THE INTERPRETATION OF EXPERIMENTS WITH POULTRY

by

MI YEON SHIM

(Under the Direction of LYNNE BILLARD)

ABSTRACT

Many agricultural scientists doing technical experiments analyze data themselves. It allows them to save high analyzing cost. However, it may cause a problem. Since scientists often have their preferred way to analyze data, they usually use the same statistical method even though the experiments are conducted with different designs. If the statistical method which they use is not appropriate for the data, the corresponding results will not be correct. Statistical analyses are important methods for interpreting results of agricultural experiments. Statistical analyses also need to be clearly communicated so that readers can properly interpret the results of experiments with poultry. Different statistical models and programming statements may lead to quite different conclusions.

INDEX WORDS: Analysis of Variance, Regression, Interaction, Statistical Analysis, Qualitative and quantitative factors, A × linear B, linear A × linear B contrasts

THE INTERPRETATION OF EXPERIMENTS WITH POULTRY

by

MI YEON SHIM

B.S., Sahm Yook University, South Korea, 2004

M.S., The University of Georgia, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2010

THE INTERPRETATION OF EXPERIMENTS WITH POULTRY

by

MI YEON SHIM

Major Professor:	LYNNE BILLARD
Committee:	ABHYUDAY MANDAL
	GENE M. PESTI

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2010

DEDICATION

I dedicate this dissertation to my husband, parents and sister. Their sincerity, patience, strength and encouragement have always inspired me to do the best that I can. Thank you.

ACKNOWLEDGEMENTS

I would first like to thank my husband, parents and sister for their unconditional love. I wish to express my deepest gratitude to my adivisors, Lynne Billard and Gene M. Pesti, for their mentoring, support and encouragement. Without them, none of this would have been possible. I would like to thank Abhyuday Mandal for serving on my committee and for his advice. Finally, I would like to thank all those individuals who helped me develop academically and personally over the course of my study in University of Georgia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Statistical analyses are important methods for interpreting results of agricultural experiments. Scientific writing of these results should clearly communicate the particulars of the research being described in a way that it can be precisely repeated. Appropriate statistical analyses are of primary importance for understanding experimental results. Probabilities (*p*-values) are often described in articles in *Poultry Science* and related journals to compare treatment means with each other and to compare regression coefficients to zero. Most published data are subjected to ANOVA (analysis of variance) or regression models using the GLM (general linear models) procedures of the SAS program (SAS Institute, 2006). The object is to test hypotheses and determine the significance levels that means are different. Different statistical models and programming statements may lead to quite different conclusions. This study compared five models to explain the influence of different statistical approaches on results of a two factor nutrition experiment with broiler chickens. The choice of an appropriate statistical model is important because conclusions from the subsequent analyses depend on the particular model used. Moreover, designs with more than two factors are needed frequently because of the complexity of modern broiler and egg production. Statistical analyses need to be clearly communicated so that readers can properly interpret the results of experiments with poultry. Designs with two or more factors are frequent players in the world of experimental design. The computational burden of the attendant analysis of variance is somewhat eased by the presence of statistical packages. Contrary to expectation, it is not clear from texts or the Manual(s) how the package(s) can be used to find components of the interaction effects, whether the factors are qualitative or quantitative factors. However, SAS can be persuaded to calculate these components (A × Linear B, etc., when A is a qualitative and B is a quantitative factor, and Linear A × Linear B, etc., when both A and B are quantitative factors). The general principle discussed and described in this work applies to many packages. However, the vehicle used here to illustrate these principles is the SAS package.

CHAPTER 2

LITERATURE REVIEW

Many agricultural scientists doing technical experiments analyze data themselves. It allows them to save high analyzing cost. However, it may cause a problem. Since scientists often have their preferred way to analyze data, they usually use the same statistical method even though the experiments are conducted with different designs. If the statistical method which they use is not appropriate for the data, the corresponding results will not be correct.

A primary object of any scientific writing should be to communicate clearly the particulars of the research being described in a way that it can be precisely repeated. Statistical analyses are often described in articles in *Poultry Science* and related journals with statements like "Differences in treatments (variables) were determined by ANOVA (analysis of variance) using the GLM (general linear models) procedures.", "Data were analyzed by using the GLM procedure of SAS (SAS Institute, 2006)" and "Data were subjected to ANOVA using the GLM procedure of SAS (SAS Institute, 2006)". These statements are from the first few papers of a recent issue of *Poultry Science*. Such statements are quite ambiguous since there are several ways to program the SAS GLM procedure. Critically, the different analyses may lead to quite different results and therefore different conclusions.

For instance, some papers showed simple or multiple regression models even though an ANOVA model is more appropriate. Dozier et al. (2010) analyzed linear and quadratic trends using PROC REG. They could have used a one-way ANOVA design with Lys (their data and included quantitative factor) as a main effect.

Some papers stopped with a one-way ANOVA model even though they measured more variables and could conduct higher order ANOVA designs. When experiments were repeated twice under the same conditions, the two experiments can be two levels of an additional factor, i.e., with "experiment" as a factor. However, some papers used pooled data from both experiments for a one-way ANOVA analysis (e.g., Atteh et al., 2007). In Saez et al. (2010), the statistical analysis states the species effect was tested by ANOVA for each age (5, 9, 12, 13, and 14 wk) and each time period (before and 1, 2, 4, and 8 h after the meal). They tested on species using only a one-way ANOVA even though they can have included age and time effects in the analysis.

One goal of the present work is to apply several different models to the same data set; see Chapter 3. Any of the models could have been used in many papers, but the details are often minimized. The models increase in complexity and ability to provide interpretable results. The last one is the most appropriate for the experiment generating the data used in Chapter 3 to illustrate the different approaches. Generic criticisms of simple models were made more than 25 years ago in plant biology (Chew, 1976; Little, 1978; Nelson and Rawlings, 1983; Swallow, 1984) that also apply to *Poultry Science*. Criticisms are equally applicable to *Poultry Science*, but have largely gone unheeded. The present analysis includes criticisms and provides an example of how to appropriately analyze and interpret data from a typical poultry science research trial.

In a different direction, some papers showed a two-way ANOVA model was fitted obtaining results for each variable (X_1 and X_2) separately as well as their interaction term effects. These papers can be extended as linear and quadratic terms if the factors are quantitative. Yadgary et al. (2010) measured parameters on two different hen ages (30 and 50 wk) and several days of incubation (0, 13, 15, 17, 19 and 21 d). They analyzed age and day as main effects using two-way ANOVA including their interaction term. Although they provided many plots which showed visually there were trends across incubation time (a quantitative factor), they did not make any statistical analysis for the presence of trends.

The availability of statistical packages has eased considerably the computational burden of many statistical analyses. Those who use them extensively are grateful. However, those same users are also painfully aware of the limitations of any particular package, limits that beguile the glossy "covers" (so–to–speak) seemingly promising so much more apparently than can be delivered, and/or limits exposed when trying to reconcile inconsistent answers generated by supposedly clear but in fact oftentimes obscure Manual instructions. This work in Appendix B focuses attention on the use of the SAS package to study trends, and in particular on an aspect of the GLM procedure as used in the analysis of experimental design data. More specifically, we consider a standard factorial design with two (or more) factors. The factors of interest are A and B. Suppose factor B is a quantitative factor. Then, among the usual quantities of interest, we can also find appropriate statistics relating to the components of B, such as Linear B, Quadratic B, etc. The GLM procedure does this and the documentation is clear on how to carry out this task. The difficulties come when we try to find components of the interaction term A × B. If A is a quantitative factor, interest centers on components Linear A × Linear B, Quadratic A × Linear B, Linear A × Quadratic B, etc. The SAS Manual provides no evidence that its GLM (or any other) procedure will calculate these components. If A is a qualitative factor, we may wish to consider components A × Linear B, A × Quadratic B, etc. Here too we are left to believe these components cannot be calculated by a SAS procedure, though there is evidence suggesting that components Linear B at a (specific) level of A, etc., can be found. Unfortunately, Manual instructions to do this are very oblique and are from a practical point of view nonexistent. Not surprisingly there is a widespread belief that SAS cannot calculate these components. This is unfortunate since the need for these components arises frequently, especially in agricultural and biological applications and in social science applications, and too often such applied researchers do not take their analyses to these extra steps because they think they “cannot” and/or “need not”.

REFERENCES

Atteh, J. O., O. M. Onagbesan, K. Tona, E. Decuypere, J. M. Geuns and J. Buyse. 2008. Evaluation of supplementary stevia *(stecia rebaudiana, bertoni)* leaves and stevioside in broiler diets: effects on feed intake, nutrient metabolism, blood parameters and growth performance. *J. Anim. Physiol. Anim. Nutr.* 92:640-649.

Chew, V. 1976. Comparison treatment means: A compendium. *HortScience* 11: 348-357.

Dozier III, W. A., A. Corzo, M. T. Kidd, P. B. Tillman, J. P. McMurtry and S. L. Branton. 2010. Digestible lysine requirements of male broilers from 28 to 42 days of age. *Poult. Sci.* 89:2173-2182.

Little, T. M. 1978. If Galileo published in HortScience. *HortScience* 13: 504-506.

Nelson, L. A. and J. O. Rawlings. 1983. Ten common misuses of statistics in agronomic research and reporting. *J. Agronomic Educ.* 12: 100-105.

Saez, G., E.Baéza, M. D. Bernadet, and S. Davail. 2010. Is there a relationship between the kinetics of liupoprotein lipase activity after a meal and the susceptibility to hepatic steatosis development in ducks? Poult. Sci. 89:2453-2460.

Swallow, W. H. 1984. Those overworked and oft-misused mean separation procedures – Duncan's, LSD, etc. *The American Phytopathological Societ*y 40:919-921.

Yadgary, L., A. Cahaner, O. Kedar and Z. Uni. 2010. Yolk sac nutrient composition and fat uptake in late-term embryos in eggs from young and old broiler breeder hens. *Poult. Sci.* 89:2441-2452.

CHAPTER 3

A COMPARISON OF STATISTICAL MODELS USING GENERAL LINEAR MODEL PROCEDURE OF SAS

[1] M. Y. Shim, L. Billard and G. M. Pesti. To be submitted to journal related to poultry.

ABSTRACT

Statistical analyses are important methods for interpreting results of agricultural experiments for scientific writing, which should clearly communicate the particulars of the research being described in a way that it can be precisely repeated. Probabilities (*p*-values) are often described in articles in *Poultry Science* and related journals to compare treatment means to each other and to compare regression coefficients to zero. Most published data are subjected to ANOVA (analysis of variance) or regression models using the GLM (general linear models) procedures of the SAS program (SAS Institute, 2006). The object is to determine the significance levels that means are different. Different statistical models and programming statements may lead to quite different conclusions. Data from an experiment with two independent variables (X_1 and X_2) and one dependent variable (Y) were analyzed. There were 6 levels of X_1 and 2 levels of X_2. Several ANOVA and regression models are reported here with or without “Class” statements in SAS. The ANOVA model requires a Class statement be included for each independent variable to signify classification variables. With the Class statement, SAS computes the Sums of Squares (SS) with n-1 degrees of freedom where n is the number of levels of each independent variable. However, without the Class statement, SAS computes the SS with only 1 degree of freedom, as in a regression model. Using either a one-way ANOVA with Duncan’s New Multiple Range Test or a two-way ANOVA, no differences between treatments were detected. When using a linear regression model, X_2 and the $X_1 \times X_2$ interaction term had significant *p*-values (.0222 and .0103, respectively). When using a second order polynomial regression model, only X_2 had a significant *p*-value (.0279). When an ANOVA with components including linear and quadratic terms was computed, the interaction term between linear X_1 and X_2 had a significant *p*-value (.0281). The choice of an appropriate statistical model is important because conclusions from the subsequent analyses depend on the particular model used.

