Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Yixuan Zhao Date

The Analysis of 2x2 Crossover Design with Repeated Baseline Measurement within Period

By

Yixuan Zhao

Master of Science in Public Health

Department of Biostatistics and Bioinformatics

_____

Renee Moore, PhD

Thesis Advisor

_____

Christina Mehta, PhD

Reader

The Analysis of 2x2 Crossover Design with Repeated Baseline Measurement within Period

By

Yixuan Zhao

Bachelor of Medicine

Peking University

2017

Thesis Committee Chair: Renee Moore, PhD

Reader: Christina Mehta, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Biostatistics and Bioinformatics
2019

Abstract

The Analysis of 2x2 Crossover Design with Repeated Baseline Measurement within Period

By Yixuan Zhao

Background: Neuroimaging study suggests precuneus is related to visuospatial imagery. Evidence suggests, the precuneus of Neanderthal is less developed than the ancestor of modern man and Neanderthal is not able to hunt prey by throwing from long distance away; while the ancestor of modern man has a more developed precuneus, and is known for throwing long-distance weapon in hunting. Thus, it is hypothesized that, the developed precuneus, provided the ancestor of modern man with the improved visuospatial skills needed in hunting prey with projectiles from a distance.

Objective: Identify that precuneus is statistical significant related to visuospatial ability, which is reflected by throw accuracy. Then the result of this study serves as evidence for the proposed hypothesis.

Methods: 2x2 crossover design was applied. Precuneus's function was oppressed by TMS treatment, while Sham treatment serves as a placebo. Throw score was chose as outcome to reflect visuospatial skill. Baseline measurement was implemented at each period before delivery of treatment. In analysis, four methods of handling baseline, modeling mean score or individual score, and various covariance structures were applied to the data. Analysis was done by applying linear mixed model.

Result: The model which handling baseline with "change from baseline" method, modeling individual change score, and assuming equal variance, cross-visit correlation is zero and within visit correlation is ρ, was chosen. The estimated difference among the effect of Sham and the effect of TMS on throw score is 0.1715, p=0.3548. After investigation into subject characteristics and experiment characteristics, temperature was adjusted in model, the estimated difference is -0.0258, p=0.8899. There is no statistical significant difference among the effects of TMS and Sham.

Conclusion: This study did not prove that precuneus function is related to visuospatial ability of human. Future study design may consider the following: 1. Indoor experiment controlling for temperature; 2. Apply randomized parallel design instead of crossover design; 3. May consider uniform MRI intensity in treatments, rather than subject specific MRI intensity.

Keywords: Crossover design, Baseline, Repeated measurement within period

The Analysis of 2x2 Crossover Design with Repeated Baseline Measurement within Period

By

Yixuan Zhao

Bachelor of Medicine

Peking University

2017

Thesis Committee Chair: Renee Moore, PhD

Reader: Christina Mehta, PhD

Investigators: James K. Rilling PhD, Charles M. Epstein MD, Lynnet Richey

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Biostatistics and Bioinformatics
2019

Acknowledgement:

I would first like to thank my thesis advisor Dr. Renee Moore of Rollins School of Public Health at Emory University. She gave me the opportunity to participate in BCC projects and gain real analysis experience. During the drafting of my thesis, her instruction and edition made the logic presenting of this thesis much more improved, and from her I also learned a lot on statistical reporting. I would also like to thank my thesis reader Dr. Christina Mehta for her kindly giving instructions on methodology part of joint modeling, and her recommendation for the re-arrangement of this thesis is enlightening. Finally, I want to express my gratitude to BIOS department as comparing to the time I graduated from Peking University, my statistical theory, programming skill and writing ability are all improved a lot through the lecturing and training here.

Table of Contents

Introduction

Crossover design is characterized by assigning subjects several treatments and randomizing subjects to receive the assigned treatments in different sequences. This design allows effects of different treatments to be compared on the same subject during different treatment periods[1]. Its key difference from parallel design is the effects of treatments are compared primarily based on within subject information. A 2x2 crossover design is the simplest one among this class of designs.

The basic concepts in crossover design are “treatment”, “sequence” and “period”. In a 2x2 crossover design, where the 2x2 represents two treatments and two periods. Annotating the two treatments to be compared as “treatment A” and “treatment B”. Subjects are randomized in 1:1 ratio to receive treatments either in arrangement AB or arrangement BA, and thus there are two sequences of receiving treatments, sequence AB (treatment A applied first then treatment B) and sequence BA (treatment B applied first then treatment A)[2]. As each treatment is applied only once, the subject will undergo two treatment periods. “Period” can be regarded as cycle of receiving the treatment and measuring the outcome.

Other key issues in crossover design are “carryover effect”, “wash out period”, and “baseline”. Carryover (or residual effect) is defined as the effect of the treatment from the previous time period on the response at the current time period. Carryover effect biases the

estimation of treatment effects, which makes it a major drawback of crossover designs[3]. Washout period is defined as the time between treatment periods. Washout period aims to diminish or eliminate the impact of carryover effects, and to allow the subjects to return to initial status. Baseline is the measurement of the interested outcome prior to the delivery of the treatment, and is often referred to as pre-treatment measures[1].

The implementation of baseline in crossover design is widely discussed, but opinions toward its role are different. It serves a method to deal with the issues regards to carryover and/or treatment by period interaction[4]. Devan V. Mehrotra[2] summarized 10 methods to handle baseline in 2x2 crossover design with one baseline and one post-treatment assessment per period, while those 10 methods can be traced back to the 4 methods summarized by Kenward et al.[5]. "Yuanyuan Liang and Keumhee Chough Carriere 2010"[6] argue baseline, either as covariate or model the changes in crossover design can achieve a better efficiency in estimation of treatment effects. However, "Yan et al. 2013"[7] favors ignore baseline and argues the implement of baseline as covariate may be harmful. Stephen Senn[11] favors "adjust baseline as covariate" over change from baseline, and is against joint modeling, insisting that baselines shouldn't be modeled as dependent variable.

The data for this thesis is a 2x2 crossover design with 20 baseline assessments and 20 post-treatment assessments per period, "Functional Basis for Precuneus Expansion". This type of design is rare and among the 33 real-data analysis of 2x2 crossover design with baseline

published in 2014-2019, none has multiple baseline measures per period. The analysis of similar data is mentioned in " Xun Chen, Zhaoling Meng and Ji Zhang 2012"[7], in which joint modeling is recommended from simulation result. Xun Chen refers to crossover design which measure baseline multiple times within each period as "crossover design with repeated baseline measurement within period", which is adapted in this thesis.

Thus, to reach a better efficiency and accuracy in estimating the difference between treatment effects in the study "Functional Basis for Precuneus Expansion", where the existence of carryover effect is not sure and confounding may exist. Four methods summarized by Kenward in 2010[5] are discussed: joint modeling, ignore baseline, change from baseline, and adjust baseline as covariate. The result of averaging the data to bring it back to simple case or not will also be compared. From the results of the four methods, this thesis will make a recommendation for the preferred method to use in the study, "Functional Basis for Precuneus Expansion".

Material and Method

1. Study Design

The data to be analyzed is from the study "Functional Basis for Precuneus Expansion," which is a 2x2 crossover design. For the study, subjects went through two interventions, transcranial magnetic stimulation (TMS) and Sham TMS, at two visits. The subjects were randomized to receive the two treatments in different sequences. During each visit the patient received one

treatment and throw scores were measured before and after treatment; for this study, each visit is a “period.” The aim of the study is to prove precuneus is related to visuospatial ability, reflected by throw accuracy. Then it serves as evidence to support the hypothesis that the developed precuneus, provided the ancestor of modern man with the improved visuospatial skills needed in hunting prey with projectiles from a distance. When the subjects underwent TMS, their precuneus function was oppressed. Sham serves as a placebo.

The original protocol plan was to enroll 30 subjects, but only 25 were actually enrolled. 25 right-handed male subjects aged between 18-40, who had baseball experience at the high school level or higher were randomized to receive treatments either in sequence TMS-SHAM or SHAM-TMS. 10 were randomized to the first sequence and 15 to the other. The subjects paid two visits to the experiment site in total. The planed washout was 1 week, but in fact the subjects did not return after a uniform schedule. The target for measuring throw accuracy was 10m away from the subjects. At the beginning of each visit, 10 warm-up throws were conducted. Then the subjects conducted 20 baseline throws. After the baseline throws, the assigned treatment was delivered and the subjects completed 20 post-treatment throws during the window period of the treatment. The experiment procedure is shown as:

Experiment procedure									
	Visit 1					Visit 2			
Sequence 1	Warm up	20 baseline throws	Deliver TMS	Window period 20 post-treatment throws	Wash out	Warm up	20 baseline throws	Deliver Sham	Window Period 20 Post-treatment throws
Sequence 2	Warm up	20 baseline throws	Deliver Sham	Window Period 20 Post-treatment throws	Wash out	Warm up	20 baseline throws	Deliver TMS	Window Period 20 Post-treatment throws

2. Data Collection:

(1) Throw score:

The throws were scored between 0 and 10, with 0.5 as smallest interval. Warm-up throws at the beginning of each visit were not recorded.

(2) Subject characteristics:

The characteristics of the subject collected by the investigators were height, weight, age, baseball level, race, and the year since the last time the subject played baseball. Based on the year since the subject last played baseball, the subjects were grouped as experienced or not.

(3) Experimental characteristics:

The experimental characteristics collected by the investigator for each treatment period were temperature, stimulation intensity, the time from receiving treatment to completing the last throwing, and whether all the post-treatment throws were completed within the window period of the treatment.

For stimulation intensity, TMS intervention had a stimulation intensity 90% of the subject's own resting motor threshold and the Sham intervention had a stimulation intensity 65% of the subject's own resting motor threshold. The stimulation intensity received by each subject at each treatment were recorded.

3. Statistical Aim

Statistically, the study compares the effect on throwing score of the two treatments, TMS, (referred as TMS) and Sham TMS (referred as Sham). Sham can be regarded as a placebo. The negative effect of TMS on throwing accuracy is expected to be stronger than the negative effect of Sham, thus, it is expected that the difference of the two treatment effects, (SHAM-TMS), is positive.

4. Statistical Method

4.1 Notation

There were 80 individual throws per subject, which included 20 baseline throws for visit 1, 20 post-treatment throws for visit 1, 20 baseline throws for visit 2 and 20 post-treatment throws for visit 2. Taking the average of the corresponding throws, for each subject, 4 means are calculated, the mean throw score for baseline throws visit 1, the mean throw score for post-treatment throws visit 1, the mean throw score for baseline throws visit 2, and the mean throw score of post-treatment throws visit 2.

For subject i (i=1 to 25), the 20 baseline throws of visit 1 are denoted by X_{i1m} (m=1 to 20),

$[X_{i1}]_{20\times1}$ is the matrix of those 20 throws, and their mean value is be $\overline{X}_{i1}$; the 20 post-treatment throws of visit 1 are denoted by Y_{i1m} (m=1 to 20), $[Y_{i1}]_{20\times1}$ is the matrix of those 20 throws, and their mean value is be $\overline{Y}_{i1}$; the 20 baseline throws of visit 2 are denoted by X_{i2m} (m=1 to 20), $[X_{i2}]_{20\times1}$ is the matrix of those 20 throws, and their mean value is $\overline{X}_{i2}$; the 20 post-treatment throws of visit 2 are denoted by Y_{i2m} (m=1 to 20), $[Y_{i2}]_{20\times1}$ is the matrix of those 20 throws, and their mean value is $\overline{Y}_{i2}$.

4.2 Assumptions

Assume:

(1) the randomization successfully balanced subject characteristics among the two sequences.

(2) no carryover effect.

(3) no treatment-by-period interaction, which in this study, is no treatment-by-visit interaction.

In the material and method part, the models presented assume these three assumptions all hold and there is no sequence effect, while in statistical analysis, factor "sequence" is added in the models for test of those assumptions. In general, the significance of variable "sequence" can be caused by either one or more of the assumptions failing[8], or there is a true sequence effect.

4.3 Methods of Handling Baseline Measurements

The methods for handling baseline considered in the analysis are:

method 1: joint modeling of baseline and post-treatment response;

method 2: ignore baseline;

method 3: change from baseline;

method 4: adjust baseline as covariate;

Modeling mean scores and individual scores will all be considered. Different covariance structures will be applied. In total, 24 models will be fit, as shown:

Summary of Models		
	Method	Covariance structure
Model 1a	Joint modeling mean scores	UN
Model 1b	Joint modeling mean scores	V1
Model 1c	Joint modeling mean score	V2
Model 1d	Joint modeling mean score	CS
Model 1e	Joint modeling individual score	V3
Model 1f	Joint modeling individual score	V2
Model 1g	Joint modeling individual score	CS
Model 1h	Alternative joint modeling individual score	V3
Model 1i	Alternative joint modeling individual score	V2
Model 1j	Alternative joint modeling individual score	CS
Model 2a	Ignore baseline mean score	CS
Model 2b	Ignore baseline individual score	V4
Model 2c	Ignore baseline individual score	V2
Model 2d	Ignore baseline individual score	CS
Model 3a	Mean change	CS
Model 3b	Mean change	VC
Model 3c	Individual change	V4
Model 3d	Individual change	V2
Model 3e	Individual change	CS
Model 3f	Individual change	V5
Model 4a	Post mean scores | baseline of same period	CS
Model 4b	Post individual scores | baseline of same period	V4
Model 4c	Post individual scores | baseline of same period	V2
Model 4d	Post individual scores | baseline of same period	CS

UN refers to “unstructured”, CS refers to “compound symmetry”, and VC refers to “variance components”. The other covariance structures will be defined in detail under the section of each methods.

4.3.1 Joint Modeling

Under this method, baseline measurements are treated as outcome without accompany any treatment effects, and compose the dependent variable together with post-treatment measurements[5]. The advantage of this method is that it uses all information of all the data collected in the experiment, and if we assume baseline outcome and post-treatment outcome homoscedasticity, this method will lead to the most precise estimation of σ^2. When with-in period correlation is stronger than between-visit correlation, joint modeling of baselines and post-treatment responses will gain efficiency[7] in the estimation of the difference between effects of treatments, under the model composition proposed by Kenward.

4.2.1.1 Modeling Mean Throw Scores

When modeling, $\overline{X}_{i1}, \overline{Y}_{i1}, \overline{X}_{i2}, \overline{Y}_{i2}$, assume that:

$$\begin{bmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \\ \overline{Y}_{i1} \\ \overline{Y}_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} u_{\overline{X}_{i1}} \\ u_{\overline{X}_{i2}} \\ u_{\overline{Y}_{i1}} \\ u_{\overline{Y}_{i2}} \end{bmatrix}, V \right), V = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{bmatrix}$$, where $\begin{bmatrix} u_{\overline{X}_{i1}} \\ u_{\overline{X}_{i2}} \\ u_{\overline{Y}_{i1}} \\ u_{\overline{Y}_{i2}} \end{bmatrix}$ is the matrix of expectations.

Covariance Structures

In the sections illustrating covariance structure, I_n indicates identity matrix with $n \times n$ dimension, and J_n indicates a matrix of ones with $n \times n$ dimension.

(1) The first covariance structure applied is $V \sim$ UN, and the model is referred as model 1a.

