ISSN 1682-8356
ansinet.org/ijps

INTERNATIONAL JOURNAL OF

POULTRY SCIENCE

ANSInet
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan
Mob: +92 300 3008585, Fax: +92 41 8815544
E-mail: editorijps@gmail.com

International Journal of Poultry Science 13 (2): 76-87, 2014
ISSN 1682-8356

Experimental Design and Analysis with Emphasis on Communicating What Has Been Done: I. A Comparison of Statistical Models Using General Linear Model Procedure of SAS

M.Y. Shim[1], L. Billard[2] and G.M. Pesti[1]
[1]Department of Poultry Science, University of Georgia, Athens, GA-30602, USA
[2]Department of Statistics, University of Georgia, Athens, GA-30602, USA

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 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) procedure 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. Illustrative 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. By 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 (0.025 and 0.014, respectively). When using a second order polynomial regression model, only X_2 had a significant p-value (0.036). When an ANOVA with components including linear and quadratic terms was computed, the interaction term between X_2 and (linear X_1) had a significant p-value (0.023). 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 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 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 will only compute regression coefficients if the "/SOLUTION" option is included with the MODEL 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.

Corresponding Author: G.M. Pesti, Department of Poultry Science, University of Georgia, Athens, GA-30602, USA

Although poultry science journals have well-defined instructions on how to describe the statistical models adequately and how they are analyzed, it is clear these instructions are not always being followed. Therefore, our aim herein is to consider just five different analyses (there are many other possible analyses and even each of these illustrative five can themselves be modified in various ways to produce yet more possible analyses) to illustrate the consequences of inadequate documentation of the underlying statistical procedures implemented. The results of a recent broiler chicken 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)". However, different statistical models can be analyzed by the GLM procedure of SAS (Fig. 1). 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. The first two are pure regression models and the last three are analysis of variance models; all use the "proc glm" procedure. 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

Suppose 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 dyschondroplasia percent incidence). There were 6 levels of X_1 (1, 3, 5, 7, 9, 11) and two levels of X_2 (absent, present, set as 0, 500, respectively), Table 1 for an illustrative dataset. There were 4 replicate observations per treatment combination. A summary of our five possible models is provided in Appendix A.

Table 1: Illustrative data set

OBS	Treatment	X_1	X_2	Y
1	1	1	0	10.0
2	2	1	500	11.1
3	3	3	0	11.1
4	4	3	500	15.4
5	5	5	0	10.0
6	6	5	500	11.1
7	7	7	0	0.0
8	8	7	500	0.0
9	9	9	0	0.0
10	10	9	500	11.1
11	11	11	0	18.2
12	12	11	500	0.0
13	1	1	0	10.0
14	2	1	500	10.0
15	3	3	0	0.0
16	4	3	500	25.0
17	5	5	0	30.0
18	6	5	500	20.0
19	7	7	0	22.2
20	8	7	500	10.0
21	9	9	0	14.3
22	10	9	500	10.0
23	11	11	0	22.2
24	12	11	500	0.0
25	1	1	0	9.1
26	2	1	500	30.0
27	3	3	0	30.0
28	4	3	500	10.0
29	5	5	0	10.0
30	6	5	500	0.0
31	7	7	0	11.1
32	8	7	500	11.1
33	9	9	0	0.0
34	10	9	500	0.0
35	11	11	0	0.0
36	12	11	500	11.1
37	1	1	0	0.0
38	2	1	500	40.0
39	3	3	0	0.0
40	4	3	500	10.0
41	5	5	0	0.0
42	6	5	500	11.1
43	7	7	0	10.0
44	8	7	500	11.1
45	9	9	0	33.3
46	10	9	500	7.033
47	11	11	0	22.2
48	12	11	500	20.0

The SAS statements were used to analyze the data in several ways. The SAS statements used for inputting data are:

