

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1994 - 6th Annual Conference Proceedings

APPLICATIONS OF ESTIMABLE FUNCTIONS IN AGRICULTURAL RESEARCH WITH SPECIAL EMPHASIS ON THE GLM PROCEDURE OF SAS.

L. Munyakazi

R. L. Hintz

B. D. Selby

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Munyakazi, L.; Hintz, R. L.; and Selby, B. D. (1994). "APPLICATIONS OF ESTIMABLE FUNCTIONS IN AGRICULTURAL RESEARCH WITH SPECIAL EMPHASIS ON THE GLM PROCEDURE OF SAS." *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1365>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

APPLICATIONS OF ESTIMABLE FUNCTIONS IN AGRICULTURAL RESEARCH WITH SPECIAL EMPHASIS ON THE GLM PROCEDURE OF SAS.

L.Munyakazi ,R.L. Hintz, and B.D. Selby
Monsanto Company

Abstract

An understanding of estimable functions is essential when using an overparameterized linear model. The most attractive features of these functions are their invariance property to the solution vector and their wide range of practical applications in agricultural research. This study reviews some of the ways that estimable functions can be used by the agricultural scientist.

(KEY WORDS: Estimable function, regression coefficients, carryover, direct and permanent effects, sum-to-zero and set-to-zero restrictions).

Introduction

Many agricultural experiments that are originally designed to generate balanced data result in collected data that are unbalanced. It is also the case that many experiments are intrinsically unbalanced due to research or financial constraints placed on data collection. Because the majority of agricultural researchers use linear models for parameter estimation and hypothesis testing of data that are often unbalanced, it is necessary that a knowledge of the use of estimable functions is cultivated.

The use of estimable functions is the main ingredient in overspecified or overparameterized linear model parameter estimation and hypothesis testing. From a well defined linear model, various statistics can be derived from the model form of estimable function, and the appropriate use of corresponding coefficients.

The definition of an estimable function is based on its invariance property to the solution vector. From Searle (1987) *any function $k'\beta$, i.e, k =any vector and β =vector of parameters, for which $k'\beta^o$, i.e, β^o =solution vector of parameter estimates, is invariant to β^o is said to be an estimable function; and is estimated by $k'\beta^o$* . Consequently, it is expected that scientists analyzing the same data for the same purpose of estimation and hypothesis testing should obtain identical results regardless the generalized inverse they use.

It is the purpose of this paper to show that estimable functions can be used in many different applications of agricultural research. Examples are provided to (1) test for heterogeneity of the random interaction variance, and (2) to relate the usual orthogonal contrasts in a class GLM and their corresponding regression coefficients. The polynomial contrasts will be used to construct a response surface through adjusted treatment means. Finally, we present useful techniques to (3) perform a single degree of freedom test for interaction in a balanced unreplicated experiment (4) and to estimate residual (first-order

carryover), direct, and permanent effects in designs balanced for residual effects and (5) coefficients to derive estimates of within and of between experimental units in a crossover experiment. In each case, an example is given in the appendixes.

1. Heterogeneity of the Interaction Variance

Often, the agricultural scientist wants to obtain valuable information about a product by repeating the same experiment in different sites or places. The idea is to recommend application of the product to a larger population of places or to measure the effect of varying external influences or conditions. Regardless the purpose of the experiment, the analysis of the results of such a series of experiments always presents statistical challenges.

The model that is commonly used to analyze such data is a mixed model of the form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (1)$$

where $\mathbf{X}\boldsymbol{\beta}$ represents the fixed effects portion of the model, $\mathbf{Z}\boldsymbol{\mu}$ represents the random portion and $\boldsymbol{\varepsilon}$ the residuals. One of the challenges when analyzing the data using model (1) with the GLM procedure of SAS is the assessment of the interaction magnitude contained in $\boldsymbol{\mu}$. More specifically, if a treatment made of 5 different levels of application, 0, 40, 80, 120 and 160 units is tested in 6 different places, the different sources of variation are: treatment (fixed effect of $\boldsymbol{\beta}$), places and treatment*places (random effects of $\boldsymbol{\mu}$), and experimental error ($\boldsymbol{\varepsilon}$).

In many instances, it is necessary to test whether the components of the interaction treatment*places variance $\sigma^2_{\tau\mu}$ are homogeneous i.e., $\sigma^2_{\tau\mu}$ is a constant. One solution is to partition the interaction into components $\sigma^2_{1\mu}, \sigma^2_{2\mu}, \dots, \sigma^2_{p\mu}$ that extract all or almost all the useful information. This approach requires use of estimable functions. We use the example of Cochran and Cox (1957, p548) to illustrate the concept.

The orthogonal partition of the interaction is obtained by multiplying the vector of the full coefficients (extended to contain interaction terms) of treatment contrasts, K_{τ} , with the full coefficients of the *random effect* contrasts, K_{μ} . More specifically, in the example of Table 2, K_{τ} represents the first four treatment contrasts (linear, quadratic and deviations contrasts), whereas K_{μ} represents the next four deviations denoted "*Places (i-last)*".

