

General Linear Models: An Integrated Approach to Statistics

Sylvain Chartier Andrew Faulkner
University of Ottawa

Generally, in psychology, the various statistical analyses are taught independently from each other. As a consequence, students struggle to learn new statistical analyses, in contexts that differ from their textbooks. This paper gives a short introduction to the general linear model (GLM), in which it is showed that ANOVA (one-way, factorial, repeated measure and analysis of covariance) is simply a multiple correlation/regression analysis (MCRA). Generalizations to other cases, such as multivariate and nonlinear analysis, are also discussed. It can easily be shown that every popular linear analysis can be derived from understanding MCRA.

The most commonly used statistical analyses are ANOVA, *t*-tests, and regressions (Cousineau, 2005). These are taught as independent modules during the training of psychology students, which sadly fragments their knowledge into pieces that seem different and disconnected from one another. Multivariate statistic books are no stranger to this practice (*e.g.* Howell, 2002; Shavelson, 1996; Stevens, 1992; Tabachnick & Fidell, 2001) – the vast majority give only a very brief introduction to the subject. This is even worse in univariate books, in which even such cursory treatment is lacking. With this is combined with the fact that mathematical reasoning is not a priority in most psychology curricula (Giguère, Hélie, & Cousineau, 2004), the net result is a general misunderstanding of statistics among many students of social science.

In addition, more time is devoted to using how-to-do-it computer tools (due to their accessibility and user-friendly interfaces), and less time is spent on understanding the

concepts – which is absolutely necessary to do any work in research. According to Tatsuoka (1988):

“... much is to be gained by the student’s going through the calculations by hand ... Students who have undergone this sort of learning experience will be more likely to develop a thorough understanding of the major steps involved in a sequence of computation than will those who, from the outset, leave all the “busy work” to the computer.”

Consequently, after their mandatory courses, students still usually have difficulties choosing the correct statistical test for their data. A student who masters software does not indicate his comprehension – all he/she shows is a bit of technical competence.

Although most psychology teachers know that ANOVA and regression are linked through the general linear model (GLM), few actually teach it in their courses. GLM offers a unique pedagogical perspective to provide such a unified view of statistical testing. To provide a more detailed explanation, it will be shown that ANOVA - as well as the *t*-test - are simply special cases of multiple correlation/regression analysis (MCRA). Table 1 shows the different analyses that can be derived from MCRA given the number of independent variables (IV), and the number of dependant variables (DV). Therefore, if the IV is nominal and there is only one continuous DV, the ANOVA is a special case of MCRA. Thus, a general framework can be

Address correspondence to Sylvain Chartier, University of Ottawa, School of Psychology, 125 University, Ottawa, Ontario, K1N 6N5, E-mail: sylvain.chartier@uottawa.ca. The authors are grateful to an anonymous reviewer and Jean-François Ferry for their help in reviewing the manuscript and Jean-François Allaire and Isabelle Smith for their comments in reviewing a previous version of the manuscript.

Table 1. Uni variate and multivariate representations of the GLM.

	<i>DV</i>	<i>Form</i>	<i>IV</i>	<i>Form</i>	Type of analysis
Univariate GLM	1	nominal	1	nominal	Phi coefficient / Chi-square
	1	continuous	1	nominal	<i>t</i> -test
	1	nominal	≥1	continuous and/or nominal	Logistic regression / Discriminant function
	1	continuous	1	continuous	Simple correlation / regression
	1	continuous	≥2	nominal	ANOVA (one-way, factorial, repeated measure)
	1	continuous	≥2	continuous and nominal	ANCOVA
	1	continuous	≥2	continuous	Multiple correlation / regression
Multivariate GLM	≥ 2	nominal	≥2	nominal	Correspondence
	≥ 2	nominal	≥2	continuous and/or nominal	Multivariate logistic regression / Discriminant functions
	≥ 2	continuous	≥2	nominal	MANOVA
	≥ 2	continuous	≥1	latent	Principal component / Factor
	≥ 2	continuous and/or nominal	≥1	latent	Multidimensional scaling
	≥ 2	continuous and/or latent	≥1	continuous and/or latent	Structural equation modeling
	≥ 2	continuous	≥2	continuous	Canonical correlation

Note: DV, dependent variable, IV, independent variable
taught in which statistical methods are viewed as a whole, which would facilitate their comprehension and application. However, to really understand statistical analysis, linear algebra must be used, which, paradoxically, is not part of the mandatory psychology curriculum in many universities. This once again reinforces the disadvantage that psychology students (unwittingly) endure in their training, compared to other sciences (Giguère, Hélie, & Cousineau, 2004). While it should be noted that linear algebra is briefly introduced in some multivariate statistics books (e.g. Tabachnick & Fidell, 2001; Tatsuoka, 1988), far stronger material is found in books about linear algebra itself (e.g. Lipschutz & Lipson, 2001; Strang, 1988).

The information in this paper is primarily a synthesis of knowledge first presented in Cohen & Cohen (1983), Tatsuoka (1988), Kutner, Nachtsheim, Neter, & Li (2005), and Morrison (1976), and it is mainly divided into three parts. The first shows that when using MCRA or any kind of

ANOVA (repeated, one-way, factorial, covariance), the same three steps are always involved: compute the appropriate coding matrix, create the SSCP matrix, and calculate the R-squared. More precisely, a description and review of multiple correlation analysis (MCRA) and simple correlation/regression analysis (SCRA) is provided. Subsequently, ANOVA, factorial ANOVA, and repeated-measures are presented as special cases of MCRA, in that order. The second part of the paper asserts the various links between MCRA and ANOVA with some numerical examples using one-way ANOVA and repeated-measures ANOVA. The final section discusses the multivariate case and nonlinear analysis (the generalized linear model).

Multiple Correlation/Regression Analysis

The purpose of MCRA is to determine the strength of correlation between a criterion (the dependent variable) and multiple predictors (the independent variables). If a

functional relationship is desired, then multiple regression analysis can be performed.

Sum of square and cross product (SSCP) matrix and the R-squared

Different variables can be expressed, using standard matrix notations, as follows. The predictor variables are defined as:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

and the criterion variable is defined as:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (2)$$

where \mathbf{X} is matrix of dimension $n \times p$, \mathbf{y} a matrix (vector) of $1 \times n$ and, n and p are the number of participants and predictors, respectively. These two matrices can be put into a single matrix :

$$\mathbf{M} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2p} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} & y_n \end{bmatrix} \quad (3)$$

where x_{ij} represents the j^{th} predictor of the i^{th} participant, y_i the criterion of the i^{th} participant. From the \mathbf{M} matrix, the sum of squares and cross product (SSCP) matrix can be computed. Useful information such as the variance, covariance, and R-squared can also be extracted from the SSCP matrix, which is obtained by:

$$\begin{aligned} \mathbf{SSCP} &= (\mathbf{M} - \bar{\mathbf{M}})^T (\mathbf{M} - \bar{\mathbf{M}}) \\ &= \mathbf{M}^T \mathbf{M} - \bar{\mathbf{M}}^T \bar{\mathbf{M}} \\ &= \mathbf{M}^T \mathbf{M} - (\mathbf{1}^T \mathbf{M})^T (\mathbf{1}^T \mathbf{M}) / n \end{aligned} \quad (4)$$

where $\mathbf{1}$ is defined as a vector (of dimension n) in which all elements are equal to 1, $\bar{\mathbf{M}}$ is a means-score matrix (of dimension $n \times p+1$) where the mean of each column is repeated over n lines, and T denotes the matrix transpose operation. If the \mathbf{SSCP} is divided by the corresponding degrees of freedom ($n-1$), then the variance/covariance matrix is obtained. Thus, the \mathbf{SSCP} is a convenient way to represent a lot of information about variability in a single matrix. Naturally, the same can be found for SCRA using Equation 4. In that case, the matrix will be reduced to a 2 by 2 format. The element found at the junction of the first row and the first column will be used to estimate the variance of the predictor, the elements at the junction of the first row and second column will be used to estimate the covariance (this information is also available at the junction of the second row and the first column), and the element at the

junction of the second row and second column will be used to estimate the criterion variance:

$$\mathbf{SSCP}_{\text{SCRA}} = \begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})^2 & \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) & \sum_{i=1}^n (y_i - \bar{y})^2 \end{bmatrix} \quad (5)$$

where \bar{x} and \bar{y} represent the mean of the predictors and of the criterion, respectively.

