X_A and X_B again separately on the criterion or dependent variable \underline{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 \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^{4,5} the extra sum of squares due to a given factor is calculated as

$$ESS = SS(F) - SS(R) \quad (18)$$

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

$$ESSA=SSR-SSA; ESSB=SSR-SSB \quad (19)$$

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

$$ESS=SS(F)-SS(R)=SST-SS(F)-SST-SS(R)=SSE(R)-SSE(F) \quad (20)$$

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.

$$ESSA = SSEA - SSE \quad (21)$$

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

$$F_A = \frac{MESA}{MSE} \quad (22)$$

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

$$MESA = \frac{ESSA}{c-1} \quad (23)$$

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

$$F_B = \frac{MESB}{MSE} \quad (24)$$

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

$$MESB = \frac{ESSB}{a-1} \quad (25)$$

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	$SSR = \underline{b}' X' \underline{y} - n \bar{y}^2$	a+c-2	$MSR = \frac{SSR}{a+c-2}$	$F = \frac{MSR}{SSR}$
Error	$SSE = \underline{y}' \underline{y} - \underline{b}' X' \underline{y}$	(a-1)(c-1)	$MSE = \frac{SSE}{(a-1)(c-1)}$	
Factor A (Subjects on block of subjects)				
Regression	$SSA = \underline{b}'_A X'_A \underline{y} - n \bar{y}^2$	a-1	$MSA = \frac{SSA}{a-1}$	
Error	$SSEA = \underline{y}' \underline{y} - \underline{b}'_A X'_A \underline{y}$	a(c-1)	$MSEA = \frac{MSA}{a(c-1)}$	
Factor B(Treatment)				
Regression	$SSB = \underline{b}'_B X'_B \underline{y} - n \bar{y}^2$	c-1	$MSB = \frac{SSB}{c-1}$	$F = \frac{MSB}{MSEB}$
Error	$SSEB = \underline{y}' \underline{y} - \underline{b}'_B X'_B \underline{y}$	c(a-1)	$MSEB = \frac{MSEB}{c(a-1)}$	
Total	$\underline{y}' \underline{y} - n \bar{y}^2$	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 (ESS=SS(F)-SS(R))	Degrees of freedom(DF)	Extra mean sum of squares (EMSA)	F-ratio
ESR=SSR	a+c-2	$EMSR = \frac{SSR}{a+c-2}$	$F = \frac{MSR}{MSE}$
ESER=SSE	(a-1)(c-1)	$EMSE = \frac{SSE}{(a-1)(c-1)}$	
Factor A			
ESSA=SSR-SSA	c-1	$EMSA = \frac{ESSA}{c-1}$	$F_A = \frac{EMSA}{MSE}$
ESSEA=SSEA-SSE=ESSA	c-1	$EMSEA = \frac{ESSEA}{c-1}$	

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	$EMS_B = \frac{ESSB}{a-1}$	$F_B = \frac{EMS_B}{MSE}$
ESSEB=SSEB-SSE=ESSB	a-1	$EMSEB = \frac{ESSEB}{a-1}$	
$\underline{y}'\underline{y} - n.\bar{y}^2$	a.c-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.⁶ 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.⁶

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

$$\beta_A \text{ dir} = \frac{dE(y_i)}{d_A} = \sum_{l=1}^{a-1} \beta_{l:A} \cdot \frac{dE(x_{l:A})}{d_A} + \sum_l \beta_{l:Z} \cdot \frac{dE(x_{l:Z})}{d_A} \quad (26)$$

or

$$\beta_A \text{ dir} = \sum_{l=1}^{a-1} \beta_{l:A} \cdot \frac{dE(x_{l:A})}{d_A} + \sin ce \sum_l \beta_{l:Z} \cdot \frac{dE(x_{l:Z})}{d_A} = 0$$

For all other independent variable 'z' in the model different from 'A'.

The weight $\alpha_{l:A} = \frac{dE(x_{l:A})}{d_A}$ is estimated by fitting a simple regression

line of dummy variable. $x_{l:A}$ 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 $x_{l:A}$ regressing on its parent independent variable 'A' is expressed as $E(x_{l:A}) = \alpha_0 + \alpha_{l:A} \cdot A$

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

$$\frac{dE(x_{l:A})}{d_A} = \alpha_{l:A} \quad (27)$$

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

$$\beta_A \text{ dir} = \sum_{l=1}^{a-1} \alpha_{l:A} \cdot \beta_{l:A} \quad (28)$$

Whose sample estimate is from Equation 8

$$\hat{\beta}_A \text{ dir} = b_A \text{ dir} = \sum_{l=1}^{a-1} \alpha_{l:A} \cdot b_{l:A} \quad (29)$$

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

$$\hat{\beta}_A = b_A \quad (30)$$

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 b_A and $b_A \text{ dir}$, that is as

$$\hat{\beta}_A \text{ indir} = b_A \text{ indir} = b_A - b_A \text{ dir} \quad (31)$$

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.