The resulting variances may be tested for their homogeneity in the usual manner. When the null hypothesis is rejected, i.e., the interaction variance is heterogeneous, then the component of the treatment sum of squares is tested against its *own* interaction with the random factor (Cochran and Cox, 1957; Kempthorne, 1952; Milliken and Ash, personal communication).

2. Orthogonal Contrasts and Regression Coefficients

The types of experiments in which the application is simple are those resulting in balanced data with the treatment levels quantitative and equally spaced. It is possible to use combinations of the orthogonal polynomials to find the corresponding regression parameter estimates *within* the same run of GLM. The example used to illustrate the mechanics of relating dose levels defined as class variables to regression coefficients is from Steel and Torrie (1980, p366). The effect of five row spacings was tested on Ottawa Mandarin soybeans in six blocks of a randomized block design (Appendix I_a).

The mathematical expression of the two-way classification model is:

$$Y_{ijk} = u + \rho_j + \alpha_i + \varepsilon_{ijk} \quad (1)$$

$i = 1, 2, \dots, t; j = 1, 2, \dots, r$

where u is the overall mean, α_i are unknown fixed parameters, ρ_j and ε_{ijk} are independent normally distributed random variables with zero mean and variances σ_ρ^2 and σ_ε^2 , respectively. It is assumed that the 't' treatment levels are equally spaced and there is one observation in each of the 'rt' cells. The null hypothesis of equal means is formulated as:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_t$$

When the null hypothesis is rejected at a specified alpha level, researchers typically investigate further for a linear and/or quadratic response in the data through the use of orthogonal contrasts. The question of relating the results of the contrasts to corresponding regression coefficients is often asked. A simple relationship between the linear trend and the regression coefficients can be formulated as:

$$L_{\beta_1} = (\lambda * C_i) / (d * \Sigma C_i^2) \quad (2)$$

$$\text{and } L_{\beta_0} = \bar{Y} - \bar{X} * L_{\beta_1} \quad (3)$$

where the values are either constants (λ , d , \bar{X}), easily estimable (the overall mean, \bar{Y}) or intrinsically estimable contrasts of the means or totals such as the orthogonal contrasts (see Table 15.11 of Steel and Torrie, 1980). Combinations of the coefficients C_i , in the manner of (2) and (3) result in $L^* = \{L_{\beta_0}, L_{\beta_1}\}$ such that $L^* \beta^\circ = \beta$ where β° is any solution of model (1) and $\beta = \{\beta_0, \beta_1\}$ is the vector of the regression coefficients.

To illustrate the mechanics of above equations, the coefficients for L_{β_0} and L_{β_1} are derived for the soybean example:

Effect Parameter	I n t e r c e p t	Block						Spacing					D i v i s o r
		I	II	III	IV	V	VI	I	II	III	IV	V	
		$L_{\beta_1} = C_i$	0	0	0	0	0	0	0	-2	-1	0	
$\bar{Y} = \text{mean yield}$	60	10	10	10	10	10	10	12	12	12	12	12	60
$\bar{X} * L_{\beta_1} = 30 * L_{\beta_1}$	0	0	0	0	0	0	0	-60	-30	0	30	60	60
$L_{\beta_0} = \bar{Y} - \bar{X} * L_{\beta_1}$	60	10	10	10	10	10	10	72	42	12	-18	-48	60

The divisor of L_{β_1} is $d * \sum_1^5 C_i = 6 * (4+1+0+1+4) = 60$. The values of λ , d , \bar{X} are 1, 6, and 30, respectively. To compute L_{β_0} , mathematical operations between vectors are performed on coefficients within the same parameter. Appendices (I_a) and (I_b) provide the derivation of L^* for various response surfaces from the simplest with one dependent variable (I_a) to the complex with two variables, including their quadratic and interactions terms (I_b). The details that lead to these coefficients can be obtained from the first author.

3. Tukey's single-degree-of-freedom test

The basic idea of Tukey's test is to evaluate the merit of adding an interaction term in a model of main effects when no replicate is available (Tukey, 1949). It is based on partitioning the experimental error of the additive model

$$Y = \mu + \tau_i + \beta_j + \gamma_{ij} \tag{4}$$

with no interaction into two components, one associated with nonadditivity sum of squares (SSN) and another associated with the remaining sum of squares (SSR).

The traditional approach requires many runs of GLM (Shiferaw and Griffin, 1986). With the use of estimable functions, the number of steps is reduced to two. First, from the additive model, one needs to derive a new solution vector that satisfy the *sum-to-zero* restriction on the parameters instead of the default *set-to-zero* restriction. The solution given by the GLM procedure has the last parameter set-to-zero ($\tau_{last} = 0$ and $\beta_{last} = 0$). To convert it to a sum-to-zero solution, each τ_i and each β_j are assigned an integer, say 1, and the others are assigned an equal weight, $-1/(t-1)$, and $-1/(b-1)$ respectively such that $\sum \tau_i = 0$ and $\sum \beta_j = 0$.

The next step is to merge the new column estimates, say τ^* and β^* , by each of the

two factors, with the existing data and run a covariance analysis with the product of the two columns $\tau^*\beta^*$ as the covariate. The test for the covariate is the Tukey's single-degree-of-freedom for non-additivity. The data for this example are from Milliken and Johnson (1989) and are reported in Appendix II.

