

A CASE STUDY IN APPLYING GENERALIZED LINEAR MIXED MODELS TO
PROPORTION DATA FROM POULTRY FEEDING EXPERIMENTS

by

CARLIE SHANNON

B.S., University of Kentucky, 2004

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Leigh Murray

Copyright

CARLIE SHANNON

2013

Abstract

This case study was motivated by the need for effective statistical analysis for a series of poultry feeding experiments conducted in 2006 by Kansas State University researchers in the department of Animal Science. Some of these experiments involved an automated auger feed line system commonly used in commercial broiler houses and continuous, proportion response data. Two of the feed line experiments are considered in this case study to determine if a statistical model using a non-normal response offers a better fit for this data than a model utilizing a normal approximation. The two experiments involve fixed as well as multiple random effects. In this case study, the data from these experiments is analyzed using a linear mixed model and Generalized Linear Mixed Models (GLMM's) with the SAS Glimmix procedure. Comparisons are made between a linear mixed model and GLMM's using the beta and binomial responses. Since the response data is not count data a quasi-binomial approximation to the binomial is used to convert continuous proportions to the ratio of successes over total number of trials, N , for a variety of possible N values. Results from these analyses are compared on the basis of point estimates, confidence intervals and confidence interval widths, as well as p -values for tests of fixed effects. The investigation concludes that a GLMM may offer a better fit than models using a normal approximation for this data when sample sizes are small or response values are close to zero. This investigation discovers that these same instances can cause GLMM's utilizing the beta response to behave poorly in the Glimmix procedure because lack of convergence issues prevent the obtainment of valid results. In such a case, a GLMM using a quasi-binomial response distribution with a high value of N can offer a reasonable and well behaved alternative to the beta distribution.

Table of Contents

List of Figures	v
List of Tables	vi
Acknowledgements.....	vii
Chapter 1 - Introduction.....	1
Chapter 2 - Brief Review of the Motivating Example.....	3
Description of Experiment 1.....	4
Chapter 3 - Overview of Linear Mixed Models and Generalized Linear Mixed Models.....	7
Correlation Structure.....	7
Linear Mixed Models.....	10
Generalized Linear Mixed Models	11
LMM and GLMM with Experiment 1	14
LMM and GLMM with Experiment 2.....	17
Chapter 4 - Results.....	22
Correlation Structure.....	22
Experiment 1 Correlation Structure	22
Experiment 2 Correlation Structure	25
Results for Experiment 1	27
Results for Experiment 2	30
Experiment 2 Objective 1	30
Experiment 2 Objective 2	34
Chapter 5 - Conclusions.....	38
References.....	41
Appendix A - SAS Programs.....	42

List of Figures

Figure 2.1 General diagram of commercial auger system feed line	4
Figure 3.1 Q-Q plot for residuals for Experiment 1 using a LMM.....	16
Figure 3.2 Histogram of distribution of residuals for Experiment 1 using a LMM.....	16
Figure 3.3 Q-Q plot for residuals for Experiment 2 Objective 1 using a LMM.....	19
Figure 3.4 Histogram of distribution of residuals for Experiment 2 Objective 1 using a LMM ..	19
Figure 3.5 Q-Q plot for residuals for Experiment 2 Objective 2 using a LMM.....	20
Figure 3.6 Histogram of distribution of residuals for Experiment 2 Objective 2 using a LMM..	21
Figure 4.1 LS-Means plot of point estimates for each Feed by Pan for Experiment 1.....	23
Figure 4.2 Plot of Covariance by Lag for Experiment 1.....	24
Figure 4.3 LS-Means plot of point estimates by Pan for Experiment 2 Objective 2.....	26
Figure 4.4 Plot of Covariance by Lag for Experiment 2 Objective 2.....	26
Figure 4.5 High PDI mean estimates with 95% confidence intervals for Experiment 1	28
Figure 4.6 Low PDI mean estimates with 95% confidence intervals for Experiment 1.....	29
Figure 4.7 Pan 1 & Pan 12 mean estimates with 95% confidence intervals for Experiment 1	29
Figure 4.8 95% confidence interval widths for mean estimates for Experiment 1.....	30
Figure 4.9 Pan 1 mean estimates with 95% confidence intervals for Experiment 2 Objective1 ..	31
Figure 4.10 Pan 6 mean estimates with 95% confidence intervals for Experiment 2 Objective 132	32
Figure 4.11 Pan 12 mean estimates with 95% confidence intervals for Experiment 2 Objective 1	32
Figure 4.12 95% confidence interval widths for mean estimates for Experiment 2 Objective1 ..	34
Figure 4.13 Pan 0 mean estimates with 95% confidence intervals for Experiment 2 Objective 235	35
Figure 4.14 Pan 0 & Pan 1 mean estimates with 95% confidence intervals for Experiment 2 Objective 2	37
Figure 4.15 Pan 0 & Pan 12 mean estimates with 95% confidence intervals for Experiment2 Objective 2	37
Figure 4.16 95% confidence interval widths for mean estimates for Experiment 2 Objective2 ..	38

List of Tables

Table 2.1 ANOVA table for Experiment 1	5
Table 2.2 ANOVA table for Experiment 2 Objective 1 by Pan	6
Table 2.3 ANOVA table for Experiment 2 Objective 2 for Pellets only	6
Table 3.1 Tests for Normality of Residuals for Experiment 1 using a LMM.....	17
Table 3.2 Tests for Normality of Residuals for Experiment 2 Objective 1 using a LMM	20
Table 3.3 Tests for Normality of Residuals for Experiment 2 Objective 2 using a LMM	21
Table 4.1 AICC values for models with various correlation structures for Experiment 1	25
Table 4.2 AICC values for models with various correlation structures for Experiment 2 Objective 2.....	27
Table 4.3 P-values for test of difference in mean percent fat in Fines and Pellets by Pan	33
Table 4.4 P-values for test of difference in mean percent fat in Pan 0 and Pans 1-12	36

Acknowledgements

I would like to thank my wonderful advisor Dr. Leigh Murray. I would never have begun, continued, or completed this program of study without her encouragement, motivation, and insights.

Chapter 1 - Introduction

The motivating example for this master's report comes from the field of poultry science. Kansas State University researchers in the Animal Science department conducted a series of experiments in 2006 involving the effectiveness of poultry feeding practices in large, commercial chicken houses. Some experiments involved an automated auger feed line system commonly used in commercial broiler houses. Three of these are described in Hancock et al. (2011). The statistical analyses described in that paper were completed with consultation assistance from the department of Statistics at Kansas State University. These initial analyses utilized linear mixed models to describe the presence of fixed and random effects with normally distributed response data. All responses were assumed to follow the normal distribution despite two of the responses being continuous values bounded between zero and one.

Statistical analysis using linear mixed models is appropriate for many experiments containing more than one random effect. Assuming normality of parameter estimates, like means, are justified by Central Limit Theorems when the number of observations in "large." Describing the response as normally distributed is desirable for a number of reasons. The normal distribution is a symmetric, two-parameter distribution that is considered mathematically well-behaved. Its mathematical properties allow for simplicity of analysis because the error term has expected value of zero, the variance of the observations is equal to the variance of the errors, and confidence intervals for estimates are all symmetric. In maximum likelihood estimation with normal probability density functions, the canonical parameter, θ , is equal to the mean and the variance function, $V(\mu)$, is equal to 1 making estimation less computationally intensive than with other two parameter distributions in the exponential family.

The assumption of normally distributed observations can be a poor approximation for responses that are 0/1, counts, or are continuous proportions between 0 and 1. In these instances, descriptive statistics may indicate that the observations are not symmetric about a mean or there may clearly be no linear relationship between the response and its expected value. Each of the three experiments discussed in Hancock et al. (2011) used a different response. One experiment measured the weight of birds after a feeding program, another experiment measured the ratio of weight of fines (crushed feed) to the total weight of feed in each identified, sequential pan along

the auger feed line after the feed was run through, and the last measured percent fat content in pellets and fines in a control pan and in pans along the auger feed line. The assumption of normally distributed response data was used in all three analyses. Analysis of residuals showed no evidence of non-normality of residuals for the weight response, but the other two residual analyses indicated definite deviations of the residuals from normality.

Historically, a transformation on the response data has been used to stabilize variance with the resulting effect that a linear relationship is created between the response data and its expected value. The transformed response data are frequently approximately normally distributed and linear model methods can be used to conduct valid analysis on the transformed data scale. When a transformation is applied, the response data are actually changed according to a function such as log, square root arc sine, etc. These functions tend to diminish the dependence of the variance on the mean creating a data set with two independent parameters. This property is consistent with the normal distribution since non-normal data typically has a variance that is a function of the mean.

Transforming the response data allows one to maintain the symmetry and mathematical simplicity of linear modeling when the data is significantly non-normal, but transforming response data actually changes the model and several unfortunate consequences can result. Point estimates may be infeasible or unrealistic when back-transformed to the original scale. For example, a mean estimate can be negative when the original data has a lower bound at zero (Gbur et al. 2012). Confidence intervals are symmetric, but they may lack substantive meaning. Back-transforming the endpoints of a confidence interval may not necessarily maintain the meaning of the interval (Stroup 2012). This is because under normal assumptions the mean and median are the same so confidence intervals for mean estimates from a log-transformed response are the same as confidence intervals for medians. When these confidence intervals are back-transformed to a skewed distribution they remain the intervals surrounding the measure of central tendency, which is the median not the mean. Confidence intervals for the difference of means for log-transformed data also lose their meaning when back-transformed (Gbur et al. 2012). Confidence intervals for differences on the log-scale back-transform to confidence intervals of the ratio of medians and provide no information about the difference of means. These issues create a number of potential problems involving interpretation of results from analyses using response transformations.

A modern approach to handling non-normal response data is the use of generalized linear models (GLM's) to describe the response data with an exponential family distribution that closely fits the actual response. Extending this idea to models with random effects allows for the use of generalized linear mixed models (GLMM's). The use of these complex models was not computationally feasible until the recent past when software advances and improvements to statistical analysis programs enabled users to easily, quickly and accurately apply GLMM's to data sets (Stroup 2012). In a GLMM, the response data is not changed as it is with a transformation; instead, a function of the expected value of the response is modeled as having a linear relationship with the explanatory variables.

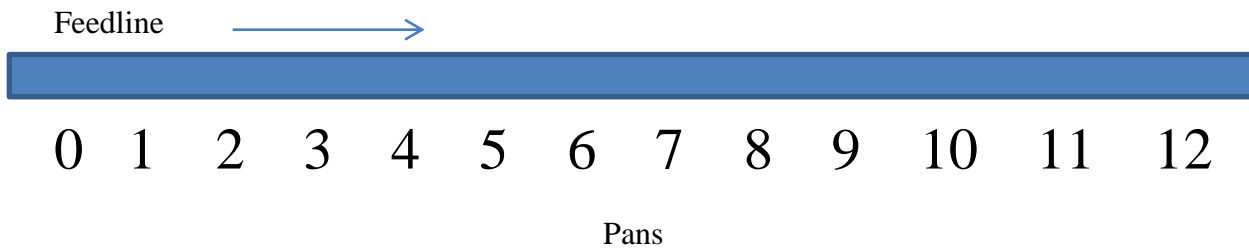
In this master's report, I will consider the two auger line poultry experiments with continuous, proportional responses described in Hancock et al. (2011). The primary objective of this paper is to determine if a generalized linear mixed model using the beta distribution or a binomial approximation to the beta offers a better fitting model than a linear mixed model. The secondary objective is to investigate the potential correlation structure between pans based on distance. These objectives will be addressed in the following way. Chapter 2 describes the design and methods used in the auger feed line experiments of Hancock et al. (2011). Chapter 3 offers an overview of the theory of correlation structures in mixed models, an overview of the theory of linear mixed models (LMM's) and GLMM's as well as describing how these models would be implemented with the auger feed line experiments. Chapter 4 first summarizes the results of correlation structure investigation and second summarizes the results of GLMM analysis compared to LMM analysis for each experiment. The final chapter offers some conclusions.

Chapter 2 - Brief Review of the Motivating Example

Complete descriptions of the two experiments under consideration here and results of the initial analysis can be found in Hancock et al. (2011). Descriptions of the two experimental designs and a brief overview of experimental methods follow. The data from the two experiments with continuous, proportional responses involve the distribution of poultry feed via an automated auger feed line used in commercial broiler houses. The objective of the experiments was to determine if the auger system was degrading the nutritional quality of feed due to the shearing and compression effects caused by standard line operation.

Both experiments utilized the same large, commercial auger system that distributes feed to 93 total feed pans. The researcher had narrowed consideration down to 12 representative pans evenly spaced along the line (Fig. 2.1). Pan 1 was the actual first pan on the line and pan 12 was the 93rd and last pan on the line. The experiments all used variations of a sorghum-soy based diet that was pelleted using standard procedures and met NRC guidelines (Hancock et al. 2011).

Figure 2.1 General diagram of commercial auger system feed line



Description of Experiment 1

In Experiment 1, two types of pelleted feed were manufactured using differing amounts of whole sorghum. This created one feed with high pellet durability index (PDI) and another with low PDI. The researcher was interested to know how great an impact this difference in PDI would have on the amount of intact, pelleted feed distributed to the different positions along the feed line. The line was run four times with the high PDI feed and four times with the low PDI feed. After each run, the total weight of feed in each of the 12 identified pans was measured. The feed in each pan was then sifted and the weight of the fines, crushed pellets that are destroyed by the feed line, was measured. The response of interest was the ratio of weight of fines to total weight of feed for each pan.

The experimental design used in Experiment 1 was a completely randomized design with a split plot. There were two fixed factors, feed with two levels (High PDI Feed, Low PDI Feed) and pan with twelve levels (1, 2, 3,... 12 locations along the feed line). There were two random effects, run with four levels (1, 2, 3, 4 runs of the feed line) and the residual term.

The ANOVA table with degrees of freedom for Experiment 1 is shown in Table 2.1. The whole plot main effect is Feed. The whole plot experimental unit and the whole plot error term is Run nested within Feed. The split plot fixed effects are Pan and the Pan*Feed interaction. The

split plot error term, Pan crossed by Run nested within Feed, is the residual term for the experiment.

Table 2.1 ANOVA table for Experiment 1

Source	Deg. Freedom
Overall Mean	1
Feed	1
Run(Feed)	6
Pan	11
Pan*Feed	11
Pan*Run(Feed)	66
Total	96

Description of Experiment 2

Experiment 2 used only one high PDI, pelleted, sorghum-soy based feed. The pellets were coated on the outside with soybean oil by means of mixing the pellets with the oil in an industrial mixer for one minute. From the mixer, a control pan (denoted Pan 0) was filled with coated pellets that would not be run through the line. The process of coating a batch of pellets with oil and removing a control pan was completed four times, once for each run of the feed line. Each batch was run through the automated auger line individually. After each run, the feed in each of the 12 identified pans was separated into pellets and fines and the fat content was measured. The response of interest was the percent fat in the pellets and the percent fat in the fines. The researcher had two objectives in Experiment 2. Objective 1 was to determine if there was a statistically significant difference between the percent fat in the pellets and in the fines for each of the 12 pans along the line. Objective 2 was to determine if there was a statistically significant difference in the percent fat of the pellets in the control pan and the pellets in the 12 pans from the run.

The experimental design for Experiment 2 is a randomized complete block design. The blocks are the four runs and are a random effect. The random runs are the replication in the

experiment. The control pan is Pan 0. Pan is the fixed treatment factor with 12 levels (1, 2, 3, ... 12 locations along the feed line) for Objective 1 and with 13 levels (the previous 12 plus the control pan) for Objective 2.

Objective 1 is to compare the percent fat for the two types of feed, pellets and fines, at each pan. Thus, the ANOVA table (Table 2.2) and associated model contain a fixed treatment effect of Type, a random block effect of Run, and a random interaction effect of Run* Type. The experimental unit is the Run*Type combination and the interaction of Run by Type serves as the residual term.

Table 2.2 ANOVA table for Experiment 2 Objective 1 by Pan

Source	Deg. Freedom
Overall Mean	1
Type	1
Run	3
Run*Type	3
Total	8

Objective 2 is a comparison of percent fat in pellets of the control pan (Pan 0) to each pan along the line (Pans 1-12). The ANOVA table for Objective 2 has the fixed effect of Pan rather than the fixed effect of Type (see Table 2.3).

Table 2.3 ANOVA table for Experiment 2 Objective 2 for Pellets only

Source	Deg. Freedom
Overall Mean	1
Pan	12
Run	3
Pan*Run	36
Total	52

Chapter 3 - Overview of Linear Mixed Models and Generalized Linear Mixed Models

There are numerous possible models for the two experiments described in Chapter 2 that would appropriately address the research objectives of the poultry scientists. The models considered here are ones that also satisfy the objectives of this report. Primary concern is on comparing LMM's to GLMM's, thus models utilizing a normal response distribution will be compared to similar models utilizing a beta response distribution and those utilizing a variation of the binomial distribution to approximate the beta distribution. Because spatial correlation between pans may exist, correlation structure will be investigated and the best fitting correlation structure will be chosen for each analysis. These results are summarized and compared in Chapter 4. The major considerations in model choice are classification of effects as fixed or random, variance-covariance structure, and response distribution.

Only mixed models will be considered because all of the experimental designs contain more than one random effect. Experiment 1 and both objectives for Experiment 2 include the random effect of Run. Run is considered random because the runs are not chosen for their individual properties, but are considered random representatives from the population of all possible runs. Since Run is a random effect, any interaction with Run is also random. Pans are a fixed effect that may be correlated as a function of distance. Mixed models accommodate complex variance-covariance structures including correlation structures commonly used with repeated measures designs as in Experiment 1.