By partitioning the \mathbf{SSCP} matrix correctly, the coefficient of determination - R-squared (R^2) - can be obtained. This is done by dividing the \mathbf{SSCP} into four sectors, which we name \mathbf{S}_{pp} , \mathbf{S}_{pc} , \mathbf{S}_{cp} , \mathbf{S}_{cc} . These are (in order), the sum of squares of the predictors alone, the sum of cross-products between the predictors and the criterion, the sum of cross-products between the criterion and the predictors (note that $\mathbf{S}_{cp} = \mathbf{S}_{pc}^T$), and finally the sum of squares of the criterion alone.

$$\mathbf{SSCP} = \begin{bmatrix} \mathbf{S}_{pp} & \mathbf{S}_{pc} \\ \mathbf{S}_{cp} & \mathbf{S}_{cc} \end{bmatrix} \quad (6)$$

Once this is drawn up, the coefficient of determination ($0 \leq R^2 \leq 1$) can be obtained with the following matrix multiplication:

$$R^2 = \mathbf{S}_{pc}^T \mathbf{S}_{pp}^{-1} \mathbf{S}_{pc} \mathbf{S}_{cc}^{-1} \quad (7)$$

In the particular case of SCRA, each element of the SSCP matrix is a scalar, and thus the R^2 will also be a scalar. It is given by the following:

$$\begin{aligned} R_{\text{SCRA}}^2 &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-1}}{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)^{-1}} \\ &= \frac{\left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \end{aligned} \quad (8)$$

Equation 8 is the standard way to obtain the coefficient of determination. If the standard bivariate correlation (R_{SCRA}) is desired, then one must find the square root of R_{SCRA}^2 , as found by equation 8, which must then be multiplied by the sign (+ or -) of \mathbf{S}_{pc} (direction of the covariance).

For an unbiased estimate of R^2 (\tilde{R}^2), the following correction must be applied:

$$\tilde{R}^2 = 1 - \frac{(1 - R^2)(n-1)}{(n-p-1)} \quad (9)$$

This is called the shrunken R-squared, or the adjusted R-squared.

Partial and semipartial coefficients

Unlike SCRA, defining the contribution of each predictor in MCRA is not straightforward. To illustrate the different ways that those relations can be computed, a Venn diagram

is used. This is illustrated in Figure 1a as an example with two predictors, in which R^2 is the sum of a, b and c areas. The total variation (Y) is equal to 1 ($a+b+c+e = 1$). The relationship between each predictor and the criterion can be expressed by taking only the area "a" for the first predictor and the area "b" for the second predictor, which is formally expressed by:

$$a = R^2 - r_{yx_2}^2 = a + b + c - (b + c) \quad (10a)$$

and:

$$b = R^2 - r_{yx_1}^2 = a + b + c - (a + c) \quad (10b)$$

where $r_{yx_1}^2$ and $r_{yx_2}^2$ represent the bivariate squared correlation between the criterion and a given predictor (x_1 or x_2). Thus, those areas represent the proportion of variation that uniquely overlaps the criterion, which is called the squared semi-partial correlation. The general formula for p predictors is given by:

$$sr_i^2 = R^2 - R_{(i)}^2 \quad (11)$$

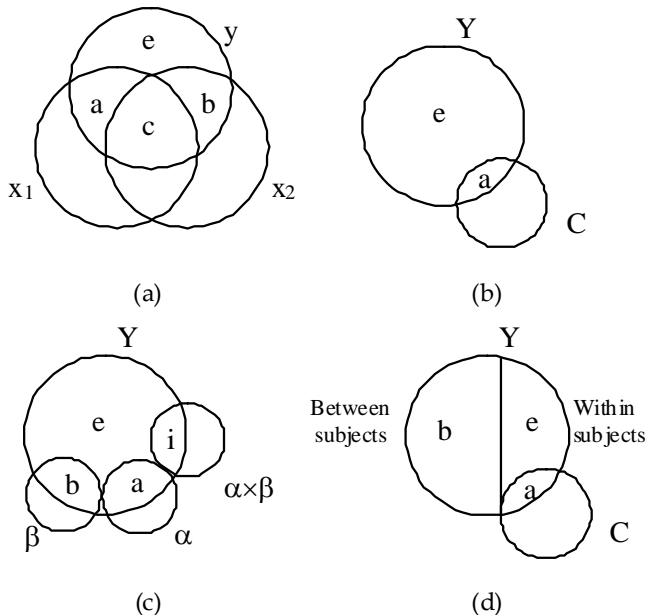
where $R_{(i)}^2$ is the strength of association between all the predictors except the i^{th} predictor. Another way to express the relationship between each predictor and the criterion is to compute the ratio of $a/(a+e)$ for the first predictor and $b/(b+e)$ for the second predictor. This is expressed by:

$$\frac{a}{a+e} = \frac{R^2 - r_{yx_2}^2}{1 - r_{yx_2}^2} = \frac{a + b + c - (b + c)}{1 - (b + c)} \quad (12a)$$

and:

$$\frac{b}{b+e} = \frac{R^2 - r_{yx_1}^2}{1 - r_{yx_1}^2} = \frac{a + b + c - (a + c)}{1 - (a + c)} \quad (12b)$$

Figure 1. Venn diagrams for a) two predictors, b) one-way analysis of variance, c) two-way ANOVA and d) repeated measures.



This relationship is called the squared partial correlation. The general formula for p predictors is expressed by:

$$pr_i^2 = \frac{R^2 - R_{(i)}^2}{1 - R_{(i)}^2} = \frac{sr_i^2}{1 - R_{(i)}^2} \quad (13)$$

Thus, the squared partial correlation is the proportion of the criterion variance that is independent of the remaining predictors ($1 - R_{(i)}^2$), which is accounted for uniquely by x_i .

Significance test

Before introducing analyses of variance using MCRA, one must perform a significance test. This is done by using Fisher distribution, based around finding on the F value, which is obtained with the following equation:

$$F = \frac{\frac{R^2}{p}}{\frac{1-R^2}{n-p-1}} = \frac{R^2(n-p-1)}{(1-R^2)p} \quad (14)$$

with $df = p$ and $n-p-1$.

This can be viewed as a ratio of explained variation in relation to the unexplained variation balanced by their respective degrees of freedom. Note that in the case of SCRA, the F ratio can be reduced to a simple t-test ($t = \sqrt{F}$, $df = n-2$).