4. Estimation of direct, residual, and permanent effect

Crossover trials are frequently used by researchers from many fields. In experiments with residual effects, partial confounding exists between effects in the model. Consequently, in the case of a model that contains all class effects, using LSMEANS statement within GLM results in non-estimability of the fixed effects. This is especially true for overparameterized models which do not yield a unique solution but rather a biased vector based on the default *set-to-zero restrictions* on certain (last) parameters. In this instance the use of estimable functions with appropriate coefficients (see appendix III_a) leads to appropriate treatment comparisons since the expected difference between any two treatments contains a component associated with the residual difference of the same effects. Kempthorne (1957) stated that this comparison is the sole advantage of balancing with regard to the preceding treatment since "*we are not measuring a linear combination of the direct effects of one treatment and the residual effects of the others.*" Another common method that also relies on restrictions imposed on the elements of the parameter vector is the *sum-to-zero* restrictions. This method yields a full rank model (see appendix III_b).

A third alternative for estimating the parameters of interest is to combine the previous two, generally by restricting only the (class) residual to sum-to-zero. This method is easier to implement, it allows the LSMEANS statement to be estimable for the direct effects, and therefore does not require to provide own coefficients, and at the same time provides correct carryover differences as slopes in the solution vector (see appendix III_c). The data used to illustrate the example is from Cochran and Cox (1957 p135). Three rations were tested for their effects on fat corrected milk yield. The milk was collected from each cow at the end of a six week period. The basic design was an orthogonal 3X3 latin square. Cows were assigned to square based on their preliminary production: cows with similar corrected milk production were grouped in sequence of three. Then within each sequence, a randomly chosen square had its columns randomly assigned to cows. Three different statistical models were used according to the restrictions on the model parameters.

5. Estimates of within and between carryover differences

When carryover designs use two sizes of experimental units (EU), two error terms, and two levels of analysis (between and within EU), the GLM procedure of SAS fails to provide an adequate analysis. Among the three possible estimates of carryover differences, two of them, the *within* and the *between* differences can easily be obtained within GLM. The estimates of the *within experimental unit* are obtained as in the section described above (Section 4).

The estimates of the *between experimental unit* carryover differences are obtained by comparison of *whole size (EU) means*. For illustration purpose, we use the data from a three-period crossover design with three treatments (Milliken and Johnson, 1984 p447). The statistical model includes sequence effect (α), experimental unit within sequence (η), treatment effect (a , b , and c), carryover effect (Φ), period effect (Π) and the residuals (ϵ).

To obtain the between-experimental-unit estimates of carryover differences, one needs to fit the experimental unit mean model. It follows that the estimates of the carryover difference between, say Φ_a and Φ_c , is given by

$$\Phi_a - \Phi_c = \{(\alpha_1 + \alpha_3) - (\alpha_4 + \alpha_6)\} * 3/2. \quad (4)$$

Since each sequence mean is estimable, by the rule of combination of estimable functions, $\Phi_a - \Phi_c$ is also estimable (Searle 1971, 1987). Other comparisons can be made in a similar fashion (see Table 1 and appendix IV). It remains however, that if the coefficients lead to estimability for the entire model including the random effect, then the standard-error of the between difference will be based solely on the estimated MSE, σ_ϵ^2 instead of the estimate of $\sigma_\epsilon^2 + \sigma_\eta^2$.

One solution is to utilize the mixed model procedure, PROC MIXED of SAS where the RANDOM statement is used to specify the random portion of the model. The coefficients corresponding to the random effects in each ESTIMATE statement will be assigned to zero automatically (broad inference space). The narrow inference space standard error is .708. The broad inference space standard error is 2.27 (Milliken and Jonhson, 1984, give a standard error of 2.34).

Authors Note: Copies of SAS output for all examples can be requested from Dr. Louis Munyakazi, Monsanto Company, Mail Zone BB2D, 700 Chesterfield Village Parkway, St. Louis, Mo. 63198.