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

where, variables with subscripts are expressed in the SAS statements as X1 = X_1, e.g., Notice that if the X_2 variable is a non-numeric variable, it can be coded as a

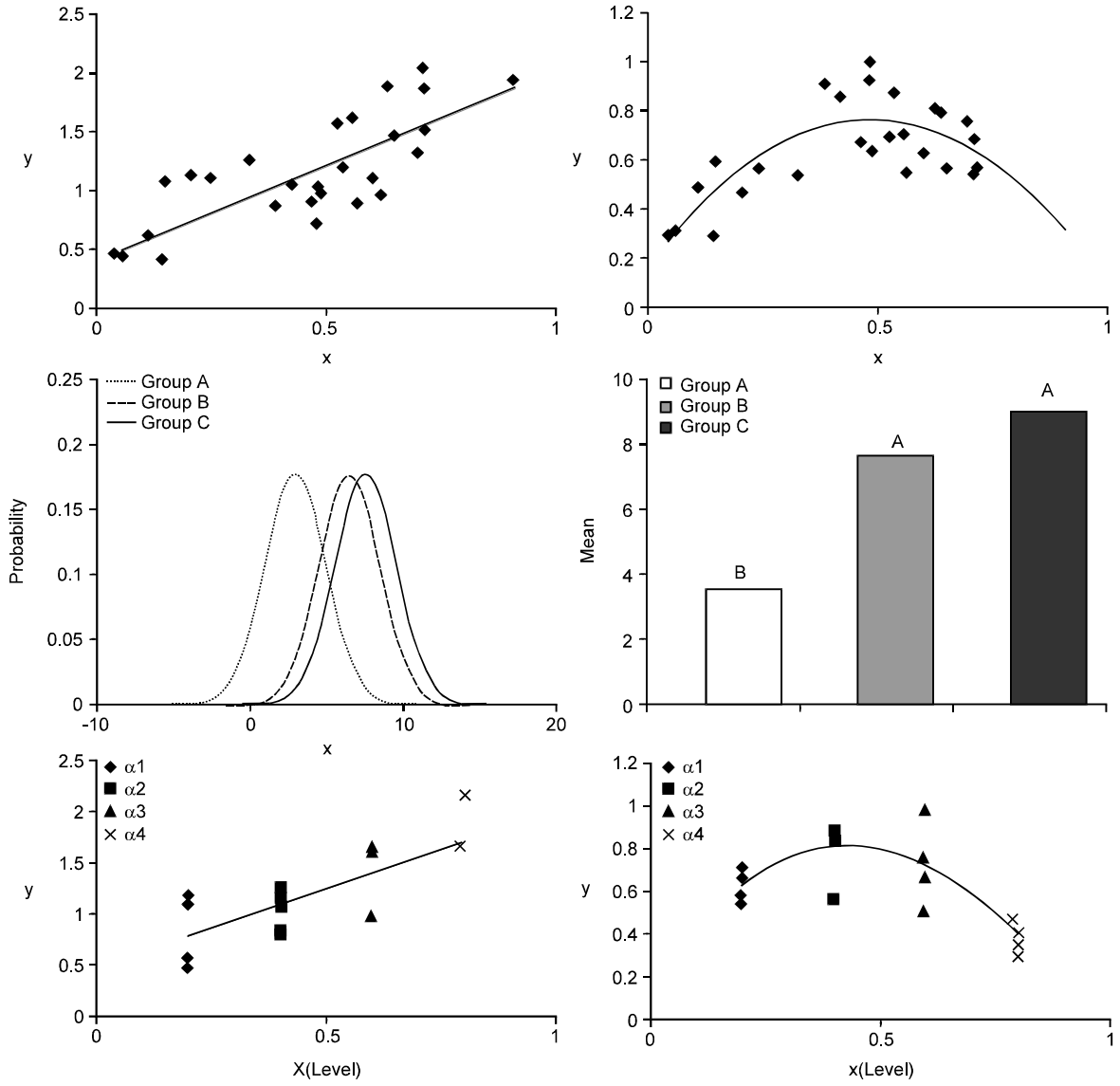

Fig. 1: Several models that could all be performed "using the GLM procedure of SAS". 1st order linear regression (Model 1), 2nd order polynomial regression (Model 2), ANOVA (Model 3-6), Multiple range test (Model 3-5), ANOVA with linear trend (Model 5) and ANOVA with quadratic trend (Model 6)