Key words: Analysis of Variance, Regression, Interaction, Statistical Analysis

INTRODUCTION

A primary object of any scientific writing should be to communicate clearly the particulars of the research being described in a way that it can be precisely repeated. Statistical analyses are often described in articles in *Poultry Science* and related journals with statements like "Differences in treatments (variables) were determined by ANOVA (analysis of variance) using the GLM (general linear models) procedures.", "Data were analyzed by using the GLM procedure of SAS (SAS Institute, 2006)" and "Data were subjected to ANOVA using the GLM procedure of SAS (SAS Institute, 2006)". These statements are from the first few papers of a recent issue of *Poultry Science*. Such statements are quite ambiguous since there are several ways to program the SAS GLM procedure. Critically, the different analyses may lead to quite different results and therefore different conclusions.

For instance, there are two possible ways to program the SAS GLM procedure when there are several levels of the independent variables (the treatments). First, one way is as an ANOVA Model in which the SAS GLM procedure requires a "Class" statement identifying each independent variable which is being used in the analysis. The SAS GLM procedure will only compute regression coefficients if the /SOLUTION option is included with the MODEL statement. The SAS program computes relevant Sums of Squares (SS) with n-1 degrees of freedom (where n = the number of levels of each independent variable). Second, a REGRESSION Model may also be used with no Class statement. The SAS GLM procedure computes the SS with 1 degree of freedom for each independent variable and automatically calculates the regression coefficients. Degrees of freedom are important because among other roles, they are a measure of the sensitivity of the attendant F-tests and their associated *p*-values.

The results of a recent experiment were analyzed by several methods that could all be included in a statement like: “Data were analyzed by using the GLM procedure of SAS (SAS Institute, 2006)”. Using different programming statements led to different results and interpretations. The present comparative analysis was done to: 1) Show how different SAS GLM programming statements lead to different interpretations of the same data; 2) Explain how the various models should be interpreted; 3) Present the most appropriate model for analyzing the illustrating data; and 4) Make suggestions on minimum terminology that should be included when describing how experiments were analyzed independent of the statistical software package that is being used.

We apply five models all of which could fit the description of “Data were analyzed using GLM procedure of SAS (SAS Institute, 2006)” to our dataset. Any of these models could have been used in many papers, but the details are often minimized to the extent that which model was actually used is unclear. The five models herein increase in complexity and ability to provide interpretable results. The last one is the most appropriate for the particular experiment that produced these data. Generic criticisms of simple models were made more than 25 years ago in plant biology (Chew, 1976; Little, 1978; Nelson and Rawlings, 1983; Swallow, 1984). Criticisms are equally applicable to *Poultry Science*, but have largely gone unheeded. The present analysis includes criticisms (advantages and disadvantages) and provides an example of how to appropriately analyze and interpret data from a typical poultry science research trial.

The general principle discussed and described in this paper applies to many packages. However, the vehicle used here to illustrate these principles is the SAS package.

MATERIALS AND METHODS

The data were generated in a chick growth trial with two independent variables, X_1 (vitamin D) and X_2 (phytase), and one dependent variable, Y (tibial dischondroplasia percent incidence). There were 6 levels of X_1 and two levels of X_2 (Table 3.1). There were either 3 or 4 replicate observations per treatment combination. The models are provided in Appendix A.

The SAS statements were used to analyze the data in several ways. The SAS statements used for inputting data were (Variables with subscripts in the SAS statements are expressed (e.g., as $X_1 \equiv$ X1):

```
data a (=data name); input X1 X2 $ Y; cards; data;
```

Model 1

The first model was a linear regression model. It was used to see if a linear relationship exists between the X variables and Y. The SAS statements are therefore:

```
proc GLM; model Y = X1 X2 X1*X2/ss3; run;
```

If the term X1*X2 is omitted, then the model excludes the possibility of the existence of an interaction between X_1 and X_2.

Model 2

Second, a second order polynomial regression model was fitted estimating the coefficient for a quadratic term in X_1 for each term in model 1, and to determine the probabilities that the coefficient is not equal to zero. The SAS statements are now:

proc GLM; model Y = X1 X2 X1*X1 X1*X2 X1*X1*X2/ss3; run;

Other models with other kinds of quadratic terms (e.g., X2* X2) could also be considered.

Model 3

Third, a one-way ANOVA model was fitted including a comparison of pairwise means by the Duncan's test (a pairwise test on means). The one-way ANOVA model analyzed all combinations of the X_1 and X_2 factors as though there was one level, referred to as "treatments" with $6 \times 2 = 12$ levels giving 11 degrees of freedom. The SAS statement to input treatments is:

data a (=data name); input treatment; cards; data;

It is necessary to include a class statement. In order to carry out Duncan's test, a means statement is required. Therefore, the SAS statements are:

proc GLM; class treatment; model Y = treatment; means treatment/Duncan; run;

Other tests on means (such as Tukey's test) could also be considered as variations of this model.

Model 4

Fourth, a two-way ANOVA model was fitted obtaining results for each variable (X_1 and X_2) separately as well as their interaction term effects. The SAS statements become:

```
proc GLM; class X1 X2; model Y = X1 X2 X1* X2/ss3; run;
```

Note that omission of the X1*X2term (inadvisable) has a consequence that interactions between X_1 and X_2 are not considered.

Model 5

Finally, this two-way ANOVA model was repeated but the analysis included looking at interaction components such as factor by linear and quadratic terms. Though not evident from its Manuals, SAS can be persuaded to calculate these components ($X_2 \times$ Linear X_1, etc.) when X_1 is a quantitative factor and X_2 is a qualitative factor (Appendix B). The procedure can be adapted to fit other packages which have provisions for contrast calculations. A perusal of the manual suggests these components (e.g., $X_2 \times$ Linear X_1) cannot be calculated by a SAS procedure. However, one factor components such as Linear X_1 at a specific level (level 1 or 2) can be calculated (e.g., Myers, 1971). Further, statistical inference may indicate that the interaction effect is not statistically significant when in fact it is significant at differing levels of the factors involved. Applying the Appendix B methodology to the current example, we can obtain these interaction components from the set of SAS statements displayed in Figure 3.1.

Likewise, SAS can also be persuaded to calculate these components ($X_2 \times$ Quadratic X_1, etc.) when X_1 is a quantitative factor and X_2 is a qualitative factor as well as the components (Linear $X_1 \times$ Linear X_2, Linear $X_1 \times$ Quadratic X_2, etc.) when both X_1 and X_2 are quantitative factors (see Appendix B).

RESULTS

Casual observation of the data suggests that there is an interaction between X_1 and X_2 with respect to how they influence Y (Figure 3.2). Three of the five SAS GLM procedures suggest to us different conclusions (Table 3.2).

For the linear regression model (**Model 1**; Figure 3.3), we see that X_2 and the $X_1 \times X_2$ interaction term had significant *p*-values (0.0222 and 0.0103, respectively). The significant interaction indicates that both X_1 and X_2 are influencing variation in Y, and the influences are interdependent.

The second model was designed to test the hypothesis that there is a second order effect of X_1 on Y, and an interaction between X_1 and X_2 with respect to Y (**Model 2**; Figure 3.4). Using a second order polynomial regression model, only X_2 had a significant *p*-value (0.0279).

Using a one-way ANOVA design with Duncan's New Multiple Range Test included (**Model 3**; Figure 3.5), we found that no differences between treatments were detected (p = 0.6709). For the 12 treatments, the means ranged from $\bar{x}_7 = 22.775, \cdots\cdots, \bar{x}_{12} = 3.700$. Thus, even $\mu_7 = \mu_{12}$ by Duncan's New Multiple Range Test, for these data, this is because the standard deviation is large.

The fourth model (**Model 4**; Figure 3.6) was a two-way ANOVA design including classification variables. From the ANOVA table for this model, there was very little indication that any of the effects (either X_1, X_2) were influencing the variation in Y since the *p*-values are all substantially greater than 0.05. This includes the interaction effect (p = 0.2913) despite any insights suggested by Figure 3.2.

The fifth model (**Model 5**; Figure 3.7) was a two-way ANOVA, as was the fourth model, but now the analysis was extended to include main effect and interaction components. In particular, since X_1 is a quantitative factor (i.e., the levels are numerical values, here, X_1 = 0, 3, 5, 7, 9, 11; Table 3.1). We can test whether or not there is a linear trend across these levels. Here, since X_2 is a qualitative factor (with or without phytase), we can calculate the component $X_2 \times$ linear X_1 (i.e., we are testing: does the linear trend across levels of X_1 differ when phytase is present or when it is not present). The interaction between linear X_1 and X_2 had a significant *p*-value (0.0281) indicating that indeed the linear trend across levels of X_1 is indeed different when phytase is present from the corresponding trend when phytase is not present. This statistically identifies the significant interaction component observed in Figure 3.2.

A summary of the analysis for each of the five models is provided in Table 3.2. From this, it is clear that different analyses have produced different results.

DISCUSSION

From a biological perspective, both X_1 and X_2 are known to influence Y, and the experiment was conducted to determine the magnitude of the responses in the range of levels studied for a particular genotype. The choice of appropriate statistical models is dependent on what the researcher hopes to learn from the experiment.

Model 1 strengths – Regression is a form of analysis in which the relationship between one or more independent variables and the dependent variable as a linear combination of one or more model predictor variables is each weighted by so-called "regression coefficients". A linear regression model is such that the dependent variable is linearly related to each of the predictor variables and represents a straight line when the predicted value is plotted against the independent predictor variable. When there is only one predictor variable under consideration, this is called a simple linear regression. This model is simple, easy to interpret and the error degrees of freedom are maximized. Maximizing the error degree of freedom results in better estimation of σ^2 and so produces a more sensitive test.

Model 1 weaknesses – One assumption for linear regression is that observations are selected at random from the population of interest; another is that the error terms follow identical and independent normal distributions, with zero mean and common variance σ^2 for all levels of the treatments. Violation of the normality assumption on the error terms is usually of no consequence unless the sample size is very small. This follows from central limit theorems (Rice, 1995) which imply that, as long as the error terms have finite variance and are not too strongly correlated, the parameter estimates will be approximately normally distributed even when the underlying errors are not. Researchers often neglect to check for common variances. Thus, violation of the common variance assumption may be considered a weakness. However, it does not have to be, because there are variance stability transformations which can be introduced to take account of this.

Model 2 strengths – Since X_1 had more than 2 levels, this model could include second order terms (e.g., X_1^2) and also the interaction of the second order of $X_1$and X_2 ($X_1^2 \times X_2$). Note these higher order terms are equivalent to additional first order variables (e.g., $X_1^2 \equiv X_3$) so that the linear regression model still pertains. The error degrees of freedom are reduced by one for each additional term; but now the possibility of an interaction between X_1 and X_2 (e.g.) is included in the model.

Model 2 weaknesses – Same as for Model I

Model 3 strengths – One-way ANOVA is used to test for differences between two or more independent factors. The investigator is often interested in determining treatment combinations of these factors that maximize or minimize responses. The Duncan's, Tukey's, or other Multiple Range tests appear to discriminate between these treatments, suggesting one treatment is better, the same, or worse than another (Duncan, 1955; Tukey, 1949; Snedecor and Cochran, 1967).

Model 3 weaknesses – Multiple Range tests (Duncan's, Tukey's, etc) are frequently used. On balance, it is inadvisable to use them because of a lack of power. Multiple Range tests result in too high an experimentwise error rate which does not control Type I error (Boardman and Moffitt, 1971). It assumes there is no order among the different levels of the independent variables, but there most often really is. That is, it assumes the different treatments could be input as A, B, C as well as 1, 2, 3 or B, A, C. In reality, a treatment factor of 2.51 may be that best response between 1.00 and 3.00. One-way ANOVA models cannot identify this, whereas a multiple regression model could. The same concerns prevail when using the least significant difference test (LSD), see Morris (1999, p.166). Furthermore, if interaction exists between the factors, fitting one-way ANOVA models on treatment combinations is unable to identify such interaction.

Model 4 strengths – The model should test for an interaction of two independent variables affects on the dependent variable.

Model 4 weakness – Type I and Type III sums of squares are not equal when the data are unbalanced. This can result in confusion as to whether Type I or III SS should be used.