References

- Cochran, W.G., Autrey, K.M, and Cannon, C.Y. (1941). A double change-over design for dairy cattle feeding experiments. *Jour. Agr. Sci.* 24 937-951.
- Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs 2nd*. New York: John Wiley & Sons.
- SAS Institute Inc. SAS[®] SAS/STAT[®] User's Guide, Version 6, Fourth Edition, Volume 1-2 , Cary, NC: SAS Institute Inc. 1989. pg. 943.
- Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data. Vol I: Designed Experiments*. Van Nostrand Reinhold Co. Inc., New York.
- Steel, R.G.D, and Torrie, J.H. (1980). *Principles and Procedures of Statistics. A Biometrical Approach*. New York. McGraw_Hill Book Co.
- Little, Joyce.E., Milliken, G.A. and Schwenke, J.R. (1988). The Analysis of Williams Square Designs with Several Subjects per Sequence. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, New Orleans, 232-234.
- McLean, R.A. Sanders, W.L and Stroup, W.W. (1991). A Unified Approach to Mixed Linear Models. *The American Statistician* 45, 54-64.
- Munyakazi, L, Hintz R.L. and Selby,B.D. (1995). Estimation of Direct, Residual and Permanent Effects in Designs Balanced for Residual Effects in the GLM. *Observations[®]*. The Technical Journal for SAS Software users (in press).
- Searle, S.R. (1971). *Linear Models*, New York: John Wiley & Sons.
- Searle, S.R. (1987). *Linear Models for Unbalanced data*, New York: John Wiley & Sons.
- Milliken, G.A. and Ash, K.A.. An Investigation of the Effects of Heterogeneity of Variance on the Combined Analysis of Data From Several Experiments (personal communication).
- Kempthorne, O (1952). *The design and Analysis of Experiments*. A Wiley Intl. Ed.
- Tukey, J.W (1949). One degree of Freedom for Non-additivity. *Biometrics* 5:232-242.
- Shirferaw, G.M and Griffin, D.W (1986). One degree of freedom for transformable nonadditivity- A generalized SAS macro. in the Preceeding of SAS Users Group International. Eleventh Annual Conference. Atlanta, Georgia. February 9-12, 1986.

Table 1. Coefficients of between carryover differences in the sequence model.

Sequence Model	Between Differences		
	Φ_a vs Φ_b	Φ_a vs Φ_c	Φ_b vs Φ_c
$\alpha_1 = \delta^* + 1/3 (\Phi_a + \Phi_b)$	0	1.5	1.5
$\alpha_2 = \delta^* + 1/3 (\Phi_a + \Phi_c)$	1.5	0	-1.5
$\alpha_3 = \delta^* + 1/3 (\Phi_a + \Phi_b)$	0	1.5	1.5
$\alpha_4 = \delta^* + 1/3 (\Phi_b + \Phi_c)$	-1.5	-1.5	0
$\alpha_5 = \delta^* + 1/3 (\Phi_a + \Phi_c)$	1.5	0	-1.5
$\alpha_6 = \delta^* + 1/3 (\Phi_b + \Phi_c)$	-1.5	-1.5	0

$\delta = \mu + \text{period} + \text{treatment} + \text{subject (sequence)} + \epsilon$ (all are means, see Milliken and Johnson, 1984)

Table 2. Coefficients for TREAT*PLACES Variance Interaction Test

Effect	Treatment Contrasts				Places (i-last)				Linear*Places				Quadratic*Places					
	Linear	Quadratic	Deviations		(1-5)	(2-5)	(3-5)	(4-5)	Linear*(i-5)				Quadratic*(i-5)					
TREAT	0	-2	2	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	40	-1	-1	2	-4	0	0	0	0	0	0	0	0	0	0	0	0	0
	80	0	-2	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
	120	1	-1	-2	-4	0	0	0	0	0	0	0	0	0	0	0	0	0
	160	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
PLACES	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0
TREAT*	0	2	-.4	.4	-.2	.2	.2	0	0	0	0	0	0	.8	0	0	0	0
PLACES	0	3	-.4	.4	-.2	.2	0	.2	0	0	0	0	0	0	.8	0	0	0
	0	4	-.4	.4	-.2	.2	0	0	.2	0	0	0	0	0	0	.8	0	0
	0	5	-.4	.4	-.2	.2	0	0	0	.2	0	0	0	0	0	0	.8	0
	0	6	-.4	.4	-.2	.2	-.2	-.2	-.2	-.2	.8	.8	.8	.8	-.8	-.8	-.8	-.8
	40	2	-.2	-.2	.4	-.8	.2	0	0	0	-.4	0	0	0	-.4	0	0	0
	40	3	-.2	-.2	.4	-.8	0	.2	0	0	0	-.4	0	0	0	-.4	0	0
	40	4	-.2	-.2	.4	-.8	0	0	.2	0	0	0	-.4	0	0	0	-.4	0
	40	5	-.2	-.2	.4	-.8	0	0	0	.2	0	0	0	-.4	0	0	0	-.4
	40	6	-.2	-.2	.4	-.8	-.2	-.2	-.2	-.2	.4	.4	.4	.4	.4	.4	.4	.4
	80	2	0	-.4	0	1.2	.2	0	0	0	0	0	0	0	0	0	0	0
	80	3	0	-.4	0	1.2	0	.2	0	0	0	0	0	0	0	-.8	0	0
	80	4	0	-.4	0	1.2	0	0	.2	0	0	0	0	0	0	0	-.8	0
	80	5	0	-.4	0	1.2	0	0	0	.2	0	0	0	0	0	0	0	-.8
	80	6	0	-.4	0	1.2	-.2	-.2	-.2	-.2	0	0	0	0	.8	.8	.8	.8
	120	2	.2	-.2	-.4	-.8	.2	0	0	0	.4	0	0	0	-.4	0	0	0
	120	3	.2	-.2	-.4	-.8	0	.2	0	0	0	.4	0	0	0	-.4	0	0
	120	4	.2	-.2	-.4	-.8	0	0	.2	0	0	.4	0	0	0	0	-.4	0
	120	5	.2	-.2	-.4	-.8	0	0	0	.2	0	0	.4	0	0	0	0	-.4
	120	6	.2	-.2	-.4	-.8	-.2	-.2	-.2	-.2	-.4	-.4	-.4	-.4	.4	.4	.4	.4
	160	2	.4	.4	.2	.2	.2	0	0	0	.8	0	0	0	.8	0	0	0
	160	3	.4	.4	.2	.2	0	.2	0	0	0	.8	0	0	0	.8	0	0
	160	4	.4	.4	.2	.2	0	0	.2	0	0	0	.8	0	0	0	.8	0
	160	5	.4	.4	.2	.2	0	0	0	.2	0	0	0	.8	0	0	0	.8
	160	6	.4	.4	.2	.2	-.2	-.2	-.2	-.2	-.8	-.8	-.8	-.8	-.8	-.8	-.8	-.8