numeric value (0, 1 for absent, present). Or, it can be coded as a non-numeric value (e.g., -, + for absent, present); in this case the "X2" in the data input line is replaced by "X2 $". The "data lines;" term can alternatively be entered as "cards;" and "data" refers to the list of data values.

Model 1: The first possible model used herein was a linear regression model. It is typically used to see if a linear relationship exists between the (X_1 and X_2) variables and Y. The SAS statements are therefore:

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

That is, each term in the model statement except the observational error term e (see Appendix A) appears in the SAS model statement line. 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. In previous versions of SAS, to obtain the Type III SS output only, "/ss3" is added to the model statement. The most recent version now has this Type III SS as the default output; we will assume this default exists for our subsequent models. A different version of this SAS statement can be used, if the analyst wants to test for the lack of fit of this model, by adding "/lackfit" at the end of the model statement.

Model 2: Second, a second order polynomial regression model was fitted estimating the coefficient for a quadratic term in X1 for each term in Model 1. The SAS statements are:

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

Other models with other kinds of quadratic terms (e.g., X2* X2) could also be considered. Indeed, an analyst may choose to add new terms to Model 1 systematically and progressively by adding only the "X1*X1" first (i.e., omitting the "X1*X1*X2" term in our illustrative Model 2), then an "X2*X2" term and so on, until reaching a model that gave an adequate fit. If the researcher's goal was to find the best regression model, then this progressive approach is one that could be used (or, the model could start with all possible polynomial terms and systematically drop terms shown to be not statistically significant). Our purpose here is not necessarily to find the best model, but to show there are many possible models and hence SAS statements, that could be used under the sweeping assertions that "using SAS, we ...."

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 x 2 = 12 levels. The model term reflecting these treatments is the T_i (Appendix A). The SAS statement to input treatments is:

data a (=data name); input treatment Y; data lines; data

For the "proc glm" procedure to run an analysis of variance, 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. (In the analysis of variance context, these "variables" are now often called "factors".) The SAS statements become:

proc GLM; class X1 X2; model Y = X1 X2 X1*X2; run

The model now replaces the treatment term T_i in Model 3 by its components A_i, B_j and $(AB)_{ij}$ (Appendix A). Note that omission of the X1* X2 term has a consequence that interactions between X_1 and X_2 are not considered and so is another possible model albeit inadvisable.

Model 5: Finally, this two-way ANOVA of Model 4 was repeated but the analysis included looking at factor components such as linear X_1 and also interaction components such as factor by linear and quadratic terms. Though not evident from its Manuals, SAS can be instructed to calculate these components (X_2 × Linear X_1, etc.) when X_1 is a quantitative factor and X_2 is a qualitative factor (Billard *et al*., 2014). A perusal of the manual suggests these components (e.g., X_2 × Linear X_1) are not calculated directly 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 Billard *et al*. (2014) methodology to the current example, we can obtain these interaction components. The SAS statements for Model 5 are presented elsewhere (Billard *et al*., 2014).

Likewise, SAS can also be instructed to calculate the components (X_2 × 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 × Linear X_2, Linear X_1 × Quadratic X_2, etc.) when both X_1 and X_2 are quantitative factors (Billard *et al*., 2014).

When there are more than two factors, e.g., X_1, X_2 and X_3, analogous model choices can be made. Thus, the SAS model statements for Model 1, Model 4 and Model 5 become:

model Y = X1 X2 X3 X1*X2 X1*X3 X2*X3

where, the class statement is omitted in Model 1 but becomes "class X1 X2 X3" in Model 4 and Model 5.