(2) The second covariance structure applied is $V \sim$ V1, where V1 is defined as Kenward[5] suggested, and the model is referred as model 1b.

$$\Sigma_{XX} = \sigma_{XX} \otimes I_2 + \eta_{XX} \otimes J_2$$

$$\Sigma_{YY} = \sigma_{YY} \otimes I_2 + \eta_{YY} \otimes J_2$$

$$\Sigma_{XY} = \sigma_{XY} \otimes I_2 + \eta_{XY} \otimes J_2$$

That is:

$$var(\overline{X}_{i1}) = var(\overline{X}_{i2}) = \sigma_{XX} + \eta_{XX},$$

$$var(\overline{Y}_{i1}) = var(\overline{Y}_{i2}) = \sigma_{YY} + \eta_{YY},$$

$$cov(\overline{X}_{i1}, \overline{Y}_{i1}) = cov(\overline{X}_{i2}, \overline{Y}_{i2}) = \sigma_{XY} + \eta_{XY}$$

$$cov(\overline{X}_{i1}, \overline{Y}_{i2}) = cov(\overline{X}_{i2}, \overline{Y}_{i1}) = \eta_{XY}$$

(3) The third covariance structure applied is V ~ V2, where V2 assumes homoscedasticity, within-visit correlation = $\rho + \gamma$ and between visit correlation = ρ, and the model is referred as model 1c. V2 is the covariance structure proposed and investigated by Xun Chen et al[7]. In this case, under V2, the covariance structure is shown as:

$$V\left(\begin{bmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \\ \overline{Y}_{i1} \\ \overline{Y}_{i2} \end{bmatrix}\right) \sim \begin{bmatrix} 1 & \rho & \rho+\gamma & \rho \\ \rho & 1 & \rho & \rho+\gamma \\ \rho+\gamma & \rho & 1 & \rho \\ \rho & \rho+\gamma & \rho & 1 \end{bmatrix} \sigma^2$$

Or, for better illustration, re-arrange it：

$$V\left(\begin{bmatrix} \overline{X}_{i1} \\ \overline{Y}_{i1} \\ \overline{X}_{i2} \\ \overline{Y}_{i2} \end{bmatrix}\right) \sim \begin{bmatrix} 1 & \rho+\gamma & \rho & \rho \\ \rho+\gamma & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho+\gamma \\ \rho & \rho & \rho+\gamma & 1 \end{bmatrix} \sigma^2$$

(4) The forth covariance structure applied is V ~CS, and the model is referred as model 1d.

$$V\left(\begin{bmatrix}\overline{X}_{i1}\\ \overline{Y}_{i1}\\ \overline{X}_{i2}\\ \overline{Y}_{i2}\end{bmatrix}\right) \sim \begin{bmatrix}1 & \rho & \rho & \rho\\ \rho & 1 & \rho & \rho\\ \rho & \rho & 1 & \rho\\ \rho & \rho & \rho & 1\end{bmatrix}\sigma^2$$

Statistical Modeling

The expectations of $\overline{X}_{i1}, \overline{Y}_{i1}, \overline{X}_{i2}, \overline{Y}_{i2}$ are shown in Table 1a. Treat the baseline time as undergoing treatment "none", so treatment has three levels: "Sham", "TMS" and "none". The expectations shown in model 3a are based on Kenward et al. 2010 suggested[5]. To better illustrate his model, define according to timeline: visit 1 baseline time as interval 1, visit 1 post-treatment time as interval 2, visit 2 baseline time as interval 3 and visit 2 baseline time as interval 4. As Kenward does not constrain the effects associated to visits to be the same at baseline measurements and post-treatment measurements, so I convert the fixed effects associated to visit 1 and visit 2 on baseline throw scores and post-treatment throw scores into corresponding fixed effects associated to intervals.

Table 1a. Expectation of Mean Throw Score—Joint Modeling

			Annotation	Sequence 1	Sequence 2
Visit 1	Baseline	Interval 1	$u_{\overline{X}_{i1}}$	$\mu + \pi_1' + \tau_{none}$	$\mu + \pi_1' + \tau_{none}$
	Post-treatment	Interval 2	$u_{\overline{Y}_{i1}}$	$\mu + \pi_1 + \tau_{tms}$	$\mu + \pi_1 + \tau_{sham}$
Visit 2	Baseline	Interval 3	$u_{\overline{X}_{i2}}$	$\mu + \pi_2' + \tau_{none}$	$\mu + \pi_2' + \tau_{none}$
	Post-treatment	Interval 4	$u_{\overline{Y}_{i2}}$	$\mu + \pi_2 + \tau_{sham}$	$\mu + \pi_2 + \tau_{tms}$

Where the terms are:

μ, an intercept, the mean of the average of 20 throw scores.

$\pi_1', \pi_1, \pi_2', \pi_2$ are the fixed effects associated to interval 1, interval 2, interval 3 and interval 4. Note that, they are actually the fixed effect associated to visit 1 on baseline throws, the fixed effect associated to visit 1 on post-treatment throws, the fixed effect associated to visit 2 on baseline throws, and the fixed effect associated to visit 2 on post-treatment throws.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham, τ_{none} is a term associated with baseline, no actual treatment is received.

Set interval 4 and treatment TMS as reference, the model for modeling 4 means per subject can be write as:

$$\overline{outcome}_{ij} = \beta_0 + \beta_1 * I(Interval = 1) + \beta_2 * I(Interval = 2) + \beta_3 * I(interval = 3)$$
$$+ \beta_4 * I(treatment = \text{Sham}) + \epsilon_{ij}, \epsilon_{ij} \sim [V].$$
$$i\ indicates\ subject(1\ to\ 25), j\ indicates\ interval(1\ to\ 4)$$

Illustration for the composition $\overline{outcome}_{ij}$ of subject i	
$\overline{outcome}_{i1}$	$\overline{X}_{i1}$
$\overline{outcome}_{i2}$	$\overline{Y}_{i1}$
$\overline{outcome}_{i3}$	$\overline{X}_{i2}$
$\overline{outcome}_{i4}$	$\overline{Y}_{i2}$

Where, β_0 is the expectation of averaging 20 throw scores at interval 4 with treatment TMS, $\beta_1 = \pi_1' - \pi_2, \beta_2 = \pi_1 - \pi_2, \beta_3 = \pi_2' - \pi_2, \beta_4 = \tau_{Sham} - \tau_{tms}$. β_1, β_2, and β_3 belong to factor "interval" and estimates the fluctuation of mean throw score. β_4 belongs to factor "treatment" and estimates the difference between the effect of Sham treatment and the effect of TMS treatment.

To justify this method, use interval 4 and treatment TMS as reference, let:

$$X_1 = \begin{cases} 1 & interval = 1 \\ 0 & else \end{cases}$$

$$X_2 = \begin{cases} 1 & interval = 2 \\ 0 & else \end{cases}$$

$$X_3 = \begin{cases} 1 & interval = 3 \\ 0 & else \end{cases}$$

$$X_4 = \begin{cases} 1 & treatment = "Sham" \\ 0 & else \end{cases}$$

$$X_5 = \begin{cases} 1 & treatment = "none" \\ 0 & else \end{cases}$$

Aside from β_0, the design matrix of a subject from sequence 1, will be:

X1	X2	X3	X4	X5
1	0	0	0	1
0	1	0	0	0
0	0	1	0	1
0	0	0	1	0

Col(X5)=Col(X1)+Col(X3)

Then the design matrix of a subject from sequence 2, will be:

X1	X2	X3	X4	X5
1	0	0	0	1
0	1	0	1	0
0	0	1	0	1
0	0	0	0	0

Col(X5)=Col(X1)+Col(X3)

So, the beta coefficient stands for $(\tau_{none} - \tau_{tms})$ has no degree of freedom, the three-level factor “treatment” has only 1 degree of freedom in test of type III SS and the beta coefficient is estimating $(\tau_{Sham} - \tau_{tms})$, the difference between the effect of Sham and the effect of TMS.

4.3.1.2 Modeling Individual Throw Scores

When modeling 80 throw scores per subject,

Assume

$$\begin{bmatrix} X_{i1} \\ Y_{i1} \\ X_{i2} \\ Y_{i2} \end{bmatrix}_{80\times 1} \sim N\left(\begin{bmatrix} \mu_{X\,i1} \\ \mu_{Y\,i2} \\ \mu_{X\,i2} \\ \mu_{Y\,i2} \end{bmatrix}, V\right), V = \begin{bmatrix} \Sigma X_1X_1 & \Sigma X_1Y_1 & \Sigma X_1X_2 & \Sigma X_1Y_2 \\ \Sigma^{\mathrm{T}} X_1Y_1 & \Sigma Y_1Y_1 & \Sigma X_2Y_1 & \Sigma Y_1Y_2 \\ \Sigma^{\mathrm{T}} X_1X_2 & \Sigma^{\mathrm{T}} X_2Y_1 & \Sigma X_2X_2 & \Sigma X_2Y_2 \\ \Sigma^{\mathrm{T}} X_1Y_2 & \Sigma^{\mathrm{T}} Y_1Y_2 & \Sigma^{\mathrm{T}} X_2Y_2 & \Sigma Y_2Y_2 \end{bmatrix}, \text{ where } \begin{bmatrix} \mu_{X\,i1} \\ \mu_{Y\,i1} \\ \mu_{X\,i2} \\ \mu_{Y\,i2} \end{bmatrix} \text{ is the}$$

matrix of expectations.

Covariance Structures

(1) The first covariance structure to be applied is V3, and the model is referred as model 1e.

In V3, it is assumed that：

$$\Sigma X_1X_1 = \Sigma Y_1Y_1 = \Sigma X_2X_2 = \Sigma Y_2Y_2 = (\eta + \sigma) \otimes J_{20} + \mathrm{TOEP}_{20}$$

$$\Sigma X_1Y_1 = \Sigma^{\mathrm{T}} X_1Y_1 = \Sigma^{\mathrm{T}} X_1Y_1 = \Sigma X_2Y_2 = (\eta + \sigma) \otimes J_{20}$$

$$\begin{bmatrix} \Sigma X_1X_2 & \Sigma X_1Y_2 \\ \Sigma X_2Y_1 & \Sigma Y_1Y_2 \end{bmatrix} = \begin{bmatrix} \Sigma^{\mathrm{T}} X_1X_2 & \Sigma^{\mathrm{T}} X_1Y_2 \\ \Sigma^{\mathrm{T}} X_2Y_1 & \Sigma^{\mathrm{T}} Y_1Y_2 \end{bmatrix} = \eta \otimes J_{40}$$

The structure constrains homoscedasticity of all throw scores, and the covariance among throw scores between visit is η; the covariance among throw scores within visit but between interval is $(\eta + \sigma)$, $\sigma>0$; For throws within visit and within interval, their covariance equal to $(\eta + \sigma)$ plus the term from TOEPLITZ structure.

This covariance structure is proposed basing on the following reasons:

a. The correlation among throws may not be uniform, so TOEP structure is considered, in this structure, the covariance between m_1^{th} throw and m_2^{th} throw depends on $|m_1-m_2|$, so covariance changes over time.

b. However, directly assume V_{80}~TOEP has problem as: the correlation between 40th and 39th or 41th throw will be constrained to be the same, as |40-39|=1 and |40-41|=1 too. However it may not hold as the 39th and 40th are from visit 1, the 41th is from visit 2.

c. Assume the 40 throws from same visit ~ TOEP has a similar problem, the correlations between 20th and 19th or 21th are constrained to be the same, which may not hold as 19th 20th are baseline 21th is after treatment.

d. Within-visit correlation may be stronger than between-visit correlation.

To accommodate those concerns, V3 is proposed, which incorporates TOEP but is more flexible to accommodate the study data.

(2) The second covariance structure to be applied is V2, which assumes homoscedasticity, within-visit correlation = $\rho + \gamma$ and between visit correlation = ρ.The model is referred as model 1f.

$$\begin{bmatrix} \Sigma X_1X_1 & \Sigma X_1Y_1 \\ \Sigma^{\mathrm{T}} X_1Y_1 & \Sigma Y_1Y_1 \end{bmatrix} = \begin{bmatrix} \Sigma X_2X_2 & \Sigma X_2Y_2 \\ \Sigma^{\mathrm{T}} X_2Y_2 & \Sigma Y_2Y_2 \end{bmatrix} = (\rho + \gamma) * \sigma^2 \otimes J_{40} + (1 - \rho - \gamma) * \sigma^2 \otimes I_{40}$$

$$\begin{bmatrix} \Sigma X_1X_2 & \Sigma X_1Y_2 \\ \Sigma X_2Y_1 & \Sigma Y_1Y_2 \end{bmatrix} = \begin{bmatrix} \Sigma^{\mathrm{T}} X_1X_2 & \Sigma^{\mathrm{T}} X_1Y_2 \\ \Sigma^{\mathrm{T}} X_2Y_1 & \Sigma^{\mathrm{T}} Y_1Y_2 \end{bmatrix} = \rho * \sigma^2 \otimes J_{40}$$

(3) The third covariance structure to be applied is V~CS, and the model is referred as model 1g.

$$\begin{bmatrix} \Sigma X_1X_1 & \Sigma X_1Y_1 & \Sigma X_1X_2 & \Sigma X_1Y_2 \\ \Sigma^{\mathrm{T}} X_1Y_1 & \Sigma Y_1Y_1 & \Sigma X_2Y_1 & \Sigma Y_1Y_2 \\ \Sigma^{\mathrm{T}} X_1X_2 & \Sigma^{\mathrm{T}} X_2Y_1 & \Sigma X_2X_2 & \Sigma X_2Y_2 \\ \Sigma^{\mathrm{T}} X_1Y_2 & \Sigma^{\mathrm{T}} Y_1Y_2 & \Sigma^{\mathrm{T}} X_2Y_2 & \Sigma Y_2Y_2 \end{bmatrix} = \rho * \sigma^2 \otimes J_{40} + (1 - \rho) * \sigma^2 \otimes I_{40}$$

Statistical Model

The expected value of individual throw, shown in Table 1b.

Table 1b. Expectation of Individual Throw Score

			Annotation	Sequence 1	Sequence 2
Visit 1	Baseline	Interval 1	$\mu_{X_{i1}}$	$\mu + \pi_1' + \tau_{none}$	$\mu + \pi_1' + \tau_{none}$
	Post-treatment	Interval 2	$\mu_{Y_{i1}}$	$\mu + \pi_1 + \tau_{tms}$	$\mu + \pi_1 + \tau_{sham}$
Visit 2	Baseline	Interval 3	$\mu_{X_{i2}}$	$\mu + \pi_2' + \tau_{none}$	$\mu + \pi_2' + \tau_{none}$
	Post-treatment	Interval 4	$\mu_{Y_{i2}}$	$\mu + \pi_2 + \tau_{sham}$	$\mu + \pi_2 + \tau_{tms}$

Where the terms are:

μ, an intercept, the mean of a throw score.

$\pi_1', \pi_1, \pi_2', \pi_2$ are the fixed effects associated to interval 1, interval 2, interval 3 and interval 4. Note that, they are actually the fixed effect associated to visit 1 on baseline throws, the fixed effect associated to visit 1 on post-treatment throws, the fixed effect associated to visit 2 on baseline throws, and the fixed effect associated to visit 2 on post-treatment throws.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham, τ_{none} is a term associated with baseline.

The model incorporating 80 individual throws per subject can be write as:

$$\text{outcome}_{\text{ijm}} = \beta_0 + \beta_1 * I(Interval = 1) + \beta_2 * I(Interval = 2) + \beta_3 * I(interval = 3)$$

$$+\beta_4 * I(treatment = \text{Sham}) + \epsilon_{ijm}, \epsilon_{ijm} \sim [V]. \, i \; indicates \; subject(1 \; to \; 25),$$

$$j \; indicates \; interval(1 \; to \; 4), m \; indicates \; throws \; (1 \; to \; 20).$$

Illustration for the composition of $\text{outcome}_{\text{ijm}}$ of subject i	
$\text{outcome}_{\text{i1m}}$(m=1 to 20)	X_{i1m}(m=1 to 20)
$\text{outcome}_{\text{i2m}}$(m=1 to 20)	Y_{i1m}(m=1 to 20)
$\text{outcome}_{\text{i3m}}$(m=1 to 20)	X_{i2m}(m=1 to 20)

outcome_{i4m}(m=1 to 20)	Y_{i2m}(m=1 to 20)

Then, β_0 is the expectation of an individual throw score at interval 4 with treatment TMS, $\beta_1 = \pi_1' - \pi_2, \beta_2 = \pi_1 - \pi_2, \beta_3 = \pi_2' - \pi_2, \beta_4 = \tau_{Sham} - \tau_{tms}$. β_1, β_2, and β_3 belong to factor "interval" and estimate the fluctuation of throw score. β_4 belongs to factor "treatment" and estimates the difference between the effect of Sham treatment and the effect of TMS treatment. As justified in section 4.3.1.1, beta coefficient stands for $(\tau_{none} - \tau_{TMS})$ has no degree of freedom.

4.3.1.3 Alternative Method of Joint Modeling

In joint modeling, there is another way to compose the model, suggested by Dr. Christina Mehta. The terms in model are: visit, treatment, type and type*treatment. Where type=0 suggests baseline, and type=1 suggests post-treatment. For baseline and post-treatment throws, variable "treatment" will all be the treatment assigned to the subject. This alternative modeling method will be illustrated by modeling individual throws.