Appendix I_a

Title 'Data is from Steel and Torrie 2nd Ed. p366';
 Form of the equations: $Y = \beta_0 + \beta_1X$ and $Y = \beta_0 + \beta_1X + \beta_2X^2$
 (X =spacing)

```
options ls=125;
data one;
do block=1 to 6;
do spacing=18 to 42 by 6;
input yield @;
output;end;end;
cards;
33.6 31.1 33.0 28.4 31.4
37.1 34.5 29.5 29.9 28.3
34.1 30.5 29.2 31.6 28.9
34.6 32.7 30.7 32.3 28.6
35.4 30.7 30.7 28.1 29.6
36.1 30.3 27.9 26.9 33.4
;
proc glm data=one outstat=s;
class block spacing;
model yield=spacing block;
random block;
*----- Computation of Overall Trends -----*;
estimate 'Linear' spacing -12 -6 0 6 12;
estimate 'Quadrat' spacing 2 -1 -2 -1 2;
estimate 'Cubic' spacing -6 12 0 -12 6;
estimate 'Quartic' spacing 6 -24 36 -24 6;
*----- Computation of Contrast Sums of Squares -----*;
contrast 'SS Linear' spacing -2 -1 0 1 2;
contrast 'SS Quadratic' spacing 2 -1 -2 -1 2;
contrast 'SS Cubic' spacing -1 2 0 -2 1;
contrast 'SS Quartic' spacing 1 -4 6 -4 1;
contrast 'Total SS' spacing -2 -1 0 1 2,
spacing 2 -1 -2 -1 2,
spacing -1 2 0 -2 1,
spacing 1 -4 6 -4 1;
estimate 'Y-mean' intercept 30 spacing 6 6 6 6 block 5 5 5 5 5
/divisor=30;
*----- Computation of Polynomial 1st degree -----*;
estimate 'Beta 0' intercept 60
spacing 72 42 12 -18 -48 block 10 10 10 10 10 /divisor=60;
estimate 'Beta 1' spacing -2 -1 0 1 2 /divisor=60;
contrast 'Beta 0' intercept 60
spacing 72 42 12 -18 -48 block 10 10 10 10 10 ;
contrast 'Beta 1' spacing -2 -1 0 1 2 ;
*----- Computation of Polynomial 2nd degree -----*;
estimate 'Beta 0' intercept 11.025
spacing 49.455 -10.395 -34.02 -21.42 27.405
block 1.8375 1.8375 1.8375 1.8375 1.8375 1.8375
/divisor=11.025;
estimate 'Beta 1' spacing -820.8 309.6 720 410.4 -619.2/divisor=3024;
estimate 'Beta 2' spacing 12 -6 -12 -6 12/divisor=3024;
contrast 'Beta 0' intercept 11.025
spacing 49.455 -10.395 -34.02 -21.42 27.405
block 1.8375 1.8375 1.8375 1.8375 1.8375 1.8375
;
contrast 'Beta 1' spacing -820.8 309.6 720 410.4 -619.2;
contrast 'Beta 2' spacing 12 -6 -12 -6 12;
run;
proc print data=s label noobs split='*';
var _source_ _type_ df ss f prob;
label
_source_='Source*Of Variation'
_type_='Type*Of Sum Squares'
df='Degrees*Of Freedom'
ss='Sum*Of Squares'
f='F_value*(MSsource/MSE)';
prob='Probability*Pr > F'
run;
```