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 (Fig. 2). Three of the five SAS GLM procedures suggest different conclusions (Table 2). Clearly, not all can be correct for the actual experiment as run to produce the data and/or for different goals. These differences become self-evident by using the different models/analyses on the same data. Therefore, it is essential that the researcher specifies completely how the experiment was conducted, what model is being used and what analysis is being implemented.

For the linear regression model (Model 1; Fig. 3), we see that X_2 and the X_1 × X_2 interaction term had significant p-values (0.025 and 0.014, respectively). The significant interaction indicates that both X_1 and X_2 are influencing variation in Y and the influences are interdependent.

Table 2: Comparison of SAS models used to analyze the same experimental data

Source	DF	Type III SS	Mean Square	F Value	Pr>F
Model 1. Linear Regression Model (proc glm with no class statement) $Y = 7.54060 + 0.64393\ X_1 + 0.02611\ X_2 - 0.00419\ X_1X_2$					
X_1	1	116.100	116.100	1.24	0.271
X_2	1	500.734	500.734	5.36	0.025
$X_1{}^*X_2$	1	615.514	615.514	6.59	0.014
Error	44	4110.483	93.420	-	-
Total	47	4820.331	-	-	-
Model 2. Second Order Polynomial Regression Model (proc glm with no class statement) $Y = 7.44826 + 0.68946\ X_1 + 0.03851\ X_2 - 0.00379\ X_1{}^2 - 0.01031\ X_1X_2 + 0.00051\ X_1{}^2X_2$					
X_1	1	7.446	7.446	0.08	0.780
X_2	1	443.492	443.492	4.70	0.036
$X_1{}^*X_1$	1	0.034	0.034	0.00	0.985
$X_1{}^*X_2$	1	208.046	208.046	2.21	0.145
$X_1{}^*X_1{}^*X_2$	1	77.538	77.538	0.82	0.370
Error	42	3959.957	94.285	-	-
Total	47	4820.331	-	-	-
Model 3. One-way ANOVA (proc glm with class statement)					
Treatment	11	898.657	81.696	0.75	0.685
Error	36	3921.674	108.935	-	-
Total	47	4820.331	-	-	-
Duncan: $\mu_2 = 22.775$ ---------- $M_{10} = 7.033$					
A ---------- A					
Treatments: 2, 11, 4, 5, 9, 7, 6, 3, 8, 12, 1, 10					
Model 4. Two-way ANOVA (proc glm with class statement)					
X_1	5	177.188	35.438	0.33	0.894
X_2	1	2.723	2.723	0.02	0.875
$X_1{}^*X_2$	5	718.746	143.749	1.32	0.278
Error	36	3921.674	108.935	-	-
Total	47	4820.331	-	-	-
Model 5. Two-way ANOVA including interaction contrast with Linear and Quadratic terms (proc glm with class statement)					
X_1	5	177.188	35.438	0.33	0.894
Lin X_1	1	91.612	91.612	0.84	0.365
Quad X_1	1	72.988	72.988	0.67	0.418
X_2	1	2.723	2.723	0.02	0.875
$X_1{}^*X_2$	5	718.746	143.749	1.32	0.278
$X_2{}^*\text{Lin}X_1$	1	615.514	615.514	5.65	0.023
$X_2{}^*\text{Quad}\ X_1$	1	77.538	77.538	0.71	0.404
Error	36	3921.674	108.935	-	-
Total	47	4820.331	-	-	-

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; Fig. 4). If we use a second order polynomial regression model, only X_2 had a significant p-value (0.036).