Model 5 strengths – Error degrees of freedom are greatly reduced compared to Model I and any interaction is with the ANOVA procedure (that was limited to the model 4 version). Significant differences between input variable levels should be detected as well as whether the differences appear to follow linear or quadratic trends, with the default being linear. Although the interaction between X_1 and X_2 may or may not be found to be significant, by testing for components of interaction, we can identify any interaction of Linear X_1 trends across the various levels of X_2, which for our data were significant. In our case, when levels of X_2 are ignored, the interaction effects effectively "cancel" out, and so the interaction ($X_1 \times X_2$) test alone suggests they are not significantly different.

Model 5 weakness – It is hard to program codes for SAS (SAS Institute, 2006) and other programs to extract these interaction components.

Which model is the most appropriate to answer the question: "Do X_1 and X_2 influence Y, and is there a significant interaction between the variables in the ranges studied?" (Had we conducted the third or fourth models first, we may well have concluded there is no effect of either X_1 or X_2 on Y). However, only the simplest regression model (Model 1) and the most complex ANOVA (Model 5) indicate that there is, indeed, a significant interaction between X_1 and X_2 with respect to Y (Table 3.1). Testing for the quadratic effect of X_1 (adding $X_1 \times X_1$ to the GLM model) obscures the $X_1 \times X_2$ interaction. The one-way and two-way ANOVA models do not indicate the presence of significant interaction components until, in model 5, the $X_2 \times$ Linear X_1 effect is factored out of the 5 df for the $X_1 \times X_2$ interaction. Whenever the interaction is significant, it is clear that all the independent variables are influencing the dependent variables even though further analyses are necessary to determine the nature of that inter-dependence. However, as seen in the present data set, significant *p*-values for X_1 and X_2 are not necessary to conclude the interacting factors are influencing Y since their influence may be as an interaction component such as the $X_2 \times$ Linear X_1 term observed in our dataset.

In a different direction, there is the issue of whether the Type I or Type III SS should be used. The Type I SS is a sequential procedure with the SS for the different effects calculated incrementally depending on the order these effects appear in the model statement. For example, when the model statement is

```
model Y = X1 X2 X1* X2;
```

the Type I SS are as shown in Table 3.3 (a). When the model statement is

```
model Y = X2 X1 X1* X2;
```

(i.e., the order of X_1 and X_2 is reversed), the Type I SS are as shown in Table 3.3 (b). Thus, the SS associated with the factor X_1 differs in the two cases. However, the sum (SS X_1 + SS X_2) is the same for each model. In contrast, the Type III SS shown in Table 3.3 (c) gives the same results regardless of the order written in the model statement. The same phenomena prevail if there is no interaction term (Table 3.4). Notice that when no interaction term is included, the error MS has a different value; and this also clearly impacts on the F- and *p*-values.

Further, when each variable in X (treatment) has different replications, i.e., when the data are unbalanced, Type I and III SS give different results. Again, when the data have different numbers of replications per cell, we should use Type III SS. Overall, Searle (1987, 1995) suggests that it is preferable to use the Type III SS exclusively rather than the Type I SS, though Nelder (1994) prefers the Type I SS approach. Clearly, when there is only one factor (as in Model 3 with output in Figure 3.5), the same result occurs for both Type I and Type III SS.

The PROC ANOVA procedure performs an analysis of variance for balanced designs (SAS Institute, 2006). We note here that (with few exceptions such as a one-factor design) to use PROC ANOVA, we must have a balanced design. The PROC GLM procedure is generally more efficient than ANOVA for these designs. The default use of PROC GLM obviates the need to be concerned with unequal replication numbers.

What terminology should be used to effectively communicate just how ANOVA were conducted and how results were calculated? Presently, complete programming statements would seem to be necessary when a package is used. As we have illustrated in this paper, in the absence of such statements, the reader cannot properly interpret the results or repeat the procedure, since accurate details of the analysis used are missing. Detailed explanations of SAS programming statements are available on the internet on an unrestricted basis. Therefore, readers practically anywhere can learn how calculations were made. Complete explanations of how the statistical packages are used should be available, if readers are to properly interpret computations that were made and correctly interpret the reported results. It would be better if computational methods could be included in manuscripts if they are not excessively long.

We reiterate the importance of the earlier papers to plant science (Chew, 1976; Little, 1978; Nelson and Rawlings, 1983; Swallow, 1984). The arguments are equally important to poultry science. Finally, the principles elucidated in the present work extend those of Morris (1983, 1999). In particular, the progression of models presented herein do not stop at just comparing treatment means, but advocate more detailed analyses by testing for responses starting with linear trends and interaction response components.

REFERENCES

Boardman, T. J., and D. R. Moffitt. 1971. Graphical Monte Carlo Type I error rates for multiple comparison procedures. *Biometrics* 27:738-744.

Chew, V. 1976. Comparison treatment means: A compendium. *HortScience* 11: 348-357.

Duncan, D. B. 1955. Multiple range and multiple F-tests. *Biometrics* 11:1.

Little, T. M. 1978. If Galileo published in HortScience. *HortScience* 13: 504-506.

Morris, T.R. 1983. The interpretation of response data from animal feeding trials. Pages 1 – 11 in Recent Developments in Poultry Nutrition, Cole, D.J.A. and Haresign, W. eds, Butterworths, London.

Morris, T.R. 1999. Experimental Design and Analysis in Animal Sciences. CABI Publishing, Wallingford, Oxon.

Myers, R. H. 1971. Response surface methodology. Aclyn × Bacon, Boston.

Nelder, I. A. 1994. The statistics of linear models: Back to basics. *Statistics and computing* 4: 221-234.

Nelson, L. A. and J. O. Rawlings. 1983. Ten common misuses of statistics in agronomic research and reporting. *J. Agronomic Educ.* 12: 100-105.

SAS Institute, 2006. SAS User's Guide: Statistics. Version 9.1.3 ed. SAS Inst. Inc., Cary, NC.

Searle, S. R. 1987. Linear models in unbalanced data, John Wiley, New York.

Searle, S. R. 1995. Comments on J. A. Nelder 'The statistics of linear models: back to basics' *Statistics and computing* 5: 103-107.

Snedecor, G. W. and W. G. Cochran. 1967. Statistical Methods, 6th ed. Iowa State University Press, Ames, Iowa.

Swallow, W. H. 1984. Those overworked and oft-misused mean separation procedures – Duncan's, LSD, etc. *The American Phytopathological Societ*y 40:919-921.

Tukey, J. W. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5: 99-114.

Rice, J. 1995. Mathematical Statistics and Data Analysis (Second ed.), Duxbury Press.

Table 3.1. Partial SAS data set

OBS	Treatment	X_1	X_2	Y
1	1	0	–	10.0
2	2	0	+	11.1
3	3	3	–	11.1
4	4	3	+	15.4
5	5	5	–	10.0
6	6	5	+	11.1
7	7	7	–	0.0
8	8	7	+	0.0
9	9	9	–	0.0
10	10	9	+	11.1
11	11	11	–	18.2
12	12	11	+	0.0
13	1	0	–	10.0
14	2	0	+	10.0
15	3	3	–	0.0
16	4	3	+	25.0
17	5	5	–	30.0
18	6	5	+	20.0
19	7	7	–	22.2
20	8	7	+	10.0
21	9	9	–	14.3
22	10	9	+	10.0
23	11	11	–	22.2
24	12	11	+	0.0
25	1	0	–	9.1
26	2	0	+	30.0
27	3	3	–	30.0
28	4	3	+	10.0
29	5	5	–	10.0
30	6	5	+	0.0
31	7	7	–	11.1
32	8	7	+	11.1
33	9	9	–	0.0
34	10	9	+	0.0
35	11	11	–	0.0
36	12	11	+	11.1
37	1	0	–	0.0
38	2	0	+	40.0
39	3	3	–	0.0
40	4	3	+	10.0

41	5	5	–	0.0
42	6	5	+	11.1
43	7	7	–	10.0
44	9	9	–	33.3

Table 3.2. Comparison of SAS models used to analyze the same experimental data

Model 1. Linear Regression Model (proc glm with no class statement)

$Y = 8.30366 + 0.46911\ X_1 + 0.02520\ X_2 - 0.00433\ X_1\ X_2$

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	1	65.0253	65.0253	0.70	0.4087
X_2	1	527.9220	527.9210	5.66	0.0222
X_1*X_2	1	676.6620	676.6620	7.26	0.0103
Error	40	3730.1243	93.2531	-	-
Total	43	4613.1716	-	-	-

Model 2. Second Order Polynomial Regression Model (proc glm with no class statement)

$Y = 7.61201 + 0.93521\ X_1 + 0.03027\ X_2 - 0.04339\ X_1^2 + 0.00780\ X_1\ X_2 + 0.00032\ X_1^2\ X_2$

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	1	23.1847	23.1847	0.24	0.6274
X_2	1	506.0086	506.0086	5.23	0.0279
X_1*X_1	1	6.3264	6.3264	0.07	0.7996
X_1*X_2	1	196.9171	196.9171	2.03	0.1620
$X_1*X_1*X_2$	1	42.5457	42.5457	0.44	0.5114
Error	38	3679.2456	96.8223	-	-
Total	43	4613.1716	-	-	-

Model 3. One-way ANOVA (proc glm with class statement)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Treatment	11	960.3716	87.3065	0.76	0.6709
Error	32	3652.8000	114.1500	-	-
Total	43	4613.1716	-	-	-

Duncan: $\mu_7 = 22.775$ ◄——————————————► $\mu_{12} = 3.700$

A ◄——————————————► A

Treatments: 7, 8, 6, 3, 5, 4, 9, 2, 1, 10, 11, 12

Model 4. Two-way ANOVA (proc glm with class statement)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	5	232.5486	46.5097	0.41	0.8400
X_2	1	0.00075	0.00075	0.00	0.9980
X_1*X_2	5	738.1765	147.6353	1.29	0.2913
Error	32	3652.8000	114.1500	-	-
Total	43	4613.1716	-	-	-

Model 5. Two-way ANOVA including interaction contrast with Linear and Quadratic terms (proc glm with class statement)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	5	232.5486	46.5097	0.41	0.8400
Lin X_1	1	197.7284	197.7284	1.73	0.1975
Quad X_1	1	16.6435	16.6435	0.15	0.7051
X_2	1	0.00075	0.00075	0.00	0.9980
X_1*X_2	5	738.1765	147.6353	1.29	0.2913

Lin X_1*X_2	1	603.970	603.970	5.29	0.0281
Quad X_1*X_2	1	54.362	54.362	0.48	0.4951
Error	32	3652.8000	114.1500	-	-
Total	43	4613.1716	-	-	-

Table 3.3. Comparison of Type I SS and Type III SS used to analyze the same experimental data

Type I SS (a)					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
X_1	1	7.5099	7.5099	0.08	0.7780
X_2	1	198.8754	198.8754	2.13	0.1520
X_1*X_2	1	676.6620	676.6620	7.26	0.0103
Error	40	3730.1243	93.2531	-	-
Total	43	4613.1716	-	-	-

Type I SS (b)					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
X_2	1	201.1134	201.1134	2.16	0.1498
X_1	1	5.2720	5.2720	0.06	0.8133
X_2*X_1	1	676.6620	676.6620	7.26	0.0103
Error	40	3730.1243	93.2531	-	-
Total	43	4613.1716	-	-	-

Type III SS (c)					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	1	65.0253	65.0253	0.70	0.4087
X_2	1	527.9220	527.9210	5.66	0.0222
X_1*X_2	1	676.6620	676.6620	7.26	0.0103
Error	40	3730.1243	93.2531	-	-
Total	43	4613.1716	-	-	-

Table 3.4. Comparison of Type I SS and Type III SS (no interaction)

Type I SS					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
X_1	1	7.5099	7.5099	0.07	0.7928
X_2	1	198.8754	198.8754	1.85	0.1812
Error	41	4406.7862	107.4826	-	-
Total	43	4613.1716	-	-	-
Type I SS					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
X_2	1	201.1134	201.1134	1.87	0.1788
X_1	1	5.2720	5.2720	0.05	0.8258
Error	41	4406.7862	107.4826	-	-
Total	43	4613.1716	-	-	-
Type III SS					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
X_1	1	5.2720	5.2720	0.05	0.8258
X_2	1	198.8754	198.8754	1.85	0.1812
Error	41	4406.7862	107.4826	-	-
Total	43	4613.1716	-	-	-