Covariance Structures

Similar as section 4.3.1.2, in the alternative method of joint modeling, apply covariance structure V3, the model is referred as 1h; apply covariance structure V2, the model is referred as 1i; apply covariance structure CS, the model is referred as 1j;

Statistical Model

$$outcome_{ijm} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treatment = Sham) + \beta_3 * I(type = 1)$$

$$+ \beta_4 * I(type = 1\ and\ treatment = sham) + \epsilon_{ijm}, \epsilon_{ijm} \sim V$$

Expectation of Throw Score			
		Sequence 1	Sequence 2
Visit 1	Baseline	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2$
	Post-treatment	$\beta_0 + \beta_1 + \beta_3$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4$
Visit 2	Baseline	$\beta_0 + \beta_2$	β_0
	Post-treatment (interval 4)	$\beta_0 + \beta_2 + \beta_3 + \beta_4$	$\beta_0 + \beta_3$

β_0 estimates the mean of throw score at visit 2 with TMS treatment being assigned. β_1 estimates the fluctuation of throw score. β_2 estimates the effect associated with treatment Sham assigned. β_3 estimates the effect of treatment on post-treatment throw scores. β_4 estimates the difference between the effect of Sham treatment and TMS treatment on post-treatment throw scores. Thus, in this model, the terms related to the effects of treatments are β_3 and β_4, rather than β_2.

This method of modeling assumes main effect associated with treatment, which applies to both baseline throw scores where the subjects hasn't receive treatments yet, and post-treatment throw scores. And it also constrains the effects associated with visit are the same for baseline and post-treatment time.To be specific, illustrate the baselines only:

Expectation of Baseline Throw Score		
	Sequence 1	Sequence 2
Visit 1 Baseline	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2$
Visit 2 Baseline	$\beta_0 + \beta_2$	β_0

This model to assumes the baseline throw scores are associated with the treatment the subject receiving, which may not be the case in practice. However, the estimation of the difference between the effect of Sham and effect of TMS in this model, is not influenced by the terms related to main effect of treatments, which in this study is the term β_2. The result of this alternative methods will also be presented.

4.3.2 Ignore Baseline:

Ignore baseline means discard all the baselines and analysis the post-treatment throws. The main reason to including this method is: it serves as a standard reference. Yan et al. 2012 shows from simulation that when most information of treatment effects is based on with-in subject information, then this method has similar efficacy as the methods adjusting for baseline[5].

4.3.2.1 Modeling Mean Throw Scores

Covariance Structures

When modeling $\begin{bmatrix}\overline{Y}_{i1}\\ \overline{Y}_{i2}\end{bmatrix}$, assume $\begin{bmatrix}\overline{Y}_{i1}\\ \overline{Y}_{i2}\end{bmatrix} \sim N\left(\begin{bmatrix}u_{\overline{Y}_{i1}}\\ u_{\overline{Y}_{i2}}\end{bmatrix}, V\right)$ where $\begin{bmatrix}u_{\overline{Y}_{i1}}\\ u_{\overline{Y}_{i2}}\end{bmatrix}$ the matrix of expectations, V=$\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\sigma^2$, CS structure. The model will be referred as model 2a.

Statistical Model

Table 2a. Expectation of Mean Throw Score—Ignore Baseline

	Annotation	Sequence 1	Sequence 2

Visit 1	Post-treatment	$u_{\overline{Y}_{i1}}$	$\mu + \pi_1 + \tau_{tms}$	$\mu + \pi_1 + \tau_{sham}$
Visit 2	Post-treatment	$u_{\overline{Y}_{i2}}$	$\mu + \pi_2 + \tau_{sham}$	$\mu + \pi_2 + \tau_{tms}$

Where the terms are:

μ, an intercept, the mean of the average of 20 throw scores;

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is:

$$\overline{Y}_{\text{ij}} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treamtent = \text{"Sham"}) + \epsilon_{ij}, \epsilon_{ij} \sim V \text{ where i indicates subject, j indicates visit}$$

Then, β_0 is the expectation of averaging 20 throw scores at visit 2 with treatment TMS.

$\beta_1 = \pi_1 - \pi_2$, $\beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of post-treatment mean score across visits, β_2 estimates the difference between the effect of Sham and the effect of TMS.

4.3.2.2 Modeling Individual Throw Scores,

When modeling $\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}$,

Assume:

$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}_{40\times1} \sim N\left(\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}, V\right)$, where $\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}$ is the matrix of expectation and

$V=\begin{bmatrix} \Sigma Y_1 Y_1 & \Sigma Y_1 Y_2 \\ \Sigma^T Y_1 Y_2 & \Sigma Y_2 Y_2 \end{bmatrix}$

Covariance Structures

Three covariance structures will be assumed.

(1) Similar to V3, in V4, assume the post-treatment throws follow homoscedasticity, and the covariance between post-treatment throw scores between-visit is η; for post-treatment throws scores within visit, their covariance equal to η plus the term from TOEP structure. This is equal to assume the between visit correlation is uniform, but within visit correlation is not and changes according to time gap. This model will be referred as model 2b.

$$\Sigma Y_1Y_1 = \Sigma Y_2Y_2 = \eta \otimes J_{20} + TOEP,\ \Sigma Y_1Y_2 = \Sigma^T Y_1Y_2 = \eta \otimes J_{20}$$

(2) Apply structure V2, which assumes homoscedasticity, within-visit correlation = $\rho + \gamma$ and between visit correlation = ρ.The model is referred as model 2c.

$$\Sigma Y_1Y_1 = \Sigma Y_2Y_2 = (\rho + \gamma) * \sigma^2 \otimes J_{20} + (1 - \rho - \gamma) * \sigma^2 \otimes I_{20}$$
$$\Sigma Y_1Y_2 = \Sigma^T Y_1Y_2 = \rho * \sigma^2 \otimes J_{20}$$

(3) Apply CS structure. The model is referred as model 2d.

$$V=\begin{bmatrix} \Sigma Y_1Y_1 & \Sigma Y_1Y_2 \\ \Sigma^T Y_1Y_2 & \Sigma Y_2Y_2 \end{bmatrix} = \rho * \sigma^2 \otimes J_{40} + (1 - \rho) * \sigma^2 \otimes I_{40}$$

Statistical Model

The expectation of throw score is:

Table 2b. Expectation of individual throw score—Ignore Baseline

		annotation	Sequence 1	Sequence 2
Visit 1	Post-treatment	μ_{Yi1m}	$\mu + \pi_1 + \tau_{TMS}$	$\mu + \pi_1 + \tau_{Sham}$
Visit 2	Post-treatment	μ_{Yi2m}	$\mu + \pi_2 + \tau_{Sham}$	$\mu + \pi_2 + \tau_{TMS}$

Where the terms are:

μ, an intercept, the mean of a throw score;

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is

$$\mathrm{Y_{ijm}} = \beta_0 + \mathrm{B_1} * I(visit = 1) + \beta_2 * I(treatment = \mathrm{Sham}) + \epsilon_{ijm}, \epsilon_{ijm} \sim V$$

Where β_0 is the expectation of a single throw score at visit 2 with treatment TMS, $\beta_1 = \pi_1 - \pi_2, \beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of post-treatment mean score across visits, β_2 estimates the difference between the effect of Sham and the effect of TMS.

4.3.3 Change from Baseline

This method is widely used in practice, and, the most advantage is the increased power of detecting carryover effect compared to other methods[9]. If assume carryover effect from the previous period remains the same at baseline time and post-treatment time of the following period, then carryover effect is eliminated[10]. The drawback is, under most covariance structures, the estimator of this method has greater theorical variance compare to "ignore baseline" and "Joint Modeling"[2,5].

Define change score and mean change score, $D_{ijm} = Y_{ijm} - \overline{X}_{ij}$, $\overline{D}_{ij} = \overline{D_{ijm}} = \overline{Y}_{ij} - \overline{X}_{ij}$. $[\mathrm{D_{i1}}]_{20\times1}$ the matrix for $D_{i1m}(\mathrm{m} = 1 \text{ to } 20)$, $[\mathrm{D_{i2}}]_{20\times1}$ the matrix for $D_{i2m}(\mathrm{m} = 1 \text{ to } 20)$.

Illustration for the composition of change score D_{ijm} of subject i	
D_{i1m}(m=1 to 20)	Y_{i1m}(m=1 to 20) - $\overline{X}_{i1}$
D_{i2m}(m=1 to 20)	Y_{i2m}(m=1 to 20) - $\overline{X}_{i2}$

Illustration for the composition of mean change score $\overline{D}_{ij}$ of subject i	
$\overline{D}_{i1}$	$\overline{Y}_{i1} - \overline{X}_{i1}$
$\overline{D}_{i2}$	$\overline{Y}_{i2} - \overline{X}_{i2}$

4.3.3.1 Modeling Mean Changes

Assume $\begin{bmatrix}\overline{D}_{i1}\\ \overline{D}_{i2}\end{bmatrix} \sim N\left(\begin{bmatrix}\mu_{\overline{D}_1}\\ \mu_{\overline{D}_2}\end{bmatrix}, V\right)$, where $\begin{bmatrix}\mu_{\overline{D}_1}\\ \mu_{\overline{D}_2}\end{bmatrix}$ is the matrix of expectations

Covariance Structures

(1) Assume V~CS. The model is referred as model 3a;

(2) Assume V~VC, correlation between $\overline{D}_{i1}$ and $\overline{D}_{i2}$ is 0. The model is referred as model 3b;

The reason why this structure is proposed:

When

$$V\left(\begin{bmatrix}\overline{X}_{i1}\\ \overline{Y}_{i1}\\ \overline{X}_{i2}\\ \overline{Y}_{i2}\end{bmatrix}\right) \sim \begin{bmatrix}1 & \rho+\gamma & \rho & \rho\\ \rho+\gamma & 1 & \rho & \rho\\ \rho & \rho & 1 & \rho+\gamma\\ \rho & \rho & \rho+\gamma & 1\end{bmatrix}\sigma^2, \text{ or}$$

$$\mathrm{V}\left(\begin{bmatrix}\overline{X}_{i1}\\ \overline{Y}_{i1}\\ \overline{X}_{i2}\\ \overline{Y}_{i2}\end{bmatrix}\right) \sim \begin{bmatrix}1 & \rho & \rho & \rho\\ \rho & 1 & \rho & \rho\\ \rho & \rho & 1 & \rho\\ \rho & \rho & \rho & 1\end{bmatrix}\sigma^2, \text{ then:}$$

$$V\left(\begin{bmatrix}\overline{Y}_{i1}-\overline{X}_{i1}\\ \overline{Y}_{i2}-\overline{X}_{i2}\end{bmatrix}\right) = \begin{bmatrix}-1 & 1 & 0 & 0\\ 0 & 0 & -1 & 1\end{bmatrix} \otimes V\left(\begin{bmatrix}\overline{X}_{i1}\\ \overline{Y}_{i1}\\ \overline{X}_{i2}\\ \overline{Y}_{i2}\end{bmatrix}\right) \otimes \begin{bmatrix}-1 & 1 & 0 & 0\\ 0 & 0 & -1 & 1\end{bmatrix}^T = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\sigma^{*2}$$

Statistical Modeling

The expectation of mean change score is:

Table 3a. Expectation of Mean Change Score (Post-Pre)

	Annotation	Sequence 1	Sequence 2
Visit 1	$\mu_{\overline{D}_1}$	$\pi_1 + \tau_{TMS}$	$\pi_1 + \tau_{Sham}$
Visit 2	$\mu_{\overline{D}_2}$	$\pi_2 + \tau_{Sham}$	$\pi_2 + \tau_{TMS}$

Where the terms are:

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2 on mean change score.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is:

$$\overline{D}_{\mathrm{ij}} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treamtent = \text{"Sham"}) + \epsilon_{ij}, \epsilon_{ij} \sim V, where\ i\ indicates\ subject, j\ indicates\ visit$$

Where β_0 is the expectation of mean change at visit 2 with TMS treatment, $\beta_1 = \pi_1 - \pi_2, \beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of mean change score across visit. β_2 estimates the difference between the effect of Sham and the effect of TMS.

4.3.3.2 Model Individual Change Score

Assume $\begin{bmatrix} D_{i1} \\ D_{i2} \end{bmatrix}_{40\times1} \sim N\left(\begin{bmatrix} \mu_{D_1} \\ \mu_{D_2} \end{bmatrix}, V\right)$, where $\begin{bmatrix} \mu_{D_1} \\ \mu_{D_2} \end{bmatrix}$ is the matrix of expectations

$$V = V\left(\begin{bmatrix} D_{i1} \\ D_{i2} \end{bmatrix}_{1\times40}\right) = \begin{bmatrix} \Sigma D_1 D_1 & \Sigma D_1 D_2 \\ \Sigma^{\mathrm{T}} D_1 D_2 & \Sigma D_2 D_2 \end{bmatrix}$$

Covariance Structures

(1) Apply V4, the model is referred as model 6a. Assume homoscedasticity of mean change score, between visit covariance is η, and within visit covariance is η plus the term from TOEP structure. This is equal to assume the between visit correlation is uniform, but within visit correlation is not and changes according to time gap. This model will be referred as model 3c.

$$\Sigma D_1 D_1 = \Sigma D_2 D_2 = \eta \otimes J_{20} + TOEP_{20}$$
$$\Sigma D_1 D_2 = \Sigma^{\mathrm{T}} D_1 D_2 = \eta \otimes J_{20}$$

(2) Apply V2, the model is referred as model 3d.

$$\Sigma D_1 D_1 = \Sigma D_2 D_2 = (\rho + \gamma) * \sigma^2 \otimes J_{20} + (1 - \rho - \gamma) * \sigma^2 \otimes I_{20}$$

$$\Sigma^{\mathrm{T}} D_1 D_2 = \Sigma D_1 D_2 = \rho * \sigma^2 \otimes J_{20}$$

Note that, if estimated $\gamma = 0$, then this structure collapse as CS; and if estimated $\rho = 0$, then this structure collapse as V5.

(3) Apply CS structure, the model is referred as model 3e.

$$\begin{bmatrix} \Sigma D_1 D_1 & \Sigma D_1 D_2 \\ \Sigma^{\mathrm{T}} D_1 D_2 & \Sigma D_2 D_2 \end{bmatrix} = \rho * \sigma^2 \otimes J_{40} + (1 - \rho) * \sigma^2 \otimes I_{40}$$

(4) Apply V5, which assumes only changes from the same visit is uniformly correlated, cross-visit correlation is 0. The model is referred as model 3f.

$$\Sigma D_1 D_1 = \Sigma D_2 D_2 = \rho * \sigma^2 \otimes J_{20} + (1 - \rho) * \sigma^2 \otimes I_{20}$$

$$\Sigma D_1 D_2 = \Sigma^{\mathrm{T}} D_1 D_2 = 0 * J_{20}$$

The reason why this structure is proposed:

When $V\left(\begin{bmatrix} X_{i1} \\ Y_{i1} \\ X_{i2} \\ Y_{i2} \end{bmatrix}\right) \sim V2\ or\ CS$,

The covariance of changes from different visit, $cov(D_{i1k}, D_{i2m}) = cov\left(Y_{ijk} - \frac{1}{20}\sum_{m=1}^{20} X_{i1m}, Y_{i2m} - \frac{1}{20}\sum_{m=1}^{20} X_{i2m}\right) = [A\ BCD] \otimes V \otimes [EFGH]', where\ [C] = [D] = [E] = [F] = [0]_{1\times 20}, [A] = [G] = \left[-\frac{1}{20}\right]_{1\times 20}$, [B] and [H] are 1×20 matrix, with only kth or mth elements=1,else elements=0.