Using a one-way ANOVA design with Duncan's New Multiple Range Test included (Model 3; Fig. 5), we found that no differences between treatments were detected ($p = 0.685$). For the 12 treatments, the means ranged from $\bar{X}_2 = 22.775$, ..., $\bar{X}_{10} = 7.033$. Thus, even μ_2 is not statistically significantly different from μ_{10} by Duncan's New Multiple Range Test, for these data. This is because the standard deviation (10.437) is large relative to the range (15.742 = 22.775-7.033) of treatment mean values (i.e., $|15.742| < (1.96)(10.437) = 20.457$).

The fourth model (Model 4; Fig. 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 ($p = 0.278$) despite any insights suggested by Fig. 2.

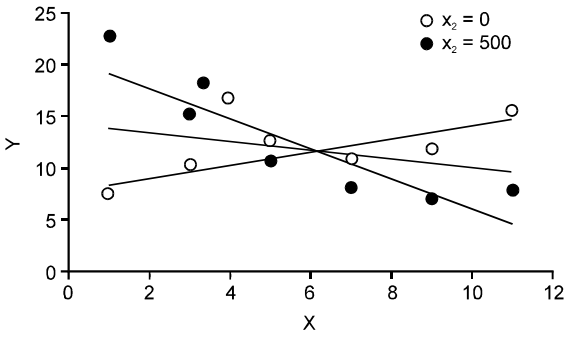

Fig. 2: Graphical representation of the means of the data

The fifth model (Model 5; Fig. 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, $X_1 = 1, 3, 5, 7, 9, 11$; Table 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

Fig. 3: Linear Regression Model (Model 1)

SAS (SAS Institute, 2006) Output for the model: Proc GLM; model Y = X1 X2 X1*X2.

The SAS System

The GLM Procedure

Dependent Variable: Y

Source	DF	Squares	Mean Square	F Value	Pr>F
Model	3	709.848711	236.616237	2.53	0.0692
Error	44	4110.482672	93.420061		
Corrected Total	47	4820.331383			

R-Square	Coeff. Var.	Root MSE	Y Mean
0.147261	83.01934	9.665405	11.64235

Source	DF	Type III SS	Mean Square	F Value	Pr>F
X1	1	116.1003214	116.1003214	1.24	0.2710
X2	1	500.7343978	500.7343978	5.36	0.0253
X1*X2	1	615.5135432	615.5135432	6.59	0.0137

Parameter	Estimate	Standard Error	t Value	Pr>|t|
Intercept	7.540595238	3.98793846	1.89	0.0652
X1	0.643928571	0.57761845	1.11	0.2710
X2	0.026114221	0.01127959	2.32	0253
X1*X2	-0.004193579	0.00163375	-2.57	0.0137

Fig. 4: Second Order Polynomial Regression Model (Model 2)

SAS (SAS Institute, 2006) 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	Squares	Mean Square	F Value	Pr>F
Model	5	860.374686	172.074937	1.83	0.1287
Error	42	3959.956697	94.284683		
Corrected Total	47	4820.331383			

R-Square	Coeff. Var.	Root MSE	Y Mean
0.178489	83.40263	9.710030	11.64235

Source	DF	Type III SS	Mean Square	F Value	Pr>F
X1	1	7.4462143	7.4462143	0.08	0.7801
X2	1	443.4923067	443.4923067	4.70	0.0358
X1*X1	1	0.0344048	0.0344048	0.00	0.9848
X1*X2	1	208.0463640	208.0463640	2.21	0.1449
X1*X1*X2	1	77.5384275	77.5384275	0.82	0.3697

Parameter	Estimate	Standard Error	t Value	Pr>|t|
Intercept	7.448258929	6.27821179	1.19	0.2421
X1	0.689464286	2.45337866	0.28	0.7801
X2	0.038512652	0.01775746	2.17	0.0358
X1*X1	-0.003794643	0.19864709	-0.02	0.9848
X1*X2	-0.010307873	0.00693920	-1.49	0.1449
X1*X1*X2	0.000509525	0.00056186	0.91	0.3697

calculate the component X_2 × linear X_1 (i.e., we are testing: does the linear trend across levels of X_1 differ when phytase is present from that when it is not present). The interaction between X_2 and (linear X_1) had a significant p-value (0.023) indicating that 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 Fig. 2. These five different statistical models for the GLM procedure using SAS are presented in Appendix A. A summary of the corresponding analysis is provided in Table 2. From this, it is clear that different analyses have produced different results, again re-iterating the necessity to be specific about what is actually being done.