Correlation Structure

Mixed models allow for specification of a complex variance-covariance structure. The variance-covariance structure of mixed models is necessarily more complex because it includes the contribution of random effects other than the "residual" or "error" term. Gbur et al. (2012, p24) describe the role of fixed and random effects in an analysis as follows, "Fixed effects determine a model for the mean of the response variable and random effects determine a model for the variance." Mixed models can include specific variance-covariance structures to model patterns of variability due to time or space. For Experiment 1 and Experiment 2 Objective 2, the spatial correlation between pans will be investigated.

In both Experiment 1 and Experiment 2 Objective 2, measurements taken from the 12 evenly spaced, sequential pans along the line are repeated measures with potential for correlated errors. If significant spatial correlation between pans exists but is ignored and the within Run errors are assumed independent, the Type I error rate will become inflated (Stroup 2012). Alternately, if the correlation structure is over-modeled the power of hypothesis tests will be negatively affected. A correlation structure that is a logical match to the data and a parsimonious fit should be identified.

In order to choose a suitable between-Pan correlation structure for the feed line experiments, a variety of reasonable structures will be compared using fit statistics and graphical methods. Specifically, AICC fit statistics, mean plots, and plots of covariance as a function of lag between pairs of pans will be used to compare models with various correlation matrices. The null model likelihood ratio test will also be used to test a model with a specified Σ correlation matrix against the model with no between-Pan correlation, $\Sigma = \sigma^2 I$. In this context, Σ is the portion of the R matrix (described in detail later) that describes the correlation among pans for each subject.

The simplest variance-covariance structure that includes correlation within subjects is compound symmetry, which is coded CS in SAS. This structure is the default Σ matrix used by proc GLIMMIX when the ‘random’ statement is used. In the feed line experiments compound symmetry would assume equal variance for all levels of fixed effects and equal correlation ρ between each pan regardless of their separation on the line. The Σ matrix using compound symmetry is

$$\Sigma_{CS} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

The next level of complexity in variance-covariance structure is the first order autoregressive model. This model, coded AR(1) in SAS, is a homogeneous variance, two parameter structure like compound symmetry, but it employs a power function so that pans closer together on the line would be more closely correlated.

$$\Sigma_{AR(1)} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^n \\ \rho & 1 & \rho & \cdots & \rho^{n-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^n & \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}$$

Even though the AR(1) structure has an additional element of complexity compared to the compound symmetry structure, it is done with the same two parameters (σ^2, ρ) used in the compound symmetry structure.

Similar to the AR(1) model, but without the restriction that correlation decreases with increasing power of ρ , is the Toeplitz model. The Toeplitz structure requires $n + 1$ parameters, where $n + 1$ is the number of rows or columns of the Σ matrix. A separate correlation parameter ρ_k is calculated for each pair of distances, but every pair of distances that have the same lag have the same correlation.

$$\Sigma_{Toeplitz} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_n \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_n & \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$

The next level of complexity in variance-covariance structure modeling is the first-order ante-dependence model coded ANTE(1) in SAS. This model assumes heterogeneous variance components and relaxes the assumption of a strict power function relationship for correlation within subjects. The R matrix for ANTE(1) is as follows

$$\Sigma_{ANTE(1)} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \cdots & \sigma_1\sigma_n \prod_{i=1}^{n-1} \rho_i \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \cdots & \sigma_1\sigma_{n-1} \prod_{i=2}^{n-2} \rho_i \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \cdots & \sigma_1\sigma_{n-2} \prod_{i=3}^{n-3} \rho_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_n \prod_{i=1}^{n-1} \rho_i & \sigma_1\sigma_{n-1} \prod_{i=2}^{n-2} \rho_i & \sigma_1\sigma_{n-2} \prod_{i=3}^{n-3} \rho_i & \cdots & \sigma_n^2 \end{bmatrix}$$

The most complex correlation structure is the unstructured model. The unstructured correlation structure involves an individually estimated variance and covariance for every element in Σ .

Linear Mixed Models

A Linear Mixed Model (LMM) is a mixed model with a normal response. LMM's will be used to model the feed line experiments and will serve as a baseline for comparisons. Using a LMM for these experiments naively ignores that the response is a continuous proportion, which may not be well approximated by the normal distribution.

The classic representation of a LMM in matrix form is written

$$Y = X\beta + Zu + e \tag{Eq. 3.1}$$

$$u \sim N(0, G) \tag{Eq. 3.2}$$

$$e \sim N(0, R) \tag{Eq. 3.3}$$

where X and Z are known design matrices for the fixed and random effects respectively, β is an unobservable vector of fixed effects, u is an unobservable vector of random effects distributed according to Eq. 3.2, and e is an unobservable vector of random error effects distributed according to Eq. 3.3 (Christensen 2002). The vectors u and e are independent of one another and both have an expectation of zero. The variance-covariance matrices for u and e are labeled G and R respectively.

The same model can be written in probability distribution form in two different but equivalent ways. The first is the *marginal* model.

$$Y \sim N(E[Y] = X\beta, V = ZGZ' + R) \tag{Eq. 3.4}$$

In the marginal model, the mean is based only on fixed effects and the parameters that describe the random effects appear in the variance-covariance matrix, V , only (Littell et al. 2006). The second model is the *conditional* model.

$$Y|u \sim N(X\beta + Zu, R) \tag{Eq. 3.5}$$

In the conditional model, u is distributed as in Eq. 3.2. For LMM's, the two models are exactly the same. If the response is modeled according to a non-normal distribution, the models are different (Stroup 2012).

Mixed model equations are used to obtain best linear unbiased estimates (BLUEs) for the fixed-effect parameter vector, β . Best linear unbiased predictors (BLUPs) are the predictors used when information on the random effects, u , needs to be incorporated. In the auger feed line experiments, BLUE's are used to obtain fixed effect parameter estimates averaged over all runs and BLUP's are used to obtain predicted values for individual runs. The estimate of the expectation of the conditional LMM (Eq. 3.5), $E(Y|u)$, is the BLUP and the estimate of the expectation of the marginal LMM (Eq. 3.4), $E(Y)$, is the BLUE.

Generalized Linear Mixed Models

GLMM's extend the theory of LMM's to responses that have a non-normal distribution. In GLMM's the response data is not transformed; instead, the explanatory variables are expressed as having a linear relationship to a *function* g of the mean of $Y|u$, the response variable conditioned on the random effects. This function is the link function and it relates the response to the explanatory variables in a linear fashion, thus allowing the use of standard LMM techniques for estimation and hypothesis testing. The *conditional linear predictor* for a GLMM is

$$\eta = g(E[Y|u]) = X\beta + Zu. \quad \text{Eq. 3.6}$$

The conditional model is used to describe GLMM's since the link function is a function of the conditional expectation.

Appropriate choice of response distribution and accompanying link function are pivotal in GLMM applications. In the feed line experiments, responses are continuous proportions between 0 and 1. The response distributions of interest for this data are the beta and binomial distributions (Stroup 2012). Both beta and binomial distributions typically use the logit link function in GLMM's.

$$\text{Logit Link:} \quad \eta = \log\left(\frac{\pi}{1-\pi}\right) \quad \text{Eq. 3.7}$$

The beta distribution is a practical choice to describe continuous response data that is bounded by zero and one (Johnson and Kotz 1970). The beta distribution can take many shapes due to its two independent parameters. This makes it ideal for modeling percent data that may create varying histogram shapes based on the correlation between subjects or other patterns specific to the data set. The beta distribution's two-parameter flexibility also makes it

computationally difficult to use in SAS analyses involving repeated measures. Stroup (2012, p435) discusses the uncertainty involving the relationship of the lowest level random effect and the conditional response in repeated measures analyses. Stroup concludes that two-parameter exponential family distributions are problematic because “we have not yet clearly established how best to model repeated measures with these data.”

The standard parameterization according to Johnson and Kotz (1970) of the beta distribution with parameters a, b has the probability density function

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad \text{where } (0 \leq x \leq 1)$$

with expected value

$$E[X] = \mu = \frac{a}{a+b}$$

and variance

$$Var[X] = \frac{ab}{(a+b+1)(a+b)^2}$$

SAS uses the parameterization of Ferrari and Cribari-Neto (2004) for the beta distribution. This alternative parameterization of the beta density function is expressed in terms of the expected value μ and the precision parameter ϕ . The precision parameter is inversely proportional to the variance of y . The reciprocal of ϕ is a dispersion parameter. (Cribari-Neto and Zeileis 2010) The SAS parameterization for $y \sim \beta(\mu, \phi)$ uses $\mu = \frac{a}{a+b}$ and $\phi = a + b$ to rewrite the density function as

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\phi\mu-1} (1-y)^{(1-\mu)\phi-1} \quad \text{where } (0 \leq y \leq 1) \quad \text{Eq. 3.8}$$

with expected value

$$E[Y] = \mu \quad \text{Eq. 3.9}$$

and variance

$$Var[Y] = \frac{\mu(1-\mu)}{(1+\phi)} \quad \text{Eq. 3.10}$$

The variance of y is a function of μ , thus this parameterization is heteroscedastic. When SAS implements a GLMM analysis in the GLIMMIX procedure, a different standard error will automatically be calculated for each estimated mean. If the same analysis were implemented with a normally distributed response, SAS would by default assume a common variance (Schabenberger 2005).

The potential problems associated with beta response models create the need for a reasonable alternative, one-parameter distribution that offers better behavior when included in a repeated measures model. The binomial distribution is mathematically and practically similar to the beta distribution but it is designed for discrete responses that are the ratio of successes to number of trials. This discrepancy can be overcome and the binomial distribution can be used for continuous proportion data if the data are forced into binomial form. Cribari-Neto and Zeileis (2010 p5) successfully use beta response regression models to model binomial data for large numbers of trials. These authors show that for large enough values of N a discrete binomial model can be approximated with a continuous beta model. It is intuitive that for equally large values of N a continuous beta model should be well approximated by a discrete binomial model. McCullagh and Nelder (1989 p124-128) present the “quasi-binomial” response model that can be substituted for the beta response model. McCullagh and Nelder (1989) show how continuous proportions can be approximated by a number of successes out of number of trials provided a large number of trials are chosen.

In order to have a one-parameter exponential family distribution model to compare beta response model results to, a quasi-binomial data set will be created for each of the feed line experiments. Once the quasi-binomial methodology is used to create new data sets, the responses can be modeled as standard binomial responses. Since McCullagh and Nelder (1989) specify “a large number of trials” a variety of values for N (number of theoretical trials) will be explored. A new set of responses Y_N (integer number of successes) will be created for each value of N used. The following equation is solved for each Y_N to create the new data set.

$$\frac{Y_N}{N} = \textit{Proportion} \quad N = 10, 25, 50, 100, 200, 300, 400, 500, 750, 1000$$

“Proportion” is the original continuous response. To ensure the Y_N values are integers, $Y_N = N * \textit{Proportion}$ will need to be rounded to the nearest integer. Rounding will have less impact on the data set for large values of N. The SAS code in Appendix A fully illustrates how the quasi-binomial data set is created for each experiment.

The standard parameterization of the binomial distribution according to Mood et al. (1974, p88-89) is

$$f_X(x; n, p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \textit{for } x = 0, 1, \dots, n \\ 0 & \textit{otherwise} \end{cases} \quad 0 \leq p \leq 1$$

with expected value

$$E[X] = \mu = np$$

and variance

$$Var[X] = npq$$

where n is the number of trials, p is the probability of success in each trial and $q = 1 - p$.

SAS parameterizes the binomial distribution in terms of the expected value μ .

$$f_Y(y; \mu) = \begin{cases} \binom{n}{r} \mu^r (1 - \mu)^{n-r} & \text{for } y = \frac{r}{n}, r = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq. 3.11}$$

with expected value

$$E[Y] = \mu \quad \text{Eq. 3.12}$$

and variance

$$Var[Y] = \frac{\mu(1-\mu)}{n} \quad \text{Eq. 3.13}$$

The scale parameter, ϕ , is equal to one for the binomial distribution and only the parameter μ is estimated.

GLMM's require an iterative procedure for parameter estimation. The two general categories of iterative procedure used by SAS are linearization and integral approximation. The GLIMMIX procedure uses the Pseudo-likelihood method for linearization and either Laplace approximation or adaptive Gauss-Hermite quadrature for integral approximation. These methods maximize log likelihoods of exponential family, non-normal distributions. Pseudo-likelihood is the default method for Proc GLIMMIX. The method used for all analysis in this report is Laplace. The Laplace method is an approximation for maximum likelihood and is computationally simpler than adaptive Gauss-Hermite quadrature. The Laplace method generally performs the same as Pseudo-likelihood but performs better when used with beta response variables (Stroup 2012, p328).

LMM and GLMM with Experiment 1

In Experiment 1, the fixed effects are Feed, Pan and Feed*Pan. The random effects are Run(Feed) and Pan*Run(Feed), the whole plot and split plot error terms, respectively. The researcher intends to make conclusions about the destructiveness of the feed line on two kinds of feed, high PDI and low PDI. The following LMM in scalar form is used to describe the experiment

$$y_{ijk} = \mu + \tau_i + r(\tau)_{ij} + \rho_k + \tau\rho_{ik} + e_{ijk} \quad \text{Eq. 3.14}$$

where $i = 1,2$ $j = 1,2,3,4$ $k = 1,2 \dots 12$

μ is the overall mean, τ_i is the effect of the i th feed, $r(\tau)_{ij}$ is the random effect of the j th run nested within the i th feed, ρ_k is the fixed effect of the k th pan, and $\tau\rho_{ik}$ is the interaction effect of the i th feed and the k th pan. The random effects are assumed to be distributed as follows given there is no between-Pan correlation structure

$$r(\tau)_{ij} \sim iid N(0, \sigma_{r(\tau)}^2) \quad e_{ijk} \sim iid N(0, \sigma_e^2)$$

with $r(\tau)_{ij}$ and e_{ijk} being independent of one another. The probability distribution of the response conditioned on the random effects is

$$y_{ijk} | r(\tau)_{ij} \sim N(\mu + \tau_i + \rho_k + \tau\rho_{ik}, \sigma_e^2)$$

The model for Experiment 1 in matrix form (Eq. 3.1) consists of: Y – a 96×1 vector of responses, X – a 96×39 design matrix for the fixed effects of Feed, Pan and Feed*Pan, β – a 39×1 vector of the fixed effect parameters μ , τ_i , ρ_k , and $\tau\rho_{ik}$, Z – a 96×8 design matrix of random Runs nested within Feed, u – a 8×1 vector of random effects $r(\tau)_{ij}$, and e – a 96×1 vector of the unobservable split plot errors. $u \sim N(0, G)$ where the G matrix contains all the variance components associated with random vector u . $e \sim N(0, R)$ where the R matrix contains the variance components for Pan and the between-Pan covariance components over all 8 Run by Feed combinations. If no correlation among Pans exists and Pan variability is constant, then $R = \sigma_e^2 I$.

When the analysis of Experiment 1 is conducted in SAS using the LMM, few lack of fit issues are evident in the residual analysis. There are minor deviations from normal at each end of the Q-Q plot of residuals shown in Fig. 3.1 and some right skew of the histogram in Fig. 3.2. The results of the Tests for Normality used by SAS in the Univariate procedure are summarized in Table 3.1. None of the tests have a p-value that would indicate the residuals differ significantly from the normal distribution.

Figure 3.1 Q-Q plot for residuals for Experiment 1 using a LMM

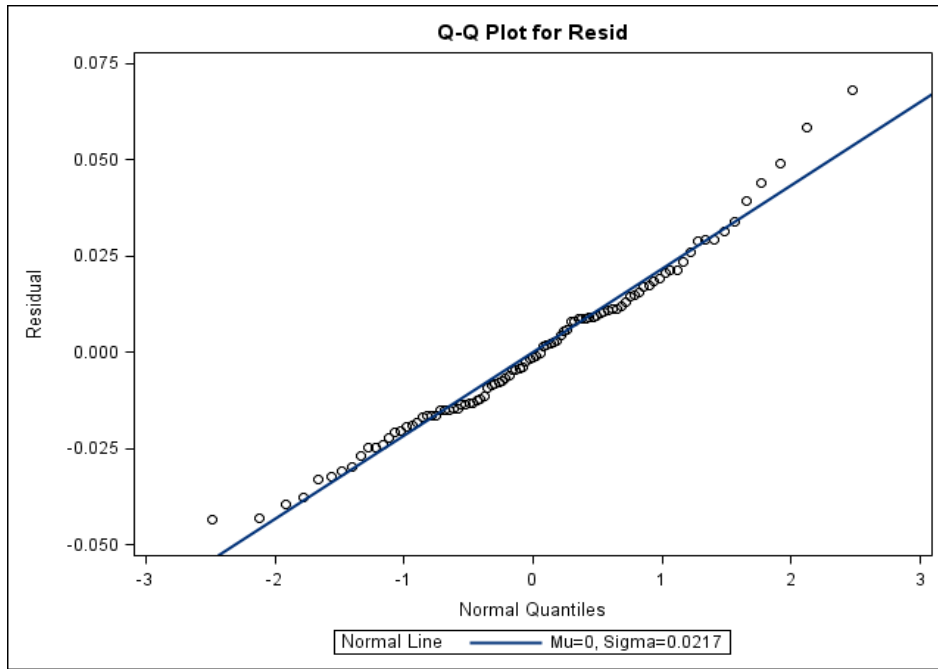


Figure 3.2 Histogram of distribution of residuals for Experiment 1 using a LMM

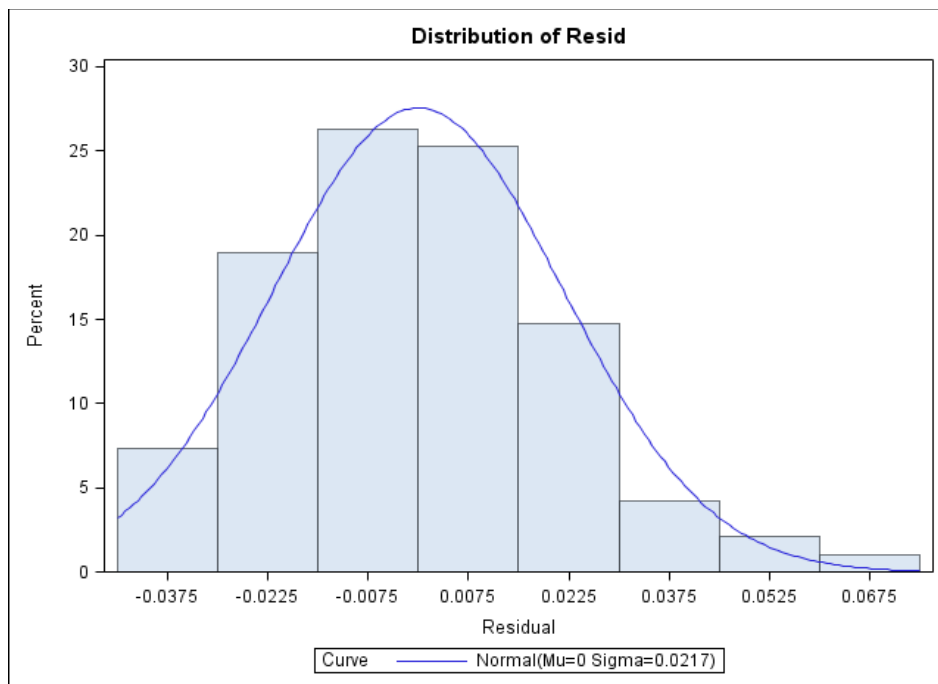


Table 3.1 Tests for Normality of Residuals for Experiment 1 using a LMM

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.981325	Pr < W	0.1940
Kolmogorov-Smirnov	D	0.060903	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.056135	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.383482	Pr > A-Sq	>0.2500

