OVERCOMING RESISTANCE TO MULTIVARIATE ANALYSIS OVER TIME

April J. Milliken

Anne M. Parkhurst

Recommended Citation

OVERCOMING RESISTANCE TO MULTIVARIATE ANALYSIS OVER TIME

April J. Milliken, M.S., and Dr. Anne M. Parkhurst
Department of Statistics, Kansas State University, and
Department of Biometry, University of Nebraska-Lincoln

One aspect of statistical consulting is assessing a clients needs. Sometimes the need for simplicity beclouds the information contained in the experiment. As an example, an experiment was performed as a multivariate study with repeated measures, yet the client preferred numerous univariate analyses that ignored time. The challenge was to show how a more sophisticated analysis provided additional insight into the biological process. Various covariance structures were employed to illustrate the usefulness of progressively more complex analyses. Multivariate methods were performed to utilize the correlation among variables to illuminate biological concepts. To complicate the whole process, an additional problem occurred where extreme variability among experimental units within treatment groups led to the identification and clumping of homogeneous units to increase precision. Power studies were performed to determine required sample sizes to minimize potential error reoccurrences in future trials.

1. <u>INTRODUCTION</u>

George Box (JASA, 1976) once said there is a "...tendency to force all problems into the molds of one or two routine techniques, insufficient thought being given to the real objectives of the investigation or to the relevance of the assumptions implied by the imposed methods." In consulting, this quote applies to those who get so used to routine statistical techniques they fail to consider the objectives of the study, and the assumptions that might be imposed by a selected analysis. One aspect of statistical consulting is assessing a client's needs, i.e., conclusions the client wishes to reach and the statistical level at which these results can be achieved. Sometimes the client's need for simplicity overshadows the information contained in the experiment and obscures evidence for the research hypothesis. Plainly, the simplest procedure is not always the best way to get to the answer. For example, an animal science physiology experiment was performed as a multivariate study with repeated measures, i.e., several response variables were measured at several time points. The client, however, requested numerous univariate analyses that ignored time as a consequential factor, ignored the multivariate structure of the data, and resisted using methods that were consistent with the original experiment design. The resulting efficiency and precision of the univariate analyses were minimal at best. In addition, the study was tainted with heterogeneous within treatment experimental units which inflated the variability. Several distinct power analyses indicated that increased sample sizes were needed to detect significant differences between treatments. Thus, the consultant's challenge was to show how more sophisticated analyses could potentially provide more insight into the biological process under study.

2.   RESEARCH BACKGROUND

This challenge arose from a "simple" consulting session with a student. The student was investigating the concept that an animal's diet influences the composition of the intestinal contents, i.e., characteristics of the stool. Many variables provide the make up of the stool, all of which need to be measured in order to understand how the colon is affected by a dietary treatment. Fifteen animals were fed one of three diets which varied in type of dietary protein: MEAT, SOY (vegetable), or CASEIN (a milk and cheese protein). Fecal material was collected every ten minutes for two hours resulting in 12 samples plus one for a saline solution control to provide a baseline. Eight biochemical and physiological characteristics were measured every collection period: base excess, bicarbonate, carbon dioxide, chloride, effluent volume, osmometer, pH, and sodium. Figure 1 shows the variability between treatments for mean bicarbonate adjusted for control over time by subtracting the control value per time from each treated value per time. The original design was considered to be multivariate in nature, with the experimental unit being the animals, and repeated measures taken over time for each biochemical variable under study. However, the analysis requested by the client did not coincide with the data collection process and experimental design. This challenge was to overcome the resistance to using multivariate analysis over time.

3.   CONCERNS REGARDING THE REQUESTED UNIVARIATE ANALYSIS

The requested analysis per biochemical variable consisted of taking an average of the 12 time measurements and looking at mean comparisons. This analysis ignores the design of the experiment and the original underlying interest in the study, that is, what happens to the biochemical composition over the 12 time periods. The client's prerogative to choose an elementary analysis led to a variety of concerns: 1) ignoring how the experiment was conducted, 2) overlooking important factors and variable relationships, 3) failure to meet assumptions required for the chosen analysis, and 4) inappropriate or non-significant results (need for power analysis). The concerns are discussed individually and jointly, as they are related to overcoming the resistance to multivariate analysis.

3a.   Ignoring How the Experiment was Conducted

Ignoring how the experiment was conducted is the first barrier to overcome. Since the client spent time and energy in conducting the experiment, it should be easy to dissipate this resistance. You might start by asking the client why they collected such an amount of data and why they want to lose so much information. The major challenge is to get the client to see exactly how much information they are losing so that they understand what their requested analysis is lacking.