Significance testing can also be applied to partial and semi-partial coefficients. The formula used is the same as Equation 14, the only difference resides in the corresponding degrees of freedom. Since semi-partial and partial coefficients are mathematically linked (as can be seen in Equation 13), they will give the same outcome. Thus, in the case of semi-partial coefficients, the F value is obtained by:

$$F_i = \frac{\frac{sr_i^2}{1}}{\frac{1-R^2}{n-p-1}} = \frac{(sr_i^2)(n-p-1)}{(1-R^2)} \quad (15)$$

Regression coefficients

Finally, if a functional linear relationship is desired, then a multiple regression equation must be used. This relationship is expressed by:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (16)$$

or:

$$\hat{y} = b_0 + \mathbf{Xb} \quad (17)$$

or:

$$\hat{y} = \mathbf{X}^+\mathbf{b}^+ \quad (18)$$

where b_i represents the regression weights, \hat{y} the criterion vector, \mathbf{b} the regression weights without the constant (b_0), \mathbf{b}^+ the regression weights including the constant, \mathbf{X} the predictor matrix (Equation 1) and \mathbf{X}^+ the predictor matrix including the unit vector ($\mathbf{1}$) in the first column. The regression weights (b_i) can be obtained by:

$$\mathbf{b}^+ = (\mathbf{X}^{+T}\mathbf{X}^+)^{-1}\mathbf{X}^{+T}\mathbf{y} \quad (21)$$

This is the general solution for any number of predictors.

Moreover, for the special case of simple regression analysis, it can be directly shown that Equation 21 reduces to the standard form for b_1 :

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{Cov_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x} \quad (22)$$

where Cov_{xy} represents the covariance between the criterion and the predictor, r_{xy} represents the bivariate correlation, and s_y and s_x represent the standard deviation for the criterion and the predictor respectively. Equation 21 can also be reduced to the following form for b_0 :

$$b_0 = \bar{y} - b_1 \bar{x} \quad (23)$$

Note that the regression weights without the constant (\mathbf{b}) can be obtained from the partitioned SSCP matrix as well:

$$\mathbf{b} = \mathbf{S}_{pp}^{-1} \mathbf{S}_{pc} \quad (24)$$

Therefore, Equation 22 is simply a special case for bivariate relationships (shown in Equation 24). To find the constant, generalization of Equation 23 is used, which gives:

$$b_0 = \bar{y} - \bar{\mathbf{x}}^T \mathbf{b} \quad (25)$$

These are all of the MCRA concepts needed for ANOVA. Now, in order to complete the picture, ANOVA will be examined under the MCRA perspective.

One-Way ANOVA

In ANOVA, each subject's score is based on the following equation:

$$y_{ij} = \mu + \tau_j + e_{ij} \quad (26)$$

where y_{ij} represents the score of the i^{th} participant in the j^{th} group, μ the grand population mean, τ_j the treatment applied to the j^{th} group, and e_{ij} the error associated with the i^{th} participant in the j^{th} group. This function can be translated into MCRA as:

$$\mathbf{y} = \mathbf{X}^+ \mathbf{b}^+ + \mathbf{e} \quad (27)$$

where, for example, $\mathbf{b}^+ = [\mu, \tau_1, \tau_2, \dots, \tau_{k-1}]^T$, \mathbf{X}^+ the predictor matrix obtained from a coding matrix, and k the number of groups.

The hardest part of linking MCRA to ANOVA is in

creating the coding matrix (\mathbf{X}) correctly. There are different coding matrices: effect, contrast, dummy, and even nonsense coding. They all give the same global significance test, but the choice of one coding matrix over another varies according to the research hypothesis, and the correct matrix must be chosen for results to be meaningful. For this reason, coding matrices encourage researchers to develop well formed hypotheses from the start.

Dummy coding

A dummy coding matrix (defined in the first column of Table 2) has to be designed in such a way that all the information belonging to a particular group is coded in a 1/0 dichotomy. Thus, for every subject, this coding will be applied. This will give a predictor matrix (Equation 1) of dimension $(n \times k-1)$ - see the numerical example section for more details. Note that there is no x_k in the table, as that entry would be redundant - it can be entirely determined by the other columns. Using that coding matrix, if the regression weights (\mathbf{b}^+) are computed according to Equation 21, the following solution will be obtained:

$$\mathbf{b}^+ = [\bar{x}_k \quad \bar{x}_1 - \bar{x}_k \quad \bar{x}_2 - \bar{x}_k \quad \dots \quad \bar{x}_{k-1} - \bar{x}_k]^T \quad (28)$$

The last regression coefficient, b_k , can be obtained from the previous other coefficient (without the constant), using:

$$b_k = \mathbf{X} \mathbf{b} = \bar{x}_k - \bar{x}_k = 0 \quad (29)$$

In this case, its value is zero. Equation 28 tells us that the constant (b_0) will represent the mean of the last group (\bar{x}_k). Every other regression coefficient (b_1, b_2, \dots, b_k) will compare its respective mean with the last group. Thus, this coding is used when a researcher wants to compare every group to a reference group. A classic example would be to compare different treatments with a control condition.

Effect Coding

The difference between effect and dummy coding is that instead of identifying the last group with all 0s, we use all -1s. This is illustrated in Table 2 (second column). When equation 21 is used to compute the regression coefficients

Table 2. Various types of coding matrices

Dummy variable coding					Effect variable coding					Contrast variable coding				
Group	x_1	x_2	\dots	x_{k-1}	Group	x_1	x_2	\dots	x_{k-1}	Group	x_1	x_2	\dots	x_{k-1}
g_1	1	0	\dots	0	g_1	1	0	\dots	0	g_1	$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,k-1}$
g_2	0	1	\dots	0	g_2	0	1	\dots	0	g_2	$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,k-1}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
g_{k-1}	0	0	\dots	1	g_{k-1}	0	0	\dots	1	g_{k-1}	$a_{k-1,1}$	$a_{k-1,2}$	\dots	$a_{k-1,k-1}$
g_k	0	0	\dots	0	g_k	-1	-1	\dots	-1	g_k	$a_{k,1}$	$a_{k,2}$	\dots	$a_{k,k-1}$

with effect coding, the following results are obtained:

$$\mathbf{b}^+ = \begin{bmatrix} \bar{x} & \bar{x}_1 - \bar{x} & \bar{x}_2 - \bar{x} & \dots & \bar{x}_{k-1} - \bar{x} \end{bmatrix}^T \quad (30)$$

Once again, the last coefficient can be computed from \mathbf{b} and \mathbf{X} , which will give:

$$b_k = \mathbf{X}\mathbf{b} = \bar{x}_k - \bar{x} \quad (31)$$

Thus, when effect coding is used, the constant (b_k) represents the unweighted grand mean (\bar{x}). Every other coefficient is compared to this grand mean. In other words, each \mathbf{b} 's coefficient will give an estimation of the treatment effect (τ).

Contrast Coding

The last coding strategy presented is the orthogonal contrast, which is a generalization of effect coding. Contrast coding in the context of regression analysis has only one difference from contrast coding for a priori testing in ANOVA, which is that each contrast within the coding matrix must be orthogonal and the entire $k-1$ contrast must be represented in the coding matrix. The contrast coding matrix is illustrated in Table 2 (third column).

A contrast is defined as:

$$c_j = a_{1j}\bar{y}_1 + a_{2j}\bar{y}_2 + \dots + a_{kj}\bar{y}_k \quad (32)$$

where a_{ij} (i.e. a_{1j}) are the elements of the coding coefficient. In addition, the contrasts must satisfy the following conditions:

- 1) $a_{1j} + a_{2j} + \dots + a_{kj} = 0$ (null hypothesis)
- 2) $j = k-1$ (linear independence and full matrix coding requirement)
- 3) $a_i^T a_j = 0$, $i \neq j$ (orthogonal requirement)

If the regression coefficients are computed using Equation 21, it will give the following result

$$\mathbf{b}^+ = \begin{bmatrix} \bar{x} & \frac{\mathbf{a}_1^T \bar{\mathbf{x}}}{\mathbf{a}_1^T \mathbf{a}_1} & \frac{\mathbf{a}_2^T \bar{\mathbf{x}}}{\mathbf{a}_2^T \mathbf{a}_2} & \dots & \frac{\mathbf{a}_{k-1}^T \bar{\mathbf{x}}}{\mathbf{a}_{k-1}^T \mathbf{a}_{k-1}} \end{bmatrix}^T \quad (33)$$

where \mathbf{a}_j represents the column vector of the contrast coding (Table 2) and $\bar{\mathbf{x}}$ the mean vector of the group. Once again, the last coefficient can be computed from \mathbf{b} and \mathbf{X} , which will give:

$$b_k = \mathbf{X}\mathbf{b} = \frac{\mathbf{a}_k^T \bar{\mathbf{x}}}{\mathbf{a}_k^T \mathbf{a}_k} \quad (34)$$

Thus, for contrast coding, the constant represents the unweighted grand mean (\bar{x}), and all other coefficients are the weight normalization of the contrast comparison.