The analysis will be re-run using GLMM's to model the response according to beta and quasi-binomial distributions. The GLMM results and residuals can then be compared to those of the LMM with the expectation that neither model will offer a significantly superior fit. The conditional model for the GLMM with beta or binomial response consists of the logit link set equal to the linear predictor

$$\eta = \log\left(\frac{\pi}{1-\pi}\right) = \mu + \tau_i + r(\tau)_{ij} + \rho_k + \tau\rho_{ik}$$

$$\text{where } i = 1,2 \quad j = 1,2,3,4 \quad k = 1,2 \dots 12$$

and response distribution for the beta is

$$y_{ijk}|r(\tau)_{ij} \sim \text{beta}(\mu + \tau_i + \rho_k + \tau\rho_{ik}, \phi)$$

and the response distribution for the binomial is

$$y_{ijk}|r(\tau)_{ij} \sim \text{binomial}(\mu + \tau_i + \rho_k + \tau\rho_{ik}, N)$$

$$N = 10, 25, 50, 100, 200, 300, 400, 500, 750, 1000$$

The binomial distribution here is the quasi-binomial distribution described earlier in Ch. 3. Once the new data set of number of successes out of trials is created as previously described, the data is modeled as a traditional binomial response data.

LMM and GLMM with Experiment 2

In Experiment 2, the fixed effects are Type and Pan for Objective 1 and 2 respectively. The random effects for both objectives are Run and the interaction of Run and the fixed effect, which serves as the error term. Objective 1 is to compare the percent fat found in the pellets to the percent fat found in the fines for each pan 1-12 on the feed line. The following LMM in scalar form is used to describe Objective 1

$$y_{ij} = \mu + \tau_i + r_j + e_{ij} \quad \text{Eq. 3.15}$$

$$\text{where } i = 1,2 \quad j = 1,2,3,4$$

μ is the overall mean, τ_i is the effect of the i th type of feed (pellet or fine), r_j is the random effect of the j th run, and e_{ij} is the term for random error associated with the Type by Run interaction. The random effects are assumed to be distributed as follows given there is no within-subject correlation structure

$$r_j \sim iid N(0, \sigma_r^2) \quad e_{ij} \sim iid N(0, \sigma_e^2)$$

with r_j and e_{ij} being independent of one another. The probability distribution of the response conditioned on the random effects is

$$y_{ij} | r_j \sim N(\mu + \tau_i + r_j, \sigma_e^2)$$

Objective 2 is to compare the percent fat found in the pellets for each Pan 1-12 along the feed line to that of the control Pan (Pan 0). The model for Objective 2 is identical to that of Objective 1 expect the term τ_i in Eq. 3.15 is replaced with ρ_i where $i = 0, 1, 2, \dots, 12$ and ρ_i is the effect of the i th pan.

The model for Experiment 2 in matrix form (Eq. 3.1) consists of: Y – a 8×1 vector of responses for Objective 1 and a 52×1 vector of responses for Objective 2, X – a 8×3 design matrix for the fixed effect Type for Objective 1 and a 52×13 design matrix for the fixed effect Pan in Objective 2, β – a 3×1 vector of the fixed effect parameters μ and τ_i for Objective 1 and a 13×1 vector of the fixed effect parameters μ and ρ_i for Objective 2, Z – a 8×4 design matrix of random Runs for Objective 1 and a 52×4 design matrix of random Runs for Objective 2, u – a 4×1 vector of random effects r_j for both objectives, and e – a 8×1 vector of unobservable errors for Objective 1 and a 52×1 vector of unobservable errors for Objective 2. $u \sim N(0, G)$ where the G matrix contains all the variance components associated with random vector u . $e \sim N(0, R)$ where $R = \sigma_e^2 I$ in Objective 1 and R contains the variance components of Pan and the between-Pan covariance components for each Run in Objective 2. If no correlation among Pans exists and Pan variability is constant, then $R = \sigma_e^2 I$ in Objective 2.

The residual analysis for Experiment 2 Objective 1 using a LMM indicates that the residuals deviate somewhat from normality, but not enough to result in p-values less than 0.05 in the Tests of Normality conducted in Proc Univariate (Table 3.2). The Q-Q Plot in Fig. 3.3 and

the histogram in Fig. 3.4 both show minor deviations of the residuals from the normal distribution.

Figure 3.3 Q-Q plot for residuals for Experiment 2 Objective 1 using a LMM

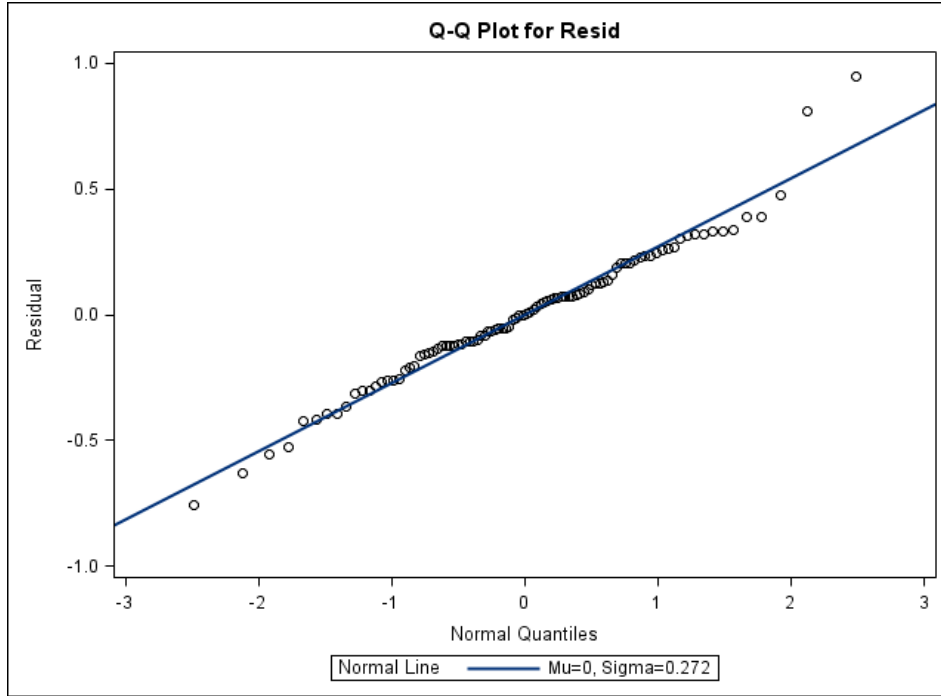


Figure 3.4 Histogram of distribution of residuals for Experiment 2 Objective 1 using a LMM

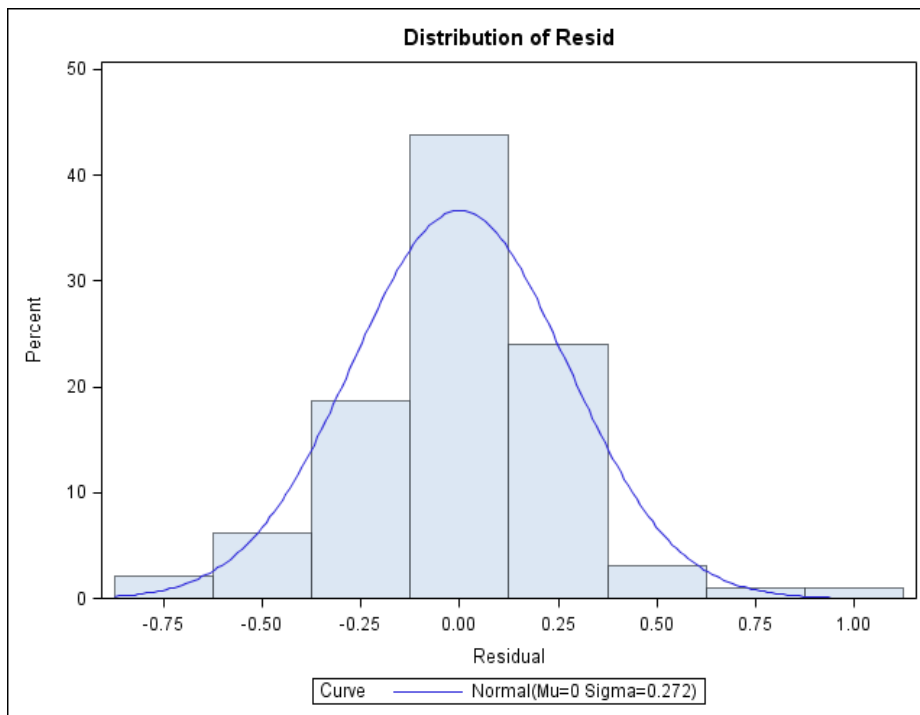


Table 3.2 Tests for Normality of Residuals for Experiment 2 Objective 1 using a LMM

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.975809	Pr < W	0.0727
Kolmogorov-Smirnov	D	0.06535	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.062801	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.487854	Pr > A-Sq	0.2261

The residual analysis for Experiment 2 Objective 2 using a LMM indicates a substantial lack of fit of the model. The Q-Q Plot for residuals in Fig. 3.5 shows deviations from the normal line and the histogram of residuals in Fig. 3.6 shows a poor match with the curve representing the normal distribution. Table 3.3 summarizes the results of four Tests for Normality performed on the residuals. Of the four tests, three tests (Shapiro-Wilk, Cramer-von Mises, and Anderson-Darling) have p-values less than 0.05 indicating the residuals follow a distribution that is non-normal.

Figure 3.5 Q-Q plot for residuals for Experiment 2 Objective 2 using a LMM

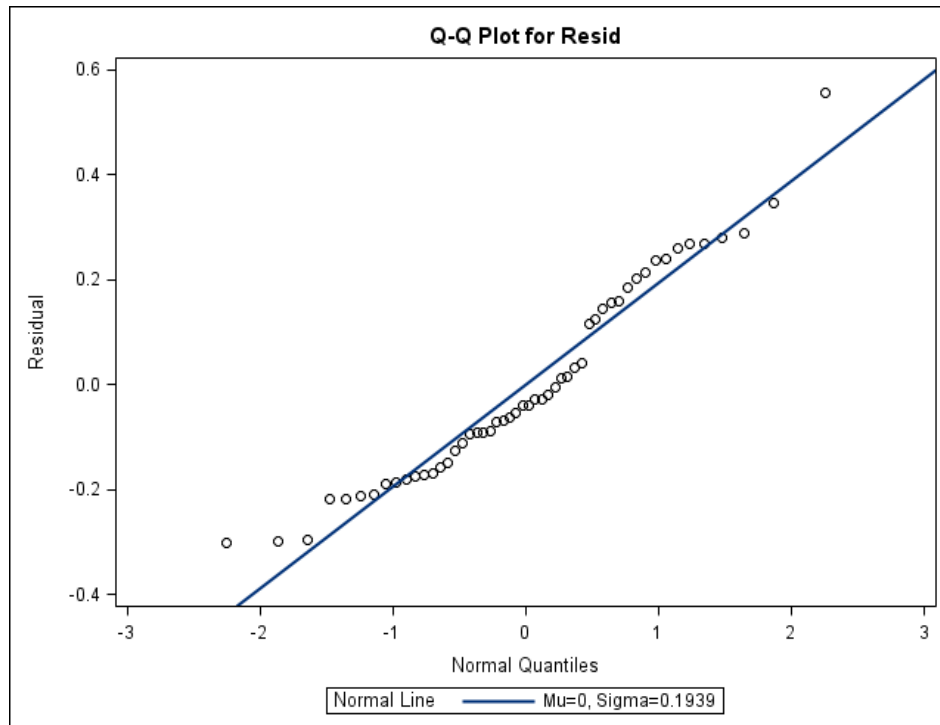


Figure 3.6 Histogram of distribution of residuals for Experiment 2 Objective 2 using a LMM

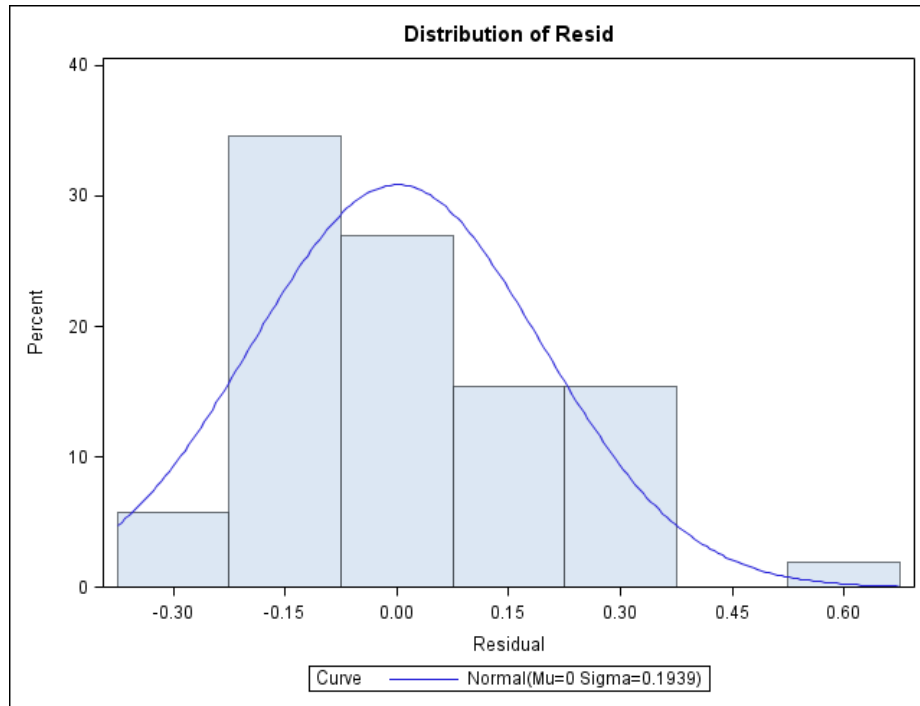


Table 3.3 Tests for Normality of Residuals for Experiment 2 Objective 2 using a LMM

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.949253	Pr < W	0.0271
Kolmogorov-Smirnov	D	0.114855	Pr > D	0.0858
Cramer-von Mises	W-Sq	0.160412	Pr > W-Sq	0.0180
Anderson-Darling	A-Sq	0.907599	Pr > A-Sq	0.0205

The analyses for Objective 1 and Objective 2 will be re-run using GLMM's to model the response according to beta and quasi-binomial distributions. The scalar models for the GLMM's will be created as was done in Experiment 1. The expectation is that a better fitting models can be obtained, which may impact the conclusions of the analysis, particularly for Objective 2.

Chapter 4 - Results

This chapter presents the results of analysis. The objectives are to determine if the conclusions drawn regarding research objectives remain consistent across the different response models under consideration and to compare the models themselves. The response is modeled as Gaussian (LMM), beta (GLMM) and with 10 variations of the quasi-binomial distribution where the sample size N varies from 10 to 1000 (GLMM). Correlational structure between Pans within Run is determined for each experiment and incorporated into the variance-covariance structure. The analysis results will allow us to determine how the GLMM models compare to the LMM and what range of count parameters for the quasi-binomial distribution offers a good approximation to the beta distribution. Models will be compared on the basis of point estimates, their confidence intervals and confidence interval widths, as well as p-values for F-tests of the fixed effects.