> **"He uses statistics as a drunk uses lamp-posts,
> for support rather than illumination."**
>
> **--- Andrew Lang**

3b.   Overlooking Important Factors and Relationships

The design required collecting data over time in the hopes of modelling the effect of treatment over time on each of the variables. However, if time was an initial factor in the design, should it not be found in the model? When is it necessary to keep time as a factor? When is the client 'correct' in ignoring it's usage? These questions can easily be answered with a simple analysis of variance that involves time and the treatment by time interaction. If the interaction is zero, then averaging over time is reasonable if time is not important to the researcher. For instance, the researcher may assume a peak at Time 4 and make all treatment comparisons at that peak time; yet, it is still important to point out how information may be lost. Referencing energy spent in data collection may help in overcoming the resistance to a different form of analysis. Alternatively, if the interaction is significant, leaving interaction out of the model is inappropriate and the risk of false conclusion increases.

The design used in this example does involve measurement of multiple variables. To illustrate the information contained in such data, a principle component analysis was performed at each time period to utilize the correlation among variables to illuminate biological concepts. For each time period, the first two principle components accounted for approximately 88% of the total variation (see Figure 2a). The first component described the moisture content of the stool and the second component described stool texture, both of which were biologically significant. Since the results of the analysis at each time period indicated that the same two principle components could be used over all time periods, the two components were modelled over time by performing a repeated measures analysis for the two principal components (Figure 2b). The results of this type of analysis are often biologically insightful, as they were in this case, and the technique allows for a different view of the data.

3c.   Dealing with Failure of Assumptions

Another problem occurs when the structure of the data (for a given univariate response) is such that the usual analysis is not appropriate. One major complication is heterogeneity of variance of experimental units within a treatment as well as between treatments. A second concern is resistance of the client to using appropriate complex models that model the heterogeneity of variance.

In the animal example, the univariate analyses as requested by the client consisted of comparing treatments for several functions of time measurements for each measured biochemical variable to form responses:

1.   mean of the 12 time measures
2.   mean of the 12 time measures subtracted from control
3.   mean of the middle 6 time measures (to allow for both acclamation and overload of the system),
4.   average of the middle 6 time measures subtracted from control

However, when expected treatment differences were not found for any of the responses of

any of the variables, possible violations of assumptions were questioned. After plotting the data for each dog within each treatment for each response (see Figure 3 for an example), it was revealed that heterogeneity of animals within treatments was present. The process of finding undue variability led to identifying and clumping the homogeneous animals, i.e., grouping the non-extreme responding animals and grouping those animals with excessive variability. After subsetting, there were only eight animals remaining, two in one treatment and three in each of the others. Any analyses performed on the original data was also performed on the subset. The focus of attention needed to be directed toward the homogeneity of the covariance matrices over time. Since the covariance structures were heterogeneous, more advanced techniques that take such structure into consideration are needed. Utilizing plots to view the evidence of heterogeneity and a thorough explanation of the how the analysis gives difference results depending on the assumptions made is a way to overcome resistance to using advanced techniques.

3d. Mixed Model Approach

The appropriate univariate repeated measures over time model is the mixed model

$$Y_{ijk} = \mu + \alpha_j + \rho_{ij} + \tau_k + (\alpha\tau)_{ik} + \epsilon_{ijk}$$

where ρ_{ij} denotes the subject within treatment error and ϵ_{ijk} denotes the time interval within subject error. It is reasonable to assume the time interval errors with a subject are correlated and the appropriate analysis would allow for the estimate of the correlation structure.

In SAS-MIXED® (SAS Institute, 1992), the mixed model analysis is performed with the following syntax:

```
PROC MIXED;
  CLASSES TRT SUBJECT TIME;
  MODEL RESPONSE = TRT TIME TRT*TIME;
  RANDOM SUBJECT(TRT);
  REPEATED / TYPE = type  SUBJECT = SUBJECT(TRT) R;
```

where "type" indicates the form of the covariance matrix of the time interval errors. The mixed model analysis provides restricted maximum likelihood (REML) estimates of the covariance matrix parameters. The number of parameter estimates correspond to the selected covariance matrix specified by the "type" choice, i.e., the expected covariance matrix structure denoted by R. For instance, the default follows the usual assumption of constant variance and zero covariance. Below, a few structure types are displayed where only three time measurements are used for simplification.