Linking ANOVA and MCRA

This section will show that whether ANOVA or MCRA is performed, the same results will be found in terms of variance evaluation. First, it will be shown that the total sum of squares of ANOVA is equivalent to S_{cc} . Second, it will be shown that the error variation (SS_W) of ANOVA is the same as the unexplained regression variation ($S_{cc}(1-R^2)$) that MCRA finds.

ANOVA is based on the following equation:

$$y_{ij} = \mu + \tau_j + e_{ij} \quad (35)$$

The y_{ij} data are presented in Table 3. The grand mean is obtained the usual way:

$$\bar{y} = \frac{k}{j=1} \sum_{i=1}^{n_j} y_{ij} \quad (36)$$

With very few manipulations, it can be shown from Equation 35 that the total sum of squares (SS_T) can be partitioned in two: the between-groups sum of square (SS_B) and the within-groups (error) sum of squares (SS_W):

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 \quad (37)$$

To show this, Equation 35 must be rewritten in terms of parameter estimation, as follows:

$$y_{ij} = \bar{y} + (\bar{y}_j - \bar{y}) + (y_j - \bar{y}) \quad (38)$$

If the results are centered (the mean is removed), Equation 37 becomes:

$$(y_{ij} - \bar{y}) = (\bar{y}_j - \bar{y}) + (y_j - \bar{y}) \quad (39)$$

Squaring both sides, we obtain:

$$(y_{ij} - \bar{y})^2 = (\bar{y}_j - \bar{y})^2 + (y_j - \bar{y})^2 + 2(y_j - \bar{y})(\bar{y}_j - \bar{y}) \quad (40)$$

If we sum this expression for all values of i and j , then Equation 40 becomes:

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_j - \bar{y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})(\bar{y}_j - \bar{y}) \\ &= \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 + \sum_{j=1}^k (y_j - \bar{y}) \sum_{i=1}^{n_j} (\bar{y}_j - \bar{y}) \end{aligned} \quad (41)$$

Since $\sum_{i=1}^{n_j} (\bar{y}_j - \bar{y})$ is zero, then the third term of the right will be zero and results will be the same as the one expressed at Equation 37.

To show that in the case of MCRA identical partitioning of the sum of squares occurs, the same procedure is applied (shown previously for ANOVA). Thus, if we centered Equation 27 we obtain:

$$\mathbf{y} - \bar{\mathbf{y}} = \mathbf{X}^+ \mathbf{b}^+ + \mathbf{e} - \bar{\mathbf{y}} \quad (42)$$

If we square both sides, the result is:

$$\begin{aligned} (\mathbf{y} - \bar{\mathbf{y}})^2 &= (\mathbf{X}^+ \mathbf{b}^+ + \mathbf{e} - \bar{\mathbf{y}})^2 \\ &= (\hat{\mathbf{y}} + \mathbf{e} - \bar{\mathbf{y}})^2 \\ &= (\hat{\mathbf{y}} + \mathbf{e} - \bar{\mathbf{y}})^T (\hat{\mathbf{y}} + \mathbf{e} - \bar{\mathbf{y}}) \\ &= \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \hat{\mathbf{y}}^T \mathbf{e} - \hat{\mathbf{y}}^T \bar{\mathbf{y}} + \mathbf{e}^T \hat{\mathbf{y}} + \mathbf{e}^T \mathbf{e} \\ &\quad - \mathbf{e}^T \bar{\mathbf{y}} - \bar{\mathbf{y}}^T \hat{\mathbf{y}} - \bar{\mathbf{y}}^T \mathbf{e} + \bar{\mathbf{y}}^T \bar{\mathbf{y}} \end{aligned} \quad (43)$$

Since the error is orthogonal to the predicted value of $\mathbf{X}^+ \mathbf{b}^+$, their scalar product (covariance) will be zero. Moreover, since $\bar{\mathbf{y}} = \mathbf{X} \mathbf{b}^+$, the scalar product between \mathbf{e} and $\bar{\mathbf{y}}$ will also be zero. Therefore, Equation 43 can be reduced to:

$$\begin{aligned}
(\mathbf{y} - \bar{\mathbf{y}})^2 &= \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \hat{\mathbf{y}}^T \bar{\mathbf{y}} + \mathbf{e}^T \mathbf{e} - \bar{\mathbf{y}}^T \hat{\mathbf{y}} + \bar{\mathbf{y}}^T \bar{\mathbf{y}} \\
&= \hat{\mathbf{y}}^T \hat{\mathbf{y}} - 2\hat{\mathbf{y}}^T \bar{\mathbf{y}} + \bar{\mathbf{y}}^T \bar{\mathbf{y}} + \mathbf{e}^T \mathbf{e} \\
&= (\hat{\mathbf{y}} - \bar{\mathbf{y}})^2 + \mathbf{e}^T \mathbf{e} \\
&= (\hat{\mathbf{y}} - \bar{\mathbf{y}})^2 + (\mathbf{y} - \mathbf{X}^+ \mathbf{b}^+)^T (\mathbf{y} - \mathbf{X}^+ \mathbf{b}^+) \\
&= (\hat{\mathbf{y}} - \bar{\mathbf{y}})^2 + (\mathbf{y} - \hat{\mathbf{y}})^2 \tag{44a} \\
&= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{44b}
\end{aligned}$$

If all columns (group) of Table 3 are aggregated to form a single vector (Equation 2), then the average of this vector will be the same as that expressed in Equation 36. Consequently, the sums of squares of Equations 37 and 44b are equivalent. In other words, the total sum of squares is the S_{cc} given by the **SSCP** (Equation 6):

$$SS_{T\ ANOVA} = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = S_{cc} = SS_{T\ MCRA} \tag{45}$$

Finally, it can be shown that $SS_w = S_{cc}(1-R^2)$ and $SS_B = S_{cc}R^2$. For brevity, only the first equality will be demonstrated (the second one can be found in a similar fashion). For the same reason, it is assumed that the data have been standardized; the constant effect is removed (in other words, \mathbf{b}^+ will equal \mathbf{b} and \mathbf{X}^+ will equal \mathbf{X}). This is demonstrated mathematically as:

$$\begin{aligned}
SS_w &= \sum (\hat{y}_i - y_i)^2 \\
&= (\hat{\mathbf{y}} - \mathbf{y})^T (\hat{\mathbf{y}} - \mathbf{y}) \\
&= \hat{\mathbf{y}}^T \hat{\mathbf{y}} - 2\hat{\mathbf{y}}^T \mathbf{y} + \mathbf{y}^T \mathbf{y} \tag{46}
\end{aligned}$$

Substituting Equation 18 into Equation 46 gives:

$$SS_w = \mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b} - 2\mathbf{b}^T (\mathbf{X}^T \mathbf{y}) + \mathbf{y}^T \mathbf{y} \tag{47}$$