Appendix I_b

```

*-----
*-- Construct a response surface through the treatment means
*-- using the concept of estimable functions within GLM procedure
*-- This is an extension of the second degree curve through the means
*-----
Form of the equation:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$ 
(X1 =rates and X2 = days)

data one;input Days Rates @;
  do blocks=1 to 4;
    input shoots @;
    output; end;
cards;
3 0 15.7 14.6 16.5 14.7
3 4 9.8 14.6 11.9 12.4
3 8 7.9 10.3 9.7 9.6
10 0 18.0 17.4 15.1 14.4
10 4 13.6 10.6 11.8 13.3
10 8 8.8 8.2 11.3 11.2
;
proc glm;class rates days blocks;
model shoots=blocks rates*days;
estimate 'Q LinRat' rates*days -1 -1 0 0 1 1;
estimate 'Q LinDays' rates*days -1 1 -1 1 -1 1;
estimate 'Q QuaRat' rates*days 1 1 -2 -2 1 1;
estimate 'Q LinRat LinDay' rates*days 1 -1 0 0 -1 1;
estimate 'Q QuRate LinDay' rates*days -1 1 2 -2 -1 1;
estimate 'Coefficient B5' rates*days -1 1 2 -2 -1 1/divisor=224;
estimate 'Coefficient B4' rates*days 3 -3 -4 4 1 -1/divisor=56;
estimate 'Coefficient B3' rates*days -6 6 0 0 0 0/divisor=42;
estimate 'Coefficient B2' rates*days 10 -3 -20 6 10 -3/divisor=224;
estimate 'Coefficient B1' rates*days -30 9 40 -12 -10 3/divisor=56;
estimate 'Coefficient B0' intercept 84 blocks 21 21 21 21
rates*days 120 -36 0 0 0 0/divisor=84;
title1 'Estimable functions and Response Surface';

```

Appendix II

```

*-----
*-- Codes for Tukey's single degree-of-freedom test for non-additivity
*-- using GLM (or MIXED) procedure of SAS
*-----
options nodate nonumber;
data one;input temp @;
    do humidity=20 to 80 by 20;
        input height @;
        output;
    end;
cards;
50 12.3 19.6 25.7 30.4
60 13.7 16.9 27.0 31.5
70 17.8 20.0 26.3 35.9
80 12.1 17.4 36.9 43.4
90  6.9 18.8 35.0 53.0
;
proc mixed;class temp humidity;
model height=temp humidity;
make 'Estimate' out=estim;
*----- The GLM SOLUTION satisfies set-to-zero -----;
*-- restriction . We need a SUM-TO-ZERO restriction solution
*--Thus we use estimable functions to reconstruct such a solution
*--The new variables are additional data to be added to the original
*----- data set (data one) -----;
*----- The following estimates are also estimable in GLM procedure
*----- We only take advantage of the 'make' statement in the MIXED -----;
estimate 'T* 50'    temp    .8 -.2 -.2 -.2 -.2;
estimate 'T* 60'    temp    -.2 .8 -.2 -.2 -.2;
estimate 'T* 70'    temp    -.2 -.2 .8 -.2 -.2;
estimate 'T* 80'    temp    -.2 -.2 -.2 .8 -.2;
estimate 'T* 90'    temp    -.2 -.2 -.2 -.2 .8;
estimate 'B* 20'    humidity .75 -.25 -.25 -.25;
estimate 'B* 40'    humidity -.25 .75 -.25 -.25;
estimate 'B* 60'    humidity -.25 -.25 .75 -.25;
estimate 'B* 80'    humidity -.25 -.25 -.25 .75;

data temp(keep=parm est) hum(keep=parm est);
    set estim;
    if parm ge 'T* 50' then output temp;
    else output hum;

data temp(rename=(est=that));set temp;
length temp 8.;
temp=substr(parm,8,2);
proc sort;by temp;
data hum(rename=(est=hhat));set hum;
length humidity 8.;
humidity=substr(parm,8,2);
proc sort;by humidity;
proc sort data=one;
    by humidity;
data two;merge one hum;
    by humidity;
proc sort;by temp;
data add(drop=parm);merge two temp;

```

```

        by temp;
proc glm; class temp humidity;
model height=temp humidity that*hhat/ss1 ss2 ss3 ss4;
estimate 'L estimate' That*hhat 1;
contrast 'SSN Non Add' That*hhat 1;
Title1 'Tukey''s single degree-of-freedom test for non-additivity';

proc glm; class temp humidity;
model height=temp humidity that*humidity/ss3;
estimate 'Mean THAT*Humidity' that*humidity .25 .25 .25 -.75;
estimate 'Gamma 1' that*humidity 1 0 0 -1;
estimate 'Gamma 2' that*humidity 0 1 0 -1;
estimate 'Gamma 3' that*humidity 0 0 1 -1;
estimate 'Estimate Gamma1' that*humidity .75 -.25 -.25 -.25;
estimate 'Estimate Gamma2' that*humidity -.25 .75 -.25 -.25;
estimate 'Estimate Gamma3' that*humidity -.25 -.25 .75 -.25;
estimate 'Estimate Gamma4' that*humidity -.25 -.25 -.25 .75;
contrast 'SSHo' that*humidity 1 0 0 -1,
            that*humidity 0 1 0 -1,
            that*humidity 0 0 1 -1;

Title1 'Mandel''s test of interaction. Lines for each level Humidity';
Title2 'That is Mandel''s model with Temperature as the Baseline';
model height=temp humidity hhat*temp/ss3;
estimate 'Mean HHAT*Temp' hhat*temp .2 .2 .2 -.8;
estimate 'Gamma 1' hhat*temp 1 0 0 0 -1;
estimate 'Gamma 2' hhat*temp 0 1 0 0 -1;
estimate 'Gamma 3' hhat*temp 0 0 1 0 -1;
estimate 'Gamma 4' hhat*temp 0 0 0 1 -1;
estimate 'Estimate Gamma1' hhat*temp .8 -.2 -.2 -.2 -.2;
estimate 'Estimate Gamma2' hhat*temp -.2 .8 -.2 -.2 -.2;
estimate 'Estimate Gamma3' hhat*temp -.2 -.2 .8 -.2 -.2;
estimate 'Estimate Gamma4' hhat*temp -.2 -.2 -.2 .8 -.2;
estimate 'Estimate Gamma5' hhat*temp -.2 -.2 -.2 -.2 .8;
contrast 'SSHo' hhat*temp 1 0 0 0 -1,
            hhat*temp 0 1 0 0 -1,
            hhat*temp 0 0 1 0 -1,
            hhat*temp 0 0 0 1 -1;

Title1 'Mandel''s test of interaction. Lines for each level Temperature';
Title2 'That is Mandel''s model with Humidity as the Baseline';
*----- Comparison between Tukey's and Mandel's tests -----;
proc glm; class temp humidity;
model height=temp humidity that*hhat temp*hhat/ss3;
title1 'Comparison between Tukey''s and Mandel''s tests';
title2 ' ';
    
```