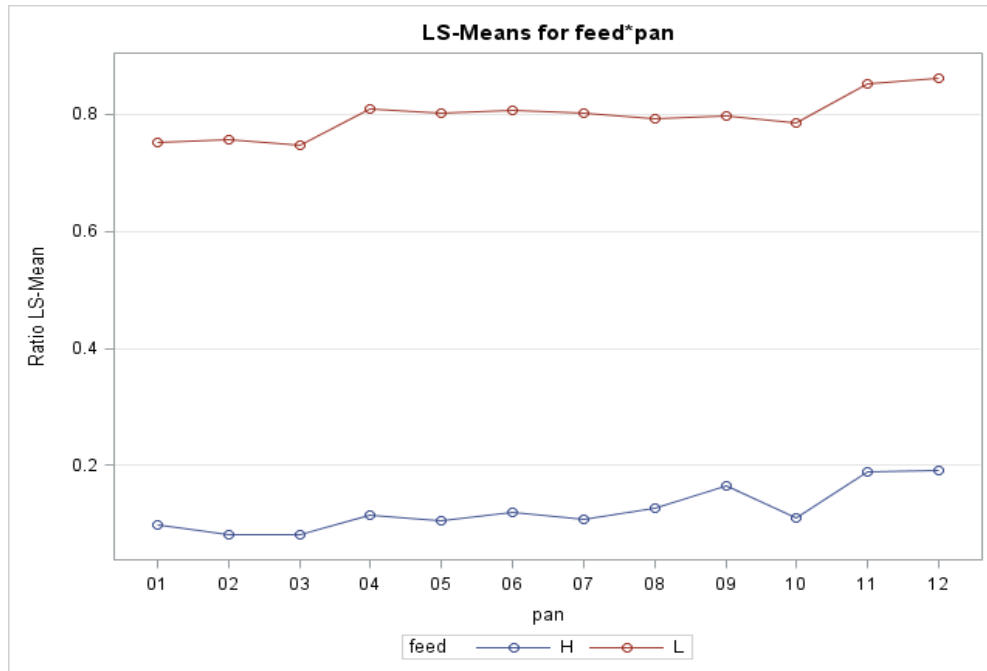
Correlation Structure

Prior to comparing response distributions, the within Run correlation between Pans will be investigated. Graphical methods will be used to visualize changes in point estimates by Pan and changes in covariance by distance between Pans. Littell et al. (2006) and Stroup (2012) recommend mean plots and plots of covariance by distance.

Experiment 1 Correlation Structure

In Experiment 1, Run(Feed) is the subject of the repeated measure Pan. The levels of Pan yield the 12 rows and 12 columns for the correlation matrix Σ . Two kinds of plots are generated in SAS to aid in the choice of reasonable correlation structure. The first is the LSMMeans plot of the response, Ratio, as a function of Pan. Fig. 4.1 indicates that the correlation between Pans within Run(Feed) is likely a function of Pan position because Pans that are far apart have greater difference in point estimates than those close together on the feed line.

Figure 4.1 LS-Means plot of point estimates for each Feed by Pan for Experiment 1



A plot of covariance as a function of lag or distance between pairs of Pans (Stroup 2012, p421) is produced in SAS to gain better understanding of the structure. This plot is shown in Fig. 4.2. It is created by using a LMM with unstructured correlation structure to analyze the data for Experiment 1 in Proc Mixed and outputting the elements of the fixed effects variance-covariance matrix, R , to be used in the plot. For the unstructured correlation structure, each element of the R matrix is individually estimated and is used as a point on the graph. Each line on the graph represents a row from the R matrix beginning with the diagonal variance component and moving right across the row. For example, the blue, square-point line in Fig. 4.2 shows the variance of Pan 2 at horizontal axis element 0, the covariance between Pan 2 and a pan that is one pan interval away at horizontal axis element 1, and so on until the covariance between Pan 2 and Pan 12 (10 pans apart) is shown at the horizontal axis element 10.

Figure 4.2 Plot of Covariance by Lag for Experiment 1

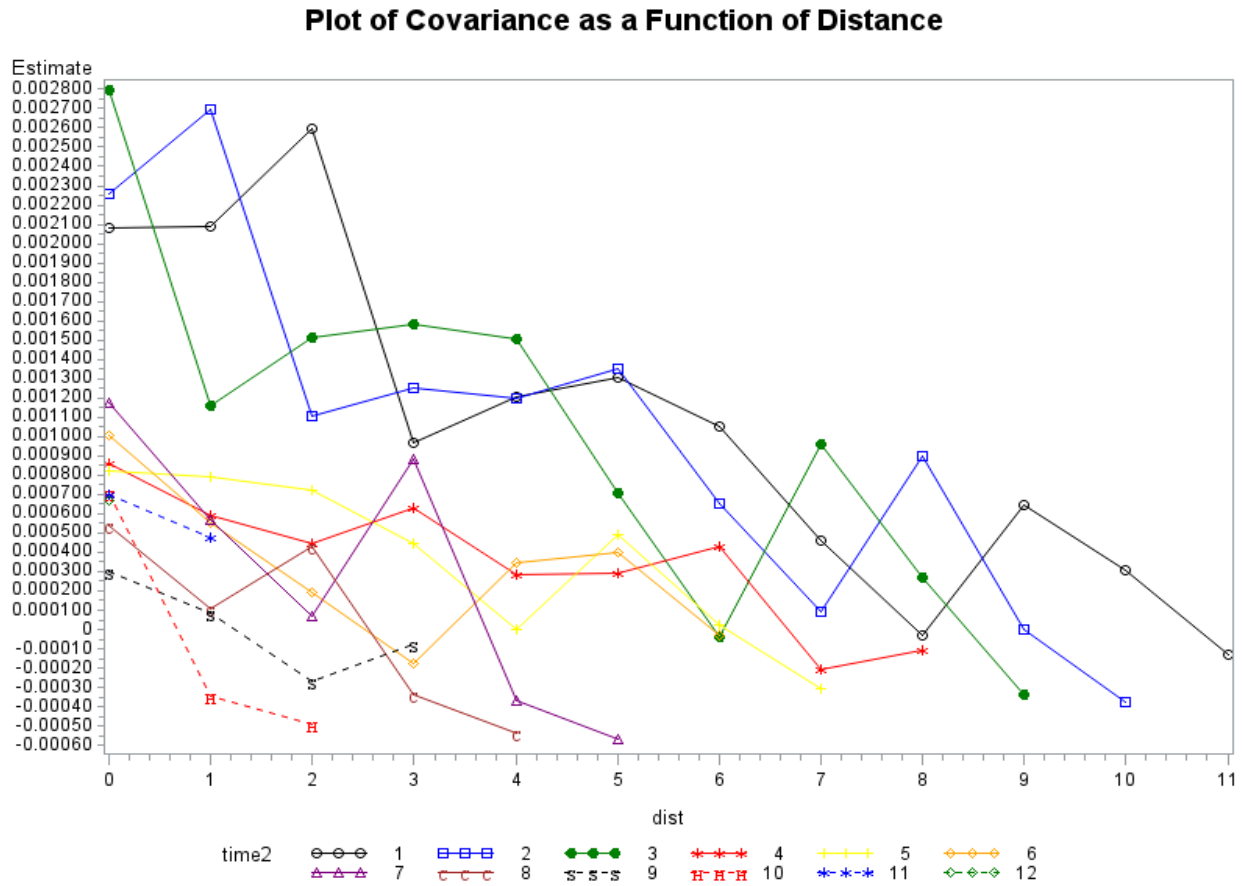


Fig.4.2 allows us to visualize the actual structure of the correlation between Pans within Run(Feed) in order to choose an appropriate structure. Fig.4.2 shows a definite decrease in covariance as distance between Pans increases. The two simplest correlation structures that fit this criteria are AR(1) and Toeplitz. A simpler structure would be compound symmetry, but we would expect the graph of covariance by lag to be flat. The graph does not show strong evidence of the amount of decrease in covariance being dependent on Pan position, thus heterogeneous variance models can be disregarded as more complex than necessary.

Fit statistics are a good method of evaluating variance-covariance structures when the response is modeled as normal and can be used to determine whether AR(1), Toeplitz, or some other structure would create a better fitting model. We will rely on fit statistics to make the final determination of correlation structure. Using a LMM in the Mixed Procedure in SAS, fit statistics for models using compound symmetry, AR(1), Toeplitz, ARH(1), and ANTE(1) correlation structures were obtained. The Null Model Likelihood Ratio Test has a p-value of

<.001 for each of the aforementioned models indicating that all models would offer significant improvement over the variance component only structure with no correlation. The Akaike Information Criterion with small sample correction (AICC) values for models fit with these correlation structures are summarized in Table 4.1. The model using Toeplitz correlation structure would be selected as a better fit based on a slightly lower AICC value.

Table 4.1 AICC values for models with various correlation structures for Experiment 1

Type	AICC
CS	-266.5
AR(1)	-283.2
Toeplitz	-287.6
ARH(1)	-263.8
ANTE(1)	-256.9

Graphical evidence and fit statistics both favor the choice of Toeplitz for the correlation structure between Pans within Run(Feed) for Experiment 1. The forthcoming results for Experiment 1 comparing the LMM to GLMM's will all use the Toeplitz structure for the Σ matrix.

Experiment 2 Correlation Structure

In Experiment 2, the correlation between Pans within Run is relevant only in Objective 2 where the percent fat in pellets along the line is compared to the percent fat of a control pan. For this objective, Run is the subject of the repeated measure Pan. Because of the addition of the Control Pan (Pan 0), the correlation matrix Σ is 13 x 13. As with Experiment 1, a LS means plot of percent fat by Pan and a plot of covariance by distance are generated in SAS to facilitate the choice of an appropriate correlation structure.

Figure 4.3 LS-Means plot of point estimates by Pan for Experiment 2 Objective 2

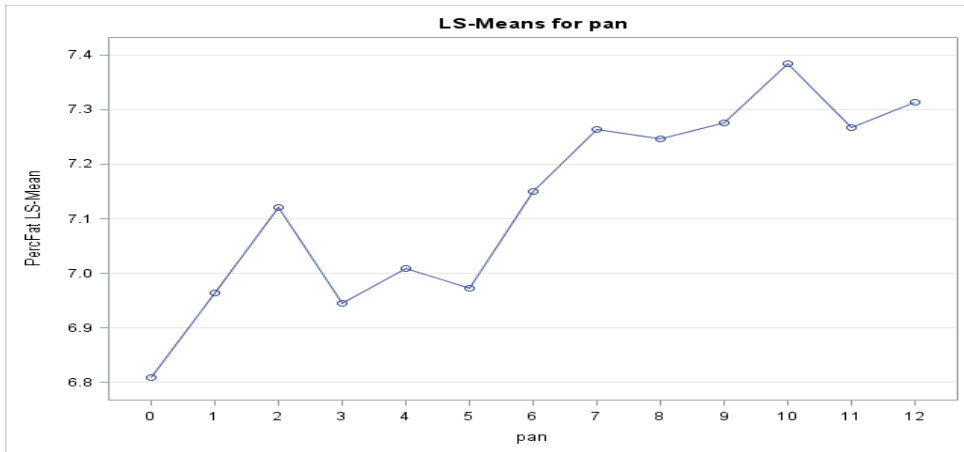


Fig. 4.3 indicates a likely positive correlation between Pan and Percent Fat in Pellets. As in Experiment 1, the plot of Correlation as a function of lag or distance (Fig. 4.4) will give a better indication of what the correlational structure is appropriate.

Figure 4.4 Plot of Covariance by Lag for Experiment 2 Objective 2

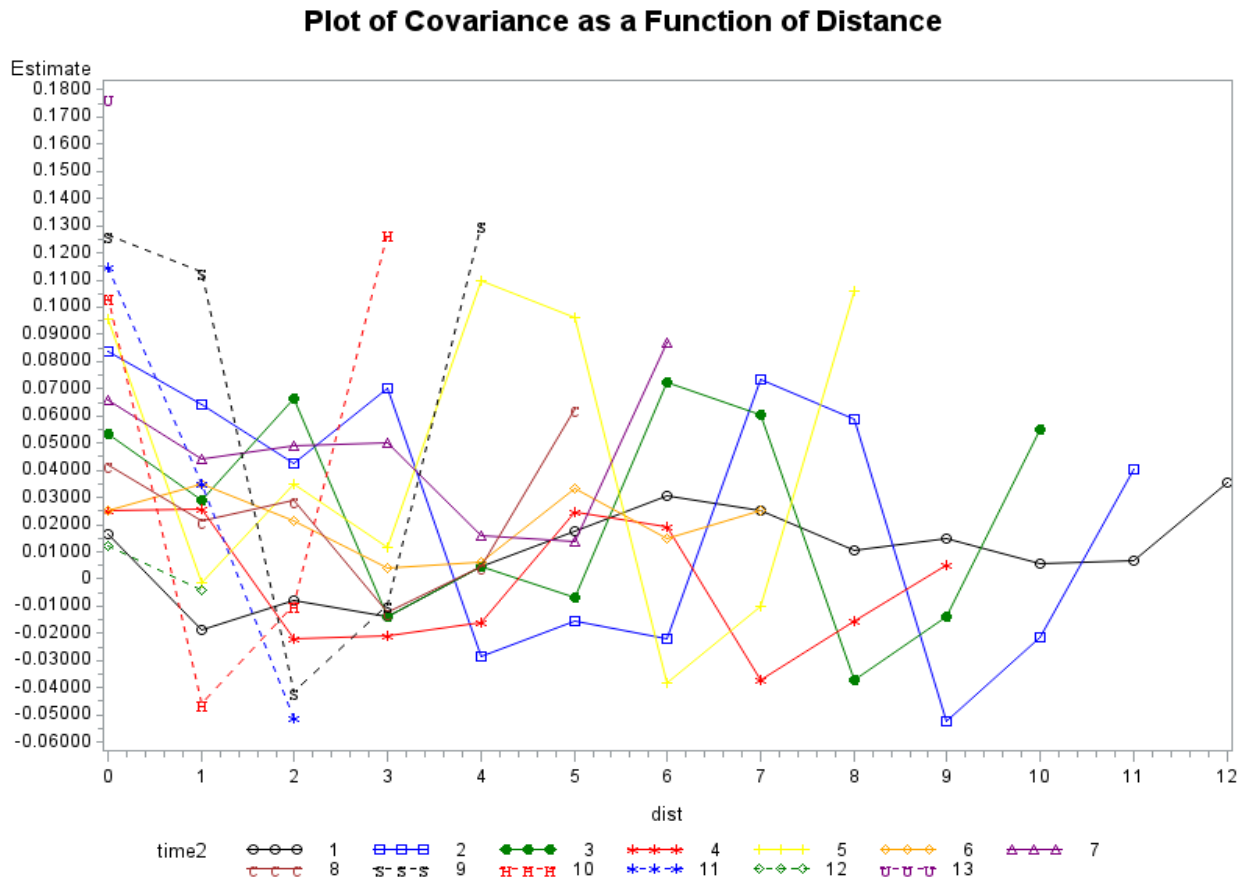


Fig. 4.4 shows a large amount of fluctuation in covariance between adjacent pans, but no overall increasing or decreasing pattern. This suggests that the Σ matrix for Experiment 2 Objective 2 should utilize Compound Symmetry structure.

To confirm this interpretation, fit statistics were generated using a LMM and the Mixed Procedure in SAS. Table 4.2 summarizes the AICC values for the LMM using variance component only, compound symmetry, AR(1), ARH(1), and ANTE(1) correlation structures. Toeplitz is not included due to lack of convergence. As expected, compound symmetry structure has the lowest AICC value and hence offers the best model fit. All forthcoming analysis results will utilize compound symmetry correlation structure for Experiment 2 Objective 2.

Table 4.2 AICC values for models with various correlation structures for Experiment 2 Objective 2

Type	AICC
VC	28.5
CS	23.5
AR(1)	26.0
ARH(1)	52.4
ANTE(1)	128.5

Results for Experiment 1

The research objectives for Experiment 1 are to determine if there is a statistically significant difference in the ratio of fines to total weight of feed in a pan for each type of feed (High PDI, Low PDI) and among the pans (Pan 1 – Pan 12) on the feed line. All models under consideration yield p-values <.0001 in the Type III Tests for Fixed Effects for Type and for Pan, but none produce a statistically significant p-value for the Feed*Pan interaction. Though the same conclusions are reached for all models, the models do not all yield the same point estimates and confidence intervals.