Substituting Equation 21 into Equation 47 gives:

$$\begin{aligned}
SS_w &= \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\
&\quad - 2\mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y}) + \mathbf{y}^T \mathbf{y} \\
&= -\mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X}^+)^{-1} (\mathbf{X}^T \mathbf{y}) + \mathbf{y}^T \mathbf{y} \\
&= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y}) (\mathbf{y}^T \mathbf{y})^{-1} \mathbf{y}^T \mathbf{y} \\
&= (1 - \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X}^+)^{-1} (\mathbf{X}^T \mathbf{y}) (\mathbf{y}^T \mathbf{y})^{-1}) \mathbf{y}^T \mathbf{y} \\
&= (1 - S_{cp} S_{pp}^{-1} S_{pc} S_{cc}^{-1}) S_{cc} \\
&= (1 - R^2) S_{cc} \tag{48}
\end{aligned}$$

Consequently, to perform ANOVA, the same steps are followed as before for MCRA: the **SSCP** matrix is computed, followed by R-squared and the F -value. This F -value will be identical to the one found performing an ANOVA. Figure 1b illustrated the variability given by the dependant variable

Table 3. Data illustration for simple k groups ANOVA

y_{11}	y_{12}	\cdots	y_{1k}
y_{21}	y_{22}	\cdots	y_{2k}
\vdots	\vdots	\ddots	\vdots
$y_{n_1,1}$	$y_{n_2,2}$	\cdots	$y_{n_k,k}$
\bar{y}_1	\bar{y}_2	\cdots	\bar{y}_k

(Y) and the condition (C) while Tables 4 shows the standard ANOVA and MCRA summary table, respectively. Thus, if a given ANOVA table is provided, it is easy to obtain the R-squared from it:

$$R^2 = \frac{SS_B}{SS_T} = \frac{1 - SS_w}{SS_T} \tag{49}$$

Although ANOVA is an MCRA, a different terminology is used to describe the ANOVA outputs. In ANOVA, the R-squared is called eta-square ($\eta^2 = R^2$) and the shrunken R-squared is called epsilon-square ($\epsilon^2 = \tilde{R}^2$). Finally, ANOVA's omega-square (ω^2) can also be computed from MCRA alone:

$$\omega^2 = \frac{SS_B - (k-1)MS_w}{SS_T + MS_w} = \frac{1 - k - (1-n)R^2}{1 - k + n - R^2} \tag{50}$$

Square partial and semi-partial coefficients are also computed the same way as before. However, their outputs and interpretations will vary according to the type of coding matrix chosen. In dummy coding, the square semi-partial coefficient is interpreted as the proportion of variance due to i - k dichotomy, and for the square partial coefficient, it is the proportion of variance due to i - k dichotomy excluding other effects. In effects coding, the square semi-partial coefficient is interpreted as the proportion of variance due to i 's effect, and for the square partial coefficient, it is the proportion of variance due to i 's effect excluding other effects. Finally, for contrast coding, the square semi-partial coefficient is interpreted as the proportion of variance due to the i^{th} contrast, and for the partial coefficient, it is the proportion of variance due to the i^{th} contrast excluding other contrasts. Naturally, for each type of coding significance, testing can be applied for both partial and semi-partial coefficients. However, since the k^{th} coefficient is not readily available from the regression coefficient, it must be determined by the other analysis results. Consequently, the F value for the k^{th} coefficient is obtained by the following:

$$F_k = \frac{\left(-k \sum_{i=1}^p b_i \right)^2}{\tilde{s}_y^2 \left(\frac{(k-1)^2}{n_k} + \sum_{i=1}^p n_i \right)} \tag{51}$$

with $df = 1, n-k-1$ where \tilde{s}_y^2 is defined as:

$$\tilde{s}_y^2 = \frac{s_y^2 (1 - R^2) n}{n - k - 1} \tag{52}$$

From Equation 51, the semi-partial and partial coefficients can be obtained by:

$$sr_k^2 = \frac{F_k (1 - R^2)}{n - k} \tag{53a}$$

$$sr_k = \text{Sign}(b_k) \sqrt{sr_k^2} \tag{53b}$$

$$pr_k^2 = \frac{F_k}{F_k + n - k} \tag{54a}$$

Table 4. ANOVA Summaries using standard equations and MCRA equations

	Source of variation	Sum of Squares (SS)	Degrees of freedom (<i>df</i>)	Mean Square (MS)	<i>F</i>
Standard ANOVA summary	Between group	$\sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2$	$k-1$	$\frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2}{k-1}$	$\frac{MS_{Between}}{MS_{Within}}$
	Within group	$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$	$n-k$	$\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n-k}$	
	Total	$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2$	$n-1$		
ANOVA summary using MCRA	Between group	$S_{cc} R^2$	$k-1=p$	$\frac{S_{cc} R^2}{p}$	$\frac{R^2(n-p-1)}{(1-R^2)p}$
	Within group	$S_{cc} (1-R^2)$	$n-k=n-p-1$	$\frac{S_{cc} (1-R^2)}{n-p-1}$	
	Total	S_{cc}	$n-1$		

$$pr_k = \text{Sign}(b_k) \sqrt{pr_k^2} \quad (54b)$$

If necessary, posthoc comparisons can be performed on the data using the standard methods. Some computation can be simplified using the equality between the MS_{w} and MCRA. Generalization to more than one factor (factorial ANOVA) is straightforward, as is shown in the next section.

Factorial ANOVA

For brevity, only a two-level factorial analysis will be considered. Of course, this can be extended to more than 2 factors. In factorial ANOVA each score is subject to a base mean (μ), some error (e_{ijk}), and hopefully some effect due to one (α) and/or two (β) factors and/or their interaction ($\alpha\beta$). This is described by the following equation:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk} \quad (55)$$

Table 5. Effect coding matrix for two-level factorial ANOVA

Group	α		β				$\alpha\beta$					
	x_1	x_2	\dots	x_{m-1}	x_m	x_{m+1}	\dots	x_{m+n-2}	x_{m+n-1}	x_{m+n}	\dots	x_{mn-1}
$\alpha_1\beta_1$	1	0	\dots	0	1	0	\dots	0	1	0	\dots	0
$\alpha_2\beta_1$	0	1	\dots	0	1	0	\dots	0	0	0	\dots	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha_m\beta_1$	-1	-1	\dots	-1	1	0	\dots	0	-1	0	\dots	0
$\alpha_1\beta_2$	1	0	\dots	0	0	1	\dots	0	0	0	\dots	0
$\alpha_2\beta_2$	0	1	\dots	0	0	1	\dots	0	0	0	\dots	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha_m\beta_n$	-1	-1	\dots	-1	-1	-1	\dots	-1	1	1	\dots	1

Again, with an appropriate coding matrix, it is possible to represent this situation in MCRA terms:

$$\mathbf{y} = \mathbf{X}^+ \mathbf{b}^+ + \mathbf{e} \quad (56)$$

where, for example,

$\mathbf{b}^+ = [\mu, \alpha_1, \alpha_2, \dots, \alpha_{m-1}, \beta_1, \beta_2, \dots, \beta_{n-1}, (\alpha\beta)_{11}, (\alpha\beta)_{12}, \dots, (\alpha\beta)_{m-1, n-1}]^T$, \mathbf{X}^+ represents the predictor matrix obtained from the coding matrix, m represents the number of groups for treatment α , and n represents the number of groups for treatment β .