For V~ V2:

$[ABCD] \otimes \mathrm{V} \otimes [EFGH]' = \left[AB\ \tilde{0}\tilde{0}\right] \otimes V \otimes \left[\tilde{0}\tilde{0}GH\right]'$ =[MN $\tilde{0}\tilde{0}$] $\otimes \left[\tilde{0}\tilde{0}GH\right]' = 0$, [M] = $\left[\frac{\rho+\gamma-1}{20}\right]_{1x20}$ $[N]_{1\times 20}: kth\ col = 1-\rho-\gamma, else = 0.$

For V~CS:

$[ABCD] \otimes V_1 \otimes [EFGH]' = \left[AB\ \tilde{0}\tilde{0}\right] \otimes V_1 \otimes \left[\tilde{0}\tilde{0}CD\right]'$ =[MN $\tilde{0}\tilde{0}$] $\otimes \left[\tilde{0}\tilde{0}CD\right]' = 0, where\ [M] = \left[\frac{\rho-1}{20}\right]_{1x20}$ $[N]_{1\times 20}: kth\ col = 1-\rho, else = 0.$

Statistical Model

Table 3b. Expectation of Change Score

	Annotation	Sequence 1	Sequence 2
Visit 1	μ_{D_1}	$\pi_1 + \tau_{TMS}$	$\pi_1 + \tau_{Sham}$
Visit 2	μ_{D_2}	$\pi_2 + \tau_{Sham}$	$\pi_2 + \tau_{TMS}$

Where the terms are:

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2.

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is:

$$\mathrm{D_{ijm}} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treamtent = \text{"Sham"}) + \epsilon_{ijm}, \epsilon_{ijm} \sim V$$

Where β_0 is the mean of change score at visit 2 with TMS treatment, $\beta_1 = \pi_1 - \pi_2, \beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of change score across visit. β_2 estimates the difference between the effect of Sham and the effect of TMS.

4.3.4 Baseline as Covariate

Similar as change from baseline study, this is another conventional way to adjust for baseline[5]. As in this study, each visit is regarded as a period, when modeling post-treatment throw scores, the baseline from the same visit is adjusted as covariate.

A critique of this method, proposed by Kenward et al[5]., illustrated by assuming the distribution of $\begin{bmatrix} \overline{X}_{i1} \\ \overline{Y}_{i1} \\ \overline{X}_{i2} \\ \overline{Y}_{i2} \end{bmatrix}$, is that:

When assuming $\begin{bmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \\ \overline{Y}_{i1} \\ \overline{Y}_{i2} \end{bmatrix} \sim N\left(\begin{bmatrix} u_{\overline{X}_{i1}} \\ u_{\overline{X}_{i2}} \\ u_{\overline{Y}_{i1}} \\ u_{\overline{Y}_{i2}} \end{bmatrix}, V \right)$, then deriving the distribution of $\left(\overline{Y}_{\mathrm{i1}}, \overline{Y}_{i2} \middle| \overline{X}_{i1}, \overline{X}_{i2}\right)$: $\left(\overline{Y}_{\mathrm{i1}}, \overline{Y}_{i2} \middle| \overline{X}_{i1}, \overline{X}_{i2}\right) \sim N(\begin{bmatrix} u_{\overline{Y}_{i1}} \\ u_{\overline{Y}_{i2}} \end{bmatrix} - \Sigma^T XY \otimes \Sigma^{-1} XX \otimes \left(\begin{bmatrix} u_{\overline{X}_{i1}} \\ u_{\overline{X}_{i2}} \end{bmatrix} - \begin{bmatrix} \overline{X}_{i1} \\ \overline{X}_{i2} \end{bmatrix} \right), V^* = \Sigma YY - \Sigma^T XY \otimes \Sigma^{-1} XX \otimes \Sigma XY)$ Under most covariance structure V, $\mathrm{E}\left(\overline{Y}_{i1} \middle| \overline{X}_{i1}, \overline{X}_{i2}\right)$ or $\mathrm{E}\left(\overline{Y}_{i2} \middle| \overline{X}_{i1}, \overline{X}_{i2}\right)$ depends on both $\overline{X}_{i1}$ and $\overline{X}_{i2}$. Thus, argues the method of adjusting only the baseline from same visit is inappropriate. In a word, if

$\begin{bmatrix}\overline{X}_{i1}\\ \overline{X}_{i2}\\ \overline{Y}_{i1}\\ \overline{Y}_{i2}\end{bmatrix} \sim N\left(\begin{bmatrix}u_{\overline{X}_{i1}}\\ u_{\overline{X}_{i2}}\\ u_{\overline{Y}_{i1}}\\ u_{\overline{Y}_{i2}}\end{bmatrix}, V\right)$, which is the assumption made in method 1 joint modeling, then under most situation of V, $\begin{bmatrix}(\overline{Y}_{i1}|\overline{X}_{i1})\\ (\overline{Y}_{i2}|\overline{X}_{i2})\end{bmatrix}$ ~MVN does not hold.

To avoid this issue, in this study, directly assuming the distribution of $\begin{bmatrix}(\overline{Y}_{i1}|\overline{X}_{i1})\\ (\overline{Y}_{i2}|\overline{X}_{i2})\end{bmatrix}$ or $\begin{bmatrix}(Y_{i1}|\overline{X}_{i1})\\ (Y_{i2}|\overline{X}_{i2})\end{bmatrix}$ follows normal distribution.

Another special concern to this method is, it strictly depends on the assumption of no carryover effect. The contamination of carryover effect makes baselines inappropriate as covariate[9]. Variable "sequence" is still added in analysis to test assumption of no treatment by period interaction and successful randomization.

4.3.4.1 Modeling Mean Throw Scores

Like in method 2, the dependent variable is composed by $\overline{Y}_{i1}$and $\overline{Y}_{i2}$.

Assume:

$\begin{bmatrix}\overline{Y}_{i1}\\ \overline{Y}_{i2}\end{bmatrix} \sim N\left(\begin{bmatrix}u_{\overline{Y}_{i1}}\\ u_{\overline{Y}_{i2}}\end{bmatrix}, V\right)$ where $\begin{bmatrix}u_{\overline{Y}_{i1}}\\ u_{\overline{Y}_{i2}}\end{bmatrix}$ the matrix of expectations. As latter on it is illustrated that $u_{\overline{Y}_{i1}}$ denpends on $\overline{X}_{i1}$ and $u_{\overline{Y}_{i2}}$ denpends on $\overline{X}_{i2}$, it is equivalent to assuming $\begin{bmatrix}\overline{Y}_{i1}|\overline{X}_{i1}\\ \overline{Y}_{i2}|\overline{X}_{i2}\end{bmatrix} \sim MVN.$

Covariance Structure:

(1) V=$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \sigma^2$, CS structure. The model is referred as model 4a.

Statistical Model

Table 4a. Expectation of post treatment mean score

	Annotation	Sequence 1	Sequence 2
Visit 1	$u_{\overline{Y}_{i1}}$	$\mu + \pi_1 + \beta * \overline{X}_{i1} + \tau_{tms}$	$\mu + \pi_1 + \beta * \overline{X}_{i1} + \tau_{sham}$
Visit 2	$u_{\overline{Y}_{i2}}$	$\mu + \pi_2 + \beta * \overline{X}_{i2} + \tau_{sham}$	$\mu + \pi_2 + \beta * \overline{X}_{i2} + \tau_{tms}$

Where the terms are:

μ, an intercept, the mean of the average of 20 post-treatment throw scores after adjusting for baseline mean score from the same visit;

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2.

β is the change in the expectation of $\overline{Y}_{ij}$ with 1 unit change in $\overline{X}_{ij}$

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is:

$$\overline{Y}_{\text{ij}} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treamtent = \text{"Sham"}) + \beta_3 * \overline{X}_{ij} + \epsilon_{ij}, \epsilon_{ij} \sim V$$

Where β_0 is the mean of averaging 20 throw scores at visit 2 treatment TMS, after adjusting for baseline mean throw score. $\beta_1 = \pi_1 - \pi_2$, $\beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of post-treatment mean score. β_2 estimates the difference between the effect of Sham and the effect of TMS. β_3 estimates the change in expectation of $\overline{Y}_{ij}$ with 1 unit

change in the baseline.

4.3.4.2 Modeling Individual Throw Scores

Similar to method 2, When modeling $\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}$,

Assume:

$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}_{40\times1} \sim N\left(\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}, V\right)$, where $\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}$ is the matrix of expectation and V=$\begin{bmatrix} \Sigma Y_1 Y_1 & \Sigma Y_1 Y_2 \\ \Sigma^T Y_1 Y_2 & \Sigma Y_2 Y_2 \end{bmatrix} \sigma^2$.

As latter on it is illustrated that μ_{i1} depends on $\overline{X}_{i1}$ and u_{i2} depends on $\overline{X}_{i2}$, it is equivalent to assuming $\begin{bmatrix} Y_{i1}|\overline{X}_{i1} \\ Y_{i2}|\overline{X}_{i2} \end{bmatrix} \sim MVN$.

Covariance Structures:

Three covariance structure will be applied.

(1) Apply V4, assume the post-treatment throws follow homoscedasticity, and the covariance between post-treatment throw scores between-visit is η; for post-treatment throw scores within visit, their covariance equal to η plus the term from TOEP structure. This is equal to assume the between visit correlation is uniform, but within visit correlation is not and changes according to time gap. The model is referred as model 4b.

$$\Sigma Y_1 Y_1 = \Sigma Y_2 Y_2 = \eta \otimes J_{20} + TOEP_{20}$$
$$\Sigma Y_1 Y_1 = \Sigma Y_2 Y_2 = \eta \otimes J_{20}$$

(2) Apply V2, which assumes homoscedasticity, within-visit correlation = $\rho + \gamma$ and between visit correlation = ρ.The model is referred as model 4c.

$$\Sigma Y_1 Y_1 = \Sigma Y_2 Y_2 = (\rho + \gamma) * \sigma^2 \otimes J_{20} + (1 - \rho - \gamma) * \sigma^2 \otimes I_{20}$$
$$\Sigma Y_1 Y_2 = \Sigma^T Y_1 Y_2 = \rho * \sigma^2 \otimes J_{20}$$

(3) Apply CS structure. The model is referred as model 4d.

$$V=\begin{bmatrix} \Sigma Y_1 Y_1 & \Sigma Y_1 Y_2 \\ \Sigma^T Y_1 Y_2 & \Sigma Y_2 Y_2 \end{bmatrix} = \rho * \sigma^2 \otimes J_{40} + (1-\rho) * \sigma^2 \otimes I_{40}$$

Statistical Model

The expectation of throw score is:

Table 4b. Expectation of Post-treatment Throw Score

	Annotation	Sequence 1	Sequence 2
Visit 1	u_{i1}	$\mu + \pi_1 + \beta * \overline{X}_{i1} + \tau_{tms}$	$\mu + \pi_1 + \beta * \overline{X}_{i1} + \tau_{sham}$
Visit 2	u_{i2}	$\mu + \pi_2 + \beta * \overline{X}_{i2} + \tau_{sham}$	$\mu + \pi_2 + \beta * \overline{X}_{i2} + \tau_{tms}$

Where the terms are:

μ, an intercept, the mean of post-treatment throw score after adjusting for baseline mean score;

π_1 , π_2 are the fixed effect associated with visit 1 and visit 2

β is the change in expectation of Y_{ijm} with 1 unit change in $\overline{X}_{ij}$

τ_{TMS} is the effect associated with TMS and τ_{sham} is the effect associated with Sham.

The model is:

$$Y_{\text{ijm}} = \beta_0 + \beta_1 * I(visit = 1) + \beta_2 * I(treamtent = \text{"Sham"}) + \beta_3 * \overline{X}_{ij} + \epsilon_{ijm}, \epsilon_{ijm} \sim V$$

Where β_0 is the mean of a throw score at visit 2 treatment TMS, after adjusting for baseline mean throw score. $\beta_1 = \pi_1 - \pi_2$, $\beta_2 = \tau_{Sham} - \tau_{TMS}$. β_1 estimates the fluctuation of post-treatment score. β_2 estimates the difference between the effect of Sham and the effect of TMS. β_3 estimates the change in expectation of Y_{ijm} with 1 unit change in the baseline.

Result

1. Descriptive Statistics

Table 5 presents the comparison of subject characteristics by sequence and overall. Subject characteristics collected by study investigators are height, weight, age, race, baseball level and the year since last time played baseball. Based on the "year since last time played baseball", subjects are divided into played within one year or longer than one year. The subjects' mean height is 5.95 feet, weight 189.04 pounds, age 22.32 years. 12% of the subjects are Asian, 28% are African American, and 60% are Caucasian. 18% of the subjects' baseball level are "high school", 20% are "high school and club". 52% are "college and/or above". 32% of the subjects had played baseline within 1 year to the starting date of the study. The randomization successfully balanced those subject characteristics among the two sequences, all p values >0.196.

Table 6 presents the summary statistics of experiment characteristics by treatment group. Experiment characteristics are time-dependent across visit, and varies across subjects. Temperature is the same within-visit at baseline throws and post-treatment throws, but different at the two visits. Temperature of this study was varying because subjects were tested on different dates outdoor. The temperatures in which subjects completed throws are significantly different between the visit to receive TMS treatment and the visit to receive Sham treatment, p=0.0008. Stimulation intensity is the MRI intensity delivered to the subject, for TMS it is 90%

of the subject's specific threshold, for Sham it is 65%. The significant difference in stimulation intensity among the treatment groups is caused by the study design. "Minute" is the time from finished treatment to the complete of the last post-treatment throw. In window period indicates whether all 20 post-treatment throws were completed during the window period of the treatment When one or more post-treatment throws were not completed in window period, the subject was labeled "not in window period" for that treatment.

Table 7 presents the summary statistic of mean throw score by visit, sequence and baseline or post-treatment. From mean values in Table 7, in Table 8 the crude estimations of $\delta = \tau_{\text{Sham}} - \tau_{\text{TMS}}$ are calculated. Without adjusting for baseline, the crude estimation of δ is 0.18 at sequence 1, 0.03 at sequence 2, 0.105 overall. When adjusting for baseline, the crude estimation of δ is 0.13 at sequence 1, 0.22 at sequence 2, 0.175 overall. The crude estimation of δ lies close to zero and thus does not suggest difference in effects of treatments. However, the crude estimation of δ varies among adjusting baseline or not, and also varies among the two sequences.

Table 5: Subject Characteristics by Randomization

Characteristic	overall (N=25)	Visit 1: TMS Visit 2: Sham (n=10)	Visit 1: Sham Visit 2: TMS (n=15)	P value
Height(feet)	5.95(0.23)	5.97 (0.23)	5.93 (0.24)	0.912
Weight(lbs)	189.04(34.05)	191.40 (32.78)	187.50 (35.92)	0.741
Age(year)	22.32(4.13)	22.20 (4.00)	22.40 (4.36)	0.823
Race				
African American	3(12)	1(10)	2(13)	0.999
Asian	7(28)	3(30)	4(26)	
Caucasian	15(60)	6(60)	9(60)	
Baseball level				
1: HS	7(18)	5(50)	2(13)	0.196
2: HS & club	5(20)	1(10)	4(26)	
3: College +	13(52)	4(40)	9(60)	
Experience				
Yes	8(32)	2(20)	6(40)	0.402
No	17(68)	8(80)	9(60)	

■For numeric characteristic, mean(sd) and p value from Wilcoxon rank-rum test are presented.

■For categorical characteristic, frequency(%) and p value from Fisher exact test are presented.

■Baseball level: 1： high school; 2: high school and club; 3: college or above.

■Experience: Yes: played baseball within a year; No: played baseball a year beyond.

■P value show that randomization successfully balanced subject characteristics.

Table 6. Experiment Characteristics

	TMS visit					Sham visit					P value
	mean	sd	minimum	median	maximum	mean	sd	minimum	median	maximum	
Temperature	65.68	9.01	50.00	63.00	83.00	59.32	11.79	39.00	58.00	84.00	0.008
Stimulation Intensity	48.96	8.07	35.10	50.40	69.30	35.36	5.83	25.35	36.40	50.05	<0.001
Minute	8.52	3.54	3.00	8.00	17.00	7.64	2.41	3.00	7.00	12.00	0.094
In window period											
Yes			23(92)					23(92)			0.999
No			2(8)					2(8)			

■Mean, sd, min, median, and max are presented for continuous characteristics, frequency(%) is presented for categorical characteristics

■Temperature, stimulation intensity and minute are tested by paired t-test

■In window period is tested by Fisher's exact test.