Figures 4.5 and 4.6 show the mean estimates and 95% confidence intervals for High PDI and Low PDI Feed respectively. Figure 4.7 compares mean estimates and 95% confidence intervals for the first and last pans on the feed line (Pan 1 and Pan12). It can be seen in all three figures that the quasi-binomial N=10 model produces very wide confidence intervals and point

Figure 4.6 Low PDI mean estimates with 95% confidence intervals for Experiment 1

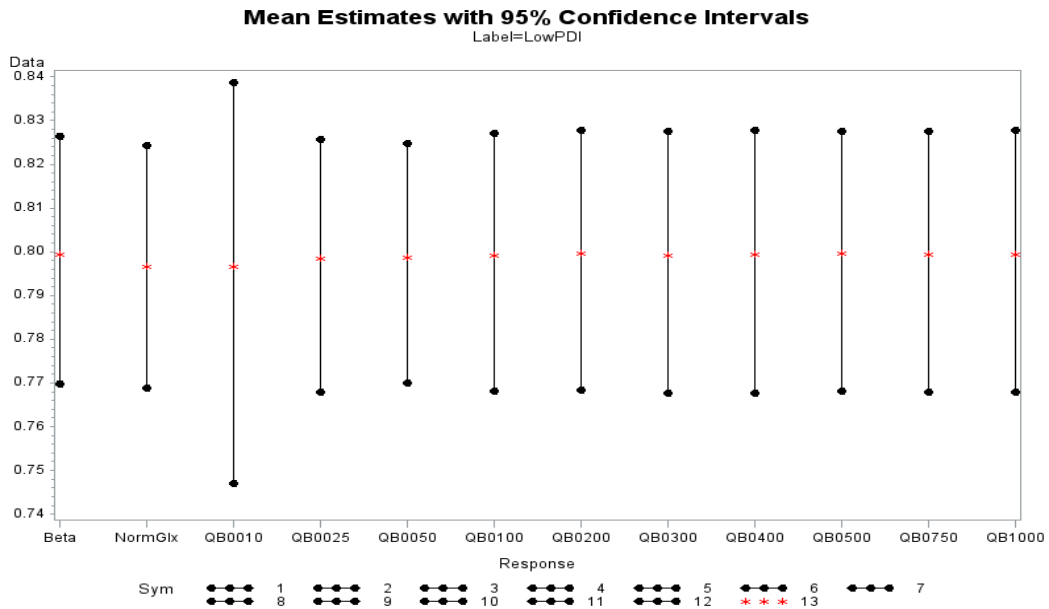


Figure 4.7 Pan 1 & Pan 12 mean estimates with 95% confidence intervals for Experiment 1

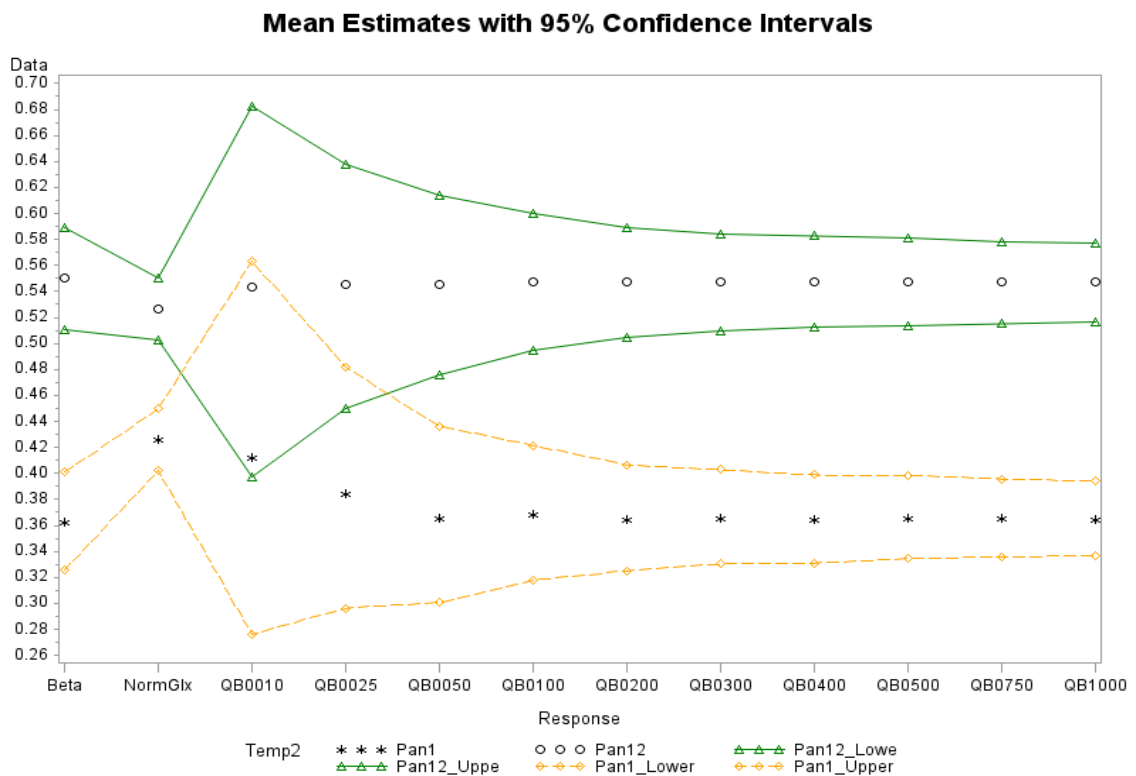
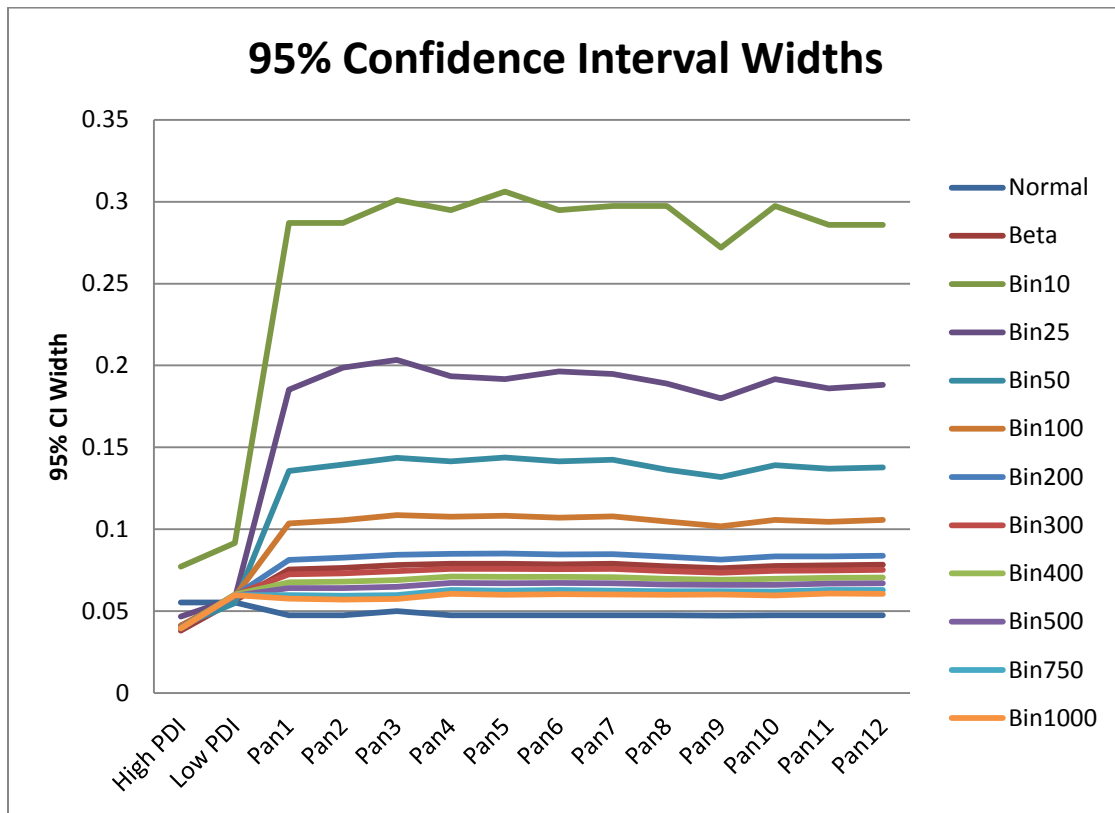


Figure 4.8 shows the widths of the 95% confidence intervals for point estimates of levels of main effects in Experiment 1. It is notable that there is a relatively small range of widths for the two levels of Feed where estimates are averages of 48 observations. There is a much greater range for the levels of Pan, where estimates are averages of only 8 observations. Overall, the normal response model produces very narrow confidence intervals and quasi-binomial response models with low values of N produce wide confidence intervals.

Figure 4.8 95% confidence interval widths for mean estimates for Experiment 1



Results for Experiment 2

Experiment 2 Objective 1

The researcher's objective in Experiment 2 Objective 1 is to determine if there is a statistically significant difference in the percent fat in the pellets and the fines for each pan along the feed line. Since this analysis is done by pan, there are only 8 degrees for each pan (Table 2.2). The combination of extremely small sample size and observations close to zero create

issues in the Glimmix procedure particularly with the beta response model. Convergence issues prevent the obtainment of confidence intervals for the beta response model, indicating that this is an instance where a good approximation to the beta response model is needed. For all pans along the feed line, the normal response model generates confidence intervals that are unrealistically narrow.

Figures 4.9, 4.10, and 4.11 show mean estimates and 95% confidence intervals for the percent fat in Fines and Pellets for Pan 1, Pan 6 and Pan 12 respectively. These figures show the first, middle, and last pans on the feed line and are representative of all the pans on the feed line. In all three figures, extremely narrow confidence intervals for the normal response models and extremely wide confidence intervals for quasi binomial response models with low values of N can be seen. These issues coupled with the results for Experiment 1, lead to the conclusion that a quasi-binomial response model with N of 200 or greater would be a good alternative to the beta response model.

Figure 4.9 Pan 1 mean estimates with 95% confidence intervals for Experiment 2

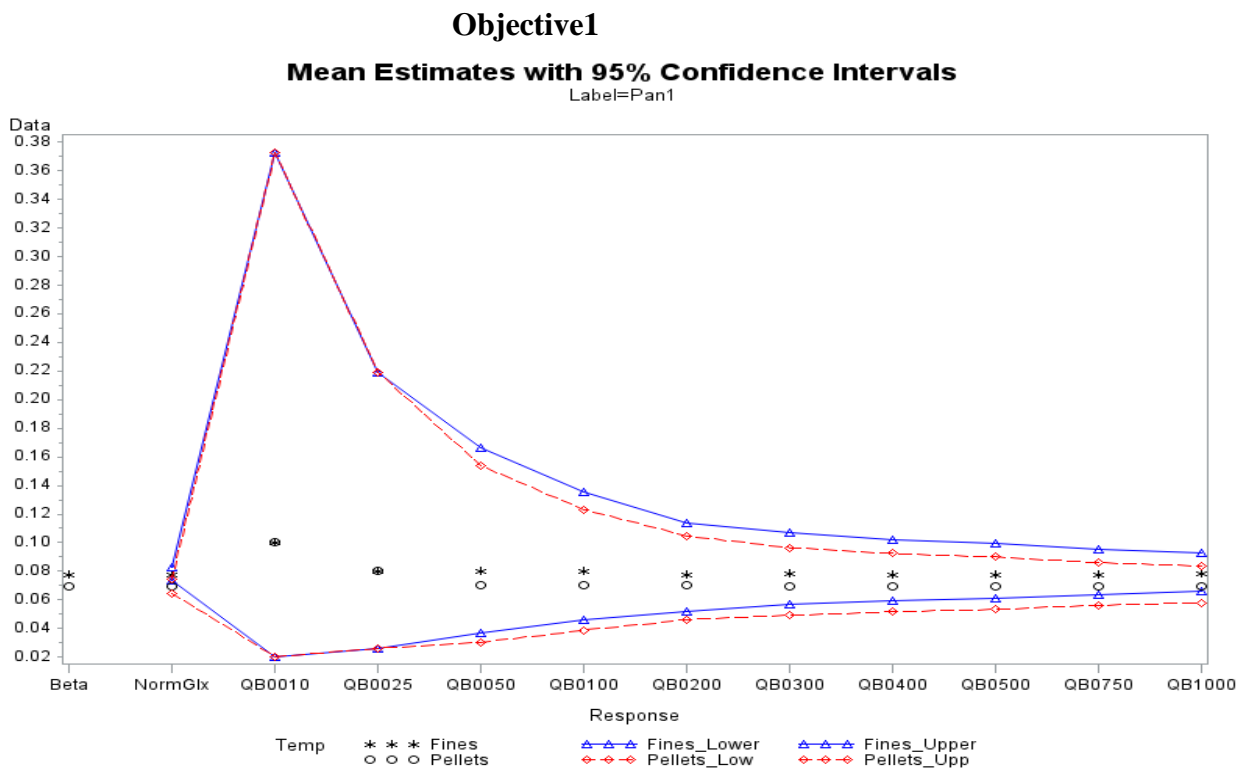


Figure 4.10 Pan 6 mean estimates with 95% confidence intervals for Experiment 2

Objective 1

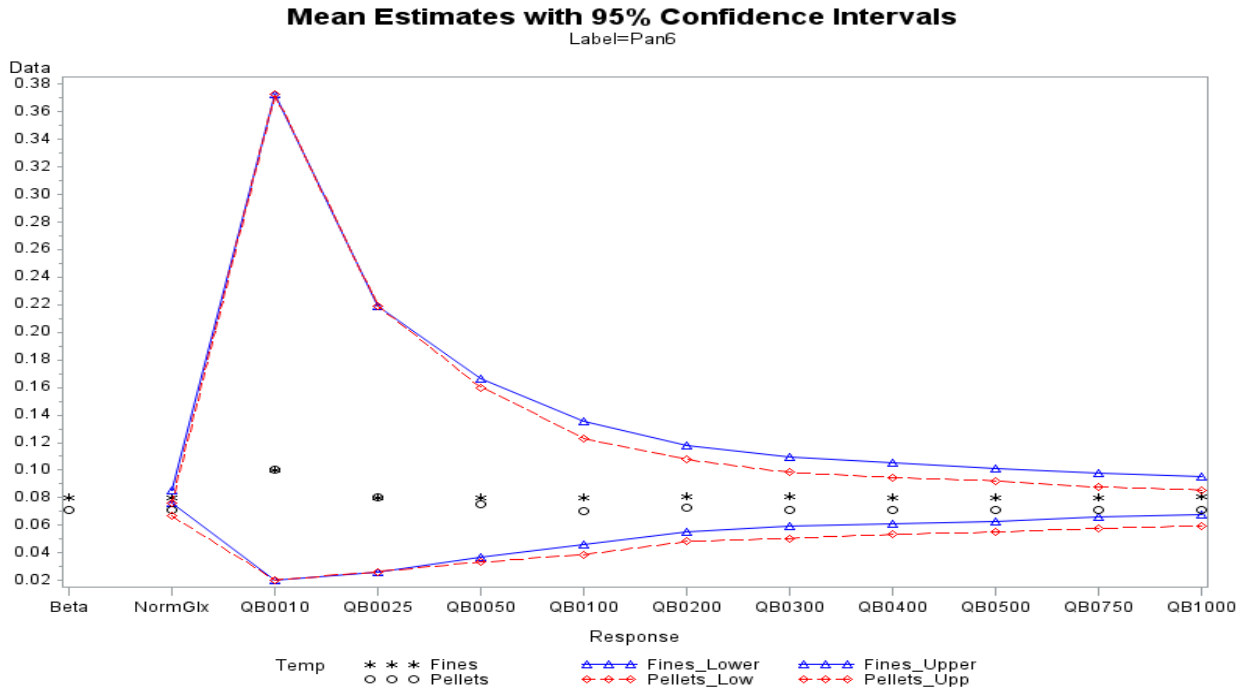
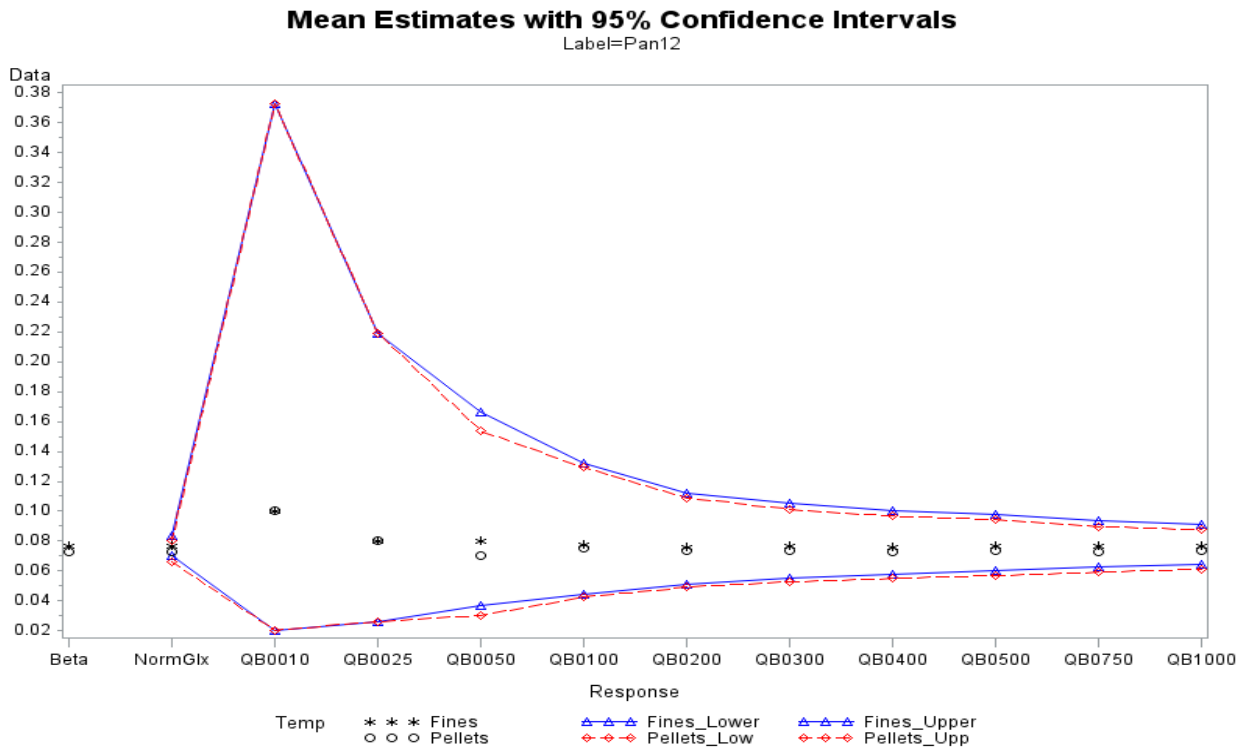


Figure 4.11 Pan 12 mean estimates with 95% confidence intervals for Experiment 2

Objective 1



The implications of response distribution are evident in the p-values for the test of difference in mean percent fat in Pellets and Fines summarized in Table 4.3. The narrow confidence intervals produced by the normal response model lead to p-values less than 0.05 for Pans 1, 2, 3, 4, 5, 6, 8, and 9. This seems unlikely based on the mean estimates, which would lead one to expect few statistically significant differences. Similar statistically significant differences are found with the beta response model, which we know to be invalid because of convergence issues. There are no statistically significant differences for Type for any of the Pans when a quasi-binomial response model is used.