DISCUSSION

From a biological perspective, both X_1 and X_2 are known to influence Y and the experiment was conducted to

Fig. 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

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	11	898.657216	81.696111	0.75	0.6848
Error	36	3921.674167	108.935394		
Corrected Total	47	4820.331383			

R-Square	Coeff. Var.	Root MSE	Y Mean
0.186431	89.64864	10.43721	11.64235

Source	DF	Type I SS	Mean Square	F Value	Pr>F
treatment	11	898.6572162	81.6961106	0.75	0.6848

Source	DF	Type III SS	Mean Square	F Value	Pr>F
treatment	11	898.6572162	81.6961106	0.75	0.6848

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	36
Error Mean Square	108.94

Number of Means	2	3	4	5	6	7	8	9	10	11	12
Critical Range	14.97	15.74	16.24	16.60	16.87	17.08	17.26	17.40	17.52	17.63	17.71

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	treatment
A	22.775	4	2
A	15.650	4	11
A	15.100	4	4
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	8.050	4	8
A	7.775	4	12
A	7.275	4	1
A	7.033	4	10

Fig. 6: Two-way ANOVA Model (Model 4)

SAS (SAS Institute, 2006) 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	898.657216	81.696111	0.75	0.6848
Error	36	3921.674167	108.935394		
Corrected Total	47	4820.331383			

R-Square	Coeff. Var.	Root MSE	Y Mean
0.186431	89.64864	10.43721	11.64235

Source	DF	Type III SS	Mean Square	F Value	Pr>F
X1	5	177.1879551	35.4375910	0.33	0.8944
X2	1	2.7231977	2.7231977	0.02	0.8753
X1*X2	5	718.7460634	143.7492127	1.32	0.2778

determine the magnitude of the responses in the range of levels studied for a particular genotype. We highlight herein some of the strengths and weaknesses of the five possible models/analyses used (in the context that

Table 3: Comparison of Type I SS and Type III SS used to analyze the same experimental data

Source	DF	Type I SS	Mean Square	F Value	Pr>F
Type I SS (a)					
X_1	1	201.113	201.113	2.16	0.150
X_2	1	5.272	5.272	0.06	0.813
X_1*X_2	1	676.662	676.662	7.26	0.010
Error	40	3730.124	93.253	-	-
Total	43	4613.172	-	-	-
Type I SS (b)					
X_2	1	7.510	7.510	0.08	0.778
X_1	1	198.875	198.875	2.13	0.152
X_2*X_1	1	676.662	676.662	7.26	0.010
Error	40	3730.124	93.253	-	-
Total	43	4613.172	-	-	-
Type III SS (c)					
X_1	1	65.025	65.025	0.70	0.409
X_2	1	527.921	527.921	5.66	0.022
X_1*X_2	1	676.662	676.662	7.26	0.010
Error	40	3730.124	93.253	-	-
Total	43	4613.172	-	-	-

Table 4: Comparison of Type I SS and Type III SS (no interaction)