Figure 3.1. SAS Code for two-way ANOVA including interaction contrast with Linear and Quadratic terms (Model 5)

```
/*QUANTITATIVE-QUALITATIVE MODEL: INTERACTION COMPONENTS*/
/*X1 is QUANTITATIVE, X2 is QUALITATIVE*/

/*To compute X2*linear X1 contrast components*/
proc glm outstat = junk; class X2 X1; model Y=X2|X1/ss3;
contrast ‘linear X1’ X1 -5 -3 -1 1 3 5;
contrast ‘lin X1@X2_1’ X1 -5 -3 -1 1 3 5 X2*X1 -5 -3 -1 1 3 5;
contrast ‘lin X1@X2_2’ X1 -5 -3 -1 1 3 5 X2*X1 0 0 0 0 0 0 -5 -3 -1 1 3 5;
run;

/*To calculate (x2*linear x1) MS*/
data three;
set junk; end=last;
title ‘X2*linear X1 contrast’;

retain div dfE dfX2 sum 0;
if _N_ = 1 then div = SS/DF;                    /* Calculates Error MS*/
else div = div+ 0;
if _N_ = 1 then dfE = DF;
else dfE = dfE + 0;                              /* Keeps Error DF as dfE*/
if _N_ = 2 then dfX2 = DF;
else dfX2 = dfX2 + 0;                            /* Keeps (X2*LinearX1) DF as dfX2*/
if _N_ < 5 then SS = 0;
else if _N_ = 5 then SS = –SS;
sum = sum + SS;                                  /* Calculates (X2*LinearX1)SS*/

/* To retain the contrast needed, and to find the F– and p-statistics */
if last then do;                                 /* Or, ‘if _N_ = 7 then do;’ */
   output;
   MSsum = sum/dfX2;                             /* Calculates (X2*LinearX1)MS */
   F = MSsum/div;                                /* Calculates (X2*LinearX1) F–value */
   p = 1 – probf(F, dfX2, dfE);                  /* Calculates (X2*LinearX1) P–value */
   file print;
   put ‘ (X2*LinearX1)SS = ‘ sum;
   put ‘ (X2*LinearX1)MS = ‘ MSsum;
   put ‘ F–value = ‘ F;
   put ‘ P–value = ‘ P;
end;
run;
```

Figure 3.2. Graphical representation of the means of the data.

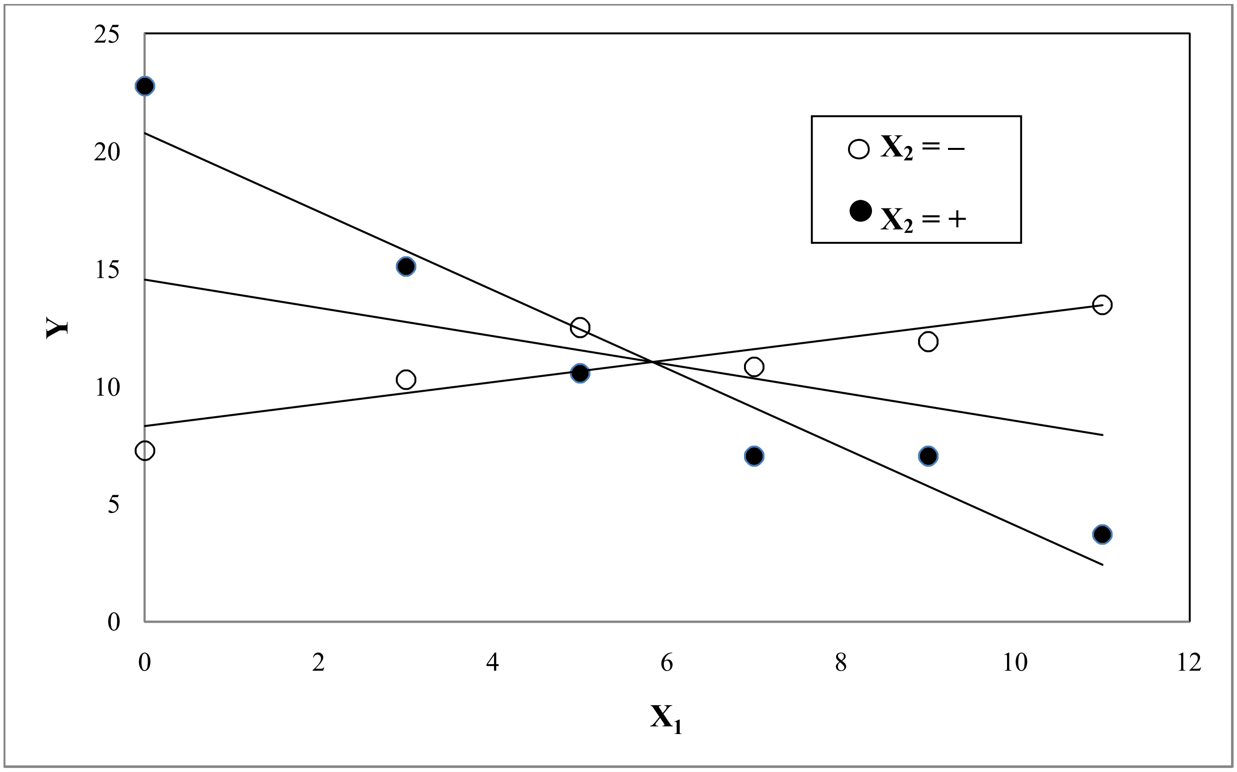

Figure 3.3. Linear Regression Model (Model 1)

SAS Output for the model: Proc GLM; model Y = X1 X2 X1*X2.

The SAS System
The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	883.047306	294.349102	3.16	0.0351
Error	40	3730.124285	93.253107		
Corrected Total	43	4613.171591			

R-Square	Coeff Var	Root MSE	Y Mean
0.191419	85.23524	9.656765	11.32955

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X1	1	65.0253449	65.0253449	0.70	0.4087
X2	1	527.9209503	527.9209503	5.66	0.0222
X1*X2	1	676.6619641	676.6619641	7.26	0.0103

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	8.303663920	3.73931509	2.22	0.0321
X1	0.469114185	0.56178338	0.84	0.4087
X2	0.025195486	0.01058937	2.38	0.0222
X1*X2	-0.004335211	0.00160937	-2.69	0.0103

Figure 3.4. Second Order Polynomial Regression Model (Model 2)

SAS Output for the model: Proc GLM; model Y = X1 X2 X1*X1 X1*X2 X1*X1*X2.

The SAS System

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	933.925967	186.785193	1.93	0.1121
Error	38	3679.245624	96.822253		
Corrected Total	43	4613.171591			

R-Square	Coeff Var	Root MSE	Y Mean
0.202448	86.85106	9.839830	11.32955

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X1	1	23.1847227	23.1847227	0.24	0.6274
X2	1	506.0086246	506.0086246	5.23	0.0279
X1*X1	1	6.3264077	6.3264077	0.07	0.7996
X1*X2	1	196.9170802	196.9170802	2.03	0.1620
X1*X1*X2	1	42.5457261	42.5457261	0.44	0.5114

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	7.612014301	4.67321885	1.63	0.1116
X1	0.935211448	1.91115739	0.49	0.6274
X2	0.030270467	0.01324121	2.29	0.0279
X1*X1	-0.043394118	0.16976176	-0.26	0.7996
X1*X2	-0.007796010	0.00546661	-1.43	0.1620
X1*X1*X2	0.000320992	0.00048423	0.66	0.5114

Figure 3.5. One way ANOVA Model (Model 3)
SAS Output for the model: Proc GLM; Class treatment; model Y = treatment; means treatment/Duncan.

```
                                The SAS System
                               The GLM Procedure
Dependent Variable: Y

                                          Sum of
        Source                    DF        Squares     Mean Square    F Value    Pr > F
        Model                     11     960.371591       87.306508       0.76    0.6709
        Error                     32    3652.800000      114.150000
        Corrected Total           43    4613.171591

                        R-Square     Coeff Var      Root MSE        Y Mean
                        0.208180      94.30299      10.68410      11.32955

        Source                    DF       Type I SS     Mean Square    F Value    Pr > F
        treatment                 11     960.3715909      87.3065083       0.76    0.6709
        Source                    DF     Type III SS     Mean Square    F Value    Pr > F
        treatment                 11     960.3715909      87.3065083       0.76    0.6709

                          Duncan's Multiple Range Test for Y

 NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error
rate.

                              Alpha                           0.05
                              Error Degrees of Freedom          32
                              Error Mean Square             114.15
                              Harmonic Mean of Cell Sizes      3.6

                               NOTE: Cell sizes are not equal.

    Number of Means      2      3      4      5      6      7      8      9     10     11     12
    Critical Range   16.22  17.05  17.59  17.97  18.26  18.49  18.68  18.83  18.95  19.06  19.15

                    Means with the same letter are not significantly different.

                  Duncan Grouping          Mean      N    treatment

                              A        22.775      4    2
                              A        15.100      4    4
                              A        13.467      3    11
                              A        12.500      4    5
                              A        11.900      4    9
                              A        10.825      4    7
                              A        10.550      4    6
                              A        10.275      4    3
                              A         7.275      4    1
                              A         7.033      3    8
```

A	7.033	3	10
A	3.700	3	12

Figure 3.6. Two-way ANOVA Model (Model 4)

SAS Output for the model: Proc GLM; Class X1 X2; model Y = X1 X2 X1*X2;

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	960.371591	87.306508	0.76	0.6709
Error	32	3652.800000	114.150000		
Corrected Total	43	4613.171591			

R-Square	Coeff Var	Root MSE	Y Mean
0.208180	94.30299	10.68410	11.32955

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X1	5	232.5485913	46.5097183	0.41	0.8400
X2	1	0.0007500	0.0007500	0.00	0.9980
X1*X2	5	738.1765324	147.6353065	1.29	0.2913

Figure 3.7. Illustration of partial SAS (SAS Institute, 2006) output generated from two-way ANOVA including interaction contrast with Linear and Quadratic terms (Model 5)

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	960.371591	87.306508	0.76	0.6709
Error	32	3652.800000	114.150000		
Corrected Total	43	4613.171591			

R-Square	Coeff Var	Root MSE	Y Mean
0.208180	94.30299	10.68410	11.32955

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X2	1	0.0007500	0.0007500	0.00	0.9980
X1	5	232.5485913	46.5097183	0.41	0.8400
X2*X1	5	738.1765324	147.6353065	1.29	0.2913

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear X1	1	197.7284444	197.7284444	1.73	0.1975
Lin X1@X2_1	1	59.5808546	59.5808546	0.52	0.4753
Lin X1@X2_2	1	742.1171463	742.1171463	6.50	0.0158

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Quad X1	1	16.64350409	16.64350409	0.15	0.7051
Quad X1@X2_1	1	5.99802647	5.99802647	0.05	0.8202
Quad X1@X2_2	1	65.00714569	65.00714569	0.57	0.4560

X2*LinearX1 Contrast

```
(X2*LinearX1)SS = 603.96955642
(X2*LinearX1)MS = 603.96955642
(X2*LinearX1)F = 5.291016701
(X2*LinearX1)P-value = 0.028105195
```

X2*QuadX1 Contrast

```
(X2*QuadX1)SS = 54.361668079
(X2*QuadX1)MS = 54.361668079
(X2*QuadX1)F = 0.476230119
(X2*QuadX1)P-value = 0.4951098148
```

CHAPTER 4

GENERAL CONCLUSION

There are multiple ways of using SAS and other statistical software packages to analyze experimental data. The approaches illustrated here are capable of extracting more information, and lead to more insightful interpretations, than are usually presented by researchers. Presently, complete programming statements would seem to be necessary when a package is used. As we have illustrated in this thesis, in the absence of such statements, the reader cannot properly interpret the results or repeat the procedure, since accurate details of the analysis used are missing. Detailed explanations of SAS programming statements are available on the internet on an unrestricted basis. Therefore, readers practically anywhere can learn how calculations were made. Complete explanations of how the statistical packages are used should be available, if readers are to properly interpret computations that were made and correctly interpret the reported results. It would be better if computational methods could be included in manuscripts if they are not excessively long.