Table 4.3 P-values for test of difference in mean percent fat in Fines and Pellets by Pan

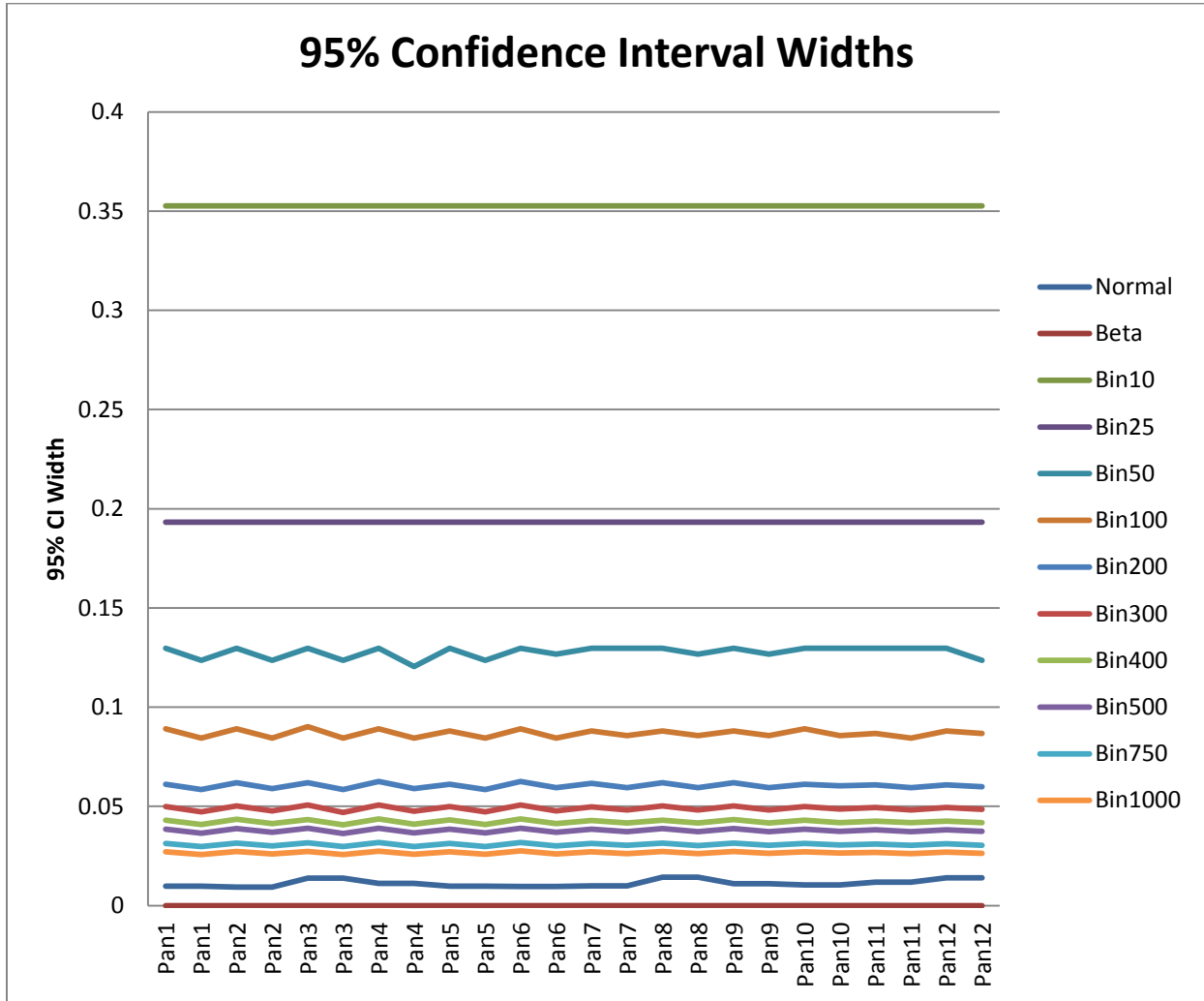
Pan	Beta +	Normal	Quasi-Bin N = 50	Quasi-Bin N = 200	Quasi-Bin N = 500	Quasi-Bin N = 1000
1	0.0036	0.0037	0.7297	0.6064	0.3796	0.2426
2	0.0004	0.0009	0.7297	0.5555	0.4091	0.2682
3	0.0409	0.0436	0.7297	0.5031	0.2775	0.1873
4	0.0041	0.0046	0.6041	0.5061	0.2975	0.1811
5	DNC	0.0222	0.7297	0.6064	0.3809	0.2431
6	0.0042	0.0046	0.8636	0.5583	0.3847	0.2169
7	DNC	0.069	1	0.6687	0.5918	0.4388
8	0.0297	0.0333	0.8636	0.6116	0.465	0.3523
9	0.0289	0.0298	0.8636	0.6116	0.4952	0.3529
10	0.1651	0.1659	1	0.8626	0.6281	0.5481
11	0.237	0.2413	1	0.7936	0.6639	0.5446
12	0.057	0.0571	0.7297	0.8616	0.7031	0.5699

+ GLMM using the Beta response did not converge for pans 5 and 7. Beta p-values are unlikely to be valid

++ For Quasi-Binomial 10 and Quasi-Binomial 25 p-values were 1 for all pans.

The differences in 95% confidence interval widths between the models can be seen in Figure 4.13. The lines for normal and beta are both approximately zero. The confidence interval widths are very similar for quasi-binomial response models with N = 200 or greater.

Figure 4.12 95% confidence interval widths for mean estimates for Experiment 2 Objective1



Experiment 2 Objective 2

The research objective for Experiment 2 Objective 2 is to determine if there is a statistically significant difference in the percent fat of the pellets in the 12 pans along the line and the percent fat of the control pan (Pan 0). The analysis for Experiment 2 Objective 2 has issues similar to Experiment 2 Objective 1. In the experimental design for Objective 2, there are 52 degrees of freedom (Table 2.3), which is substantially higher than the 8 degrees of freedom in Objective 1, but response values are still proportions close to 0. As in Objective 1 analysis, the Glimmix procedure is unable to calculate standard errors for point estimates when the beta distribution is used. The normal response produces unrealistically narrow confidence intervals

and the quasi-binomial response with high N values produces extremely wide confidence intervals. Again, using the quasi-binomial response with N greater than 100 produces consistent point estimates and useable confidence interval widths.

Figure 4.13 shows mean estimates with 95% confidence intervals for the control pan (Pan 0), Convergence issues prevent any confidence interval from being generated for the beta response and the confidence interval for the normal response is unrealistically narrow. As a result, no comparisons between Pan 0 and the other Pans along the feed line can be made for the beta response and p-values for differences in Pans for the normal response are expected to be artificially low. This is seen in Table 4.4 where p-values for differences between Pan 0 and the other 12 pans are summarized. The quasi-binomial response models with large values of N offer better results. For these models, p-values stabilize for N greater than 100 and lead to the conclusion that there are no statistically significant differences between Pan 0 and any Pan 1-12.

Figure 4.13 Pan 0 mean estimates with 95% confidence intervals for Experiment 2

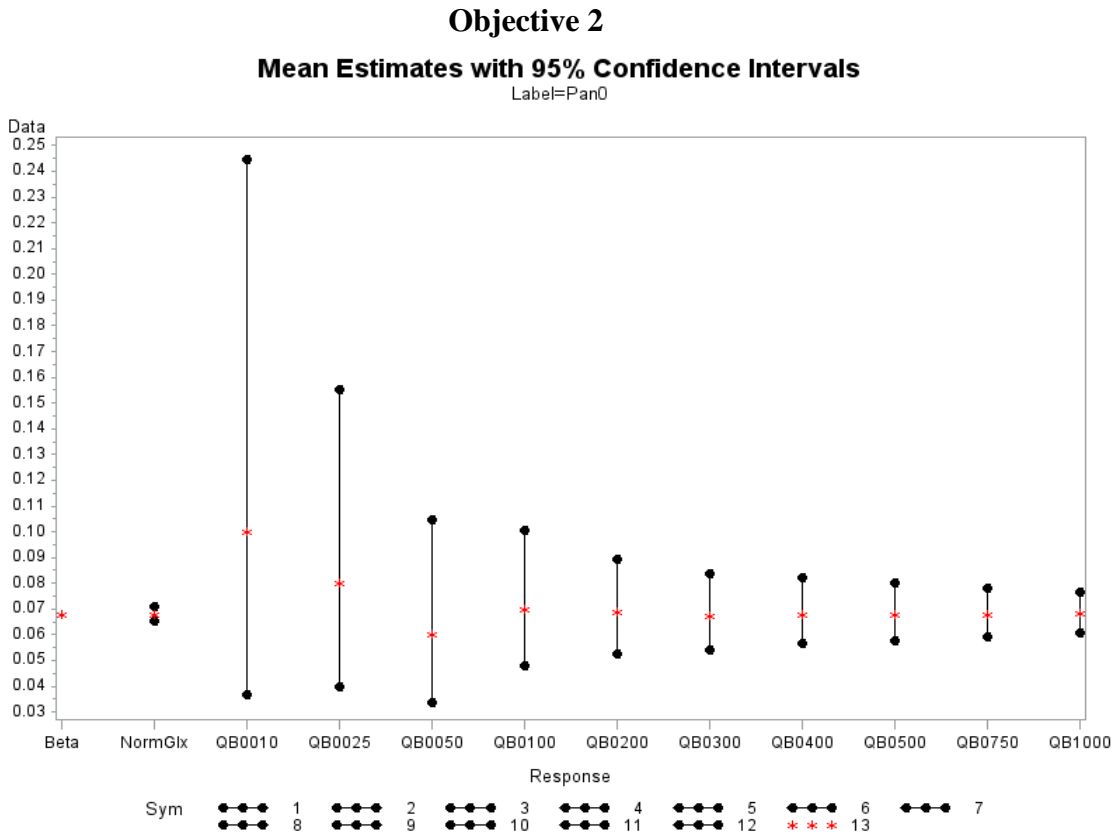


Table 4.4 P-values for test of difference in mean percent fat in Pan 0 and Pans 1-12

Pan 0 vs Pan	Beta +	Normal	Quasi-Bin N = 50	Quasi-Bin N = 200	Quasi-Bin N = 500	Quasi-Bin N = 1000
1	.	0.3467	0.6877	0.9222	0.8524	0.7928
2	.	0.0631	0.6877	0.8457	0.6665	0.6017
3	.	0.4101	0.6877	0.9222	0.9011	0.8265
4	.	0.2261	0.8376	0.8457	0.8044	0.7267
5	.	0.3207	0.6877	0.9222	0.8044	0.7595
6	.	0.0423	0.5544	0.7714	0.6665	0.5722
7	.	0.008	0.4397	0.7714	0.5408	0.4623
8	.	0.0106	0.5544	0.7714	0.5812	0.4623
9	.	0.0066	0.5544	0.7714	0.5408	0.437
10	.	0.0011	0.4397	0.6314	0.502	0.3239
11	.	0.0077	0.4397	0.7714	0.5408	0.4623
12	.	0.0036	0.6877	0.6998	0.502	0.389

+ P-values for differences are unobtainable for the beta distribution. No standard errors for estimates could be generated due to convergence issues

++ For Quasi-Binomial 10 and Quasi-Binomial 25 p-values were 1 for all pans.

The conclusion that there are no statistically significant differences between pans is verified graphically in Figure 4.14 and Figure 4.15. These figures show mean estimates and 95% confidence intervals for all models under consideration. Figure 4.14 shows Pan 0 and the first pan along the feed line (Pan 1) and Figure 4.15 shows Pan 0 and the last pan along the feed line (Pan12). In both figures, there is little difference between means for any model other than quasi-binomial responses with very low values of N and clearly overlapping confidence intervals for all models.

Figure 4.14 Pan 0 & Pan 1 mean estimates with 95% confidence intervals for Experiment 2
Objective 2

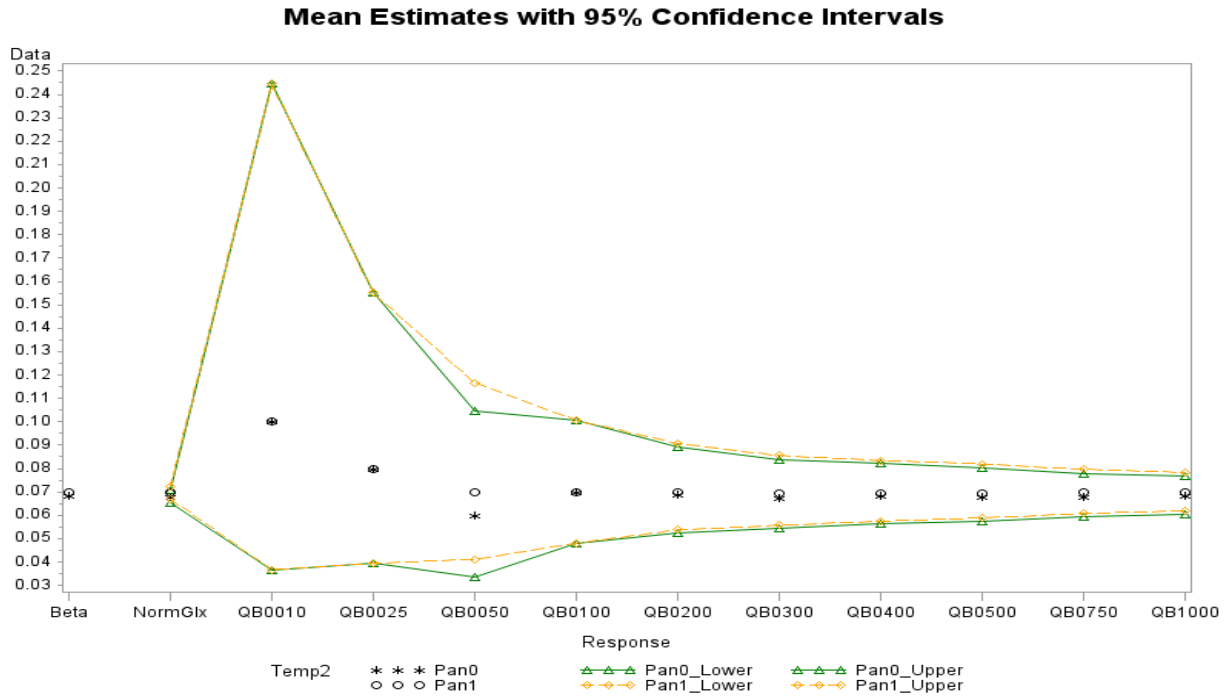
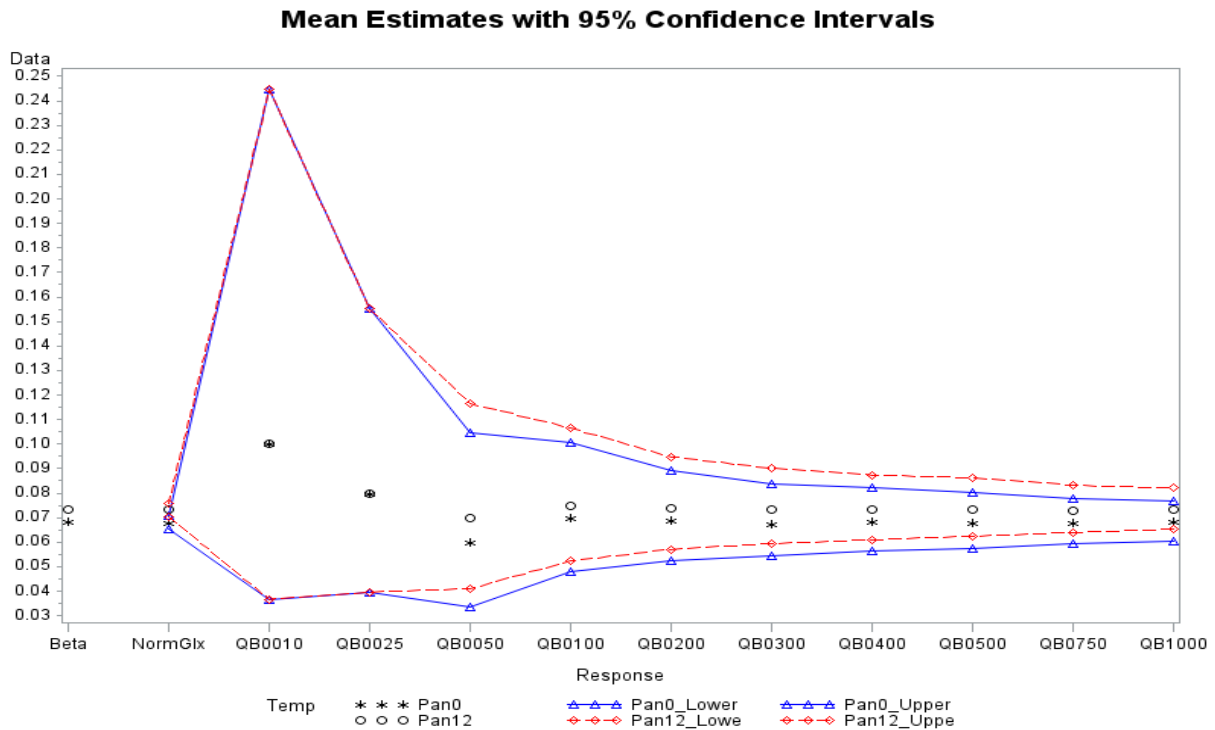
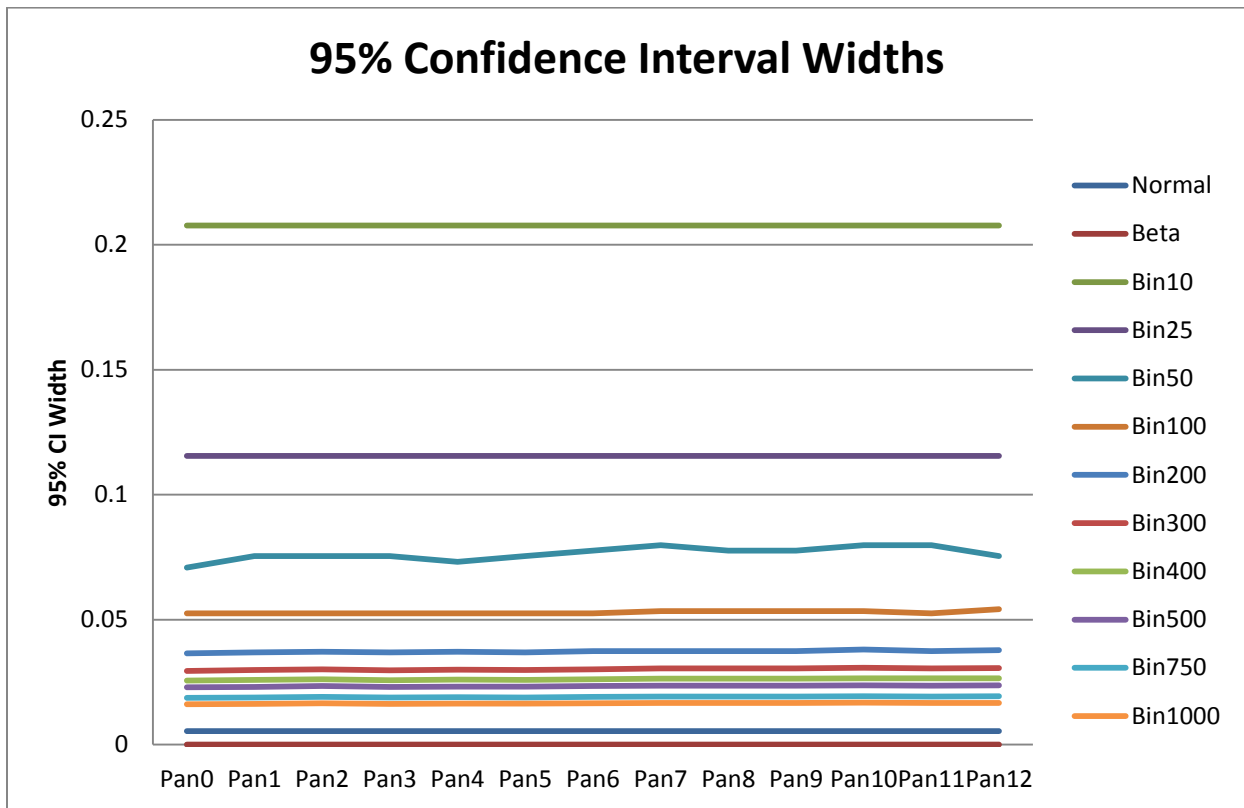


Figure 4.15 Pan 0 & Pan 12 mean estimates with 95% confidence intervals for Experiment2 Objective 2



A comparison of 95% confidence interval widths is found in Figure 4.16. The beta and normal response generate confidence intervals that are approximately zero and quasi-binomial response with N less than 100 generate confidence intervals that are extremely wide. Quasi-binomial response models with N greater than 100 have 95% confidence intervals widths of similar size.