Source	DF	Type I SS	Mean Square	F Value	Pr>F
Type I SS					
X_1	1	201.113	201.113	1.87	0.179
X_2	1	5.272	5.272	0.05	0.826
Error	41	4406.786	107.483	-	-
Total	43	4613.172	-	-	-
Type I SS					
X_2	1	7.510	7.510	0.07	0.793
X_1	1	198.875	198.875	1.85	0.181
Error	41	4406.786	107.483	-	-
Total	43	4613.172	-	-	-
Type III SS					
X_1	1	198.875	198.875	1.85	0.181
X_2	1	5.272	5.272	0.05	0.826
Error	41	4406.786	107.483	-	-
Total	43	4613.172	-	-	-

the model, e.g., regression as opposed to analysis of variance, was appropriate). Thus, for these data, if the object was to identify the relationship that exists between vitamin D and/or phytase on tibial dischrondroplasia incidence, the Model 1 (or Model 2 and its variations) is the best model to pursue; but if the object is to ascertain the effects of different levels of vitamin D and phytase on tibial dischrondroplasia incidence, then the analyses of Model 4 or 5 are better. That is, as elucidated earlier the choice of appropriate statistical model and analysis is dependent on what the researcher hopes to learn from the experiment and crucially on how the experiment was run. Further, notice that the model statement in the SAS code is identical for Models 1, 4 and 5. However, in Model 1, there is no "class" statement, with the result that SAS carries out a regression analysis; whereas with the "class" statement in Models 4-5, SAS gives an analysis of variance result (which for this experiment is the correct approach).

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 and easy to interpret.

Model 1 weaknesses A basic weakness of this model is this simplicity, if no attempt is made by the analyst to consider other fit options such as those delineated e.g., in Model 2 above; this includes the failure to include a lack-of-fit calculation.

In a different direction, 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. Rather than a weakness of the model, this is really a weakness of the analysis itself. Since our goal here is to show how different analyses can lead to different results (including a failure to check the normality and variance assumptions), we note only that these same comments apply to all models considered here and so will not be repeated.

Model 2 strengths Since X_1 had more than 2 levels, this model could include a second order term (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., $X_1^2 = X_3$) so that the linear regression model still pertains and as formulated here it is still relatively simple. 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. If the analyst includes the option to investigate yet further higher order terms, then this becomes a strength.

Model 2 weaknesses. The weaknesses are the same as for Model I. If the analyst does not include the option to investigate the inclusion of other higher order terms and/or the non-significance of any of the lower order terms (of Model 1), then this can be seen as a weakness. On the other hand, trying to consider all these options can be time consuming. Further, it can be difficult to interpret (some) higher order terms biologically.

Model 3 strengths One-way ANOVA is used to test for differences between two or more independent factors. In theory, these tests can be used on any kind of treatments (all qualitative, all quantitative, mixture of qualitative and quantitative as in our case); see any introductory text on design (e.g., Steel and Torrie, 1960). 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 experiment-wise 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 (especially for quantitative factors). 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), Morris (1999, 1983). Furthermore, if interaction exists between the factors, fitting one-way ANOVA models on treatment combinations is unable to identify such interaction. When, as for our data, there are quantitative factors involved, a far better way, without loss of sensitivity, is to calculate the orthogonal contrasts (i.e., comparisons) of Model 5.

Model 4 strengths When there is more than one factor (as for our case), then this model (unlike the "treatments" of Model 3) separates out the various factors and therefore can test for an interaction of independent variables effects on the dependent variable. It also allows for the contrast calculations such as linear and quadratic trends (or the interaction contrasts of Model 5).

Model 4 weakness A weakness is that the analysis can be non-informative unless the contrast components (see Model 5) are calculated. For example, for our data, the analysis suggests there is no significant interaction (between X1 and X2) whereas in fact there are significant interactions (Fig. 2).

Model 5 strengths 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. Billard *et al.* (2014) gives some guidance on these calculations.