APPENDIX A: Models in Chapter 3

The linear regression model (**Model 1**):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + e_i$$

The second order polynomial regression model (**Model 2**):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1}^2 X_{i2} + e_i$$

The one-way ANOVA model (**Model 3**):

$$Y_{ij} = \mu + \tau_i + e_{ij} \quad i = 1, \ldots, t; \quad j = 1, \ldots, r_i$$

where μ = overall mean
$\tau_i = i^{th}$ treatment effect
e_{ij} = observational error for $(ij)^{th}$ observation
Y_{ij} = observation for j^{th} replication on treatment

The two-way ANOVA model (**Model 4**):

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r$$

where μ = overall mean
$A_i = i^{th}$ A factor effect
$B_j = j^{th}$ B factor effect
AB_{ij} = interaction between factor A and B effect
e_{ijk} = observational error for $(ijk)^{th}$ observation
Y_{ijk} = observation for k^{th} replication on factor A and B effect

The two-way ANOVA model including linear and quadratic terms (**Model 5**):

$$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r$$

where μ = overall mean
$A_i = i^{th}$ A factor effect
$B_j = j^{th}$ B factor effect
AB_{ij} = interaction between factor A and B effect
e_{ijk} = observational error for $(ijk)^{th}$ observation
Y_{ijk} = observation for k^{th} replication on factor A and B effect

APPENDIX B: CALCULATING INTERACTION CONTRASTS WITH SAS[1]

[1]L. Billard, M. Y. Shim and G. M. Pesti. To be submitted to journal related to poultry.

ABSTRACT

Appropriate statistical analyses are of primary importance for understanding experimental results. Chapter 3 detailed the influence of different statistical approaches on results of a two factor nutrition experiment with broiler chickens. However, frequently designs with more than two factors are needed because of the complexity of modern broiler and egg production. Statistical analyses need to be clearly communicated so that readers can properly interpret the results of experiments with poultry. Designs with two or more factors are frequent players in the world of experimental design. The computational burden of the attendant analysis of variance is somewhat eased by the presence of statistical packages. Contrary to expectation, it is not clear from texts or the Manual(s) how the package(s) can be used to find components of the interaction effects, whether the factors are qualitative or quantitative factors. We show how SAS can be persuaded to calculate these components (A × Linear B, etc., when A is a qualitative and B is a quantitative factor, and Linear A × Linear B, etc., when both A and B are quantitative factors). The procedure can be adapted to fit other packages which have provision for contrast calculations. The results presented here extend and clarify the analyses of Chapter 3 on the advantages and disadvantages of various techniques for analyzing results from experiments on poultry with more than two factors.

Keywords: Qualitative and quantitative factors, A × linear B, linear A × linear B contrasts.

INTRODUCTION

Chapter 3 pointed out clear ambiguities in the way statistical procedures are presented in *Poultry Science* (and many other journals). Statements appearing in Poultry Science like *"Data were analyzed by using the GLM procedure of SAS (SAS Institute, 2006)"* make it impossible for readers to understand, and especially repeat, the methods that were applied. Chapter 3 demonstrated the influence of several different statistical analytical approaches to a data set from an experiment with broiler chickens. They showed how different models may lead to different conclusions from the same experimental results. They listed possible advantages and disadvantages of the several statistical techniques that may be applied to a relatively simple two factor experimental design. In this paper, we extend the results of Chapter 3 and show how experiments with two or more than factors, some quantitative and other qualitative, may be analyzed as they apply to experiments with poultry.

The availability of statistical packages has eased considerably the computational burden of many statistical analyses. Those who use them extensively are grateful. However, those same users are also painfully aware of the limitations of any particular package, limits that beguile the glossy "covers" (so–to–speak) seemingly promising so much more apparently than can be delivered, and/or limits exposed when trying to reconcile inconsistent answers generated by supposedly clear but in fact oftentimes obscure Manual instructions. This note focuses attention on the use of the SAS package, and in particular on an aspect of the GLM procedure as used in the analysis of experimental design data. More specifically, we consider a standard factorial design with two (or more) factors. The factors of interest are A and B. Suppose factor B is a quantitative factor. Then, among the usual quantities of interest, we can also find appropriate statistics relating to the components of B, such as Linear B, Quadratic B, etc. The GLM procedure does this and the documentation is clear on how to carry out this task. The difficulties come when we try to find components of the interaction term A × B. If A is a quantitative factor, interest centers on components Linear A × Linear B, Quadratic A × Linear B, Linear A × Quadratic B, etc. The SAS Manual provides no evidence that its GLM (or any other) procedure will calculate these components. If A is a qualitative factor, we may wish to consider components A × Linear B, A × Quadratic B, etc. Here too we are left to believe these components cannot be calculated by a SAS procedure, though there is evidence suggesting that components Linear B at a (specific) level of A, etc. can be found. Unfortunately, Manual instructions to do this are very oblique and are from a practical point of view nonexistent. Not surprisingly there is a widespread belief that SAS cannot calculate these components. This is unfortunate since the need for these components arises frequently, especially in agricultural and biological applications and in social science applications including in particular poultry scientists, and too often such applied researchers therefore do not take their analyses these extra steps because they think they "cannot" and/or "need not".

However, in fact, SAS can be persuaded to yield calculations on these interaction components. Our purpose here is to indicate how this can be done. Thus, we consider the case that both factors are quantitative, and we look at interaction components when one factor is qualitative and one is quantitative. We also draw attention to a related issue. Throughout, we will assume there are only two factors, A and B, with replications. Generalization to more than two factors, with or without replications, follows readily, and is considered briefly below. While the vehicle to develop these results is that for the SAS package, the principles described herein can be adapted to fit other software packages which allow the calculation of basic contrast components.

Our approach will be developed by way of an illustrative example using the data of Table 4.1. These data were extracted from the results of an experiment reported in Chamruspollert et al. (2002).

MATERIALS AND METHODS

Both Factors Quantitative

The vehicle for illustrating the methodology is a factorial design investigating the influence of two quantitative factors A (Arg) and B (Met) on the response variable (Average body weight gain in 14 d) of chickens.

Any analysis starts by entering the data appropriately, typically by using an INFILE statement or a DATALINE (or CARDS) statement followed by the actual data. Table 4.2(i) shows one version. We note that in this example factor A has four levels (1.52, 2.02, 2.52, 3.52%), factor B has three levels (0.35, 0.45, 0.55%), and there are three replications. This Table 4.2(i) also shows SAS statements asking the procedure GLM to execute the standard analysis which produces the usual statistics associated with A, B, and the interaction A × B. Also, it is reasonably straight–forward to calculate a linear component of A using SAS (SAS Institute, 2006). Thus, we include a CONTRAST statement contrast 'Linear A' A –3 –1 1 3,

after the MODEL statement. Likewise, the statement

contrast 'Linear B' B –1 0 l,

will calculate the linear component of B. These determinations are easily made by following appropriate guidelines such as those in the SAS Manual. The numbers (–3, –1, 1, 3) used in the Linear A contrast are those weights needed to construct a linear function across the levels of A. We recall that, in general, a contrast across "treatments" $T_1, \ldots, T_k$ is defined as, for equal replications per treatment,

$$z = \sum_{i=1}^{k} \omega_i T_i \text{ with } \sum_{i=1}^{k} \omega_i = 0. \tag{1}$$

In this context, the different levels of A constitute the treatments. Let us represent the weights as the vector

$$\mathbf{W} = (\omega_1, \ldots, \omega_k). \tag{2}$$

Thus, the vector of weights associated with the Linear A and Linear B contrasts in our example are written, respectively, as

$$\mathbf{A}_l = (-3, -1, 1, 3) \text{ and } \mathbf{B}_l = (-1, 0, 1).$$

Suppose now we wish to calculate the Linear A × Linear B component of the interaction A × B. This is achieved by inserting the statement

contrast 'Linear A × Linear B' A * B 3 0 –3 1 0 –1 –1 0 1 –3 0 3; (3)

between the MODEL and RUN statements. Similarly, to calculate the Linear A × Quadratic B, Quadratic A × Linear B, and Quadratic A × Quadratic B components, we use the statements

contrast 'Linear A × Quadratic B' A * B –3 6 –3 –1 2 –1 1 –2 1 3 –6 3;

contrast ‘Quadratic A × Linear B’ A * B –1 0 1 1 0 –1 1 0 –1 –1 0 1;

contrast ‘Quadratic A × Quadratic B’ A * B 1 –2 1 –1 2 –1 –1 2 –1 1 –2 1;

respectively, likewise, for other components of the A × B interaction.

These interaction component contrasts are but examples of the basic contrast definition in equation (1), where now the treatments correspond to the twelve (= 4 × 3) linear–linear levels of A × B. Formally, the weights are given by the vector

$$\mathbf{C} = vec(\mathbf{A}_l{}'\#\mathbf{B}_l)' \quad (4)$$

where if the column vector $\mathbf{A}'$ (of dimension a) has elements a_i and the row vector $\mathbf{B}$ (of dimension b) has elements b_j , then the matrix $\mathbf{D} = (\mathbf{A}'\#\mathbf{B})$ is of dimension ab and has elements $d_{ij} = a_i b_j$, and where $vec(\mathbf{D}')$ is the vector obtained by listing out the row elements of $\mathbf{D}'$ in order. For example, Table 4.3(a) gives the matrix elements found from evaluating (–3, –1, 1, 3)′#(–1, 0, 1). Hence, the weights for the Linear A × Linear B contrast and in the order they are to be used become readily apparent. Table 4.3 also provides the weights and their ordering for the Linear A × Quadratic B, Quadratic A × Linear B, and Quadratic A × Quadratic B contrast statements. The complete set of PROC GLM statements for these linear and quadratic contrasts is displayed in Table 4.2(ii), and its output is given in Table 4.2(iii).

Qualitative and Quantitative Factors

Let us now consider the case when one factor (A) is qualitative and one factor (B) is quantitative. First, the contrasts over B are evaluated at each level of A, separately. Thence, the final interaction contrast is subsequently calculated. Let us denote the levels of A by $A_1, \ldots, A_4$

Therefore, to calculate the A × Linear B contrast, we first calculate the contrasts Linear B at the level A_i, $i = 1, \ldots, 4$. For example, the Linear B at A_1 contrast is evaluated by inserting the statement

contrast 'Linear B at A_1' B –1 0 1 A * B –1 0 1 0 0 0 0 0 0 0 0 0; (5)

between the MODEL and RUN statements. Since weights not specified at the end of a weight vector are automatically set at zero, we can write this CONTRAST statement more simply as

contrast 'Linear B at A_1' B –1 0 1 A * B –1 0 1;

In contrast to the case when both factors are quantitative (where only weights for A * B were required, see, e.g., (3)), note there are two parts to this statement, one with weights appropriate to the Linear B component, viz., $\mathbf{B}_l = (-1, 0, 1)$ and one appropriate to the A×B component. The weights associated with A * B are as given by the general formula of equation (4), but now the weights for $A_{(\cdot)}$ equate to the vector of 0's except that a weight 1 appears in the i^{th} place when dealing with the i^{th} level of A. See Table 4.4 for the ($\mathbf{A}_2\#\mathbf{B}_l$) matrix for use in calculating the Linear B at the second level of component A, for example.

Therefore, to determine Linear B at A_2, we include the statements

contrast 'Linear B at A_2' B –1 0 1 A * B 0 0 0 –1 0 1;

and similarly for levels A_3 and A_4. We also need to calculate separately the Linear B component of the main effect of B. The complete set of SAS statements for the PROC GLM part of the program is displayed in Table 4.5(i), and the output is shown in Table 4.5(ii).

The completed sum of squares (SS) value is then readily found from

$$(\text{A} \times \text{Linear B})\text{SS} = \sum_{j=1}^{4} (\text{Linear B at A}_j)\text{SS} - (\text{Linear B})\text{SS} \qquad (6)$$

“by hand” if need be, but see below. Thence, for the data of our example,

(A × Linear B)SS = (552.25 + … + 8.00) – 18.0625 = 884.1875.