Figure 4.16 95% confidence interval widths for mean estimates for Experiment 2 Objective2



Chapter 5 - Conclusions

The mathematically logical choice of probability distribution for response data that is a continuous proportion bounded by 0 and 1 is the beta distribution. We would like to use a GLMM with a beta response for experiments with multiple random effects and a proportion response in SAS Glimmix procedure, but characteristics of the data set or experimental design

may make this impossible. The beta response model in the Glimmix procedure is known to perform inconsistently in repeated measures analyses (Stroup 2012, p.435). The beta response model also had more lack of convergence issues than other models in this series of experiments where the response proportions are close to zero.

The issues with using a beta response model in Proc Glimmix create a need for a reasonable alternative. The traditional solutions are to use a square root arc sine transformation on the response data or to rely on the normal distribution to closely approximate the beta distribution. Transformations do not always yield feasible point estimates and, as seen in Experiment 2, the normal distribution can be a poor alternative. The results of this investigation show that a quasi-binomial response model with a high value of N (greater than 100) offers a good approximation to the beta response model that behaves well with GLMM's in SAS Glimmix procedure.

Experiment 1 is a situation where the beta response model, the normal response model, and quasi-binomial models with values of N that were not extremely low are all good options for analysis. This experiment had a relatively large sample size and had response values that were not near the boundaries of the 0, 1 interval. Mean estimates and 95% confidence intervals for quasi-binomial response models with $N = 200$ or greater very closely approximate the beta response model.

Experiment 2 is an example of a situation where the beta response model is not a feasible option due to lack of convergence. In both Objective 1 and Objective 2, response values are close to zero and sample sizes are small, this is particularly true in Objective 1 where the analysis is done by Pan. In the analyses for Experiment 2, the normal response model is a poor choice because it produces extremely narrow confidence intervals. As a result, the normal response model yields a statistically significant difference in percent fat in Pellets and Fines for many of the Pans along the feed line even when there is little difference in the estimated mean values. A high type I error rate is likely for this model. The quasi-binomial response models perform consistently and produce reasonable confidence interval widths for models with $N = 200$ or greater in Experiment 2. These models were also shown to approximate the beta response model well in Experiment 1. This evidence leads to the conclusion that a quasi-binomial response model with a high value of N is a good alternative to the beta response model when sample sizes are small and response data is a continuous proportion close to zero.

The use of the quasi-binomial model as an alternative to the normal model will be recommended to researchers involved with the motivating poultry feed line experiments (Hancock et al. 2012). The quasi-binomial model is a good option for analysis in the series of experiments considered in this case study, but more investigation is needed to determine if this alternative performs well with a greater variety of data sets. It is left to other researchers to consider this question and to perform simulation studies to determine the robustness of the conclusions found in this case study.

References

- Christensen, R. (2002), *Plane Answers to Complex Questions: The Theory of Linear Model*, 3rd Ed. Springer Science+Business Media, Inc., New York, NY.
- Cribari-Neto, F., & Zeileis, A. (2010), "Beta Regression in R," *Journal of Statistical Software*, 34, 1-24.
- Ferrari, S.L., & Cribari-Neto, F. (2004), "Beta Regression for Modeling Rates and Proportions," *Journal of Applied Statistics*, 31, 799-815.
- Gbur, E.E., Stroup, W.W., McCarter, K.S., Durham, S., Young, L.J., Christman, M., West, M., & Kramer, M. (2012), *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*, ASA, CSSA, SSSA, Madison, WI.
- Hancock, C., Beyer, R.S., Rierison, R., Rude, C.M., Graham, N., Barrios, M., & Shannon, C. (2011), "The Effects of an Auger Feed Line on Nutrient Distribution of Pelleted Feeds Made with Whole Grains," *IPSF 2011*, Atlanta, GA.
- Johnson, N.L. & Kotz, M. (1970), *Distributions in Statistics Continuous Univariate Distributions, Vol. 2*, Houghton Mifflin, Boston, MA.
- Kincaid, C. (2005), "Guidelines for Selecting the Covariance Structure in Mixed Model Analysis," *Proceedings of the Thirtieth Annual SAS Users Group International Conference*, No. 198-30, April 2005.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., & Schabenberger, O. (2006), *SAS for Mixed Models*, Ed. SAS Institute, Cary, NC.
- McCullagh, P. & Nelder, J.A. (1989), *Generalized Linear Models*. 2nd Ed. Chapman & Hall.
- Milliken, G.A., & Johnson, D.E. (2009), *Analysis of Messy Data Volume I: Designed Experiments*, 2nd Ed. CRC Press, Boca Raton, FL
- Mood, A.M., Graybill, F.A., & Boes, D.C. (1974), *Introduction to the Theory of Statistics*, 3rd Ed. McGraw-Hill, Inc.
- Schabenberger, O. (2005), "Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models." *NESUG 18*, p.1.
- Stroup, W.W. (2012), *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*, CRC Press.

Appendix A - SAS Programs

This appendix contains the SAS 9.3 programs used to obtain the reported results.

Experiment 1

The following data input is used for each analysis performed for Experiment 1.

```
data feedline;
    input pan $   feed $ run $ Fines Pellets      Total      Ratio;
    N10 = 10;
    Y10 = round(Ratio*10);
    N25 = 25;
    Y25 = round(Ratio*25);
    N50 = 50;
    Y50 = round(Ratio*50);
    N100 = 100;
    Y100 = round(Ratio*100);
    N200 = 200;
    Y200 = round(Ratio*200);
    N300 = 300;
    Y300 = round(Ratio*300);
    N400 = 400;
    Y400 = round(Ratio*400);
    N500 = 500;
    Y500 = round(Ratio*500);
    N750 = 750;
    Y750 = round(Ratio*750);
    N1000 = 1000;
    Y1000 = round(Ratio*1000);
datalines;
01  H    1    141    633    774    0.182170543
02  H    1    121    660    781    0.154929577
03  H    1    129    682    811    0.159062885
04  H    1    126    668    794    0.158690176
05  H    1    116    663    779    0.148908858
06  H    1    143    667    810    0.17654321
07  H    1    109    735    844    0.129146919
08  H    1    102    689    791    0.128950695
09  H    1    133    664    797    0.166875784
10  H    1    95     699    794    0.119647355
11  H    1    192    692    884    0.21719457
12  H    1    120    457    577    0.20797227
01  L    1    634    264    898    0.706013363
02  L    1    576    261    837    0.688172043
03  L    1    552    256    808    0.683168317
04  L    1    631    180    811    0.778051788
05  L    1    612    186    798    0.766917293
06  L    1    671    177    848    0.791273585
07  L    1    629    225    854    0.736533958
```

08	L	1	613	205	818	0.749388753
09	L	1	642	173	815	0.787730061
10	L	1	589	209	798	0.738095238
11	L	1	707	92	799	0.88485607
12	L	1	518	55	573	0.904013962
01	H	2	56	686	742	0.075471698
02	H	2	42	682	724	0.05801105
03	H	2	30	501	531	0.056497175
04	H	2	75	708	783	0.095785441
05	H	2	85	691	776	0.109536082
06	H	2	101	730	831	0.121540313
07	H	2	89	689	778	0.114395887
08	H	2	109	685	794	0.137279597
09	H	2	121	685	806	0.150124069
10	H	2	94	682	776	0.121134021
11	H	2	143	668	811	0.176325524
12	H	2	127	647	774	0.164082687
01	L	2	596	165	761	0.783180026
02	L	2	585	161	746	0.784182306
03	L	2	624	170	794	0.785894207
04	L	2	668	179	847	0.78866588
05	L	2	660	145	805	0.819875776
06	L	2	664	138	802	0.827930175
07	L	2	647	148	795	0.813836478
08	L	2	641	150	791	0.810366625
09	L	2	621	180	801	0.775280899
10	L	2	670	179	849	0.789163722
11	L	2	696	103	799	0.871088861
12	L	2	659	119	778	0.847043702
01	H	3	51	697	748	0.068181818
02	H	3	41	657	698	0.058739255
03	H	3	27	442	469	0.057569296
04	H	3	80	693	773	0.103492885
05	H	3	65	694	759	0.085638999
06	H	3	73	707	780	0.093589744
07	H	3	76	707	783	0.09706258
08	H	3	103	678	781	0.131882202
09	H	3	148	653	801	0.184769039
10	H	3	84	682	766	0.109660574
11	H	3	146	662	808	0.180693069
12	H	3	108	486	594	0.181818182
01	L	3	577	187	764	0.755235602
02	L	3	584	172	756	0.772486772
03	L	3
04	L	3	661	124	785	0.842038217
05	L	3	633	143	776	0.815721649
06	L	3	604	159	763	0.791612058
07	L	3	632	146	778	0.812339332
08	L	3	638	142	780	0.817948718
09	L	3	658	142	800	0.8225
10	L	3	747	194	941	0.793836344

11	L	3	663	152	815	0.813496933
12	L	3	665	127	792	0.839646465
01	H	4	53	693	746	0.071045576
02	H	4	41	703	744	0.055107527
03	H	4	30	545	575	0.052173913
04	H	4	82	694	776	0.105670103
05	H	4	57	698	755	0.075496689
06	H	4	65	697	762	0.085301837
07	H	4	72	697	769	0.093628088
08	H	4	86	685	771	0.11154345
09	H	4	128	659	787	0.162642948
10	H	4	54	532	586	0.092150171
11	H	4	148	659	807	0.183395291
12	H	4	120	451	571	0.210157618
01	L	4	578	177	755	0.765562914
02	L	4	592	164	756	0.783068783
03	L	4	579	178	757	0.764861295
04	L	4	641	137	778	0.823907455
05	L	4	621	151	772	0.804404145
06	L	4	703	163	866	0.811778291
07	L	4	517	96	613	0.843393148
08	L	4	581	156	737	0.788331072
09	L	4	652	165	817	0.798041616
10	L	4	662	143	805	0.822360248
11	L	4	678	131	809	0.838071693
12	L	4	739	125	864	0.855324074

;

```
proc print data=feedline;
```

```
run;
```

LMM analysis with residual analysis

```
title 'Normal Response Proc MIXED';
proc glimmix data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/cl;
output resid student pred out=means;
run;
```

```
proc print data=means;
```

```
run;
```

```
proc univariate data=means normal;
```

```

var resid;
qqplot resid/ normal (mu=0 sigma=0.02172);
histogram resid/ normal;
run;

```

```

proc univariate data=means normal;
var student;
qqplot student/ normal (mu=0 sigma=0.02172);
histogram student/ normal;
run;

```

Variance Component Analysis

```

title 'Mean Plots for Ratio';
proc glimmix data=feedline;
class feed run pan;
model ratio=feed|pan;
random intercept/ subject=run(feed);
lsmeans feed*pan/
    plot=meanplot(sliceby=feed join);
output out=predicted_ratio pred=p resid=r;
run;

```

```

title 'Plot of Covariance as a Function of Distance';
proc mixed data=feedline;
class feed run pan;
model Ratio=feed|pan;
repeated/type=un sscp subject=run(feed);
ods output covparms=cov;

```

```

data times;
do time1=1 to 12;
do time2=1 to time1;
dist=time1-time2;
output;
end;
end;

```

```

data covplot;
merge times cov;

```

```

symbol color=black interpol=join line=1 v=circle;
symbol2 color=blue interpol=join line=1 v=square;
symbol3 color=green interpol=join line=1 v=dot;
symbol4 color=red interpol=join line=1 v=star;
symbol5 color=yellow interpol=join line=1 v=plus;
symbol6 color=orange interpol=join line=1 v=diamond;

```

```

symbol7 color=purple interpol=join line=1 v=triangle;
symbol8 color=brown interpol=join line=1 v=club;
symbol9 color=black interpol=join line=2 v=spade;
symbol10 color=red interpol=join line=2 v=heart;
symbol11 color=blue interpol=join line=2 v=star;
symbol12 color=green interpol=join line=2 v=diamond;

proc gplot data=covplot;
  plot estimate*dist=time2;
run;

title 'Normal Response Proc MIXED with Repeated Pan default VC';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) rcorr;
run;

title 'Normal Response Proc MIXED with Repeated Pan Compound Symmetry';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) type=CS rcorr;
run;

title 'Normal Response Proc MIXED with Repeated Pan AR(1)';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) type=AR(1) rcorr;
run;

title 'Normal Response Proc MIXED with Repeated Pan Toeplitz';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) type=toep rcorr;
run;

title 'Normal Response Proc MIXED with Repeated Pan ARH(1)';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) type=ARH(1) rcorr;
run;

title 'Normal Response Proc MIXED with Repeated Pan ANTE(1)';
proc mixed data=feedline;
class pan feed run;
model ratio = feed pan pan*feed;
repeated pan/subject=run(feed) type=ANTE(1) rcorr;

```

```
run;
```

GLMM Analysis

```
title 'Normal Response Proc MIXED';  
proc mixed data=feedline;  
class pan feed run;  
model ratio = feed pan pan*feed;  
random intercept/subject=run(feed) type=toep;  
lsmeans feed|pan/cl;  
run;
```

```
title 'Normal Response Proc GLIMMIX';  
proc glimmix data=feedline;  
class pan feed run;  
model ratio = feed pan pan*feed;  
random intercept/subject=run(feed) type=toep;  
lsmeans feed|pan/cl;  
run;
```

```
title 'Beta Response';  
proc glimmix data=feedline method=laplace;  
class pan feed run;  
model ratio = feed pan pan*feed/dist=beta;  
random intercept/subject=run(feed) type=toep;  
lsmeans feed|pan/ilink cl;  
run;
```

```
title 'Quasi-Binomial Response N=10';  
proc glimmix data=feedline method=laplace;  
class pan feed run;  
model Y10/N10 = feed pan pan*feed;  
random intercept/subject=run(feed) type=toep;  
lsmeans feed|pan/ilink cl;  
run;
```

```
title 'Quasi-Binomial Response N=25';  
proc glimmix data=feedline method=laplace;  
class pan feed run;  
model Y25/N25 = feed pan pan*feed;  
random intercept/subject=run(feed) type=toep;  
lsmeans feed|pan/ilink cl;  
run;
```

```

title 'Quasi-Binomial Response N=50';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y50/N50 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=100';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y100/N100 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=200';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y200/N200 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=300';
proc glimmix data=feedline method=laplace ;
class pan feed run;
model Y300/N300 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=400';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y400/N400 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=500';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y500/N500 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;

```

```

lsmeans feed|pan/ilink cl;
run;

title 'Quasi-Binomial Response N=750';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y750/N750 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

```

title 'Quasi-Binomial Response N=1000';
proc glimmix data=feedline method=laplace;
class pan feed run;
model Y1000/N1000 = feed pan pan*feed;
random intercept/subject=run(feed) type=toep;
lsmeans feed|pan/ilink cl;
run;

```

Experiment 2

The following data input is used for each analysis performed for Experiment 2.