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 1). Testing for the quadratic effect of X_1 (adding

Fig. 7: Illustration of partial SAS 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	898.657216	81.696111	0.75	0.6848
Error	36	3921.674167	108.935394		
Corrected Total	47	4820.331383			

R-Square	Coeff. Var.	Root MSE	Y Mean
0.186431	89.64864	10.43721	11.64235

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X2	1	2.7231977	2.7231977	0.02	0.8753
X1	5	177.1879551	35.4375910	0.33	0.8944
X2*X1	5	718.7460634	143.7492127	1.32	0.2778

Contrast	DF	Contrast SS	Mean Square	F Value	Pr>F
Linear X1	1	91.6119696	91.6119696	0.84	0.3652
Lin X1@X2_1	1	116.1003214	116.1003214	1.07	0.3088
Lin X1@X2_2	1	591.0251914	591.0251914	5.43	0.0256

Contrast	DF	Contrast SS	Mean Square	F Value	Pr>F
Quad X1	1	72.9875478	72.9875478	0.67	0.4184
Quad X1@X2_1	1	0.0344048	0.0344048	0.00	0.9859
Quad X1@X2_2	1	150.4915705	150.4915705	1.38	0.2476

X2*LinearX1 Contrast

(X2*LinearX1)SS = 615.51354322
(X2*LinearX1)MS = 615.51354322
(X2*LinearX1)F = 5.6502622639
(X2*LinearX1)P-value = 0.0228891549

X2*QuadX1 Contrast

(X2*QuadX1)SS = 77.538427513
(X2*QuadX1)MS = 77.538427513
(X2*QuadX1)F = 0.7117836087
(X2*QuadX1)P-value = 0.4044236385

X1 × X2 to the GLM model in Model 2) obscures the significance of 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 when data has unequal replications. To illustrate, observation 44, 46, 47 and 48 were deleted in this case to give unequal replications (Table 1). 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 (a). In contrast, 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 (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 (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 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 a different number of 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

Appendix A: Five different models for GLM procedure using SAS

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):

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

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 and Model 5):

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

where μ = overall mean

$A_i = i^{th}$ A factor effect

$B_j = j^{th}$ B factor effect

Ab_j = 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) is the same as Model 4, but the analysis is extended to include linear and quadratic contrasts. The extended linear and quadratic contrasts statements are presented in Billard *et al.* (2014).

In each of these models, e refers to the observational error, Y is the response variable (tibial dyschondroplasia percent incidence, in our case), X_1 (vitamin D) and X_2 (phytase) are the regression variables for regression models (as in Models 1-2), with A and B being the factors (corresponding to our vitamin D and phytase) of an analysis of variance model and X_1X_2 (or, AB) are interaction terms

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 Fig. 6), 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 is PROC 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 was used 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 (SAS Institute, 2006) 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 does not stop at just comparing treatment means, but we advocate more detailed analyses by testing for responses starting with linear trends and interaction response components.

The Poultry Science Instructions for Authors state: "Statistical methods commonly used in the animal sciences need not be described in detail, but adequate references should be provided." However, it is necessary to know the details of how an experiment was conducted and hence how statistical analyses are performed to come to the same conclusions from even the same data (as detailed above). We therefore believe that statistical methods must be described in detail, including programming statements to avoid any possible ambiguity. A statement of how statistical package results were interpreted should also be included. The statement should be found in Materials and Methods. Also, it is particularly important to know whether probabilities that are presented are based on Type I or Type III sums of squares.

REFERENCES

Billard, L.M.Y. Shim and G.M. Pesti, 2014. Experimental design and analysis with emphasis on communicating what has been done: II) Calculating interaction contrasts with SAS. Int. J. Poult. Sci. 13: 88-96.

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

Chew, V., 1976. Comparison treatment means: A compendium. Hort Sci., 11: 348-357.

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

Little, T.M., 1978. If Galileo published in HortScience. Hort Sci., 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. Stat. and Computing, 4: 221-234.

Nelson, L.A. and J.O. Rawlings, 1983. Ten common misuses of statistics in agronomic research and reporting. J. Agron. Ed., 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.

Steel, R.D.G. and J.H. Torrie, 1960. Principles and Procedures of Statistics. McGraw-Hill, Toronto.

Swallow, W.H., 1984. Those overworked and oft-misused mean separation procedures Duncan's, LSD, etc. The Am. Phytopath. Soc., 40: 919-921.

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

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