Compound symmetry (1) is a plausible covariance matrix structure. There is an assumed common variance and common covariance among times within subjects.

$$(1) \quad CompoundSymmetry = \begin{bmatrix} \sigma^2 + \sigma^2_1 & \sigma^2_1 & \sigma^2_1 \\ \sigma^2_1 & \sigma^2 + \sigma^2_1 & \sigma^2_1 \\ \sigma^2_1 & \sigma^2_1 & \sigma^2 + \sigma^2_1 \end{bmatrix}$$

The first-order autoregressive covariance structure (2) is another two parameter model. It assumes a lag relationship across time among the subjects within a treatment. As the distance between time measurements increases, the correlation between the values decreases.

$$(2) \quad 1^{st}\text{-}order\ AR = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

A progressively more advanced covariance structure involves many more parameters. The Toeplitz (3), assumes a common variance, but the covariance depends on the distance the measurements are apart. There is not necessarily any relationship between σ_1 and σ_2, making this form different from the first-order autoregressive structure.

$$(3) \quad Toeplitz = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

The most general covariance matrix form is the unstructured (4), as there are no commonalities, the only requirement is that the resulting estimates provide a positive definite matrix. Consequently, this structure involves the most parameter estimates.

$$(4) \quad Unstructured = \begin{bmatrix} \sigma^2_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma^2_{22} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma^2_{33} \end{bmatrix}$$

Other calculations obtained from PROC MIXED consist of tests of the fixed effects and model fitting information. Likelihood ratio tests are given to test the fit of the given model and it's covariance structure. Model comparisons can be made by utilizing the -2 log REML function. By calculating the deviance likelihoods of two models, the best fitting covariance

structure can be established. Several structures should be examined to determine the most appropriate structure with the minimal number of parameters to estimate.

Various covariance structures were employed to illustrate the usefulness of progressively less complex analyses. Modelling started with the unstructured covariance, then the Toeplitz (reducing the number of parameters to estimate), first-order autoregressive, compound symmetry, and finally the default (independent equal variance errors). Table 1 lists likelihood ratio information regarding one of the variables of interest, Bicarbonate concentration in the colon.

The results in Table 1 indicate one commonality for the data sets: TYPE=UN is omitted because it failed to work in either case due to the number of samples only slightly outweighing the number of time measurements. A larger sample size is required in order for the unstructured parameter estimates to be calculated. Model fitting results were found by calculating the deviance between two successive models, using the difference in degrees of freedom, and using a Chi-Squared table to determine the observed significance level.

For the data set with all 15 animals, the first-order autoregressive model fits. From Table 2, it is obvious that the time effect is significant, but treatment is not. However, no matter the model chosen, the expected result of treatment differences does not occur. For the subset of data, the Toeplitz fits. Based on this model, the time by treatment interaction is significant. No model with fewer parameters was able to find the interaction significant. The concept that different covariance matrix structures can lead to different conclusions about treatment differences is important. The varying results regarding significance of treatment differences may sway the researcher and help in overcoming the resistance to this analysis.

4. NONSIGNIFICANT RESULTS (need for Power Analysis)

The reason for the statistical analysis of data from many experiments is to try to obtain evidence to establish that treatments are significantly different. The concerned researcher will go to any measure to make sure that the "best" atmosphere for finding significant results is available. However, not all experiments are well planned and often times expected results are not found. An appropriate design is the first step. A pre-experiment power analysis is an important aid in determining this design. If the proposed analysis of the experiment includes hypothesis testing, then a power analysis may be used to help choose the appropriate design.

The statistician's preference is a pre-experiment power analysis. However, this is often not the case as the data is usually presented to the statistician after the study is completed and the researcher is ready to start the analysis. If the analysis does not detect significant differences, power studies can be utilized to determine if the number of repetitions was adequate. Improving the precision and power of a given design will lead to a more accurate look at the treatments under study. Failure to do so may result in unsatisfactory findings, the inability to reject hypotheses about treatment differences that, in theory, should be prevalent.

Two possible ways of computing the power will be examined:

1) power calculations for the analysis of variance F-test based on sums of squares of the hypothesis

2) power calculations for general linear multivariate models, including repeated measures applications

Each method is used for different design aspects, depending on what is involved in the analysis. The usage for each method will be discussed in general and it's use illustrated for this experiment.

4a. Power Analysis based on the Sums of Squares for Hypothesis

The power analysis based on the sums of squares for hypothesis can be used to answer several questions in reference to the planning of experiments. The power calculation is based on the testing the equality of treatment means. A computer code to carry out the power analysis is described in Lohr and O'Brien. This program not only calculates the power, it also allows the researcher to find out if a specific power, using different n values, is possible and computes the sample size per treatment based on that power requirement.

The SAS program needs only the number of treatment levels for the effect being tested, the sums of squares and degrees of freedom for the hypothesis, the population variance, a desired significance level, and desired power level.

Tables 3a and 3b give results of power calculations and sample size determinations for the client's requested analyses. All sums of squares here are for the hypothesis that all treatment effects are equal. Thus, the table consists of post-experiment power calculations with sample size recommendations for future experiments.