The resulting design matrix 'X' for the full model is presented in Table 3 where $x_{1;A}$ 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.

$$\hat{y}_i = 2.302 - 0.593x_{1;A} + 3.175x_{2;A} + 0.212x_{3;A} - 2.023x_{4;A} - 1.491x_{5;A} + 0.352x_{6;A} - 1.219x_{7;A} + 0.123x_{8;A} - 2.185x_{9;A} - 0.094x_{1;B} - 0.235x_{2;B} + 2.329x_{3;B} - 0.029x_{4;B}$$

Now to estimate the total or absolute effect of type of weighing machine (treatment), 'B' or body weight y of broilers, we regress y_i on 'B' represented by codes to obtain $\beta_B = b_B = 0.054$. The weights $\alpha_{j;B}$ 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_{ji} of 1s and 0s on its parent

$$\hat{\beta}_B \text{dir} = b_B \text{dir} = (-0.094 \times -0.2) + (-0.235 \times -0.10) + (0.00 \times 2.329) + (-0.029 \times 0.10) \hat{\beta}_B \text{dir} = b_B \text{dir} = 0.0394$$

Hence the corresponding indirect effect is estimated using Equation 6 as

$$\hat{\beta}_B \text{indir} = b_B \text{indir} = 0.0146.$$

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.

$$\chi^2 = \frac{12}{rc(c+1)} \sum_{j=1}^c R_j^2 - 3r(c+1) = \frac{12(13^2 + 33^2 + 27^2 + 40.5^2 + 36.5^2)}{(10)(5)(5+1)} - 3(10)(5+1) = 198.38 - 180 = 17.38$$

Which with $c-1=5-1=4$ degrees of freedom is statistically significant ($\chi_{0.99;4}^2 = 13.277$), indicating that weighing machines probability differ in

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

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

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

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 (yi)	$x_{l8;A}$													
		1	2	3	4	5	6	7	8	9	1	2	3	4	
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	1	0	0
3	2.1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
4	2.1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
5	1.9	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	1.7	1	0	1	0	0	0	0	0	0	0	1	0	0	0
7	2	1	0	1	0	0	0	0	0	0	0	0	1	0	0
8	1.8	1	0	1	0	0	0	0	0	0	0	0	0	1	0
9	2.1	1	0	1	0	0	0	0	0	0	0	0	0	0	1

Table Continued

S/no (I)		Body weight (yi)	$x_{i8:A}$													
			1	2	3	4	5	6	7	8	9	1	2	3	4	
10	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	
11	1.9	1	0	0	1	0	0	0	0	0	0	1	0	0	0	
12	2.2	1	0	0	1	0	0	0	0	0	0	0	1	0	0	
13	1.9	1	0	0	1	0	0	0	0	0	0	0	0	1	0	
14	2.2	1	0	0	1	0	0	0	0	0	0	0	0	0	1	
15	2.2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	
16	1.8	1	0	0	0	1	0	0	0	0	0	1	0	0	0	
17	2.2	1	0	0	0	1	0	0	0	0	0	0	1	0	0	
18	2.1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	
19	2	1	0	0	0	1	0	0	0	0	0	0	0	0	1	
20	2.1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	
21	1.9	1	0	0	0	0	1	0	0	0	0	1	0	0	0	
22	1.8	1	0	0	0	0	1	0	0	0	0	0	1	0	0	
23	1.9	1	0	0	0	0	1	0	0	0	0	0	0	1	0	
24	2.2	1	0	0	0	0	1	0	0	0	0	0	0	0	1	
25	2.1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	
26	1.8	1	0	0	0	0	0	1	0	0	0	1	0	0	0	
27	2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	
28	2.1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	
29	2.1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	
30	2.1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	
31	1.8	1	0	0	0	0	0	0	1	0	0	1	0	0	0	
32	2.1	1	0	0	0	0	0	0	1	0	0	1	1	0	0	
33	1.9	1	0	0	0	0	0	0	1	0	0	0	0	1	0	
34	2.2	1	0	0	0	0	0	0	1	0	0	0	0	0	1	
35	2	1	0	0	0	0	0	0	1	0	0	0	0	0	0	
36	1.7	1	0	0	0	0	0	0	0	1	0	1	0	0	0	
37	2.1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	
38	1.9	1	0	0	0	0	0	0	0	1	0	0	0	1	0	
39	1.9	1	0	0	0	0	0	0	0	1	0	0	0	0	1	
40	2.1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	
41	1.8	1	0	0	0	0	0	0	0	0	1	1	0	0	0	
42	1.9	1	0	0	0	0	0	0	0	0	1	0	1	0	0	
43	2	1	0	0	0	0	0	0	0	0	1	0	0	1	0	
44	2.1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	
45	2.1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	
46	2	1	0	0	0	0	0	0	0	0	0	1	0	0	0	
47	2.1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	
48	2	1	0	0	0	0	0	0	0	0	0	0	0	1	0	
49	2.1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	
50	2.1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	

Table 4 Ranks of body weights of broilers in Table I

Broiler(Subject)	Body Weight(Treatment)				
	1	2	3	4	5
1	1.5	3	4.5	4.5	1.5
2	1	3.5	2	5	3.5
3	1.5	4	1.5	4	4
4	1	5	3.5	2	3.5
5	2.5	1	2.5	5	4
6	1	2	4	4	4
7	1	4	2	5	3
8	1	4.5	2.5	2.5	4.5
9	1	2	3	4.5	4.5
10	1.5	4	1.5	4	4
Total	13	33	27	40.5	36.5

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.

Acknowledgement

None.

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.