The coding matrix is not very different from those defined in the previous section. This matrix is constructed by building one coding matrix for each effect (α and β), then multiplying each column of the first factor with each column of the second factor to give the interaction coding matrix ($\alpha\beta$). Effect coding is shown in Table 5. Although effect coding was used, this can be done with any other type of coding, including any type of mixed coding (e.g. α : contrast; β : dummy). From that matrix, all needed information can be obtained using the same equations illustrated by Figure 1c. In this case, the total explained variance is the sum of the three areas ($R^2 = a+b+i$), a relationship formally expressed by:

$$R^2 = R_\alpha^2 + R_\beta^2 + R_{\alpha\beta}^2 \quad (57)$$

Thus, each R_i^2 can be computed independently by partitioning the coding matrix into functions of each effect. Like in standard factorial ANOVA, F values for the main effect (α and β) will be a function of the type of effect presented in the study: for

Table 6. Factorial ANOVA summary using MCRA

Source of variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F
Between	$S_{cc}R^2$	$k-1$	SS_B/df_B	MS_B/MS_W
α	$S_{cc}R_{\alpha}^2$	$k_{\alpha}-1$	SS_{α}/df_{α}	$MS_{\alpha}/(MS_W \text{ or } MS_{\alpha\beta})$
β	$S_{cc}R_{\beta}^2$	$k_{\beta}-1$	SS_{β}/df_{β}	$MS_{\beta}/(MS_W \text{ or } MS_{\alpha\beta})$
$\alpha\beta$	$S_{cc}R_{\alpha\beta}^2$	$(k_{\alpha}-1)(k_{\beta}-1)$	$SS_{\alpha\beta}/df_{\alpha\beta}$	$MS_{\alpha\beta}/MS_W$
Within	$S_{cc}(1-R^2)$	$n-k$	SS_W/df_W	
Total	S_{cc}	$n-1$		

random effects the F value will be the mean square of the main effect divided by the interaction mean square; while for the fixed effect the F will be the mean square of the main effect divided by the within-groups mean square. Table 6 shows the factorial ANOVA summary using MCRA.

Sometimes, the summary table includes partial eta-square (ANOVA terminology). This information can be obtained from the MCRA outputs:

$$\text{partial } \eta_i^2 = \text{partial } R_i^2 = \frac{R_i^2}{R_i^2 + 1 - R^2} \quad (58)$$

where i represents the type of effect (α , β or $\alpha\beta$). Usually, when the interaction is found to be significant, then simple effects are analyzed. This is done in similar fashion as in one-way ANOVA; thus using coding matrices.

Numerical example of one-way ANOVA performed with MCRA

In order to put into practice the theory explained in the preceding sections - and thus facilitate its comprehension - let us present a numerical example. In this fictitious case, there are 4 types of nonlinear models (A, B, C, D) that were tested on a given classification task. The data are given in Table 7. In this example, since there is no particular group, nor any interesting grouping properties, effect coding (Table 3) is used. The coding matrix is thus given by:

Group	x_1	x_2	x_3
A	1	0	0
B	0	1	0
C	0	0	1
D	-1	-1	-1

From this matrix, we learn that the \mathbf{M} matrix (Equation 3) is given in the first column of the first matrix in the Appendix. Using \mathbf{M} , the \mathbf{SSCP} matrix (Equation 4) is then:

$$\mathbf{SSCP} = \begin{bmatrix} 10.96 & 4.92 & 4.96 & 27.00 \\ 4.92 & 11.83 & 4.92 & 110.00 \\ 4.96 & 4.92 & 10.96 & 127.00 \\ \hline 27 & 110.00 & 127.00 & 4538.00 \end{bmatrix}$$

This is partitioned (Equation 6) to give R^2 (Equation 7), resulting in:

$$R^2 = [27 \ 110 \ 127] \begin{bmatrix} 0.1256 & -0.035 & -0.041 \\ -0.035 & 0.114 & -0.035 \\ -0.041 & -0.035 & 0.126 \end{bmatrix} \begin{bmatrix} 27 \\ 110 \\ 127 \end{bmatrix} [0.0002]$$

$$= 0.45$$

The \bar{R}^2 (or ε^2), using Equation 9, gives:

$$\bar{R}^2 = 1 - \frac{(1-0.45)(24-1)}{(24-3-1)} = 0.36$$

Thus, 36% of the variance in the classification performance is due to the variation of the different groups (A, B, C and D).

And the ω^2 is obtained by Equation 50:

Table 7. Data used for the numerical example

	A	B	C	D
	85	95	81	64
	60	78	86	74
	75	72	79	45
	45	74	88	51
	79	68	90	65
	55	91	75	
		77		
Mean	66.50	79.29	83.17	59.80
Standard deviation (SD)	15.54	10.00	5.78	11.65
Grand Mean	73			
Grand SD	14.05			

Table 8. Summary of ANOVA using MCRA

Source of variation	Sum of Squares (SS)	Degrees of freedom (<i>df</i>)	Mean Square (MS)	<i>F</i>	<i>Prob.</i>
Between group	2021.44	3	673.813	5.355	0.0072
Within group	2516.56	20	125.828		
Total	4538	23			

$$\omega^2 = \frac{1 - 4 - (1 - 24)0.45}{1 - 4 + 24 - 0.45} = 0.35$$

The *F* value for the R^2 is given by Equation 14:

$$F = \frac{0.45(24 - 3 - 1)}{(1 - 0.45)(3)} = 5.36$$

Results found using Table 4 are summarized in Table 8. Regression weights can be obtained from Equation 21:

$$\mathbf{b}^+ = [72.19 \quad -5.69 \quad 7.10 \quad 10.98]^T$$

Thus, the constant (\mathbf{b}_1^+), 72.19, is the same as the unweighted grand mean. The remaining vector elements are the distance between each group average and this value. The regression coefficient for the last group (D) can be obtained using Equation 31:

$$\mathbf{b}_5^+ = \begin{bmatrix} -5.69 \\ 7.10 \\ 10.98 \end{bmatrix} [-1 \quad -1 \quad -1] = -12.39$$

The square semi-partial and partial coefficients for the first 3 groups are obtained using Equations 11 and 13 respectively. Their *F* value can be found using Equation 15. For the last group (D), its *F* value is calculated with Equation 51, and its (square) semi-partial and partial coefficients are given by equations (53a) 54b, and (54a) 55b, respectively. All of these results are summarized in Table 9. Interpretations for the third group (C) are given as follows. An sr_3^2 of 0.212 indicates that 21% of the variance for the classification task can be accounted for by the distinction between the C and the 3 remaining algorithms (A, B, D). In other words, about 21% of the variance for the classification task is explained by the “eccentricity” of the C algorithm. The sign of sr_3 is positive, indicating that its distinction is that of the grand mean. A $pr_3^2 = 0.276$ indicates that about 28% of the variance can be explained by the C algorithm, excluding the “eccentricity” the remaining groups (A, B, D). The sign (+ or -) of pr_3 indicates the direction of the relation:

in this case it is positive (above the grand mean). The difference between the grand mean and the C’s mean (10.98) is statistically significant ($F = 7.63$, $p \approx 0.01$). Interpretations for the remaining groups are similar. Finally, if posthoc comparisons are done, they can be performed in the way standard for ANOVA.

The last topic is the repeated measures ANOVA (or matched subject design). In this context, the computation involved is slightly different.

Repeated measures ANOVA

This section presents a simple case of repeated measures subjects by conditions, which can be generalized to more complex design. For example, Chartier & Cousineau (in press) described a two-factor mixed design (split plot) constructed using GLM approaches.