4b. Power Analysis for General Linear Multivariate Models

Using the general linear multivariate model to calculate the power comes closest to the design of the study being illustrated, and the computer code can be referenced in the Muller article. The technique is multivariate in nature allowing for the use of a repeated measures design. Thus, the variance-covariance matrix is used instead of the variance. The time interval covariance matrix is used instead of the sample variance which is the R matrix as produced by PROC MIXED. For N sampling units, p response variables (repeated), and q parameters (treatments), the model is $Y_{N*p} = X_{N*q} B_{q*p} + E_{N*p}$ with each $row_i(E) \sim N_p(0, \Sigma)$. The usual null hypothesis in the multivariate model involves the parameter $\Theta = CBU$ where $H_0: \Theta = \Theta_0$. Here each row of C defines a row of Θ and corresponds to a contrast among parameters, treatments in this case, referred to as a between-subject contrast. Each column of U delineates a column on Θ and corresponds to a transformation of the time responses, referred to as a within-subject contrast. In order to compute the approximate power the following need to be specified:

Applied Statistics in Agriculture

1. alpha level (0.05 is the default)
2. Σ, p by p covariance matrix (the covariance matrix among response contrasts which are within-subject)
3. design matrix X
4. B matrix of parameter estimates
5. C vector or matrix (between-subject)
6. U vector of matrix (within-subject), identity matrix is the default.

For example, an experiment with three treatments and four time periods have matrices

$$B = \begin{bmatrix} \tau_{11} & \tau_{21} & \tau_{31} & \tau_{41} \\ \tau_{12} & \tau_{22} & \tau_{32} & \tau_{42} \\ \tau_{13} & \tau_{23} & \tau_{33} & \tau_{43} \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

$$U = \begin{bmatrix} -3 & -1 & 3 & 1 \end{bmatrix}_{linear}$$

or

$$U = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}_{quadratic}$$

For this illustration, the estimate of the covariance matrix can be acquired through the use of PROC DISCRIM, where it was verified that the pooled covariance matrix was acceptable as an estimate of Σ. Through individual univariate analyses, the estimate of the B matrix was created. The C matrix tests the hypothesis that all treatment effects are equal. The U vector used for the results in Table 4 tests the adequacy of the simple linear regression model by using the linear orthogonal polynomial coefficients.

For each of the data sets and alpha values chosen, three total sample sizes and their corresponding power values are listed. The actual experiment consisted of only 15 animals, and other sample sizes are included for comparison. The calculations in Table 4 show the experiment would need to be ten times larger in order to detect a significant difference between treatments. However, if the animals are more homogeneous, a much smaller sample. approximately 16 animals, is required. These calculations can be used to convinces the researcher that repeated analysis over time can be used to improve the design of the experiment in future trials.

5. **CONCLUSION**

Every consulting project should start with an explanation of the assumptions being made. A realistic assessment may lead to the ability to find differences that could not be

found by glossing over the assumptions. The best way to get the client to overcome resistance to advanced techniques is to contrast the results of requested and suggested analyses. In this case, showing the client several possibilities, per response (using a mixed model) or per time (principle component analysis), enabled him to see how complex methods could provide him with more insight into the data. Nevertheless, clients will not always follow your advice or seek to understand the advantages of your suggestions, but the point is well made.