Hence, the F– and *P*– statistics, etc., can be evaluated.

Suppose further we wish to instruct SAS to carry out the calculations of equation (6).

This is achieved by changing the PROC GLM, statement to the statement

proc glm data = <datafilename> outstat = <filename>;

(where in our case the datafile name is “one” and the outstat file name is "junk", see Table 4.6), and by adding the set of statements as provided in Table 4.6. Before elaborating on this, it may be instructive to look more closely at what SAS is doing internally.

The OUTSTAT option allows us to keep (for subsequent use) internal SAS (SAS Institute, 2006) output not automatically printed in the standard output. To see the contents of this OUTSTAT data set, we can print them in the usual way. Thus, Table 4.7(i) gives the SAS statements needed to affect this, with the printed output shown in Table 4.7(ii). [Since we will only be using the information in the degrees of freedom (DF) and sum of squares (SS) columns, we may prefer to delete the other information. In this case, the program statements of Table 4.7(iii) can be used giving the printed output of Table 4.7(iv) instead.] Critically, the structure of this OUTSTAT data set instructs us on how to write our program for the calculation of the sum of squares of the A × Linear B contrast of (6). More specifically, we want to add the SS terms for "observations" OBS = 6, . . . , 9 and to subtract that for OBS = 5.

Most importantly at this stage is the realization that there is a hidden DO loop, with the consequence that the intuitive step of doing a natural DO loop on the OBS variable does not work. This is circumvented by the IF/ELSE statements on the (also hidden) automatic variable _N_ (SAS Language Manual; SAS Institute, 2006). Thus, to calculate the (A × Linear B)SS, we use the SAS (SAS Institute, 2006) statements

```
retain sum 0;
if _N_ < 5 then SS = 0;
else if _N_ = 5 then SS = –SS;
sum = sum + SS;
```

as in Table 4.6. Running totals are still automatically retained, as is illustrated by the output shown in Table 4.8(i). The required answer is the last value calculated, in this example, (A × Linear B)SS = 884.1875. Suppression of all but this last summation can be incorporated into the program [by asking for output only at the end, e.g.,. if last then output,].

In like manner, with appropriate use of the automatic variable _N_, the Error MS and hence the F– and *p*– statistics can be calculated. These have been done in the SAS statements of Table 4.6. The corresponding output is shown in Table 4.8(ii).

The complete SAS program for obtaining the A × Linear B contrast statistics as well as those for the A × Quadratic B contrast, is provided in Appendix B(i), and the output is shown in Appendix B(ii).

Not to Confuse the Issue ... But

In qualitative and quantitative factors, we developed program statements that would instruct the SAS package to calculate the A × Linear B, etc., contrast statistics. In particular, appropriate weights to insert into a CONTRAST statement, such as those in Table 4.5(i), were determined. It is critical to note that the order in which the factors A and B are inserted into the CLASS statement is also important. Reversing the order from (A B) to (B A) necessitates changes in the CONTRAST statements.

To illustrate, let us suppose that now factor A is quantitative and factor B is qualitative, and suppose we wish to calculate the (B × Linear A)SS. A set of CLASS and CONTRAST statements to be used are given in Table 4.9. Thus, we use

class B A;

contrast ‘Linear A at B_1’ A –3 –1 1 3 A * B –3 –1 1 3;

contrast ‘Linear A at B_2’ A –3 –1 1 3 A * B 0 0 0 0 –3 –1 1 3;

and so on. Or, we can use

class A B;

contrast ‘Linear A at B_1’ A –3 –1 1 3 A * B –3 0 0 –1 0 0 1 0 0 3 0 0; (7)

contrast ‘Linear A at B_2’ A –3 –1 1 3 A * B 0 –3 0 0 –1 0 0 1 0 0 3 0;

and so on. However, the following will not work:

class A B;

contrast ‘Linear A at B_1’ A –3 –1 1 3 A * B –3 –1 1 3;

since the order of (A B) in the class statement is incorrect for this format of the CONTRAST weights. To see this, we refer to the matrix of weights appropriate to the Linear B at A_1 contrast in Table 4.10, when the “class A B,” statement is used. It is immediately clear that the row vector **C** of weights (from equation (4)) produces the CONTRAST statement (7) above.

How a reversal of the factors in the CLASS statement affects the program when both factors are quantitative, if at all, is left as an exercise for the reader.

Two or More Factors

The same principles used in the previous sections apply when there are three or more factors, with each factor either qualitative or quantitative. We illustrate this briefly for the case where all factors are quantitative, and where one factor is qualitative and two factors are quantitative factors. Suppose all factors have three levels.

When all three factors are quantitative, the methods of both factors quantitative apply. Suppose we want to find the contrast Linear A × Linear B × Linear C. Thus, we need to calculate $\mathbf{A}'_l\#\mathbf{B}_l\#\mathbf{C}_l$. The weights for Linear A × Linear B are first calculated, as shown in both factors quantitative, i.e., $\mathbf{A}'_l\#\mathbf{B}_l = (1, 0, –1, 0, 0, 0, –1, 0, 1)$. These in turn are multiplied by the linear C weights $\mathbf{C}_l = (–1, 0, 1)$ again as shown in Section 2; see Table 4.11. Therefore, the CLASS and CONTRAST statements are

```
class A B C;
contrast 'Linear B × Linear C' A * B * C –1 0 1 0 0 0 1 0 –1
0 0 0 0 0 0 0 0 0 1 0 –1 0 0 0 –1 0 1;
```

Consider now the case where factor A is qualitative and each of B and C is a quantitative factor. Suppose in particular we are want to calculate the A × Linear B × Linear C contrast. Then, the appropriate linear weights are $\mathbf{B}_l = \mathbf{C}_l = (–1, 0, 1)$. Hence, first, from (4), the weights for Linear B × Linear C become $\mathbf{B}'_l\#\mathbf{C}_l = (1, 0, –1, 0, 0, 0, –1, 0, 1)$ in the analogous manner to that described in both factors quantitative and Table 4.3(a). Since A is a qualitative factor, then these weights are applied at each level of A analogously to that described in qualitative and quantitative factors.

The CLASS and CONTRAST statements become

```
class A B C;
contrast 'Linear  B × Linear C' B * C 1 0 –1 0 0 0 –1 0 1;
contrast 'Linear B × Linear C @ A1' B * C 1 0 –1 0 0 0 –1 0 1
                A * B * C 1 0 –1 0 0 0 –1 0 1;
contrast 'Linear B × Linear C @ A2' B * C 1 0 –1 0 0 0 –1 0 1
                A * B * C 0 0 0 0 0 0 0 0 0 1 0 –1 0 0 0 –1 0 1;
contrast 'Linear B × Linear C @ A3' B * C 1 0 –1 0 0 0 –1 0 1
                A * B * C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 –1 0 0 0 –1 0 1;
```

Then, the different interaction component contrasts are used to find the overall (A × Linear B × Linear C)SS, and hence the relevant F– and *p*– values as shown in qualitative and quantitative factors.

RESULTS AND DISCUSSION

The interpretation of the various interaction components can be facilitated by reference to Figure 4.1 which shows a surface plot of the means at each combination of the levels of A (Arg) and B (Met). For example, the analysis of the contrast interaction A × Linear B revealed this to be a significant component ($p = 0.0014$, see Table 4.8 (ii)). This tells us that there is a significant linear trend in B (Met) across the different levels of A (Arg) and that this trend is different for different levels of A (Arg). These differences for different A values are clearly evident in Figure 4.1. Note that depending on the data, it can be that there is no significant linear trend in B, but there are significant A × Linear B components (e.g., Chapter 3). Likewise, were we to consider the factor A as a quantitative factor and B as a qualitative factor (see not to confuse the issue … but), then the dotted lines corresponding to the three levels of B suggest that there is a different trend line across the levels of A. Indeed, in this case, the B × Linear A component has a significant value ($p = 0.0002$) and also there is a significantly different quadratic trend across A for the differing levels of B ($p = 0.0425$). That is, the linear trend of Arg across the levels of Met is significant and the quadratic trend across Arg for the differing levels of Met is significant at $p < 0.05$, and different for differing levels of Met.

When there are one qualitative and two quantitative variables, the surfaces will be as in the example of Figure 4.2. The data for this design were extracted from Chamruspollert et al. (2004), and consists of two levels (25, 35ºC) of a qualitative factor A (Temperature) and three levels (1.52, 2.52, 3.52% and 0.35, 0.55, 0.75%) for each of quantitative factors B (Arg) and C (Met). For the purposes of this illustration, it is assumed the design here is a standard factorial design. For these data, the visual suggestion that the Linear B × Linear C interaction differs for the two levels of A is corroborated by the statistical analysis for which $p < 0.0001$. In these kinds of designs, the surfaces correspond to the different level of A (A_i, $i = 1, \ldots, r$). There are linear surfaces across the levels of Arg and Met combinations, but this surface is different for different values of temperature. In this case, if there is a significant difference ($p < 0.05$) in the A × Linear B × Linear C interaction component, then these surfaces will assume different 'shapes'.

As illustrated in Chapter 3, there are multiple ways of using SAS and other statistical software packages to analyze experimental data. The approaches illustrated here are capable of extracting more information, and lead to more insightful interpretations, than are usually presented by the researcher. Complex experiments with multiple input factors are becoming increasingly important as poultry producers seek to balance multiple factors to maximize performance and profits while trying to minimize environmental impacts. Going the extra steps illustrated here should aid researchers and producers in properly interpreting trials where multiple factors influence productivity.

Finally, while this paper has illustrated how the SAS (SAS Institute, 2006) package can be adapted to obtain these interaction components, where possible the same principles can be applied to adapt other packages appropriately.

REFERENCES

Chamruspollert, M., G. M. Pesti, and R. I. Bakalli. 2002. Dietary interrelationships among arginine, methionine, and lysine in young broiler chicks. Br. J. Nutr. 88:655–660.

Chamruspollert, M., G. M. Pesti, and R. I. Bakalli. 2004. Chick responses to dietary arginine and methionine levels at different environmental temperatures. Br. Poult. Sci. 45:93–100.

Graybill, F. A. 1961. An Introduction to Linear Statistical Models, McGraw Hill.

SAS Institute, 2006. SAS User's Guide: Statistics. Version 9.1.3 ed. SAS Inst. Inc., Cary, NC.