■Minute: time from finishing treatment to complete last post-treatment throw

■In window period: Yes: all post-treatment throws were completed in window period of treatment. No: at least one post-treatment throw was not completed in window period. The subjects not in window period of TMS treatment are not the same as the subjects not in window period of Sham treatment

Table 7. Summary Statistics of Mean Throw Scores

Type	visit	sequence	annotation	mean	sd	min	median	max
baseline	1	1	$\overline{X}_{i1}$	3.06	1.13	0.85	3.10	4.38
baseline	1	2		3.21	0.94	1.65	2.85	4.80
baseline	2	1	$\overline{X}_{i2}$	3.11	1.20	0.58	3.24	4.53
baseline	2	2		3.40	0.89	1.63	3.38	4.53
post-treatment	1	1	$\overline{Y}_{i1}$	2.80	1.35	0.03	3.16	4.08
post-treatment	1	2		3.19	0.83	2.08	3.40	4.53
post-treatment	2	1	$\overline{Y}_{i2}$	2.98	1.18	0.18	3.08	4.43
post-treatment	2	2		3.16	0.71	1.73	3.25	4.20

Table 8. Crude Estimation of $\boldsymbol{\delta = \tau_{sham} - \tau_{tms}}$

		Denotation	Sequence 1		Sequence 2		Overall
			treatment	mean	treatment	mean	
Visit 1	Baseline	$\overline{X}_{i1}$	---	3.06	---	3.21	3.135
	Post-treatment	$\overline{Y}_{i1}$	TMS	2.8	Sham	3.19	2.995
	Mean change	$\overline{D}_{i1}$	TMS	-0.26	Sham	-0.02	-0.14
Visit 2	Baseline	$\overline{X}_{i2}$	---	3.11	---	3.40	3.255
	Post-treatment	$\overline{Y}_{i2}$	Sham	2.98	TMS	3.16	3.07
	Mean change	$\overline{D}_{i2}$	Sham	-0.13	TMS	-0.24	-0.185
$\hat{\delta}$				0.18		0.03	0.105
$\hat{\delta}$ (baseline adjust)				0.13		0.22	0.175

■ $\hat{\delta} = \text{mean}(\overline{Y}_{i2}) - \text{mean}(\overline{Y}_{i1})$ for sequence 1, $\text{mean}(\overline{Y}_{i1}) - \text{mean}(\overline{Y}_{i2})$ for sequence 2.

■ $\hat{\delta}(\text{baseline adjust}) = (\text{mean}(\overline{Y}_{i2}) - \text{mean}(\overline{X}_{i2})) - (\text{mean}(\overline{Y}_{i1}) - \text{mean}(\overline{X}_{i1}))$ for sequence 1.

$(\text{mean}(\overline{Y}_{i1}) - \text{mean}(\overline{X}_{i1})) - (\text{mean}(\overline{Y}_{i2}) - \text{mean}(\overline{X}_{i2}))$ for sequence 2.

■ "Overall" shows the unweighted mean among the two sequences.

■ Mean of $\overline{D}_{ij}$ can be regarded crude estimation of the effect of Sham or the effect of TMS

2. Model Result

In modeling, variable “sequence” is added in to test whether the assumptions required for crossover design all hold.

Table 9a is the result of method 1 joint modeling, with mean throw score as dependent variable, using the model recommended by Kenward. The factors in model are “sequence”, “interval” and “treatment”. Factor “sequence” tests whether the assumptions for crossover design, mentioned in section 4.2 hold. Factor interval estimates whether mean throw score fluctuates across intervals. Factor “treatment” estimates the difference among effect of Sham and effect of TMS. For model 1b with the covariance structure recommended by Kenward[5], which assumes heteroscedasticity for baseline and post-treatment measures and within-visit covariance is stronger than between -visit association, SAS warns G matrix is not positive definite. The models shown in this table are consistent that: the non-significant of factor “sequence” indicates the assumptions hold; the mean score does not statistical significantly fluctuate across interval; there is no statistical significant difference among the effect of Sham and the effect of TMS.

Table 9b is the result of method 1 joint modeling, with individual throw score as dependent variable, using the model Kenward recommended. Factors in model are the same as those in Table 9a. For model 1e with covariance structure V3, which assumes between visit covariance is η, within-visit yet not within interval covariance is $\eta + \sigma$, and within interval covariance is

$\eta + \sigma$ plus the terms from TOEP structure, SAS stops and warns too much likelihood evaluations. The models shown in this table are consistent that: the non-significant of factor "sequence" indicates the assumptions hold; the throw score does not statistical significantly fluctuate across interval; there is no statistical significant difference among the effect of Sham and the effect of TMS.

Table 9c is the result of method 1 joint modeling, with individual throw score as dependent variable, and apply the alternative model. The terms in model are different from the model presented in Table 9a and Table 9b. The factors in model are "sequence", "interval", "treatment", "type" and "type by treatment interaction". Factor "sequence" test whether the assumptions for crossover design, factor "visit" estimates whether throw score fluctuates across visits, factor "type" tests whether post-treatment score is different from baseline score, and "type by treatment interaction" estimates the difference on the effect of receiving treatment between Sham and TMS. For model 1h with covariance structure V3, SAS stops and warns too much likelihood evaluations. The models shown in this table are consistent that: the non-significant of factor "sequence" indicates the assumptions hold; the non-significant of type by treatment interaction suggests there is no statistical significant difference among the effect of Sham and the effect of TMS.

Table 10 is the result of method 2 ignore baseline, with dependent variable as mean post-treatment scores or individual post-treatment scores. The factors in model are "sequence", "visit"

and “treatment”. Factor “sequence” test whether the assumptions for crossover design all holds. Factor “visit” estimates whether throw score or mean throw score fluctuates across visits. Factor “treatment” estimates the difference among effect of Sham and effect of TMS. The models shown in this table are consistent that: the non-significant of factor “sequence” indicates the assumptions hold; the throw score or mean score does not statistical significantly fluctuate across visits; there is no statistical significant difference among the effect of Sham and the effect of TMS.

Table 11 is the result of method 3 change from baseline, with dependent variable as mean change score or change score. Factors in model are the same as those in Table 10. Model 3a which assumes CS structure of the mean changes collapses into model 3b, as the estimated correlation between $\overline{D}_{i1}$ and $\overline{D}_{i2}$ is 0. Model 3c, which assume the covariance between-visit is η, and within-visit covariance is η plus the terms from TOEP structure, estimated $\eta = 0$ and SAS warns G matrix is not positive definite. Model 3d, which assume between visit correlation is ρ and within visit correlation is $\rho + \gamma$ collapses into model 3f, as estimated $\rho = 0$. The estimated covariance structures uniformly suggest, for this study, when incorporating change score as dependent variable, between-visit covariance or correlation is 0. The models shown in this table are consistent that: the non-significant of factor “sequence” indicates the assumptions hold; the change score or mean change score does not statistical significantly fluctuate across visits; there is no statistical significant difference among the effect of Sham and the effect of TMS.

Table 12 is the result of method 4 visit specific baseline as covariate, with dependent variable as mean post-treatment scores or individual post-treatment scores. Besides the factors which are the same as Table 10, variable “baseline” is added in model. The beta coefficient of baseline estimates the change in expectation of throw score or mean throw score with 1 unit change in baseline mean score from the same visit. The models shown in this table are consistent that: the non-significant of factor “sequence” indicates assumption (1) and (3) hold; the throw score or mean throw score does not statistical significantly fluctuate across visits; there is no statistical significant difference among the effect of Sham and the effect of TMS. Post-treatment mean throw score or throw score are significantly correlated to baseline mean score from the same visit.

Table 9a. Statistical Result for Method 1—Joint Modeling

Dependent variable	Model 1a $[\overline{X}_{i1} \ \overline{Y}_{i1} \ \overline{X}_{i2} \ \overline{Y}_{i2}]_{1\times4}^{T}$	Model 1b $[\overline{X}_{i1} \ \overline{Y}_{i1} \ \overline{X}_{i2} \ \overline{Y}_{i2}]_{1\times4}^{T}$	Model 1c $[\overline{X}_{i1} \ \overline{Y}_{i1} \ \overline{X}_{i2} \ \overline{Y}_{i2}]_{1\times4}^{T}$	Model 1d $[\overline{X}_{i1} \ \overline{Y}_{i1} \ \overline{X}_{i2} \ \overline{Y}_{i2}]_{1\times4}^{T}$
Covariance Structure	UN	V1	V2	CS
AIC	210.3	205.6	202.0	200.3
P value of variables in model				
Sequence	0.5309	0.5355	0.5106	0.5106
Interval	0.4580	0.4736	0.4211	0.4534
Treatment	0.4078	0.4303	0.4497	0.4301
Estimated least square means				
Sham(SE)	3.0820(0.1996)	3.0808(0.2033)	3.0824(0.2050)	3.0853(0.2051)
TMS(SE)	2.9907(0.2052)	2.9903(0.2033)	2.9883(0.2050)	2.9853(0.2051)
Estimated δ				
$\hat{\delta}(SE)$	0.0913(0.1083)	0.0905(0.1239)	0.0942(0.1239)	0.1000(0.1260)
P value	0.4078	0.4303	0.4497	0.4301

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Interval: 1: visit 1 baseline; 2: visit 1 post-treatment; 3: visit 2 baseline; 4: visit 2 post-treatment.
- Treatment: 0: baseline, no actual treatment received; 1: TMS received; 2: Sham received.
- Model 1b's estimated G matrix is not positive definite, which suggests the covariance structure proposed does not fit the data.
- In Model 1c, estimated within-visit correlation is 0.8017, between-visit correlation is 0.8167.
- In Model 1d, estimated correlation is 0.8116.
- Least square means for Sham is estimated by contrast: intercept 1 sequence 0.5 0.5 interval 0 0.5 0 0.5 treatment 0 1 0;

Least square means for Sham is estimated by contrast: intercept 1 sequence 0.5 0.5 interval 0 0.5 0 0.5 treatment 0 0 1;

Table 9b. Statistical Result for Method 1—Joint Modeling

	Model 1e	Model 1f	Model 1g
Dependent variable	$[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$	$[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$	$[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$
Covariance Structure	V3	V2	CS
AIC	\	8287.0	8285.0
P value of variables in model			
Sequence	\	0.5086	0.5086
Interval	\	0.4106	0.4222
Treatment	\	0.4120	0.4100
Estimated least square means			
Sham(SE)	\	3.0847(0.2039)	3.0853(0.2041)
TMS(SE)	\	2.9860(0.2039)	2.9853(0.2041)
Estimated δ			
$\hat{\delta}(SE)$	\	0.0987(0.1202)	0.1000(0.1213)
P value	\	0.4120	0.4100

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Interval: 1: visit 1 baseline; 2: visit 1 post-treatment; 3: visit 2 baseline; 4: visit 2 post-treatment.
- Treatment: 0: baseline, no actual treatment received; 1: TMS received; 2: Sham received.
- Model 1e: SAS stopped because of too many likelihood evaluations.
- In Model 1f, estimated within-visit correlation is 0.1888, between-visit correlation is 0.1895.
- In Model 1g, estimated correlation is 0.1892.
- Least square means for Sham is estimated by contrast: intercept 1 sequence 0.5 0.5 interval 0 0.5 0 0.5 treatment 0 1 0;

Least square means for Sham is estimated by contrast: intercept 1 sequence 0.5 0.5 interval 0 0.5 0 0.5 treatment 0 0 1;

Table 9c. Statistical Result for Method 1—Joint Modeling

Dependent variable	Model 1h $[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$	Model 1i $[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$	Model 1j $[X_{i1} \quad Y_{i1} \quad X_{i2} \quad Y_{i2}]_{1\times 80}{}^{T}$
Covariance Structure	V3	V2	CS
AIC	\	8287.0	8285.0
P value of variables in model			
Sequence	\	0.5086	0.5086
Visit	\	0.2421	0.2475
Treatment	\	0.8666	0.8681
Type	\	0.0664	0.0664
Type*Treatment	\	0.1682	0.2813
Estimated least square means			
Sham baseline(SE)	\	3.1516(0.2038)	3.1516(0.2039)
Sham post-treatment(SE)	\	3.0878(0.2038)	3.0878(0.2039)
TMS baseline(SE)	\	3.2279(0.2038)	3.2279(0.2039)
TMS post-treatment(SE)	\	2.9829(0.2038)	2.9829(0.2039)
Estimated δ			
$\hat{\delta}(SE)$	\	0.1812(0.1682)	0.1812(0.1681)
P value	\	0.2813	0.2813

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Visit: 1: visit 1; 2: visit 2.
- Treatment: 1: TMS; 2: Sham.
- Type: 0: baseline; 1: post-treatment.

- Model 1h: SAS stopped because of too many likelihood evaluations.
- In Model 1i, estimated within-visit correlation is 0.1890, between-visit correlation is 0.1894.
- In Model 1j, estimated correlation is 0.1892.
- δ is estimated by contrast: type*treatment -1 1 1 -1.

Table 10. Statistical Result for Method 2—Ignore Baseline

	Model 2a	Model 2b	Model 2c	Model 2d
Dependent variable	$[\ \overline{Y}_{i1}\quad \overline{Y}_{i2}]_{1\times 2}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$
Covariance Structure	CS	V4	V2	CS
AIC	115.6	4190.6	4162.2	4160.6
P value of variables in model				
Sequence	0.4746	0.4481	0.4674	0.4674
Visit	0.5026	0.4907	0.4960	0.5357
Treatment	0.3732	0.3621	0.3640	0.4090
Estimated least square means				
Sham(SE)	3.0821(0.2030)	3.0615(0.2020)	3.0821(0.2030)	3.0821(0.2046)
TMS(SE)	2.9821(0.2030)	2.9598(0.2020)	2.9821(0.2030)	2.9821(0.2046)
Estimated δ				
$\hat{\delta}(sd)$	0.1000(0.1101)	0.1017(0.1116)	0.1000(0.1101)	0.1000(0.1211)
P value	0.3732	0.3621	0.3640	0.4090

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Visit: 1: visit 1; 2: visit 2.
- Treatment: 1: TMS received; 2: Sham received.
- In Model 2a, estimated correlation is 0.8529.
- In Model 2c, estimated within-visit correlation is 0.1871, between-visit correlation is 0.1942.
- In Model 2d, estimated correlation is 0.1907.

Table 11. Statistical Result for Method 3—Change from Baseline

	Model 3a	Model 3b	Model 3c	Model 3d	Model 3e	Model 3f
Dependent variable	$[\ \overline{D}_{i1}\ \ \overline{D}_{i2}]_{1\times 2}{}^{T}$	$[\ \overline{D}_{i1}\ \ \overline{D}_{i2}]_{1\times 2}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$
Covariance Structure	CS	VC	V4	V2	CS	V5
AIC	\	103.1	4177.4	\	4165.4	4149.8
P value of variables in model						
Sequence	\	0.7266	0.6874	\	0.6963	0.7251
Visit	\	0.7947	0.8067	\	0.6954	0.7935
Treatment	\	0.3594	0.3555	\	0.1664	0.3548
Estimated least square mean change						
Sham(SE)	\	-0.0752(0.1310)	-0.0969(0.1304)		-0.0752(0.1039)	-0.0752(0.1310)
TMS(SE)	\	-0.2467(0.1310)	-0.2673(0.1304)		-0.2467(0.1039)	-0.2467(0.1310)
Estimated δ	\					
$\hat{\delta}(sd)$	\	0.1715(0.1853)	0.1704(0.1844)	\	0.1715(0.1238)	0.1715(0.1852)
P value	\	0.3594	0.3555	\	0.1664	0.3548

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Visit: 1: visit 1; 2: visit 2.
- Treatment: 1: TMS received; 2: Sham received.
- In Model 3a, estimated correlation is 0, so the model collapse as Model 3b.
- In Model 3c, estimated G matrix is not positive definite, which suggests the covariance structure proposed does not fit the data.
- In Model 3d, estimated within-visit correlation is 0.0624, between-visit correlation is 0, so the model collapse as Model 3f.
- In Model 3e, estimated correlation is 0.0200.
- In Model 3f, estimated within-visit correlation is 0.0624.