> **"They that will not be counselled, cannot be helped.**
> **If you do not hear reason, she will rap you on the knuckles."**
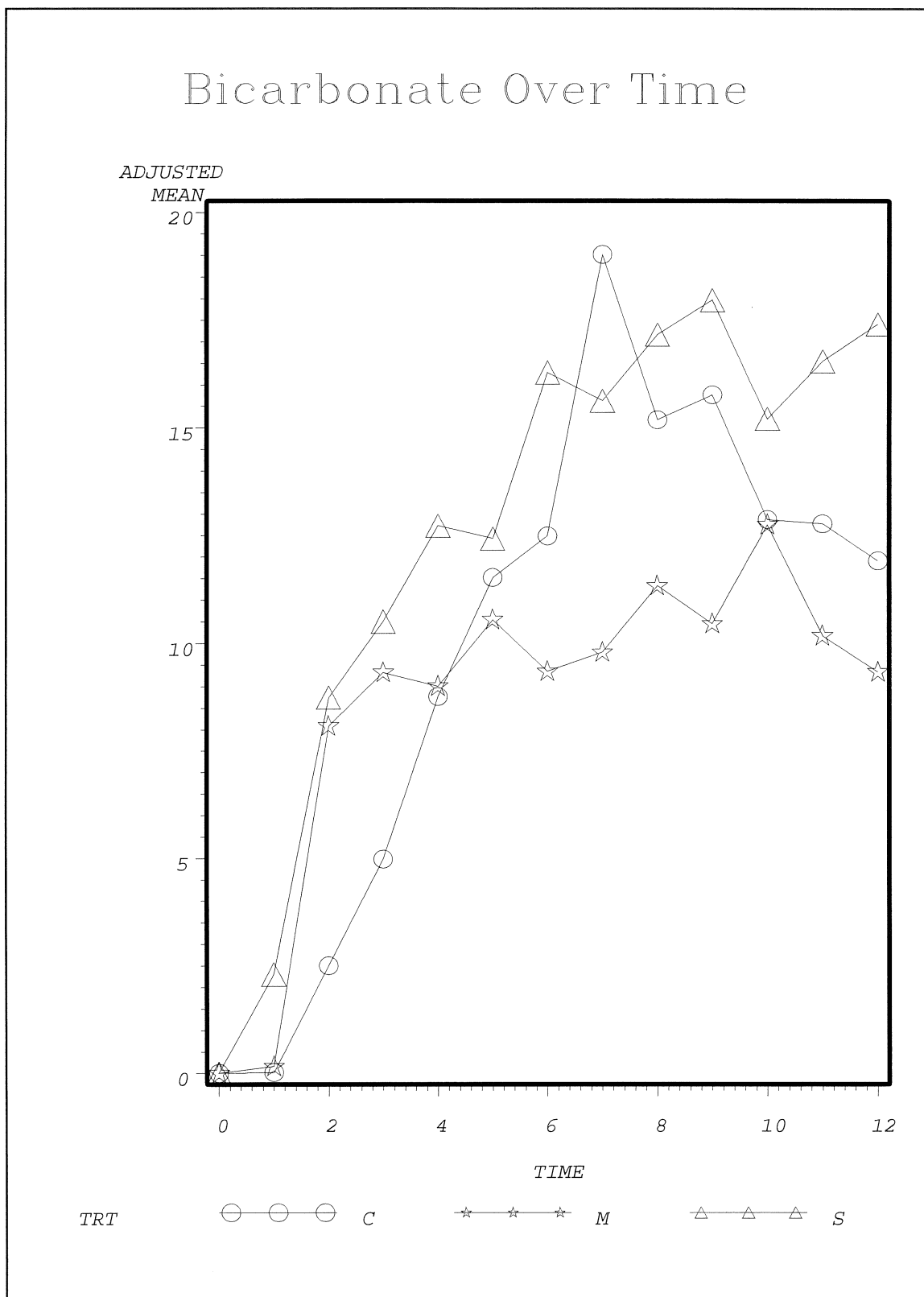>
> **--- Ben Franklin**