Table B.1: Illustrative Data

	B_1			B_2			B_3		
A_1	327.63	308.13	320.63	386.13	372.50	372.00	345.00	389.00	381.00
A_2	278.75	264.38	211.36	363.63	359.88	345.75	331.93	349.38	352.00
A_3	254.25	191.50	206.00	314.75	355.25	338.13	313.63	355.75	418.75
A_4	181.50	144.50	157.50	176.63	240.50	290.50	369.00	336.50	385.86

Table B.2(i): Basic SAS Program

```
/* QUANTITATIVE × QUANTITATIVE MODEL. INTERACTION COMPONENTS */

options ls=72 nodate pageno=1 formdlim=' '; /* List desired options */
title 'Quantitative/Quantitative Example';
data one;
do A = 1 to 4;
	do B = 1 to 3;
		do rep = 1 to 3;
			input y@@;
			output;
		end;
	end;
end;
datalines; /* Or, 'cards,' */
327.63 308.13 320.63 386.13 372.50 372.00 345.00 389.00 381.00
278.75 264.38 211.36 363.63 359.88 345.75 331.93 349.38 352.00
254.25 191.50 206.00 314.75 355.25 338.13 313.63 355.75 418.75
181.50 144.50 157.50 176.63 240.50 290.50 369.00 336.50 385.86
;
proc glm;
class A B;
model y = A|B /ss3;
run;
```

Table B.2(ii): Contrast Statements

```
/*Contrast statements*/

proc glm;
class A B;
model y = A|B /ss3;
contrast 'Linear A × Linear B' A * B 3 0 –3 1 0 –1 –1 0 1 –3 0 3;
contrast 'Linear B × Quadratic B' A * B –3 6 –3 –1 2 –1 1 –2 1 3 –6 3;
contrast 'Quadratic A × Linear B' A * B –1 0 1 1 0 –1 1 0 –1 –1 0 1;
contrast 'Quadratic A × Quadratic B' A * B 1 –2 1 –1 2 –1 –1 2 –1 1 –2 1;
run;
```

Table B.2(iii): SAS Output

Quantitative x Quantitative Factors

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	173003.6059	15727.6005	17.62	<.0001
Error	24	21419.8748	892.4948		
Corrected Total	35	194423.4807			

R-Square	Coeff Var	Root MSE	Y Mean
0.889829	9.698180	29.87465	308.0439

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	48038.59334	16012.86445	17.94	<.0001
B	2	97474.62894	48737.31447	54.61	<.0001
A*B	6	27490.38357	4581.73060	5.13	0.0016

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Lin A X Lin B	1	18882.98497	18882.98497	21.16	0.0001
Lin A X Quad B	1	3501.57656	3501.57656	3.92	0.0592
Quad A X Lin B	1	109.52554	109.52554	0.12	0.7292
Quad A X Quad B	1	4947.30623	4947.30623	5.54	0.0271

Table B.3: Both A and B Quantitative Factors

	(a) Linear B				(b) Quadratic B		
Linear A	–1	0	1		1	–2	1
–3	3	0	–3		–3	6	–3
–1	1	0	–1		–1	2	–1
1	–1	0	1		1	–2	1
3	–3	0	3		3	–6	3
	$\mathbf{A}'_l\#\mathbf{B}_l$				$\mathbf{A}'_l\#\mathbf{B}_q$		

	(c) Linear B				(d) Quadratic B		
Quadratic A	–1	0	1		1	–2	1
1	–1	0	1		1	–2	1
–1	1	0	–1		–1	2	–1
–1	–1	0	–1		–1	2	–1
1	–1	0	1		1	–2	1
	$\mathbf{A}'_q\#\mathbf{B}_l$				$\mathbf{A}'_q\#\mathbf{B}_q$		

Table B.4: A Qualitative and B Quantitative Factors
Weight Matrix. Linear B at A_2

		Linear B		
		–1	0	1
A_1	0	0	0	0
A_2	1	–1	0	1
A_3	0	0	0	0
A_4	0	0	0	0

Table B.5(i): Contrast Statements

```
/* QUALITATIVE × QUANTITATIVE MODEL. INTERACTION COMPONENTS */
/* A is QUALitative, B is QUANTitative*/

[Data input, etc., statements]

proc glm;
class A B;
model y = A|B /ss3;
contrast 'Linear B' B –1 0 1;
contrast 'Linear B at A1' B –1 0 1 A * B –1 0 1;
contrast 'Linear B at A2' B –1 0 1 A * B 0 0 0 –1 0 1;
contrast 'Linear B at A3' B –1 0 1 A * B 0 0 0 0 0 0 –1 0 1;
contrast 'Linear B at A4' B –1 0 1 A * B 0 0 0 0 0 0 0 0 0 –1 0 1;
run;
```

Table B.5(ii): SAS Output

Qualitative x Quantitative Factors

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	173003.6059	15727.6005	17.62	<.0001
Error	24	21419.8748	892.4948		
Corrected Total	35	194423.4807			

R-Square	Coeff Var	Root MSE	Y Mean
0.889829	9.698180	29.87465	308.0439

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	48038.59334	16012.86445	17.94	<.0001
B	2	97474.62894	48737.31447	54.61	<.0001
A*B	6	27490.38357	4581.73060	5.13	0.0016

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear B	1	91472.74954	91472.74954	102.49	<.0001
Linear B @ A1	1	4192.85535	4192.85535	4.70	0.0403
Linear B @ A2	1	12956.76540	12956.76540	14.52	0.0009
Linear B @ A3	1	31737.91740	31737.91740	35.56	<.0001
Linear B @ A4	1	61582.29660	61582.29660	69.00	<.0001

Table B.6: Computing the A × Linear B Contrast Statistics

```
/* QUALITATIVE × QUANTITATIVE MODEL. INTERACTION COMPONENTS */
/* A is QUALitative, B is QUANTitative*/

[Data input, etc., statements, see Table 2(i)]

proc glm data=one outstat=junk;

[Other statements for A × Linear B contrast components, see Table 5(i)]

/* To compute (A × Linear B)MS */

data three;
set junk end=last;
title ‘A × Linear B Contrast’;

retain div dfE dfA sum 0;
if N = 1 then div = SS/DF;                    /* Calculates Error MS */
else div = div +0;
if N = 1 then dfE = DF;
else dfE =dfE + 0;                            /* Keeps Error DF as dfE */
if N = 2 then dfA = DF;
else dfA =dfA + 0;                            /* Keeps A DF as dfA */
if N < 5 then SS = 0;
else if N = 5 then SS = –SS;
sum = sum + SS;                               /* Calculates (A × Linear B)SS */

/* To retain the contrast needed, and to find the F– and p– statistics */
if last then do;                              /* Or, 'if N = 9 then do,' */
      output;
      MSsum = sum/dfA;                        /* Calculates (A × Linear B)MS */
      F = MSsum/div;                          /* Calculates (A × Linear B) F–value */
      p = 1 – probf(F,dfA,dfE) ;              /* Calculates (A × Linear B) p–value */
      file print;
      put  ’ (A × Linear B)SS = ’ sum 10.4;     /*Keep 4 decimal places */
      put  ’ (A × Linear B)MS = ’ MSsum 10.4;
      put  ’ F–value = ’ F 8.4;
      put  ’ p–value = ’ p 6.5;
end;
run;
```

Table B.7(i): OUTSTAT Data Program

```
/* QUALITATIVE × QUANTITATIVE MODEL. INTERACTION COMPONENTS */
/* A is QUALitative, B is QUANTitative*/
[Data input, and PROC GLM statements]
/* To print outstat data */
proc print data=junk,
title 'Outstat data from PROC GLM',
run,
```

Table B.7(ii): SAS OUTSTAT Output

Outstat data from PROC GLM - [_N_ NAME SOURCE TYPE]

OBS	NAME	SOURCE	TYPE	DF	SS	F	PROB
1	Y	ERROR	ERROR	24	21419.87	.	.
2	Y	A	SS3	3	48038.59	17.942	0.000003
3	Y	B	SS3	2	97474.63	54.608	0.000000
4	Y	A × B	SS3	6	27490.38	5.134	0.001612
5	Y	Linear B	CONTRAST	1	91472.75	102.491	0.000000
6	Y	Linear B at A1	CONTRAST	1	4192.86	4.698	0.040343
7	Y	Linear B at A2	CONTRAST	1	12956.77	14.517	0.000850
8	Y	Linear B at A3	CONTRAST	1	31737.92	35.561	0.000004
9	Y	Linear B at A4	CONTRAST	1	61582.30	69.000	0.000000

Table B.7(iii): OUTSTAT Data Program – DF and SS only

```
/* To keep and print DF and SS data */

data two;
set junk;
keep DF SS;
proc print data=two;
title 'A × Linear B. DF/SS data only';
run,
```

Table B.7(iv): Portion SAS OUTSTAT Output

A × Linear B: DF/SS data only

OBS	DF	SS
1	24	21419.87
2	3	48038.59
3	2	97474.63
4	6	27490.38
5	1	91472.75
6	1	4192.86
7	1	12956.77
8	1	31737.92
9	1	61582.30

Table B.8(i) Contrast SS: Output as Running Totals

A × Linear B Contrast

(A × Linear B)SS = 0
(A × Linear B)SS = 0
(A × Linear B)SS = 0
(A × Linear B)SS = 0
(A × Linear B)SS = –91472.750
(A × Linear B)SS = –87279.894
(A × Linear B)SS = –74323.129
(A × Linear B)SS = 18997.085
(A × Linear B)SS = 18997.085

Table B.8(ii): SAS Output for A × Linear B Contrast Statistics

A × Linear B Contrast

(A × Linear B)SS = 18997.085
(A × Linear B)MS = 6332.362
F = 7.0951
P = 0.00141

Table B.9: Reversing the CLASS Statements

```
/* QUALITATIVE × QUANTITATIVE MODEL. INTERACTION COMPONENTS */
/* A is QUALitative and B is QUANTitative */

[Data input, etc. statements]

/* To compute the B × Linear B */

proc glm data=one outstat=junkl;
class A B;
model y = A|B /ss3;
contrast ‘Linear A’ A –3 –1 1 3;
contrast ‘Linear A at Bl’ A –3 –1 1 3 A * B –3 0 0 –1 0 0 1 0 0 3 0 0;
contrast ‘Linear A at B2’ A –3 –1 1 3 A * B 0 –3 0 0 –1 0 0 1 0 0 3 0;
contrast ‘Linear A at B3’ A –3 –1 1 3 A * B 0 0 –3 0 0 –1 0 0 1 0 0 3;
run;

/*OR, Alternatively: */

proc glm data=one outstat=junkl;
class B A;
model y = A|B /ss3;
contrast ‘Linear A’ A –3 –1 1 3;
contrast ‘Linear A at Bl’ A –3 –1 1 3 A * B –3 –1 1 3;
contrast ‘Linear A at B2’ A –3 –1 1 3 A * B 0 0 0 0 –3 –1 1 3;
contrast ‘Linear A at B3’ A –3 –1 1 3 A * B 0 0 0 0 0 0 0 0 –3 –1 1 3;
run;

/* To compute B*QuadraticA contrast components*/

proc glm data=one outstat=junk2;
title 'Qualitative × Quantitative Example';
class A B;
model y = A|B /ss3;
contrast ‘Quad A’ A –1 1 1 –1;
contrast ‘Quad A at Bl’ A –1 1 1 –1 A * B –1 0 0 1 0 0 1 0 0 –1 0 0;
contrast ‘Quad A at B2’ A –1 1 1 –1 A * B 0 –1 0 0 1 0 0 1 0 0 –1 0;
contrast ‘Quad A at B3’ A –1 1 1 –1 A * B 0 0 –1 0 0 1 0 0 1 0 0 –1;
run,
```

Table B.10: Factor A Quantitative and Factor B Qualitative Weight Matrix. Linear A at B_1

	B_1	B_2	B_3
Linear A	1	0	0
–3	–3	0	0
–1	–1	0	0
1	1	0	0
3	3	0	0
	A'_l#B_l		

Table B.11: Factors A, B, C Quantitative Weight Matrix. Linear A × Linear B × Linear C

		Linear A × Linear B								
Linear C		1	0	–1	0	0	0	–1	0	1
C_1	–1	–1	0	1	0	0	0	1	0	–1
C_2	0	0	0	0	0	0	0	0	0	0
C_3	1	1	0	–1	0	0	0	–1	0	1
		A'_l#B_l#C_l								