Table 12. Statistical Result for Method 4—Period Specific Baseline as Covariate

	Model 4a	Model 4b	Model 4c	Model 4d
Dependent variable	$[\ \overline{Y}_{i1}\quad \overline{Y}_{i2}]_{1\times 2}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\quad Y_{i2}]_{1\times 40}{}^{T}$
Covariance Structure	CS	V4	V2	CS
AIC	102.0	4176.0	4148.7	4151.9
P value of variables in model				
Sequence	0.5250	0.4746	0.5185	0.4630
Visit	0.9059	0.9268	0.9048	0.9916
Treatment	0.3842	0.3683	0.3748	0.2441
$\overline{X}_{\mathrm{ij}}$ (Baseline)	<.0001	<.0001	<.0001	<.0001
Estimated least square means				
Sham(SE)	3.1269(0.1240)	3.1023(0.1231)	3.1269(0.1240)	3.1165(0.1207)
TMS(SE)	2.9713(0.1239)	2.9474(0.1230)	2.9713(0.1239)	2.9738(0.1206)
Estimated δ				
$\hat{\delta}(sd)$	0.1556(0.1752)	0.1548(0.1721)	0.1556(0.1752)	0.1427(0.1224)
P value	0.3842	0.3683	0.3748	0.2441

- The composition of dependent variable on one subject is illustrated.
- Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.
- Visit: 1: visit 1; 2: visit 2.
- $\overline{X}_{\mathrm{ij}}$ the mean of baseline throws from the same visit.
- Treatment: 1: TMS received; 2: Sham received.
- In Model 4a, estimated correlation is 0.0014.
- In Model 4c, estimated within-visit correlation is 0.0515, between-visit correlation is 0.0001.
- In Model 4d, estimated correlation is 0.0453.

Discussion

1. The Recommended Model

From the result tables, although in table 8 the crude estimations of δ varies among sequences, which suggests carryover effect and/or visit by treatment interaction, it is uniform that the test of assumptions of crossover design is not significant and does not suggest those assumptions are violated. Across all the models, it also suggests that there is no significant difference among the effect of Sham treatment and TMS treatment.

1.1 Model Mean Outcome or Model Individual Outcome

According to the residual diagnostic plots at the appendix, for this study, on general, for each method of handling baseline, the normality and homoscedasticity is better when modeling individual outcome than modeling mean outcome, so models modeling individual throw score or change score are preferred.

In each method of handling baseline, the model modeling individual outcome with smallest AIC is chosen. Those models are model 1g for method 1 Kenward model, model 1j for method 1 alternative model, model 2d for method 2, model 3f for method 3 and model 4c for method 4. One of those models will be recommended. Model 1g and model 1j apply CS structure on the 80 throw scores from the same subject. Model 2d applies CS structure to the 40 post-

treatment throw scores from the same subject. Model 3f applies the covariance structure on the 40 change scores from the same subject with V5, while V5 can be derived when assuming the 80 throw scores following CS. Those four models all suggests that CS structure is the best fit covariance structure of the 80 throw scores from the same subject.

1.2 The Chosen of Method of Handling Baseline

1.2.1 Ability of Test Carryover Effect

In method 4, an extra risk is faced that, unlike in other methods where the existence of carryover effect can be tested in model with variable "sequence", this method shouldn't be applied at all when there is carryover effect, because carryover effect contaminates the baseline of second visit ($\overline{X}_{i2}$) and makes it unsuitable to be adjusted as covariate[9]. When there is carryover effect, the estimation of beta coefficient, $\beta * \overline{X}_{ij}$ is wrong because half of $\overline{X}_{ij}$ is contaminated by carryover effect, and after adjusting for baseline under the wrong beta coefficient, variable "sequence" in the model cannot serve as test for carryover effect. That is, applying this method strictly requires it is sure there is no carryover effect because the opportunity of test for it in modeling is lost. In Table 8 "crude estimation of δ", no matter adjust baseline or not, the estimation among the two sequence varies, which cannot rule out carryover effect. As a result, model 4c with method 4 is not considered to be the best fit for this study.

1.2.2 Interpretative of the Model

The alternative method of joint-modeling, model 1j, estimates δ in a similar manner to method 3 change from baseline, and the estimated δ is 0.1812, which is in consistent with the weighted

mean of the crude estimation of δ among the two sequences, adjusting for baseline, $\frac{2}{5} \times 0.13 + \frac{3}{5} \times 0.22 = 0.184$. The drawback of this joint model method, is the term "treatment" stands for main effect of treatment which applies to both baseline throw where the treatments hadn't been given to the subjects and post-treatment throw, thus this term cannot be aliased to true effects existing in practice. Other methods will not have issue in interpretation the terms in model. Considering this drawback, this alternative method of joint modeling is not chosen.

1.2.3 Consistency with the Crude Estimation

Presenting the result of model 1g, model 2d and model 3f:

In model 1g, the result is:

Result—Model 1g		
Treatment		P value=0.4100
	Sham	3.0853 (0.2041)
	TMS	2.9853(0.2041)
		(Sham-TMS)=0.1000, SE=0.1211
Interval		P value=0.4222
	1	3.1208(0.2037)
	2	2.9978(0.2041)
	3	3.2586(0.2037)
	4	3.0728(0.2041)
Sequence		P value=0.5086
	1	2.9867(0.2947)
	2	3.2383(0.2407)

■ P value for each factor in model is presented.
■ Mean(SE) of least squares mean at each factor level is presented.
■ For "treatment", difference (Sham-TMS) of least squares mean and SE are presented.
■ The p value of "sequence" does not suggest the assumptions of crossover design are violated.

The relative part in Table 8:

Relevant Part in Table 8. Crude Estimation of $\delta = \tau_{sham} - \tau_{tms}$							
		Denotation	Sequence 1		Sequence 2		Overall
			treatment	mean	treatment	mean	
Visit 1	Interval 1	$\overline{X}_{i1}$	---	3.06	---	3.21	3.135
	Interval 2	$\overline{Y}_{i1}$	TMS	2.8	Sham	3.19	2.995
Visit 2	Interval 3	$\overline{X}_{i2}$	---	3.11	---	3.40	3.255
	Interval 4	$\overline{Y}_{i2}$	Sham	2.98	TMS	3.16	3.07
$\hat{\delta}$				0.18		0.03	0.105

- unweighted mean after TMS is 1/2*(2.8+3.16)=2.98
- unweighted mean after Sham is 1/2*(2.98+3.19)=3.085
- unweighted mean at sequence 1 is 1/4*(3.06+2.80+3.11+3.98)=2.988
- unweighted mean at sequence 2 is 1/4*(3.21+3.19+3.40+3.16)=3.24

The result is in consistent with the part in Table 8 related to this model, the estimation of model 1g is in consistent with the crude estimation. Least square means are consistent to the unweighted means too.

In model 2d, the result is:

Result—Model 2d				
Treatment	P value=0.4090			
	Sham	3.0821 (0.2046)	TMS	2.9821 (0.2046)
	(Sham-TMS)=0.1000, SE=0.1211			
Visit	P value=0.5357			
	1	2.9946(0.2046)	2	3.0696(0.2046)
Sequence	P value=0.4674			
	1	2.8900(0.3028)	2	3.1742(0.2472)

- P value for each factor in model is presented.
- Mean(SE) of least squares mean at each factor level is presented.
- For “treatment”, difference (Sham-TMS) of least squares mean and SE are presented.
- The p value of “sequence” does not suggest the assumptions of crossover design are violated.

The relevant part in Table 8:

Relevant Part in Table 8. Crude Estimation of $\delta = \tau_{sham} - \tau_{tms}$							
		Denotation	Sequence 1		Sequence 2		Overall
			treatment	mean	treatment	mean	
Visit 1	Post-treatment	$\overline{Y}_{i1}$	TMS	2.8	Sham	3.19	2.995
Visit 2	Post-treatment	$\overline{Y}_{i2}$	Sham	2.98	TMS	3.16	3.07
$\hat{\delta}$				0.18		0.03	0.105

■ $\hat{\delta} = \text{mean}(\overline{Y}_{i2}) - \text{mean}(\overline{Y}_{i1})$ for sequence 1, $\text{mean}(\overline{Y}_{i1}) - \text{mean}(\overline{Y}_{i2})$ for sequence 2.

■ mean of $\overline{Y}_{ij}$ after TMS is 1/2*(2.8+3.16)=2.98

■ mean of $\overline{Y}_{ij}$ after Sham is 1/2*(2.98+3.19)=3.085

■ mean of $\overline{Y}_{ij}$ at sequence 1 is 1/2*(2.98+2.80)=2.89

■ mean of $\overline{Y}_{ij}$ at sequence 2 is 1/2*(3.19+3.16)=3.175

The result is in consistent with the part in Table 8 related to this model, the estimation of model 2d is in consistent with the crude estimation. Least squares means are consistent to the unweighted means too.

In model 3f, the result is:

Result Model 3f				
Treatment		P value=0.3548		
	Sham	-0.0752 (0.1310)	TMS	-0.2467 (0.1310)
		(Sham-TMS)=0.1715, SE=0.1853		
Visit		P value=0.7947		
	1	-0.1367(0.1310)	2	-0.1852(0.1310)
Sequence		P value=0.7251		
	1	-0.1935(0.1435)	2	-0.1283(0.1172)

■ P value for each factor in model is presented.

■ Mean(SE) of least squares mean at each factor level is presented.
■ For "treatment", difference (Sham-TMS) of least squares mean and SE are presented.
■ The p value of "sequence" does not suggest the assumptions of crossover design are violated.

The relevant part in Table 8:

Relevant Part in Table 8. Crude Estimation of $\delta = \tau_{sham} - \tau_{tms}$

		Denotation	Sequence 1		Sequence 2		Overall
			treatment	mean	treatment	mean	
Visit 1	Post-treatment	$\overline{D}_{i1}$	TMS	-0.26	Sham	-0.02	-0.14
Visit 2	Post-treatment	$\overline{D}_{i2}$	Sham	-0.13	TMS	-0.24	-0.185
$\hat{\delta}$				0.13		0.22	0.175

■ $\hat{\delta} = \text{mean}(\overline{D}_{i2}) - \text{mean}(\overline{D}_{i1})$ for sequence 1, $\text{mean}(\overline{D}_{i1}) - \text{mean}(\overline{D}_{i2})$ for sequence 2.
■ mean of $\overline{D}_{ij}$ after TMS is 1/2*(-0.26-0.24)= -0.25
■ mean of $\overline{D}_{ij}$ after Sham is 1/2*(-0.02-0.13)= -0.075
■ mean of $\overline{D}_{ij}$ at sequence 1 is 1/2*(-0.26-0.13)= -0.195
■ mean of $\overline{D}_{ij}$ at sequence 2 is 1/2*(-0.02+-0.24)= -0.13

The result is in consistent the part in Table 8 related to this model, the estimation of model 4c is in consistent with the crude estimation. Least squares means are consistent to the un-weighted means too.

Model 1g, model 2d and model 3f are all in consistent with the crude estimations shown in Table 8.

1.2.4 Residual Diagnostic

The residual diagnostics for model 1g are in Figure 5 series, for model 2d are in Figure 9 series,

and for model 3f are in Figure 13 series. Model 1g and model 2d applies CS structure, so this is equivalent to add in a random intercept of "subject" in model. Their conditional residual does not suggest any violation on normality or homoscedasticity. Model 3f's residual diagnostics does not suggest violation on normality or homoscedasticity. Diagnostic result of all these three models does not suggest evidence against them in terms of homoscedasticity and normality.

1.2.5 Final Recommended Model: 3f

The model recommended for joint modeling, model 1g assumes the throws scores follow CS structure, and between-visit correlation is the same as within-in visit correlation. It does not satisfy the criteria proved by Xun Chen et al. -- when within-visit correlation is stronger than cross-visit correlation, joint modeling is more efficient in estimation of δ, so model 1g is not recommended as the best fit model. Between model 2d and model 3f, model 3f is recommended as method 2 mainly serves as a reference and it is not appropriate to totally discard the baseline measures when there is no evidence against them.

Recommendation of Model	
Model	Rank
1g	2
2d	3
3f	1

Summary of Models

Model	Outcome	Method	AIC	Reason why it is not recommended
Model 1a	Mean	1	210.3	Diagnostic plots (Figure 1 series) suggest normality assumption may be violated.
Model 1b	Mean	1	205.6	G matrix is not positive definite.
Model 1c	Mean	1	202.2	Not the smallest AIC among comparable models, and the covariance structure estimated is actually close to Model 1d
Model 1d	Mean	1	200.3	No evidence directly against this model, but under CS structure, literature does not recommend joint modeling over other methods, and modeling individual outcome is preferred.
Model 1e	Individual	1	\	SAS stops because too many likelihood estimations.
Model 1f	Individual	1	8287.0	Not the smallest AIC among comparable models, and the covariance structure estimated is actually close to Model 1g
Model 1g	Individual	1	8285.0	Recommended rank 2, rank 2 because under CS structure, literature does not recommend joint modeling over other methods.
Model 1h	Individual	1	\	SAS stops because too many likelihood estimation, and interpretability of terms in model is not as clear as other methods.
Model 1i	Individual	1	8287.0	Interpretability of terms in model is not as clear as other methods.
Model 1j	Individual	1	8285.0	Interpretability of terms in model is not as clear as other methods.
Model 2a	Mean	2	115.6	Diagnostic plots (Figure 6 series) suggest normality assumption may be violated.
Model 2b	Individual	2	4190.6	Not the smallest AIC among comparable models.
Model 2c	Individual	2	4162.2	Not the smallest AIC among comparable models, and the covariance structure estimated is actually close to Model 2d.
Model 2d	Individual	2	4160.6	Recommended rank 3. Rank 3 because in study design with baseline implemented in each period, ignore baseline serves as a reference only.
Model 3a	Mean	3	\	Collapse into model 3b
Model 3b	Mean	3	103.1	Diagnostic plots (Figure 10 series) suggest normality assumption may be violated.

Model 3c	Individual	3	4177.4	G matrix is not positive definite.
Model 3d	Individual	3	\	Collapse into model 3f
Model 3e	Individual	3	4165.4	Not the smallest AIC among comparable models, and model 3d collapse into model 3f rather than model 3e.
Model 3f	Individual	3	4149.8	Recommended rank 1.
Model 4a	Mean	4	102.0	Unable to test for carryover effect, and diagnostic plots (Figure 13 series) suggest homoscedasticity assumption may be violated.
Model 4b	Individual	4	4176.0	Unable to test for carryover effect.
Model 4c	Individual	4	4147.7	Unable to test for carryover effect.
Model 4d	Individual	4	4151.9	Unable to test for carryover effect.

2. Covariate Adjustment

The next step is to figure out whether it is necessary to consider adjusting any subject level characteristics or experiment characteristics in modeling.

Regarding to the estimation of δ, as the study is a crossover design completed without missing, δ is estimated mainly basing on within-subject information, so subject level characteristics which associate with the outcome, yet do not interact with effects associated with treatments, do not have to be adjusted in model[9]. A subject level characteristic should be adjusted if it interacts with the effects associated with treatments[9]. As this design incorporates baseline measure in each period, crude estimations of $\tau_{\mathrm{TMS}}, \tau_{SHam}$ and δ are available on each subject. Table 13 presents the correlation or association between subject level characteristics and crude estimations of $\tau_{\mathrm{TMS}}, \tau_{SHam}$ and δ. No subject level characteristic statistical significantly associated with any of the crude estimations. Then, it is not necessary to adjust them in model.

Few reference can be found regarding to time dependent variables in crossover design. This is partially due to the fact that crossover designs are experimental study so that ideally time dependent characteristics are controlled. Moreover, for crossover design which does not measure baseline at each period, those time depended characteristics which varies across time and subjects are unable to investigate into when considering between subject variability. This design incorporates baseline measurement in each period, thus makes it possible to investigate

into time-dependent characteristics through descriptive statistics. As on subject level, time-dependent characteristics can associate to the outcome, or influence the effects associated to treatments, thus can bias or interacts with the estimation of δ. In this study, three of four time dependent experiment characteristics, stimulation intensity of the treatment, window period and time from receiving treatment to complete the last post-treatment throw, are characteristics of the treatment which are only possible to influence the effects associated with treatments. Temperature is repeated only twice on each subject, the data is in sufficient to distinguish on subject level that whether it associates with throw score. As a result, it is hypothesized that temperature may interact with effects associated with treatments. Table 14 presents the correlation or association between experiment characteristics and crude estimation of τ_{TMS} and τ_{Sham}. Temperature is significantly correlated with the crude estimation of τ_{TMS}.