Figure 1.   This plot shows the variability between treatments Casein, Meat, and Soy for bicarbonate adjusted for control by subtracting the mean for control from each treatment mean per time period.
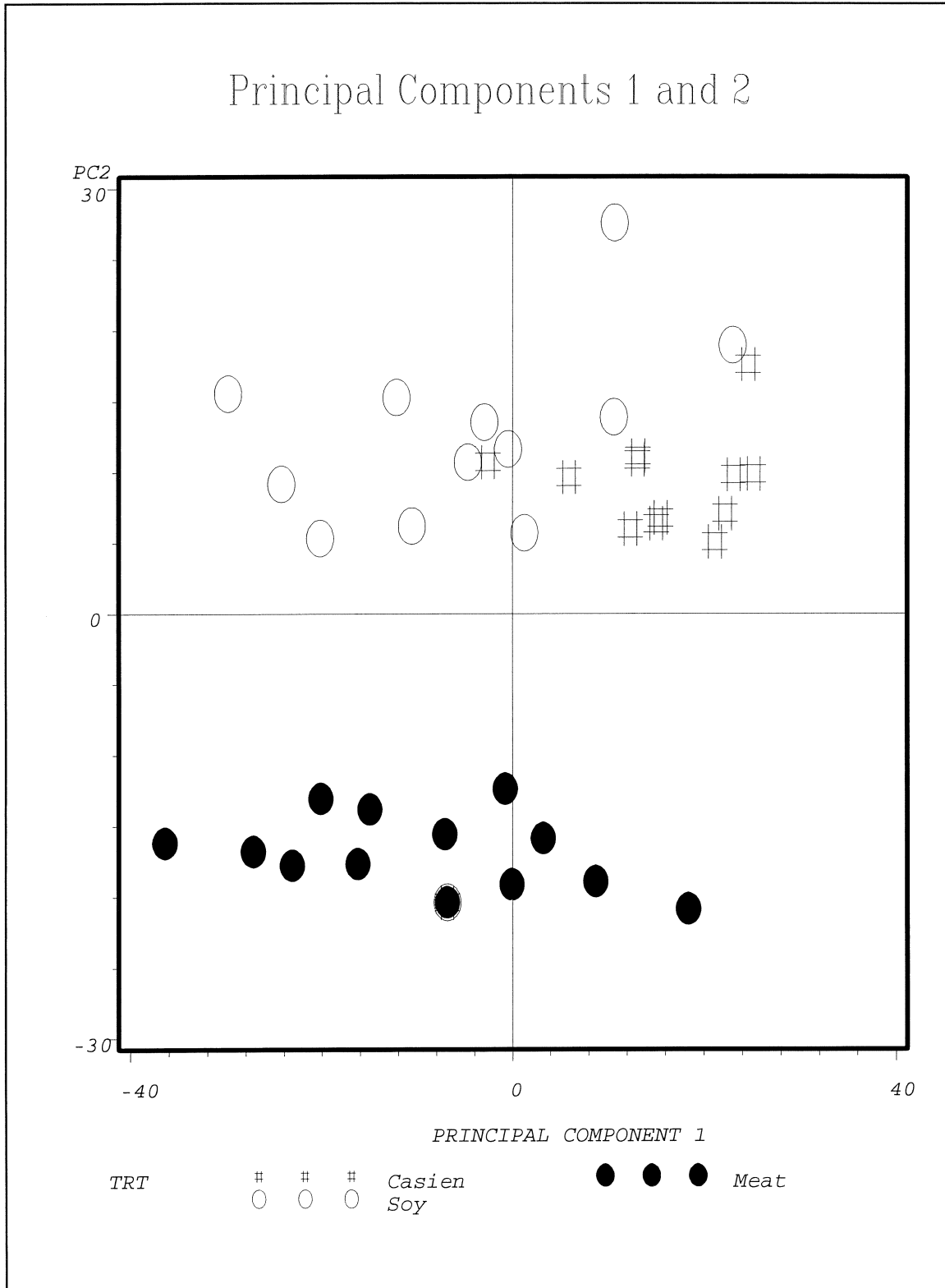
Principal Components 1 and 2