Appendix III_a

Codes for Estimation of Direct, Residual and Permanent Effects in GLM
 Data are from Cochran and Cox 1957 (p135)';
 'Set-to-zero restrictions'

```
proc glm;class square trt cow period resid;
model milk=square cow(square) period(square) trt resid/ss3 solution;
estimate 'Direct 1' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 18 0 0
resid 6 4 4 4/divisor=18;
estimate 'Direct 2' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 0 18 0
resid 6 4 4 4/divisor=18;
estimate 'Direct 3' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 0 0 18
resid 6 4 4 4/divisor=18;
estimate 'Residu (1+Mean)'intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 6 6 6
resid 6 12 0 0/divisor=18;
estimate 'Residu (2+Mean)'intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 6 6 6
resid 6 0 12 0/divisor=18;
estimate 'Residu (3+Mean)'intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 6 6 6
resid 6 0 0 12/divisor=18;
estimate 'Overall Mean' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 6 6 6
resid 6 4 4 4/divisor=18;
estimate 'Residu 1 alone' resid 0 12 -6 -6/divisor=18;
estimate 'Residu 2 alone' resid 0 -6 12 -6/divisor=18;
estimate 'Residu 3 alone' resid 0 -6 -6 12/divisor=18;
estimate 'Permanent 1' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 18 0 0
resid 6 16 -2 -2/divisor=18;
estimate 'Permanent 2' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 0 18 0
resid 6 -2 16 -2/divisor=18;
estimate 'Permanent 3' intercept 18 square 9 9 cow(square) 3 3 3 3 3
period(square) 3 3 3 3 3 trt 0 0 18
resid 6 -2 -2 16/divisor=18;
estimate 'Perm. 1-2' trt 18 -18 0 resid 0 18 -18 0/divisor=18;
estimate 'Direct(1-2)-Res(1-2)' trt 18 -18 0 resid 0 18 -18 0/divisor=18;
```

Appendix III_b
 CODES FOR SUM-TO-ZERO RESTRICTIONS

```
proc glm;
model milk=seq1 cow1 cow2 cow4 cow5 per1 per2 per4 per5 trt1 trt2
      resid1 resid2/solution;
estimate 'Overall'      intercept 18 seq1 9 cow1 3 cow2 3 cow4 3 cow5 3
                        per1 3 per2 3 per4 3 per5 3 trt1 6
                        trt2 6/divisor=18;

estimate 'Direct 1'      intercept 1 trt1 1;
estimate 'Direct 2'      intercept 1 trt2 1;
estimate 'Direct 3'      intercept 1 trt1 -1 trt2 -1;
estimate 'Direct 1 - 2'  trt1 1 trt2 -1;
estimate 'Direct 1 - 3'  trt1 2 trt2 1;
estimate 'Direct 2 - 3'  trt1 1 trt2 2;
estimate 'Residu 1'      resid1 1;
estimate 'Residu 2'      resid2 1;
estimate 'Residu 3'      resid1 -1 resid2 -1;
estimate 'Res 1 - 2'     resid1 1 resid2 -1;
estimate 'Res 1 - 3'     resid1 2 resid2 1;
estimate 'Res 2 - 3'     resid1 1 resid2 2;
estimate 'Permanent 1'  intercept 1 trt1 1 resid1 1;
estimate 'Permanent 2'  intercept 1 trt2 1 resid2 1;
estimate 'Permanent 3'  intercept 1 trt1 -1 trt2 -1 resid1 -1 resid2 -1;
estimate 'Perm 1-2'     trt1 1 resid1 1 trt2 -1 resid2 -1;
estimate 'Perm 1-3'     trt1 2 trt2 1 resid1 2 resid2 1;
estimate 'Perm 2-3'     trt1 1 trt2 2 resid1 1 resid2 2;
```