From the result of correlation and association, it is hypothesized that temperature interacts with effects associated with treatments, and it shall be adjusted in modeling. To better understand its role, adjust temperature in all the three models 1g, 2d, and 3f rather than only model 3f. Temperature and its interaction with treatment are added in model. If the interaction term is not significant while main effect is significant, then it is more likely that temperature is actually associated with throw score rather than interacts with treatments. That is, temperature biases the estimation of δ, rather than interacts with it. Result is shown in Table 15a.

In the models except joint modeling, the insignificant of the interaction between temperature

and treatments suggests temperature do not interacts with effects associates with treatment. Investigates into the parameters from joint modeling, model 1g in Table 15a in detail: temperature is only significantly correlated to baseline scores, and after receiving treatments, temperature no longer correlates with throw score. The result of model 1h is actually in consistent with model 2d and model 3f.

Estimated Effect of Temperature from Model 1g			
Temperature	Baseline Throws	Throws after TMS	Throws after Sham
Estimated effect(p)	0.01789 (0.0275)	-0.00769(0.5375)	-0.00989(0.3410)

From the result shown in Table 15a and the beta estimated from model 1g. There is an interaction between temperature and baseline or post-treatment. Temperature influences baseline throw scores yet not post-treatment throw scores. This also can explain why temperature is associated the crude estimation in Table 14, and why in Table 8 the crude estimation of δ for varies among adjusting baseline or not. Table 15b presents the result of deleting the interaction term in model 2d and model 3f.

As suggested, temperature may influence baseline throw score yet not post-treatment throw score, then, it shall be adjusted in the model picked up previously, model 3f. However, at each temperature point, there is actually few subjects, so the estimation of the effect of temperature on change score will be unprecise. Another way to deal with this problem, is to use model 2d from method 2 ignore baseline, as there is evidence that baselines may be contaminated by temperature, but post-treatment throws are not.

Re-check if any subject characteristic is associated to the crude estimation of δ without adjusting the baseline. Table 16 shows no subjects level characteristic is associated with the crude estimation of δ without adjusting for baseline. From the result of Table 16, it is reasonable to use model 2d shown in Table 10 for the analysis of this data, without adjusting for any other covariates.

Table 13. Correlation or Association between Subject Characteristics and Mean Change in Throw Score

Correlation(r)	$\hat{\tau}_{TMS}$ $r(p\ value)$		$\hat{\tau}_{Sham}$ $r(p\ value)$		$\hat{\delta}$ $r(p\ value)$	
Height	-0.328(0.110)		-0.037(0.864)		0.256(0.219)	
Weight	-0.161(0.444)		-0.256(0.216)		0.017(0.936)	
Age	0.051(0.812)		0.073(0.730)		0.075(0.723)	
	$\hat{\tau}_{TMS}$		$\hat{\tau}_{Sham}$		$\hat{\delta}$	
Association	**mean(sd)**	**p value**	**mean(sd)**	**p value**	**mean(sd)**	**p value**
Race(n)						
African American(3)	0.050(0.28)	0.258	-0.383(0.58)	0.626	-0.425(0.86)	0.345
Asian(7)	-0.586(0.57)		-0.043(0.60)		0.543(0.93)	
Caucasian(15)	-0.145(0.72)		-0.010(0.61)		0.135(1.00)	
Baseball level(n)						
High School(7)	-0.054(0.87)	0.653	-0.182(0.70)	0.825	-0.129(1.04)	0.598
High School+club(5)	-0.400(0.95)		0.021(0.48)		0.421(1.26)	
College/above(13)	-0.288(0.41)		-0.033(0.61)		0.256(0.85)	
Experience(n)						
No(17)	-0.199(0.77)	0.620	-0.069(0.64)	0.952	0.130(1.06)	0.707
Yes(8)	-0.344(0.37)		-0.053(0.54)		0.291(0.79)	

■ Spearman correlation is presented for continuous characteristics as r(p value).

■ The mean and sd of $\hat{\tau}_{TMS}$, $\hat{\tau}_{Sham}$ **and** $\hat{\delta}$ **in** each level of categorical characteristics is presented, p value for association is acquired from one-way ANOVA.

■Experience: yes: played baseball within a year; no: played baseball a year beyond.

■Notation:

$\hat{\tau}_{TMS}$: $(\overline{Y}_{i1} - \overline{X}_{i1})$ for subjects in sequence 1; $(\overline{Y}_{i2} - \overline{X}_{i2})$ for subjects in sequence 2.

$\hat{\tau}_{Sham}$: $(\overline{Y}_{i2} - \overline{X}_{i2})$ for subjects in sequence 1; $(\overline{Y}_{i1} - \overline{X}_{i1})$for subjects in sequence 2.

$\hat{\delta}$: $(\hat{\tau}_{Sham} - \hat{\tau}_{TMS})$.

Where, for subject i, $\overline{X}_{i1}$ is the mean of baseline throw scores of visit 1; $\overline{Y}_{i1}$ is the mean of post-treatment throw scores of visit 1; $\overline{X}_{i2}$ is the mean of baseline throw scores of visit 2; $\overline{Y}_{i2}$ is the mean of post-treatment throw scores of visit 2.

■Conclusion: No significant association between mean change of throw score with subject characteristics. No significant association between difference in mean change of throw score (Sham-TMS) with subject characteristics.

Table 14. Correlation or Association between Experiment Characteristics and Mean Change of Throw Score

Correlation(r)	$\hat{\tau}_{\mathbf{TMS}}$ $r(p\ value)$	$\hat{\tau}_{Sham}$ $r(p\ value)$
Temperature	-0.606(0.001)	-0.339(0.097)
Stimulation Intensity	-0.002(0.994)	0.101(0.636)
Minute	-0.211(0.316)	-0.119(0.576)

	$\hat{\tau}_{\mathbf{TMS}}$			$\hat{\tau}_{Sham}$		
Association	n	mean(sd)	P value	n	mean(sd)	p value
In window period						
Yes	23	-0.29(0.67)	0.268	23	-0.08(0.62)	0.693
No	2	0.26(0.27)		2	0.10(0.07)	

■Spearman correlation is presented for continuous characteristics as r(p value).

■The mean and sd of $\hat{\tau}_{TMS}$, $\hat{\tau}_{Sham}$ in each level of categorical characteristics is presented, p value for association is acquired from one-way ANOVA.

■Minute: time from finishing treatment to complete last post-treatment throw.

■In window period: yes: all post-treatment throws were completed in window period of treatment; no: at least one post-treatment throw was not completed in window period.

■Notation:

$\hat{\tau}_{TMS}$: $(\overline{Y}_{i1} - \overline{X}_{i1})$ for subjects in sequence 1; $(\overline{Y}_{i2} - \overline{X}_{i2})$ for subjects in sequence 2.

$\hat{\tau}_{Sham}$: $(\overline{Y}_{i2} - \overline{X}_{i2})$ for subjects in sequence 1; $(\overline{Y}_{i1} - \overline{X}_{i1})$for subjects in sequence 2.

Where, for subject i, $\overline{X}_{i1}$ is the mean of baseline throw scores of visit 1; $\overline{Y}_{i1}$ is the mean of post-treatment throw scores of visit 1; $\overline{X}_{i2}$ is the mean of baseline throw scores of visit 2; $\overline{Y}_{i2}$ is the mean of post-treatment throw scores of visit 2.

■Conclusion: Temperature is significantly associated with mean change of throw score in TMS treatment.

Table 15a. Statistical Result for Adjusting Temperature

	Model 1g	Model 2d	Model 3f
Dependent variable	$[\ X_{i1}\ \ Y_{i1}\ \ X_{i2}\ \ Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\ \ Y_{i2}]_{1\times 40}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$
Covariance structure	CS	CS	V5
AIC	8295.7	4171.6	4148.7
P value of variables in model			
Sequence	0.4436	0.7767	0.1471
Visit or Interval	0.6479	0.2566	0.9122
Treatment	0.8444	0.5650	0.4115
Temperature	0.9889	0.1634	0.0004
Temperature*Treatment	0.0041	0.5666	0.4188
Estimated δ			
$\hat{\delta}(sd)$	0.1891(0.9635)	0.6110(1.0616)	-0.9892(1.1929)
P value	0.8444	0.5650	0.4188

■ The composition of dependent variable on one subject is illustrated.

■ Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.

■ Visit: 1: visit 1; 2: visit 2.

■ Interval: 1: visit 1 baseline; 2: visit 1 post-treatment; 3: visit 2 baseline; 4: visit 2 post-treatment.

■ Treatment:

In Model 2c: 0: baseline, no actual treatment -received; 1: TMS received; 2: Sham received. Treatment=”TMS” is set to be reference.

In other models: 1: TMS received; 2: Sham received.

Table 15b. Statistical Result for Adjusting Temperature

	Model 1g	Model 2d	Model 3f
Dependent variable	$[\ X_{i1}\ Y_{i1}\ \ X_{i2}\ \ Y_{i2}]_{1\times 40}{}^{T}$	$[\ Y_{i1}\ \ Y_{i2}]_{1\times 40}{}^{T}$	$[\ D_{i1}\ \ D_{i2}]_{1\times 40}{}^{T}$
Covariance structure	CS	CS	V5
AIC	8295.7	4171.6	4144.7
P value of variables in model			
Sequence	0.4436	0.7722	0.1322
Visit or Interval	0.6479	0.3267	0.7550
Treatment	0.8444	0.9561	0.8809
Temperature	0.9889	0.1291	0.0005
Temperature*Treatment	0.0041	-	-
Estimated δ			
$\hat{\delta}(sd)$	0.1891(0.9635)	0.0075(0.1354)	-0.0258(0.1714)
P value	0.8444	0.9561	0.8809

■ The composition of dependent variable on one subject is illustrated.

■ Sequence: 1: receive TMS in first visit, Sham in second visit; 2: receive Sham in first visit, TMS in second visit.

■ Visit: 1: visit 1; 2: visit 2.

■ Interval: 1: visit 1 baseline; 2: visit 1 post-treatment; 3: visit 2 baseline; 4: visit 2 post-treatment.

■ Treatment:

In Model 2c: 0: baseline, no actual treatment -received; 1: TMS received; 2: Sham received. Treatment= "TMS" is set to be reference.

In other models: 1: TMS received; 2: Sham received.

Table 16. Correlation or Association between Subject Characteristics and Crude Estimation of δ without Adjusting Baseline

		$r(p\ value)$
Correlation		
Height		0.273(0.188)
Weight		-0.100(0.639)
Age		0.341(0.096)
Association	**mean(sd)**	**P value**
Race(n)		
African American(3)	-0.033(0.38)	0.648
Asian(7)	0.246(0.25)	
Caucasian(15)	0.033(0.65)	
Baseball level(n)		
High School(7)	-0.150(0.65)	0.316
High School+club(5)	0.320(0.73)	
College/above(13)	0.121(0.35)	
Experience(n)		
No(17)	0.131(0.61)	0.542
Yes(8)	-0.013(0.33)	

■ $\hat{\delta} = \overline{Y}_{i2} - \overline{Y}_{i1}$ for subjects in sequence 1, $\overline{Y}_{i1} - \overline{Y}_{i2}$ for subjects in sequence 2.
■ Spearman correlation is presented for continuous characteristics as $r(p\ value)$
■ Mean(sd) of $\hat{\delta}$ is presented for each level of categorical characteristics, and p value is acquired from one-way ANOVA
■ No significant correlation or association between subject characteristics and crude estimation of δ

Conclusion

For this study, considering that temperature contaminates baseline yet not post-treatment throws, there is two way to deal with it. The first one is to use model 3f with temperature adjusted, the other is to use model 2d that baseline is discard and therefore no need to adjust for temperature. This study by design, incorporates baseline measurements, thus it is inappropriate to not utilize baseline. Model 3f with adjustment of temperature is recommended, though risk that the

estimation of temperature's effect may be unprecise. In analysis, the final model is written as:

$$D_{ijm} = \beta_0 + \beta_1 * I(sequence = 1) + \beta_2 * I(visit = 1) + \beta_3 * I(treamtent = \text{"Sham"}) + \beta_4 * temperature + \epsilon_{ijm}, \epsilon_{ijm} \sim V_5$$

Result of final model			
Beta coefficient	**Estimation**	**SE**	**P value**
β_1(sequence)	0.2884	0.1881	0.1322
β_2(visit)	-0.0530	0.1654	0.7500
β_3(treatment)	-0.0258	0.1714	0.8809
β_4(temperature)	-0.0342	0.0009	0.0005
Least square mean change			
Sham	-0.1385	0.1166	-
TMS	-0.1126	0.1207	-
Estimated δ	-0.0258	0.1714	0.8899

The insignificant of "sequence" suggests the assumptions all hold, p=0.1322; Change score does not significantly fluctuates across visit, p=0.7500; there is no statistical significant difference among the effect of Sham and the effect of TMS, and the estimated difference -0.0342, close to zero, p=0.8809; temperature is significantly associated to change score, the expectation of change score decreases -0.0342 with 1 °F increase in temperature, p=0.0005.

Future Study

Basing on this study, there are some recommendations to future study investigate on the effects of TMS and Sham on throw accuracy:

1. Clinical meaningful difference among effects of Sham and TMS on throw score shall be

defined prior to data collection, that is a δ on throw score that is large enough to be clinical meaningful before data collection.

2. Exclusion criteria requires to be defined before data collection, in this study, one subject had a proportion of score 0 much more higher than other subjects, but he is not excluded from analysis as the no exclusion criteria is pre-defined.

3. The outcome of this study, throw score, is conscious physical performance of human, so psychological factor, which can cause carryover and/or period*treatment interaction but cannot be eliminated by washout period, can play a role. Although in this analysis, there is no statistical evidence suggesting carryover effect, but in the final model, after adjusting for temperature, the p=0.1322 for "sequence" rises cautious. Recommended by Kenward[9], the test for carryover-effect and/or period by treatment interaction is at significance level 0.10. If possible, randomized parallel design may be more suitable than crossover design.

4. If possible, the experiment can be done indoor and control for temperature and other factors that may relate to throw accuracy, like light condition.

5. Although the investigation shown in table 14 does not suggest stimulation intensity plays any role, in this study, stimulation intensity of the two treatments are based on the subject's own Resting Motor Threshold. Give subject-specific stimulation intensity, implies that, take TMS for example, those different subject-specific stimulation intensities will result in a uniform, or relatively uniform treatment effect on the subjects. If there is doubt in this implication, then at design stage, under each treatment, given a uniform stimulation intensity on all subjects may be favorable if it can result in treatment effects more uniform subjects than subject-specific

stimulation intensity.

Reference

1. Armitage, P. and M. Hills (1982). "THE TWO-PERIOD CROSSOVER TRIAL." The Statistician **31**: 119-131.

2. Mehrotra, D. V. (2014). "A recommended analysis for 2 _ 2 crossover trials with baseline measurements." Pharmaceutical Statistics **13**: 376-387.

3. Laird, N. M., et al. (1992). "AN ANALYSIS OF TWO-PERIOD CROSSOVER DESIGNS WITH CARRY-OVER EFFECTS." STATISTICS IN MEDICINE **11**: 1967-1979.

4. KENWARD, M. G. and B. JONES (1987). "THE ANALYSIS OF DATA FROM 2 x 2 CROSS-OVER TRIALS WITH BASELINE MEASUREMENTS." STATISTICS IN MEDICINE **6**: 911-926.

5. Kenward, M. G. and B. Jones (1987). "THE ANALYSIS OF DATA FROM 2 x 2 CROSS-OVER TRIALS WITH BASELINE MEASUREMENTS." STATISTICS IN MEDICINE **6**: 911-926.

6. Liang, Y. and K. C. Carriere (2010). "On the Role of Baseline Measurements for Crossover Designs under the Self and Mixed Carryover Effects Model." Biometrics **66**: 140–148.

7. Xun Chen, Z. M., Ji Zhang (2012). "Handling of baseline measurements in the analysis of crossover trials." STATISTICS IN MEDICINE **31**: 1791-1803.

8. Chow, S. and J. Liu (1994). "RECENT STATISTICAL DEVELOPMENTS IN BIOEQUIVALENCE TRIALS -A REVIEW OF THE FDA GUIDANCE." Drug Information Journal **28**: 851-864.

9. Jones, B. and M. G. Kenward (2010). Design and Analysis of Cross-Over Trials Chapman and Hall.

10. William, A. R. and J. L. Pater (1986). "Using Baseline Measurements in the Two-Period Crossover Clinical Trial." Controlled Clinical Trials **7**: 282-289.