For the simple case of repeated measures, the participant’s score is coded as follows:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1c} \\ y_{21} & y_{22} & \cdots & y_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nc} \end{bmatrix} \quad (59)$$

where n represents the number of subjects and c the number of times a subject is measured. Variance partitioning for the subjects by condition is illustrated in Figure 1d. Thus, total variation of the criterion, Y (not to be confused with Equation 59), is composed of two sources: the between-subject variation (b), and the within-subject variation ($a+e$). The purpose of the analysis is to determine if the ratio a/e is significant. To do this, variation between subjects (R_s^2) must first be estimated, and then removed from the total variation. This estimation is the ratio of the discard condition variation (s_v^2) in relation to the total variation (s_y^2):

Table 9. Summary of partial and semi-partial analysis

	sr_i	sr_i^2	pr_i	pr_i^2	<i>F</i>	<i>Prob</i>
A	-0.238	0.057	-0.305	0.0930	2.047	0.166
B	0.313	0.098	0.387	0.150	3.522	0.073
C	0.460	0.212	0.525	0.276	7.627	0.011
D	-0.477	0.228	-0.540	0.291	8.216	0.009

Table 10. Repeated measure ANOVA summary table using MCRA

Source of variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F
Within Subjects	$S_{cc}R_S^2$	$n - 1$		
Between (B)	$S_{cc}R_C^2$	$c - 1$	$\frac{S_{cc}R_C^2}{c-1}$	$\frac{(n-1)R_C^2}{(1-R_C^2-R_S^2)}$
Within (W)	$S_{cc}(1-R_C^2-R_S^2)$	$(c-1)(n-1)$	$\frac{S_{cc}(1-R_C^2-R_S^2)}{(c-1)(n-1)}$	
Total	S_{cc}	$cn - 1$		

$$R_s^2 = \frac{s_{\bar{y}_s}^2}{s_y^2} = \frac{(\bar{\mathbf{y}}_s - \bar{\mathbf{y}})^T (\bar{\mathbf{y}}_s - \bar{\mathbf{y}})}{\mathbf{S}_{cc}} c \quad (60)$$

where \mathbf{S}_{cc} represents the sum of squares of the criterion, and $\bar{\mathbf{y}}$ represents the mean vector. $\bar{\mathbf{y}}_s$ is defined as follows:

$$\bar{\mathbf{y}}_s = \frac{\sum_{i=1}^c \mathbf{Y}_i}{c} \quad (61)$$

(where $s = 1, 2, \dots, n$) $\bar{\mathbf{y}}_s$ is thus the average vector of the \mathbf{Y} column. Therefore, information about the conditions is discarded.

Now that the between-subject variation (R_s^2) has been estimated, it is possible to compute the ratio a/e. The estimation of "a" (R_C^2), is made by discarding the repeated measure information (all groups are treated as independent), and then using a standard coding matrix (e.g. Table 2) for one-way ANOVA (Figure 1b). The number of observations is (in that case) $c*n$ and the number of groups remains c . It is now possible to estimate the error ("e"). Figure 1d shows that $e = 1 - (a+b)$, which is formally:

$$e = 1 - (R_C^2 + R_S^2) \quad (62)$$

Once this is found, the F ratio is modified accordingly:

$$F = \frac{R_C^2}{1 - (R_C^2 + R_S^2)} (n-1) \quad (63)$$

(with $df = c-1$ and $(n-1)(c-1)$).

Table 10 shows the repeated measure ANOVA using MCRA.

Numerical example of repeated measures ANOVA performed with MCRA

In this fictitious example, 10 participants have been selected for a study about chess playing performance when trained in using the "checkmate" strategy. All participants were tested before (pretest), after (posttest) and 2 months later (follow-up). The data are given in Table 11 as are the means for each of the 3 tests. For the analysis to take place, three F ratios must be found: group, condition and interaction effects.

To start, subject variations (condition) are computed. The

general coding matrix is then expressed as:

Condition	x_1	x_2
Pretest	1	0
Posttest	0	1
Follow-up	-1	-1

From the condition coding matrix, the \mathbf{M} matrix (Equation 3) is given in the second column of the Appendix. Using \mathbf{M} , the \mathbf{SSCP} matrix (Equation 4) is then:

$$\mathbf{SSCP} = \begin{bmatrix} 10 & 5 & | & -64.6 \\ 5 & 10 & | & -34.6 \\ \hline -64.6 & -34.6 & | & 474.83 \end{bmatrix}$$

which is partitioned (Equation 6) to give R_C^2 (Equation 7):

$$R_C^2 = 0.88$$

For the group effect, $\bar{\mathbf{y}}_s$ must be computed following Equation 61:

$$\bar{\mathbf{y}}_s = \frac{\sum_{i=1}^3 \mathbf{Y}_i}{3} = [75.4 \quad 77.6 \quad 75.6 \quad 73.8 \quad 75.2]^T$$

Finally, R_s^2 is obtained from Equation 60:

$$R_s^2 = \frac{s_{\bar{y}_s}^2}{s_y^2} = \frac{(\bar{\mathbf{y}}_s - \bar{\mathbf{y}})^T (\bar{\mathbf{y}}_s - \bar{\mathbf{y}})}{\mathbf{S}_{cc}} c = \frac{7.19}{474.83} \times 3 = 0.0454$$

From those results, the F -ratio (Equation 63) can be obtained:

$$F = \frac{R_C^2}{1 - R_S^2 - R_C^2} (n-1) = \frac{0.88}{1 - 0.0454 - 0.88} (5-1) = 47.441$$

All the results are summarized in Table 12

Table 11. Data used for the repeated measures numerical example

	Moment		
	Pretest	Posttest	Follow-up
	70.5	73.9	81.7
	72.0	77.4	83.3
	68.9	76.5	81.4
	64.4	73.0	84.1
	70.2	75.2	80.1
Mean	69.2	75.2	82.12

Table 12. Repeated measures ANOVA summary table using MCRA

Source of variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F
Within Subjects	$S_{\alpha}R_S^2 = 474.83 \times 0.0454 = 21.563$	$n-1=5-1=4$		
Between (B)	$S_{\alpha}R_C^2 = 474.83 \times 0.88 = 418.021$	$c-1=3-1=2$	$\frac{S_{\alpha}R_C^2}{c-1} = \frac{418.021}{3-1} = 209.011$	$\frac{(n-1)R_C^2}{(1-R_C^2-R_S^2)} = \frac{(5-1)0.88}{(1-0.88-0.0454)} = 47.4413$
Within (W)	$S_{\alpha}(1-R_C^2-R_S^2) = 474.83(1-0.88-0.0454) = 35.25$	$(c-1)(n-1) = 2 \times 4 = 8$	$\frac{S_{\alpha}(1-R_C^2-R_S^2)}{(c-1)(n-1)} = \frac{35.25}{(3-1)(5-1)} = 4.406$	
Total	$S_{\alpha} = 474.83$	$cn-1 = 3 \times 5 - 1 = 14$		

Discussion

Since it has been shown that ANOVA and MCRA are the same analysis, we can treat every quantitative method as part of the same module – rather than seeing them as separate, they can be seen as variations of MCRA. In fact, the only things needed to accomplish the different ANOVA procedures (one-way, two-way, repeated measures, covariance, etc.) are the appropriate coding matrices. Using the proper coding matrix, the SSCP matrix and the R^2 can be computed. Coding matrices have the advantage of encouraging the researcher to think about the type of hypothesis he wants to verify before any analysis is performed.

Since the purpose of this paper was to show the link between ANOVA and MCRA, some analyses were left aside (e.g. confidence intervals, power, standardized weights, etc.). Concerning power, Chartier & Allaire (2008) present this concept as applied to the multivariate scheme. Although some interesting properties of special cases (e.g. when all groups have an equal number of subjects) were left aside for brevity, more information can be found in Cohen and Cohen (1983). Analysis of covariance (ANCOVA) has not yet been covered, as it is a mix of continuous and nominal independent variables. This analysis can be performed in the way described in this paper, using the coding matrix shown in Table 13. Also, note that this paper considered only the full model. However, testing of partial models (both hierarchical and nonhierarchical) can be done using model selection (e.g. Hélie, 2006; McCullagh & Nelder, 1989).