Figure 2a.   Principle component 1 refers to Moisture Content and principle component 2 refers to Stool Texture.   Principle component analysis indicates distinct differences in treatments, with Meat showing a different texture than the other treatments.   Also, moisture content for Casein and Soy differs slightly.
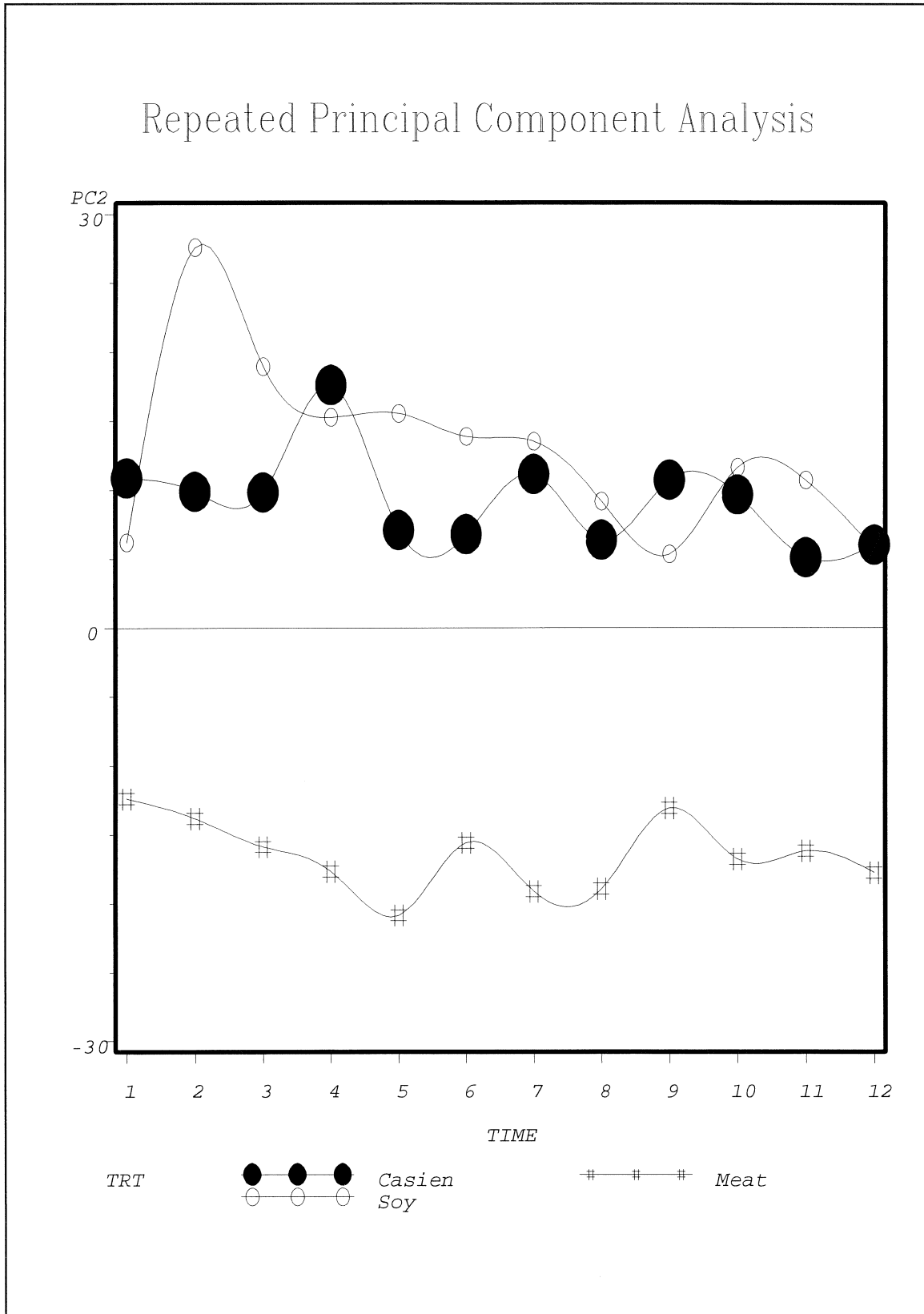
Figure 2b.   The results for a principle components analysis over time for Moisture Content indicate that the composition of Meat is consistently different from the other treatments.
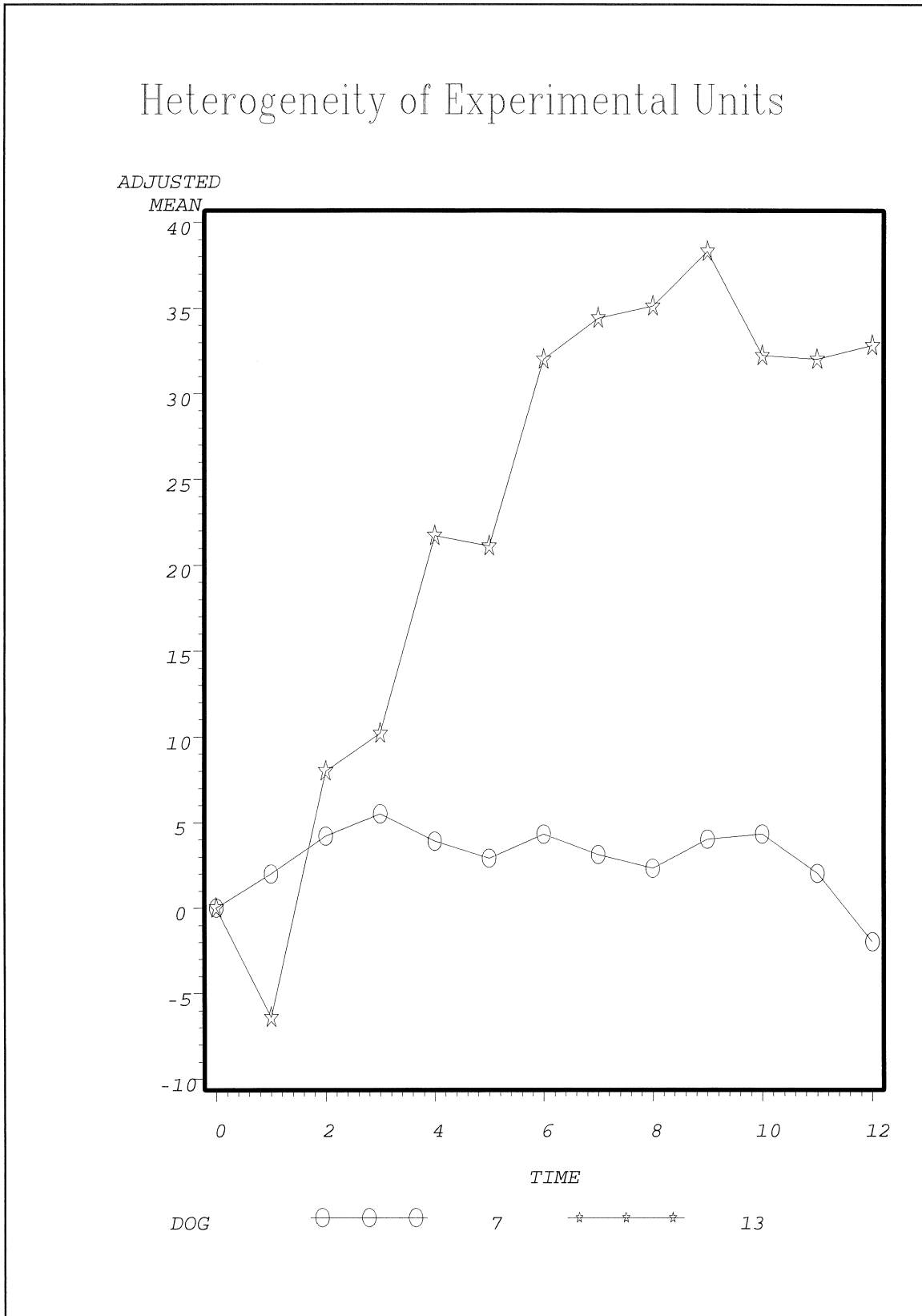
Figure 3.   This plot exemplifies variability of response for bicarbonate in this data set, where Dog 7 and Dog 13 received the same dietary treatment.