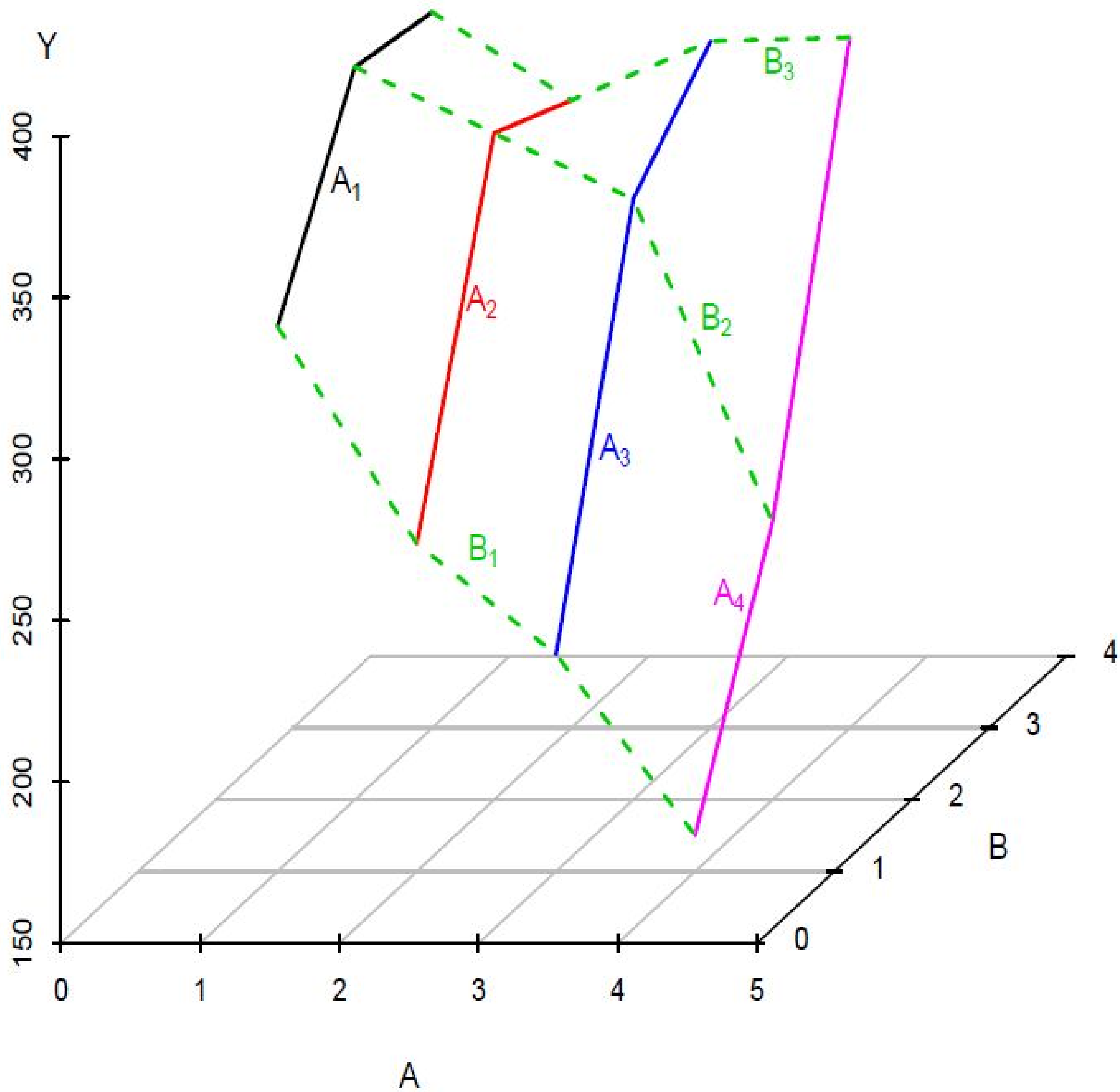

Figure B.1: Response Surface for Factors A and B

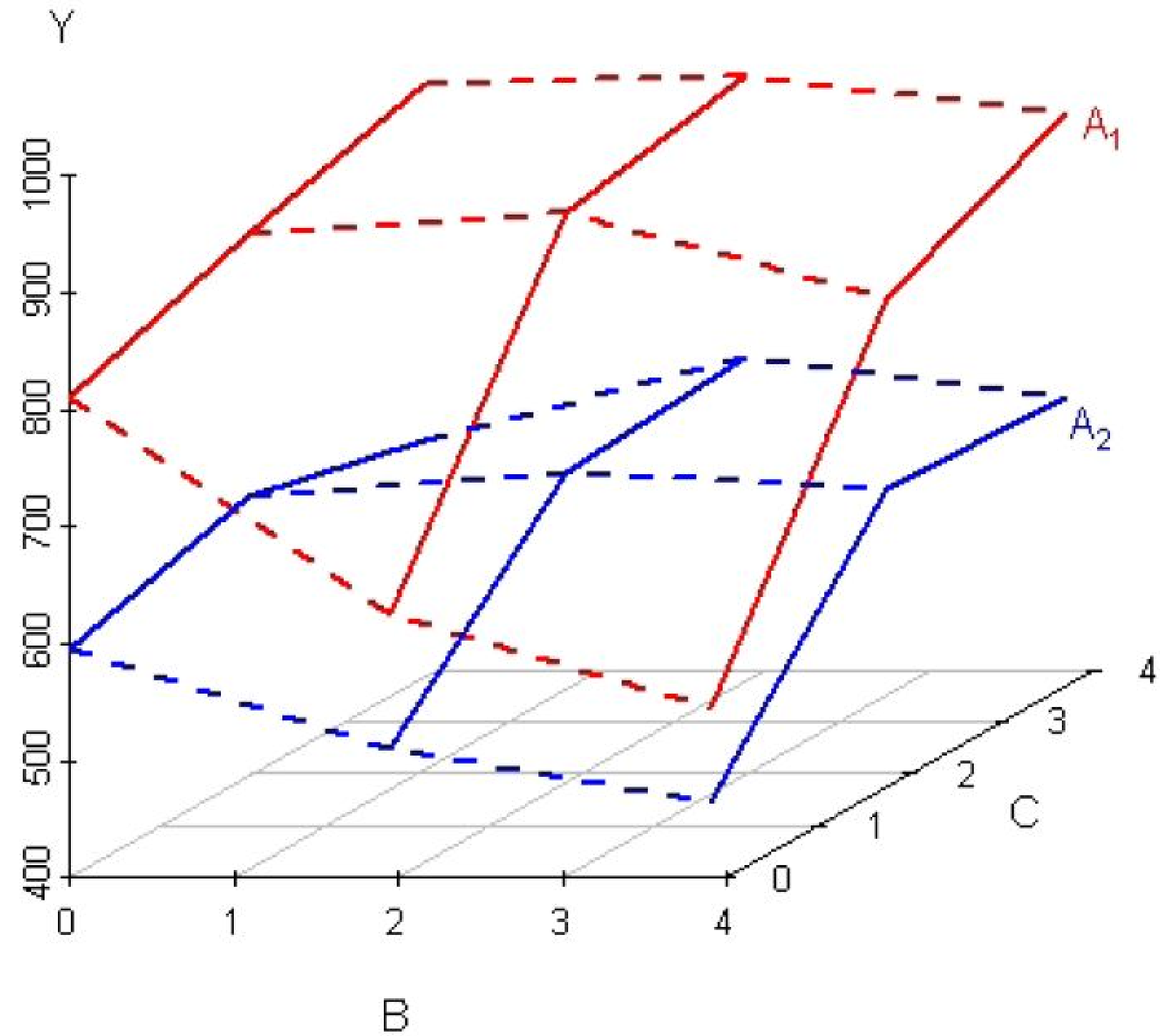

Figure B.2: Response Surfaces for B and C at A_i, i = 1, 2

APPENDIX C(i): SAS Program. A × Linear B and A × Quadratic B Contrasts

---

```
/* QUALITATIVE × QUANTITATIVE MODEL....INTERACTION COMPONENTS */
/* A is QUALitative and B is QUANTitative */

dm 'output; clear; log; clear';
options ls=72 nodate formdlim=' ' ;
title 'Qualitative × Quantitative Example';
data one;
infile '(data directory path) dataname.dat';
input ABy@@;
output;
run;

/* To compute A × Linear B contrast components */

proc glm data=one outstat=junkl;
class A B;
model y = A|B /ss3;
contrast 'Linear B' B –1 0 1;
contrast 'Linear B at Al' B –1 0 1 A × B –1 0 1;
contrast 'Linear B at A2' B –1 0 1 A × B 0 0 0 –1 0 1;
contrast 'Linear B at A3' B –1 0 1 A × B 0 0 0 0 0 0 –1 0 1;
contrast 'Linear B at A4' B –1 0 1 A × B 0 0 0 0 0 0 0 0 0 –1 0 1;
run;

/* To calculate (A × Linear B)SS */

data two;
set junkl end=last;
title 'A × Linear B Contrast';

/*Note. There is a hidden DO loop.
Therefore the output is a running total,
the required answer is the last one given.*/

retain div dfA dfE sum 0;
if _N_ = 1 then div = SS/DF;
else div = div + 0;
if _N_ = 1 then dfE = DF;
else dfE = dfE + 0;
if _N_ = 2 then dfA = DF;
```

```
else dfA = dfA + 0;
if _N_ < 5 then SS = 0;
else if _N_ = 5 then SS = –SS;
sum = sum + SS;

/* To retain the contrast needed, and to find the F– and p– statistics */

if last then do;
        output;
        MSsum = sum/dfA;
        F = MSsum/div;
        p = l–probf(F,dfA,dfE),
        file print,
        put /’ (A × Linear B)SS = ’ sum ’(A × Linear B)MS = ’ MSsum;
        put /’ F = , F ’ p–value = , p1;
end;
run;

/* To compute A × Quadratic B contrast components */

proc glm data=one outstat=junk2;
class A B;
model y = A|B /ss3;

contrast ‘Quadratic B’ B 1 –2 1;
contrast ‘Quadratic B at A1’ B 1 –2 1 A * B 1 –2 1;
contrast ‘Quadratic B at A2’ B 1 –2 1 A * B 0 0 0 1 –2 1;
contrast ‘Quadratic B at A3’ B 1 –2 1 A * B 0 0 0 0 0 0 1 –2 1;
contrast ‘Quadratic B at A4’ B 1 –2 1 A * B 0 0 0 0 0 0 0 0 0 1 –2 1;
run;

/* To calculate (A × Quadratic B)SS */
/*Note the program code is similar to that for the (A × Linear B)SS, except that
the <outstat> dataset differs since the contrast statements differ.
Marco statements can be used instead.*/
/* Note. There is a hidden DO loop*/

data three;
set junk2 end=last;
title ‘A × Quadratic B Contrast’;

retain div dfA dfE sum 0;
```

```
if _N_ = 1 then div = SS/DF;
else div = div + 0;
if _N_ = 1 then dfE = DF;
else dfE = dfE + 0;
if _N_ = 2 then dfA = DF;
else dfA = dfA + 0;
if _N_ < 5 then SS = 0;
else if _N_ = 5 then SS = –SS;
sum = sum + SS;
if _N_ = 9 then do;
        output;
        MSsum = sum/dfA;
        F = MSsum/div;
        p = 1–probf(F,dfA,dfE);
        file print;
        put /’ (A × Quadratic B)SS = ’ sum ’(A × Quadratic B)MS ’ MSsum;
        put /’ F = ’ F ’ p–value = ’ p;
end;
run;
```

APPENDIX C(ii): SAS Output: A × Linear B and A × Quadratic B Contrasts

Qualitative x Quantitative Example

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	173003.6059	15727.6005	17.62	<.0001
Error	24	21419.8748	892.4948		
Corrected Total	35	194423.4807			

R-Square	Coeff Var	Root MSE	Y Mean
0.889829	9.698180	29.87465	308.0439

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	48038.59334	16012.86445	17.94	<.0001
B	2	97474.62894	48737.31447	54.61	<.0001
A*B	6	27490.38357	4581.73060	5.13	0.0016

A x LinearB Contrasts

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear B	1	91472.74954	91472.74954	102.49	<.0001
Linear B @ A1	1	4192.85535	4192.85535	4.70	0.0403
Linear B @ A2	1	12956.76540	12956.76540	14.52	0.0009
Linear B @ A3	1	31737.91740	31737.91740	35.56	<.0001
Linear B @ A4	1	61582.29660	61582.29660	69.00	<.0001

(A*LinearB Contrast)SS = 18997.0852 (A*LinearB Contrast)MS = 6332.3617
F = 7.0951246503 p-value = 0.0014129691

A x QuadraticB Contrasts

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Quadratic B	1	6001.879401	6001.879401	6.72	0.0159
Quadratic B @ A1	1	2002.812050	2002.812050	2.24	0.1472
Quadratic B @ A2	1	6833.584356	6833.584356	7.66	0.0107
Quadratic B @ A3	1	4243.661356	4243.661356	4.75	0.0393
Quadratic B @ A4	1	1415.120000	1415.120000	1.59	0.2201

(A*QuadraticB Contrast)SS = 8493.2984 (A*QuadraticB Contrast)MS = 2831.0995
F = 3.1721187688 p-value = 0.0425295631