MCRA is not the most generalized method one can use. Generalization to multivariate cases is done through canonical correlation analysis (CCA; Thompson, 1984). Multivariate and univariate cases are described in Table 1. A

CCA approach has the advantage of covering both multivariate analysis of variance, and methods dealing with latent variables (principal component analysis, multidimensional scaling, structural equation modeling, etc.). This analysis and its different links, however, are beyond the scope of this paper.

In addition, only linear methods have been presented. However, extension to generalized linear models (GLM) (e.g. McCullagh & Nelder, 1989) can take into account nonlinear requirements:

$$\hat{\mathbf{y}} = \mathbf{X}^+\mathbf{b}^+ \quad (\text{general linear model}) \quad (64)$$

$$f(\hat{\mathbf{y}}) = \mathbf{X}^+\mathbf{b}^+ \quad (\text{generalized linear model}) \quad (65)$$

where $f(\bullet)$ is called the link function. From this generalization, the GLM is no longer tied to the normal distribution, but is open to other distributions (Poisson, Binomial, Gamma, etc.). For example, when $\hat{\mathbf{y}}$ is binary, the associated distribution is generally Binomial. In this case, logistic regression can be done using the following link function:

$$f(\hat{\mathbf{y}}) = \log\left(\frac{\hat{\mathbf{y}}}{1-\hat{\mathbf{y}}}\right) \quad (66)$$

Or, if $\hat{\mathbf{y}}$ can be counted, the associated distribution is generally Poisson. In this case Poisson regression can be done using the following link function:

$$f(\hat{\mathbf{y}}) = \log(\hat{\mathbf{y}}) \quad (67)$$

Generalized linear models could, however, be the subject of another paper.

Conclusion

This paper has shown that performing one-way, factorial, repeated measure ANOVA is no different from standard MCRA. To perform the various analyses, including ANCOVA, we need only the appropriate coding matrix (this varies as a function of research objectives). From that coding

Table 13. Effect coding matrix for ANCOVA

Group	α				β				$\alpha\beta$			
	x_1	x_2	\dots	x_{m-1}	x_m	x_{m+1}	\dots	x_{m+n-2}	x_1x_m	x_1x_{m+1}	\dots	$x_{m-1}x_{m+n-2}$
$\alpha_1\beta_1$	$c_{1,1}$	$c_{1,2}$	\dots	$c_{1,m-1}$	1	0	\dots	$c_{1,m+n-2}$	$c_{1,m+n-1}$	0	\dots	0
$\alpha_2\beta_1$	$c_{2,1}$	$c_{2,2}$	\dots	$c_{2,m-1}$	1	0	\dots	0	0	0	\dots	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha_m\beta_1$	$c_{m,1}$	$c_{m,2}$	\dots	$c_{m,m-1}$	1	0	\dots	0	$c_{m,m+n-1}$	0	\dots	0
$\alpha_1\beta_2$	$c_{m+1,1}$	$c_{m+1,2}$	\dots	$c_{m+1,m-1}$	0	1	\dots	0	0	0	\dots	0
$\alpha_2\beta_2$	$c_{m+2,1}$	$c_{m+2,2}$	\dots	$c_{m+2,m-1}$	0	1	\dots	0	0	0	\dots	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$\alpha_m\beta_n$	$c_{m,1}$	$c_{m,2}$	\dots	$c_{m,m-1}$	-1	-1	\dots	-1	$-c_{m,m+n-1}$	$-c_{m,m+n}$	\dots	$-c_{m,m+n-1}$

matrix, the SSCP matrix and the R-squared can be obtained, which are all that are needed to complete the ANOVA's summary table. Therefore, MCRA has a clear advantage over ANOVA, as it presents statistics as a whole, which would prevent the terminology confusions, and knowledge fragmentations that are so commonly seen in today's students.

References

- Chartier, S., & Allaire, J.-F. (2008). Power Estimation in Multivariate Analysis of Variance. *Tutorial in Quantitative Methods for Psychology*, 3(2), 70-78.
- Chartier, S., & Cousineau, D. (In press). Computing Mixed Design (split-plot) ANOVA. *The Mathematica Journal*.
- Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, N.J.: Erlbaum.
- Cousineau, D. (2005). The rise of quantitative methods in psychology. *Tutorial in Quantitative Methods for Psychology*, 1(1), 1-3.
- Desjardins, J. (2005). L'analyse de régression multiple. *Tutorial in Quantitative Methods for Psychology*, 1(1), 35-41.
- Giguère, G., Hélie, S., & Cousineau, D. (2004). Manifeste pour le retour des sciences en psychologie. *Revue Québécoise de Psychologie*, 25, 117-130.
- Hélie, S. (2006). An introduction to Model Selection: Tools and Algorithms, *Tutorial in Quantitative Methods for Psychology*, 2(1), 1-10.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove: Duxbury thomson learning.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models with Student CD* (5th ed.). Boston: McGraw-Hill Irwin.
- Lipschutz, S., & Lipson, M. (2001). *Schaum's Outline of Linear Algebra* (3rd ed.): McGraw-Hill Companies, inc.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.
- Morrison, D. F. (1976). *Multivariate statistical methods* (2nd ed.). New York: McGraw-Hill.
- Shavelson, R. C. (1996). *Statistical reasoning for the behavioral sciences*. Boston: Allyn and Bacon.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.
- Strang, G. (1988). *Linear algebra and its applications*. Orlando: Harcourt Brace Jovanovich.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston ; London: Allyn and Bacon.
- Tatsuoka, M. M. (1988). *Multivariate analysis : techniques for educational and psychological research* (2nd ed.). New York: Macmillan.
- Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Beverly Hills: Sage.

Manuscript received 1 January 2007

Manuscript accepted 21 September 2008

Appendix follows

Appendix: M matrices for the numerical examples

One-way ANOVA

example

$$\mathbf{M} = \begin{bmatrix}
 1 & 0 & 0 & 85 \\
 1 & 0 & 0 & 60 \\
 1 & 0 & 0 & 75 \\
 1 & 0 & 0 & 45 \\
 1 & 0 & 0 & 79 \\
 1 & 0 & 0 & 55 \\
 0 & 1 & 0 & 95 \\
 0 & 1 & 0 & 78 \\
 0 & 1 & 0 & 72 \\
 0 & 1 & 0 & 74 \\
 0 & 1 & 0 & 68 \\
 0 & 1 & 0 & 91 \\
 0 & 1 & 0 & 77 \\
 0 & 0 & 1 & 81 \\
 0 & 0 & 1 & 86 \\
 0 & 0 & 1 & 79 \\
 0 & 0 & 1 & 88 \\
 0 & 0 & 1 & 90 \\
 0 & 0 & 1 & 75 \\
 -1 & -1 & -1 & 64 \\
 -1 & -1 & -1 & 74 \\
 -1 & -1 & -1 & 45 \\
 -1 & -1 & -1 & 51 \\
 -1 & -1 & -1 & 65
 \end{bmatrix}$$

Repeated measures ANOVA

example

$$\mathbf{M} = \begin{bmatrix}
 1 & 0 & 70.2 \\
 1 & 0 & 72 \\
 1 & 0 & 68.9 \\
 1 & 0 & 64.4 \\
 1 & 0 & 70.2 \\
 0 & 1 & 73.9 \\
 0 & 1 & 77.4 \\
 0 & 1 & 76.5 \\
 0 & 1 & 73 \\
 0 & 1 & 75.2 \\
 -1 & -1 & 81.7 \\
 -1 & -1 & 83.3 \\
 -1 & -1 & 81.4 \\
 -1 & -1 & 84.1 \\
 -1 & -1 & 80.1
 \end{bmatrix}$$