Table 1. Model Fitting information for Bicarbonate								
Set	All 15 Animals				Subset of 8 Animals			
	-2 REML Log L.	Ho: LRT X²	Ho: LRT DF	Ho: LRT p-value	-2 REML Log L.	Ho: LRT X²	Ho: LRT DF	Ho: LRT p-value
TOEP	1040.7	159.77	13	0.0000	426.36	35.876	13	0.0006
AR(1)	1045.9	154.55	2	0.0000	449.04	35.876	2	0.0014
CS	1108.7	91.74	2	0.0000	453.30	8.933	2	0.0115
Simple	1108.7	91.74	1	0.0000	453.30	8.933	1	0.0028

Table 2. P-values for Tests of Fixed Effects for Various Covariance Structures						
Set	All 15 Animals			Subset of 8 Animals		
	TRT	TIME	TRT*TIME	TRT	TIME	TRT*TIME
TOEP	0.8685	0.0000	0.7423	0.1311	0.0000	0.0000
AR(1)	0.6257	0.0001	0.7996	0.0547	0.0001	0.2436
CS	0.6663	0.0000	0.8964	0.0660	0.0000	0.1447
Simple	0.6663	0.0000	0.8964	0.0660	0.0000	0.1447

Table 3a. F-test based Power Calculations and Sample Sizes for Requested Analyses, given SSH and $\hat{\sigma}^2$, $\alpha=0.05$					
Data Set	Response Variable	SSH	$\hat{\sigma}^2$	n per Treatment	Power
All	DIFFER 6	73.788	88.5874	20	0.9573
	DIFFER 12	49.891	59.3750	20	0.9559
	MID 6	67.145	92.8610	23	0.9554
	ALL 12	71.203	62.7657	15	0.9546

Table 3b. F-test based Power Calculations and Sample Sizes for Requested Analyses, given SSH and $\hat{\sigma}^2$, α=0.05

Data Set	Response Variable	SSH	$\hat{\sigma}^2$	n per Treatment	Power
Subset	DIFFER 6	116.873	26.2768	5	0.9666
	DIFFER 12	133.424	13.5758	3	0.9631
	MID 6	176.685	33.0920	5	0.9865
	ALL 12	195.704	19.2645	3	0.9657

Table 4. Power Calculations for Repeated Measures Analysis of Bicarbonate, α=0.05

Data Set	Total Sample Size	Power
All Dogs	15	0.1182
	75	0.5127
	150	0.8324
Subset of Dogs	8	0.5658
	16	0.9657
	24	0.9985

References and Acknowledgements

Box, George E. P. (1976), "Science and Statistics," Journal of the American Statistical Association, Vol. 71, No. 356, pp. 791-799.

Lohr, Virginia I., and Ralph G. O'Brien (1984), "Power Analysis for Univariate Linear Models: The SAS System Makes it Easy," SAS Users Group International Conference Proceedings, SAS Institute Inc., pp. 847-852.

Milliken, George A., and Dallas E. Johnson (1984), Analysis of Messy Data Vol. I: Designed Experiments, Von Nostrand Reinhold, New York.

Muller, Keith E., Lisa M. LaVange, Sharon Landesman Ramey, and Craig T. Ramey (1992), "Power Calculation for General Linear Multivariate Model Including Repeated Measures Applications," Journal of the American Statistical Association, Vol. 87, No. 420, pp. 1209-1226.

SAS Institute Inc., SAS® Technical Report P-229, SAS/STAT® Software: Changes and Enhancements, Release 6.07, Cary, NC: SAS Institute Inc., 192, 620 pp.

SAS Institute Inc., SAS Users Guide, Version 6.07, Cary NC: SAS Institute Inc., 1992.