Appendix III_c

CODES FOR COMBINATION OF SET-TO-ZERO & SUM-TO-ZERO RESTRICTIONS

```

proc sort;by cow;
data two;set one;
retain last_trt;
resid1=0;
resid2=0;
if period ne 1 then do;
    if last_trt=1 then resid1=1;
    if last_trt=2 then resid2=1;
    if last_trt=3 then do;
        resid1=-1;resid2=-1;
    end;end;

last_trt=trt;
drop last_trt persqu;

title1 'Sum-to-Zero Constraints on the residual effect';
title2 'Set-to-Zero constraints on the other effects';
title3 'This corresponds to an analysis of covariance';
proc glm;class square trt cow period ;
model milk=square cow(square) period(square) trt resid1 resid2/solution;
lsmeans trt; * Direct effects are estimable ;
estimate 'Residu 1'      resid1 1;
estimate 'Residu 2'      resid2 1;
estimate 'Residu 3'      resid1 -1 resid2 -1;
estimate 'Residu 1 - 2'  resid1 1 resid2 -1;
estimate 'Residu 1 - 3'  resid1 2 resid2 1;
estimate 'Residu 2 - 3'  resid1 1 resid2 2;
contrast 'Carryover SS' resid1 2 resid2 1, resid1 1 resid2 2;
    
```

Appendix IV

```

-----
*--- Approximated degrees of freedom and 'naive' standard errors
*--- in mixed models with special application to specific contrasts
*--- in carryover designs and other contrasts that have at least two
*--- components of variance in their expected mean squares.
*-- Data from a Three-Period Crossover Design with Three Treatments
*-- Source: Milliken and Johnson (1984) p 447
-----
data one;input period seq trt $ resid unit1-unit6;
  drop unit1-unit6;
  array n unit1-unit6;
  do over n;
    resp=n; unit=_I_;
  output;end;
cards;
1 1 a 0 20.1 23.3 23.4 19.7 19.2 22.2
2 1 b 1 20.3 24.8 24.8 21.3 20.9 22.0
3 1 c 2 25.6 28.7 28.3 25.7 25.9 26.2
1 2 a 0 24.7 23.8 23.6 20.2 19.8 21.5
2 2 c 1 29.4 28.7 26.4 26.2 23.7 25.5
3 2 b 3 27.5 24.1 25.0 21.4 23.3 20.8
1 3 b 0 24.3 26.4 19.9 23.9 20.5 21.8
2 3 a 2 23.2 26.4 23.7 26.8 23.2 23.6
3 3 c 1 30.1 32.3 25.5 30.8 26.3 29.1
1 4 b 0 20.9 21.9 22.0 23.3 18.8 24.6
2 4 c 2 27.5 28.6 27.4 30.7 27.9 29.8
3 4 a 3 24.3 23.1 24.5 26.6 24.6 26.6
1 5 c 0 24.0 25.9 25.5 27.9 25.3 25.7
2 5 a 3 21.8 23.7 22.0 25.4 26.4 24.7
3 5 b 1 21.6 23.9 23.4 24.4 25.8 24.9
1 6 c 0 23.2 23.9 28.0 24.6 27.7 21.5
2 6 b 3 18.9 21.5 25.3 22.7 23.5 18.1
3 6 a 2 23.8 25.4 28.1 23.8 25.6 22.8
;
data one;
  set one end=eof;
  numb+1;
  output;
* The created design matches the sequence of effects in proc mixed ;
* and in proc glm ;
proc glmmod outdesign=x noprint;
  class seq trt unit period resid;
  model resp=resid period trt seq unit(seq);

  title1 'Within and Between Experimental-Units';
  Title2 ' Analysis of Variance';
proc glm data=one;
  class seq trt unit period resid;
  model resp=resid period trt seq unit(seq)/ss3;
*----- Within Experimental-Units Carryover Estimates -----
estimate 'Within Resid A-B' resid 0 1 -1;
estimate 'Within Resid A-C' resid 0 1 0 -1;
estimate 'Within Resid B-C' resid 0 0 1 -1;
*----- Between Experimental-Units Carryover Estimates begins -----
*----- summerized form -----

```

```

*--- Differences between 'sequences' represent 2/3 of the actual -----;
*---      carryover or residual difference -----;
*--- Then the desired differences are 3/2 times sequence contrasts-----;
estimate 'Between Resid A-B' seq 0 18 0 -18 18 -18
      unit(seq) 0 0 0 0 0 0 3 3 3 3 3 3 0 0 0 0 0 0
      -3 -3 -3 -3 -3 -3 3 3 3 3 3 3 -3 -3 -3 -3 -3 -3
      resid 0 12 -12
      /divisor=12;
estimate 'Between Resid A-C' seq 18 0 18 -18 0 -18
      unit(seq) 3 3 3 3 3 3 0 0 0 0 0 0 3 3 3 3 3 3
      -3 -3 -3 -3 -3 -3 0 0 0 0 0 0 -3 -3 -3 -3 -3 -3
      resid 0 12 0 -12
      /divisor=12;
estimate 'Between Resid B-C' seq 18 -18 18 0 -18
      unit(seq) 3 3 3 3 3 3 -3 -3 -3 -3 -3 -3
      3 3 3 3 3 3 0 0 0 0 0 0 -3 -3 -3 -3 -3 -3
      resid 0 0 12 -12
      /divisor=12;
    
```