```

data PercentFat;
input Sample $ DM BeakerNumb $ EmptyBkrWt SampleWt ExtBkrWt PercFat count
Type $;

if mod (_N_,4)=1 then run=1;
if mod (_N_+1,4)=1 then run=4;
if mod (_N_+2,4)=1 then run=3;
if mod (_N_+3,4)=1 then run=2;

if count>0 and count<5 then pan=0;
if count>4 and count<9 then pan=01;
if count>8 and count<13 then pan=02;
if count>12 and count<17 then pan=03;
if count>16 and count<21 then pan=04;
if count>20 and count<25 then pan=05;
if count>24 and count<29 then pan=06;
if count>28 and count<33 then pan=07;
if count>32 and count<37 then pan=08;
if count>36 and count<41 then pan=09;
if count>40 and count<45 then pan=10;
if count>44 and count<49 then pan=11;
if count>48 and count<53 then pan=12;
if count>52 and count<57 then pan=01;
if count>56 and count<61 then pan=02;

```



```

if count>60 and count<65 then pan=03;
if count>64 and count<69 then pan=04;
if count>68 and count<73 then pan=05;
if count>72 and count<77 then pan=06;
if count>76 and count<81 then pan=07;
if count>80 and count<85 then pan=08;
if count>84 and count<89 then pan=09;
if count>88 and count<93 then pan=10;
if count>92 and count<97 then pan=11;
if count>96 and count<101 then pan=12;

```

```
drop count DM;
```

```
datalines;
```

CONTROL	100	15	62.2027	2.0058	62.3387	6.78	1	C
CONTROL	100	17	63.1907	2.0057	63.3273	6.81	2	C
CONTROL	100	18	63.0344	2.0017	63.1742	6.98	3	C
CONTROL	100	9B	66.8418	2.0025	66.9753	6.666666667	4	C
1-1-P	100	9B	67.008	2.0082	67.1439	6.767254258	5	P
2-1-P	100	10	60.9826	2.0071	61.1295	7.319017488	6	P
3-1-P	100	14B	63.0365	2.0026	63.1705	6.691301308	7	P
4-1-P	100	3	65.0005	2.0135	65.143	7.077228706	8	P
1-2-P	100	6B	61.7599	2.0197	61.9004	6.956478685	9	P
2-2-P	100	6	62.2958	2.0076	62.4453	7.44670253	10	P
3-2-P	100	2B	65.9223	2.0153	66.0625	6.956780628	11	P
4-2-P	100	11	61.558	2.007	61.7008	7.12	12	P
1-3-P	100	12B	64.8981	2.018	65.0357	6.818632309	13	P
2-3-P	100	11	62.0479	2.0061	62.1898	7.073426051	14	P
3-3-P	100	10B	67.975	2.0057	68.1113	6.795632448	15	P
4-3-P	100	18B	63.5884	2.0115	63.731	7.089236888	16	P
1-4-P	100	1	61.9094	2.006	62.0456	6.789631107	17	P
2-4-P	100	8B	62.9336	2.0078	63.0835	7.465883056	18	P
3-4-P	100	7B	64.24	2.0076	64.379	6.923689978	19	P
4-4-P	100	11B	62.4212	2.0056	62.5587	6.85580375	20	P
1-5-P	100	16	62.4971	2.0024	62.6384	7.06	21	P
2-5-P	100	6B	61.0583	2.0128	61.1984	6.9604531	22	P
3-5-P	100	3B	60.2876	2.0027	60.4301	7.115394218	23	P
4-5-P	100	4B	61.7532	2.0019	61.8884	6.753584095	24	P
1-6-P	100	13	64.9092	2.0035	65.0517	7.11	25	P
2-6-P	100	3B	64.9107	2.0146	65.0576	7.291770078	26	P
3-6-P	100	4B	63.3905	2.0009	63.5384	7.391673747	27	P
4-6-P	100	1B	62.8896	2.002	63.0259	6.808191808	28	P
1-7-P	100	14	67.6607	2.0014	67.8033	7.13	29	P
2-7-P	100	17B	67.6618	2.0093	67.808	7.276165829	30	P
3-7-P	100	16B	62.0979	2.0107	62.2497	7.549609589	31	P
4-7-P	100	16B	61.8557	2.0025	61.9979	7.101123596	32	P
1-8-P	100	2	62.5727	2.0031	62.7127	6.989166791	33	P
2-8-P	100	15B	63.3307	2.0125	63.4869	7.761490683	34	P
3-8-P	100	5B	61.4147	2.0058	61.5592	7.204108087	35	P

4-8-P	100	3	62.4056	2.0246	62.5479	7.03	36	P
1-9-P	100	4	62.058	2.0133	62.1988	6.99349327	37	P
2-9-P	100	15B	62.4992	2.0056	62.6539	7.713402473	38	P
3-9-P	100	11B	64.8469	2.0057	64.9937	7.31914045	39	P
4-9-P	100	10B	62.2041	2.0057	62.3461	7.079822506	40	P
1-10-P*	100	13	64.3848	2.0088	64.5429	7.87037037	41	P
2-10-P	100	7B	63.5169	2.0216	63.6636	7.256628413	42	P
3-10-P	100	14B	62.4149	2.0173	62.5626	7.321667575	43	P
4-10-P	100	13B	61.9529	2.0046	62.095	7.088695999	44	P
1-11-P	100	8B	61.8326	2.0086	61.9812	7.398187792	45	P
2-11-P	100	17	64.9196	2.0148	65.0652	7.226523724	46	P
3-11-P	100	13B	61.5954	2.0083	61.7421	7.304685555	47	P
4-11-P	100	8	62.4826	2.0116	62.6262	7.138596142	48	P
1-12-P	100	9	60.9319	2.0098	61.0714	6.940989153	49	P
2-12-P	100	18	65.675	2.004	65.8306	7.764471058	50	P
3-12-P*	100	12B	65.4949	2.0077	65.6471	7.580813867	51	P
4-12-P	100	5B	61.5591	2.007	61.699	6.97060289	52	P
1-1-F	100	7	61.055	2.0036	61.2066	7.57	53	F
2-1-F	100	7B	66.532	2.0025	66.698	8.289637953	54	F
3-1-F	100	18B	61.3026	2.0104	61.4583	7.744727417	55	F
4-1-F	100	18	64.8443	2.0914	65.0046	7.664722196	56	F
1-2-F	100	8B	64.9148	2.0087	65.0681	7.631801663	57	F
2-2-F	100	1	63.5125	2.0189	63.6824	8.42	58	F
3-2-F	100	2	62.4143	2.0075	62.5702	7.77	59	F
4-2-F	100	15	62.2948	2.0074	62.4541	7.935638139	60	F
1-3-F	100	5	62.8889	2.08	63.0459	7.548076923	61	F
2-3-F	100	2B	64.7879	2.0005	64.9466	7.933016746	62	F
3-3-F	100	10	61.8529	2.0064	62.0303	8.84	63	F
4-3-F	100	15	62.5271	2.0478	62.6829	7.60816486	64	F
1-4-F	100	7	67.0064	2.0648	67.1629	7.579426579	65	F
2-4-F	100	17B	62.2901	2.0225	62.46	8.400494438	66	F
3-4-F	100	11B	62.0581	2.0072	62.2257	8.349940215	67	F
4-4-F	100	14	61.3048	2.0077	61.4632	7.889624944	68	F
1-5-F	100	11	62.4136	2.0383	62.564	7.378697935	69	F
2-5-F	100	14B	65.0005	2.0222	65.1665	8.208881416	70	F
3-5-F	100	6	66.8395	2.0026	67.003	8.16	71	F
4-5-F	100	14	61.7602	2.0015	61.9128	7.624281789	72	F
1-6-F	100	8	67.9721	2.0863	68.1324	7.683458755	73	F
2-6-F	100	16B	65.6713	2.0298	65.8416	8.389989161	74	F
3-6-F	100	9	61.7504	2.0073	61.9168	8.29	75	F
4-6-F	100	10	63.3893	2.0799	63.5527	7.85614693	76	F
1-7-F	100	9B	60.9295	2.0045	61.0814	7.577949613	77	F
2-7-F	100	5	62.9305	2.0072	63.0975	8.32	78	F
3-7-F	100	10B	62.2925	2.0188	62.4494	7.771943729	79	F
4-7-F	100	17	61.594	2.0271	61.7446	7.429332544	80	F
1-8-F	100	13B	62.0405	2.0062	62.1873	7.317316319	81	F
2-8-F	100	5B	62.4785	2.0012	62.6452	8.330001999	82	F
3-8-F	100	12	63.5871	2.0027	63.7546	8.36	83	F
4-8-F	100	16	60.5294	2.003	60.6818	7.608587119	84	F

1-9-F	100	4B	62.5718	2.0031	62.7203	7.413509061	85	F
2-9-F	100	12B	60.9797	2.0014	61.1403	8.024382932	86	F
3-9-F	100	8	61.9501	2.0064	62.1166	8.30 87	F	
4-9-F	100	12	64.795	2.0008	64.9543	7.961815274	88	F
1-10-F	100	15B	61.9035	2.0161	62.0549	7.509548137	89	F
2-10-F	100	6B	60.527	2.0057	60.6813	7.693074737	90	F
3-10-F	100	13	61.4145	2.0896	61.5868	8.245597243	91	F
4-10-F	100	7	66.5353	2.0022	66.6909	7.771451403	92	F
1-11-F	100	9	65.9217	2.0318	66.0682	7.21035535	93	F
2-11-F	100	3B	65.683	2.0129	65.8305	7.327736102	94	F
3-11-F	100	12	64.8953	2.0367	65.0654	8.351745471	95	F
4-11-F	100	16	63.3307	2.0181	63.4863	7.710222486	96	F
1-12-F	100	1B	64.3788	2.0008	64.5204	7.077169132	97	F
2-12-F	100	4	65.4907	2.0065	65.6504	7.96 98	F	
3-12-F	100	6	60.2871	2.0828	60.456	8.109275975	99	F
4-12-F	100	5	65.6849	2.0049	65.8377	7.621327747	100	F

;

```
proc sort data=PercentFat;
by run;
run;
```

```
proc print data=PercentFat;
run;
```

Objective 1 LMM analysis with residual analysis

```
data FinesVPellets;
    set PercentFat;
    if Type eq 'C' then delete;
run;

proc print data=FinesVPellets;
run;

title 'Normal Objective 1 by PAN';
proc glimmix data=FinesVPellets;
class run Type;
model percfat = Type;
random intercept/ subject=run;
lsmeans Type/cl;
output resid student pred out=OBJ1;
run;

proc print data=OBJ1;
run;

proc univariate data=OBJ1 normal;
```

```

var resid;
qqplot resid/ normal (mu=0 sigma=0.272);
histogram resid/ normal;
run;

proc univariate data=OBJ1 normal;
var student;
qqplot student/ normal (mu=0 sigma=0.272);
histogram student/ normal;
run;

```

Objective 2 LMM analysis with residual analysis

```

data PelletsVControl;
    set PercentFat;
if Type eq 'F' then delete;
run;

proc print data=PelletsVControl;
run;

title 'Normal Objective 2 Control';
proc glimmix data=PelletsVControl;
class run pan;
model percfat = pan;
random intercept/ subject=run;
output resid student pred out=OBJ2;
run;

proc print data=OBJ2;
run;

proc univariate data=OBJ2 normal;
var resid;
qqplot resid/ normal (mu=0 sigma=0.1939);
histogram resid/normal;
run;

proc univariate data=OBJ2 normal;
var student;
qqplot student/ normal (mu=0 sigma=0.1939);
histogram student/normal;
run;

```

Objective 2 Variance Component Analysis

```
data PelletsVControl;
    set PercentFat;
if Type eq 'F' then delete;
run;

proc print data=PelletsVControl;
run;

title 'Mean Plots for Ratio';
proc glimmix data=PelletsVControl;
class run pan;
model percfat=pan;
random intercept/ subject=run;
lsmeans pan/
    plot=meanplot(join);
output out=predicted_ratio pred=p resid=r;
run;

title 'Plot of Covariance as a Function of Distance';
proc mixed data=PelletsVControl;
class run pan;
model percfat=pan;
repeated/type=un sscp subject=run;
ods output covparms=cov;

data times;
do time1=1 to 13;
do time2=1 to time1;
dist=time1-time2;
output;
end;
end;

data covplot;
merge times cov;

symbol color=black interpol=join line=1 v=circle;
symbol2 color=blue interpol=join line=1 v=square;
symbol3 color=green interpol=join line=1 v=dot;
symbol4 color=red interpol=join line=1 v=star;
symbol5 color=yellow interpol=join line=1 v=plus;
symbol6 color=orange interpol=join line=1 v=diamond;
symbol7 color=purple interpol=join line=1 v=triangle;
symbol8 color=brown interpol=join line=1 v=club;
symbol9 color=black interpol=join line=2 v=spade;
symbol10 color=red interpol=join line=2 v=heart;
symbol11 color=blue interpol=join line=2 v=star;
symbol12 color=green interpol=join line=2 v=diamond;
```

```
symbol13 color=purple interpol=join line=2 v=u;
```

```
proc gplot data=covplot;  
  plot estimate*dist=time2;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=VC r rcorr;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=CS r rcorr;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=AR(1) r rcorr;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=Toep r rcorr;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=ARH(1) r rcorr;  
run;
```

```
proc mixed data=PelletsVControl;  
class run pan;  
model percfat = pan;  
repeated pan/subject=run type=ANTE(1) r rcorr;  
run;
```

Objective 1 GLMM analysis

```
data FinesVPellets;
    set PercentFat;
if Type eq 'C' then delete;
run;

proc print data=FinesVPellets;
run;

proc sort data=FinesVPellets;
by pan;
run;

Title 'Proc Mixed Normal Exp2 Obj1';
proc mixed data=FinesVPellets;
by pan;
class run Type;
model PercFat = Type;
random intercept/ subject=run;
lsmeans Type/ cl pdiff;
run;

Title 'Proc Glimmix VC Normal Exp2 Obj1';
proc glimmix data=FinesVPellets;
by pan;
class run Type;
model PercFat = Type;
random intercept/ subject=run type=VC;
lsmeans Type/cl pdiff;
run;

Title 'Beta VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model PercFat = Type/dist=beta;
random intercept/ subject=run type=VC;
lsmeans Type/ilink cl pdiff;
run;

Title 'QB10 VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model Y10/N10 = Type;
random intercept/ subject=run type=VC;
```

```
lsmeans Type/ilink cl pdiff;  
run;
```

```
Title 'QB25 VC Laplace Exp2 Obj1';  
proc glimmix data=FinesVPellets method=laplace;  
by pan;  
class run Type;  
model Y25/N25 = Type;  
random intercept/ subject=run type=VC;  
lsmeans Type/ilink cl pdiff;  
run;
```

```
Title 'QB50 VC Laplace Exp2 Obj1';  
proc glimmix data=FinesVPellets method=laplace;  
by pan;  
class run Type;  
model Y50/N50 = Type;  
random intercept/ subject=run type=VC;  
lsmeans Type/ilink cl pdiff;  
run;
```

```
Title 'QB100 VC Laplace Exp2 Obj1';  
proc glimmix data=FinesVPellets method=laplace;  
by pan;  
class run Type;  
model Y100/N100 = Type;  
random intercept/ subject=run type=VC;  
lsmeans Type/ilink cl pdiff;  
run;
```

```
Title 'QB200 VC Laplace Exp2 Obj1';  
proc glimmix data=FinesVPellets method=laplace;  
by pan;  
class run Type;  
model Y200/N200 = Type;  
random intercept/ subject=run type=VC;  
lsmeans Type/ilink cl pdiff;  
run;
```

```
Title 'QB300 VC Laplace Exp2 Obj1';  
proc glimmix data=FinesVPellets method=laplace;  
by pan;  
class run Type;  
model Y300/N300 = Type;  
random intercept/ subject=run type=VC;  
lsmeans Type/ilink cl pdiff;
```



```

run;

Title 'QB400 VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model Y400/N400 = Type;
random intercept/ subject=run type=VC;
lsmeans Type/ilink cl pdiff;
run;

Title 'QB500 VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model Y500/N500 = Type;
random intercept/ subject=run type=VC;
lsmeans Type/ilink cl pdiff;
run;

Title 'QB750 VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model Y750/N750 = Type;
random intercept/ subject=run type=VC;
lsmeans Type/ilink cl pdiff;
run;

Title 'QB1000 VC Laplace Exp2 Obj1';
proc glimmix data=FinesVPellets method=laplace;
by pan;
class run Type;
model Y1000/N1000 = Type;
random intercept/ subject=run type=VC;
lsmeans Type/ilink cl pdiff;
run;

```

Objective 2 GLMM analysis

```

data PelletsVControl;
    set PercentFat;
if Type eq 'F' then delete;

```

```

run;

proc print data=PelletsVControl;
run;

Title 'Proc Mixed Normal Exp2 Obj2';
proc mixed data=PelletsVControl;
class run pan;
model PercFat = pan;
random intercept/ subject=run type=CS;
lsmeans pan/cl pdiff=control adjust=dunnett;
run;

Title 'Proc Glimmix CS Normal Exp2 Obj2';
proc glimmix data=PelletsVControl;
class run pan;
model PercFat = pan;
random intercept/ subject=run type=CS;
lsmeans pan/cl pdiff=control adjust=dunnett;
run;

Title 'Beta CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model PercFat = pan/dist=beta;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;

Title 'QB10 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y10/N10 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;

Title 'QB25 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y25/N25 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;

```

```
Title 'QB50 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y50/N50 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;
```

```
Title 'QB100 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y100/N100 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;
```

```
Title 'QB200 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y200/N200 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;
```

```
Title 'QB300 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y300/N300 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;
```

```
Title 'QB400 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y400/N400 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
run;
```

```
Title 'QB500 CS Laplace Exp2 Obj2';
proc glimmix data=PelletsVControl method=laplace;
class run pan;
model Y500/N500 = pan;
random intercept/ subject=run type=CS;
lsmeans pan/ilink cl pdiff=control adjust=dunnett;
```

```
run;
```

```
Title 'QB750 CS Laplace Exp2 Obj2';  
proc glimmix data=PelletsVControl method=laplace;  
class run pan;  
model Y750/N750 = pan;  
random intercept/ subject=run type=CS;  
lsmeans pan/ilink cl pdiff=control adjust=dunnett;  
run;
```

```
Title 'QB1000 CS Laplace Exp2 Obj2';  
proc glimmix data=PelletsVControl method=laplace;  
class run pan;  
model Y1000/N1000 = pan;  
random intercept/ subject=run type=CS;  
lsmeans pan/ilink cl pdiff=control adjust=dunnett;  
run;
```