11. Senn, S. (2006). "Change from baseline and analysis of covariance revisited." STATISTICS IN MEDICINE **25**: 4334-4344.

Appendix

General Advantage and Drawback			
criteria	When carryover exist, can it be tested in model?	How is the power of the method to detect carryover effect?	The variance of the estimator of δ
Methods			
1. Joint Modeling	yes	Did not find material in this topic, but personal opinion it is similar to method 2.	Can be smaller when within visit correlation is strong, other situation, similar
2. Ignore baseline	yes	reference	reference
3. Change from baseline	yes	Powerful than method 2	In practice, almost always much more larger
4. Baseline as covariate	no	---	Can be smaller when within visit correlation is strong, other situation, similar

For covariance structure which equivalent to involving a random effect of subjects, diagnostic plots for residual and conditional residual are presented. For covariance structure does not involve random effect of subjects, diagnostic plots for residual is presented. For model which collapses into other model or SAS gives warning, its diagnostic plots are not presented. The residual plots for alternative joint modeling method (model 1h, 1i and 1j) are not presented. The diagnostic plots are presented for 16 of 24 models, model 1a, 1c, 1d, 1f, 1g, 2a, 2b, 2c, 2d, 3b, 3e, 3f, 4a,4b, 4c, 4d.

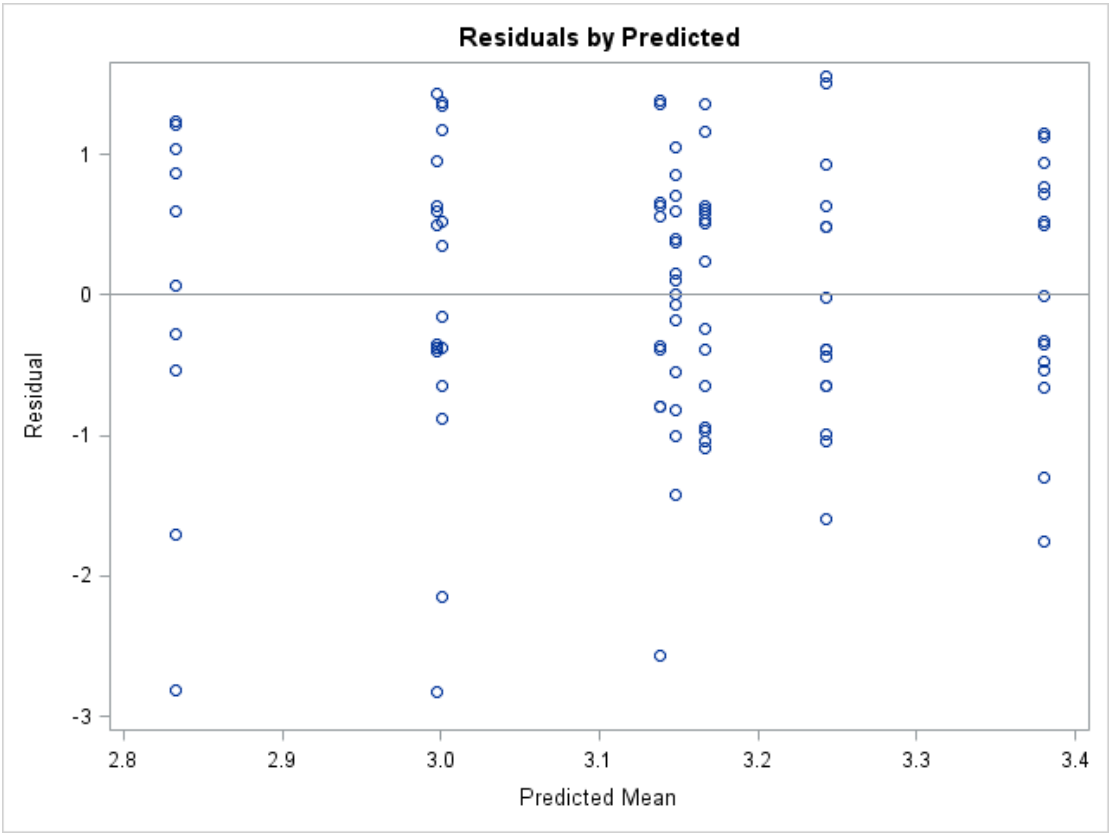

Figure 1a. "Residual by Predicted" Plot for Model 1a

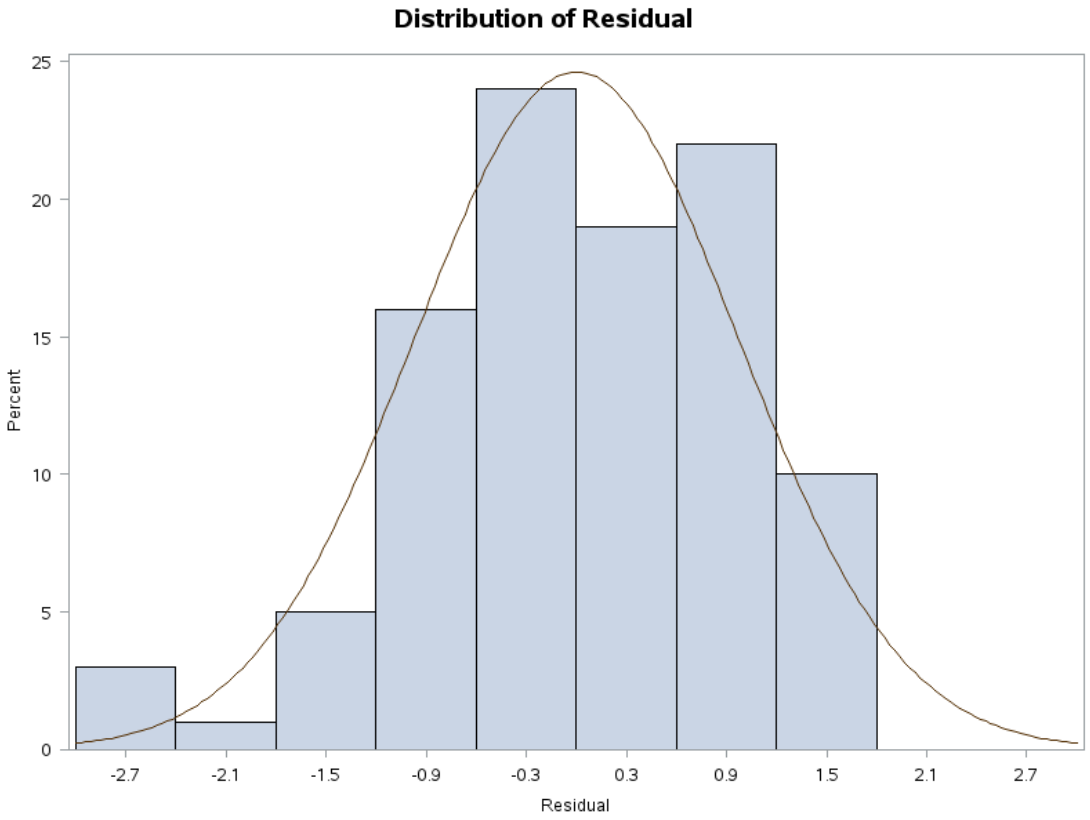

Figure 1b. "Distribution of Residual" Plot for Model 1a

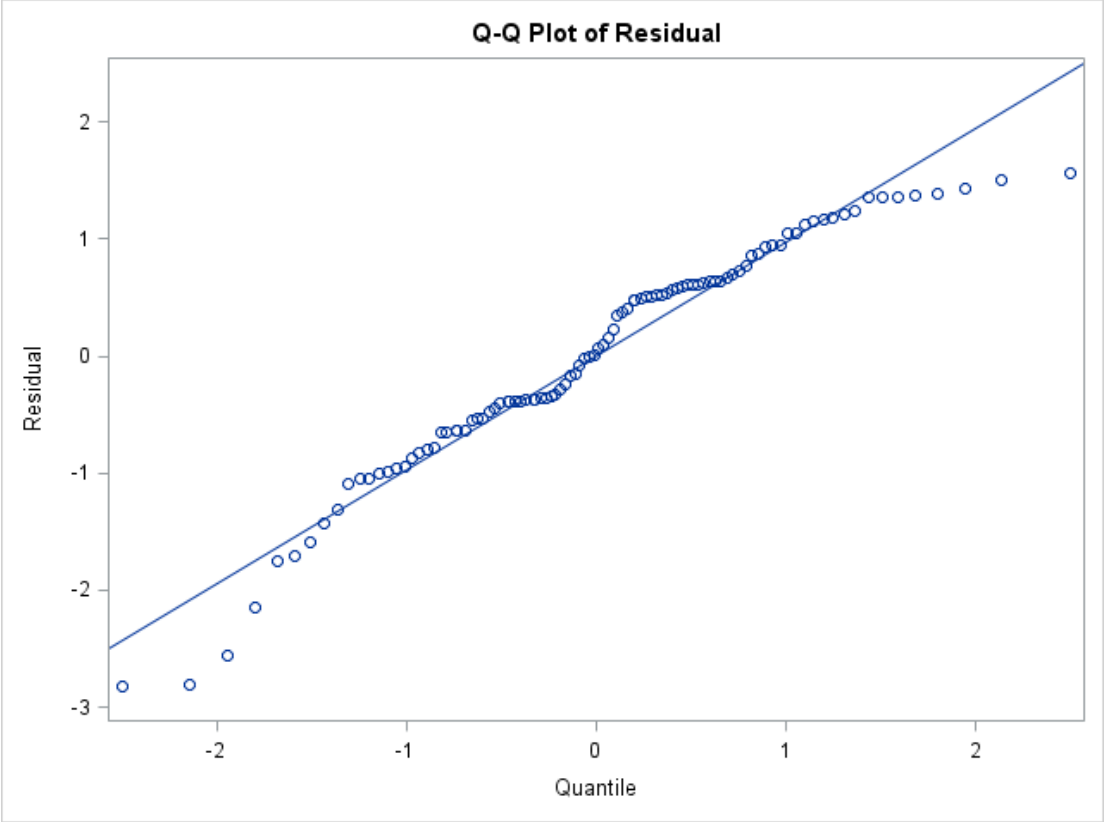

Figure 1c. "Q-Q Plot of Residual" Plot for Model 1a

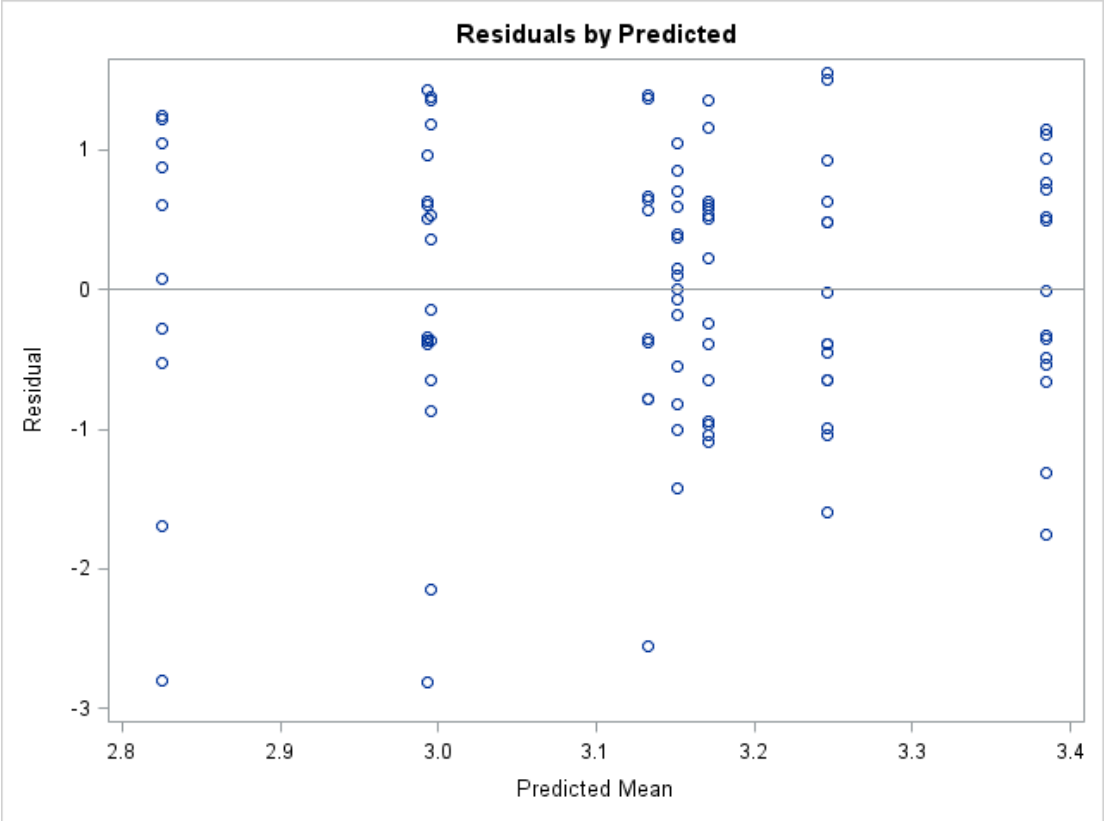

Figure 2a. "Residual by Predicted" Plot for Model 1c

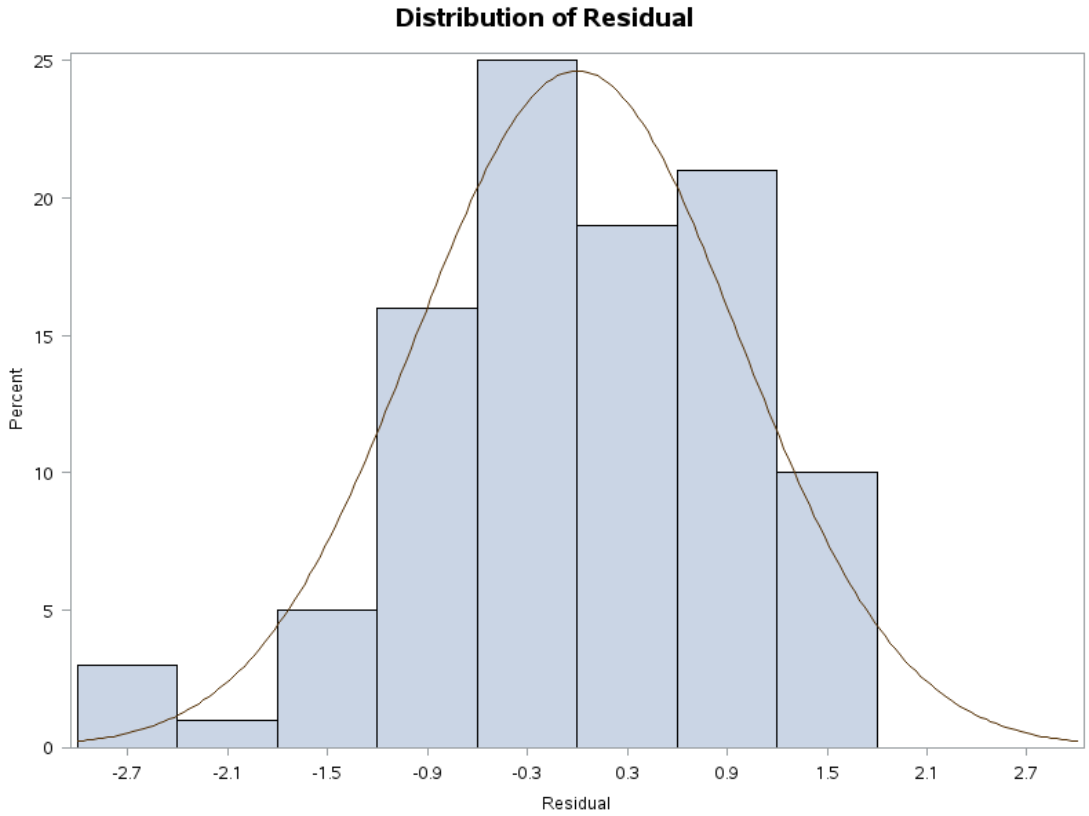

Figure 2b. "Distribution of Residual" Plot for Model 1c

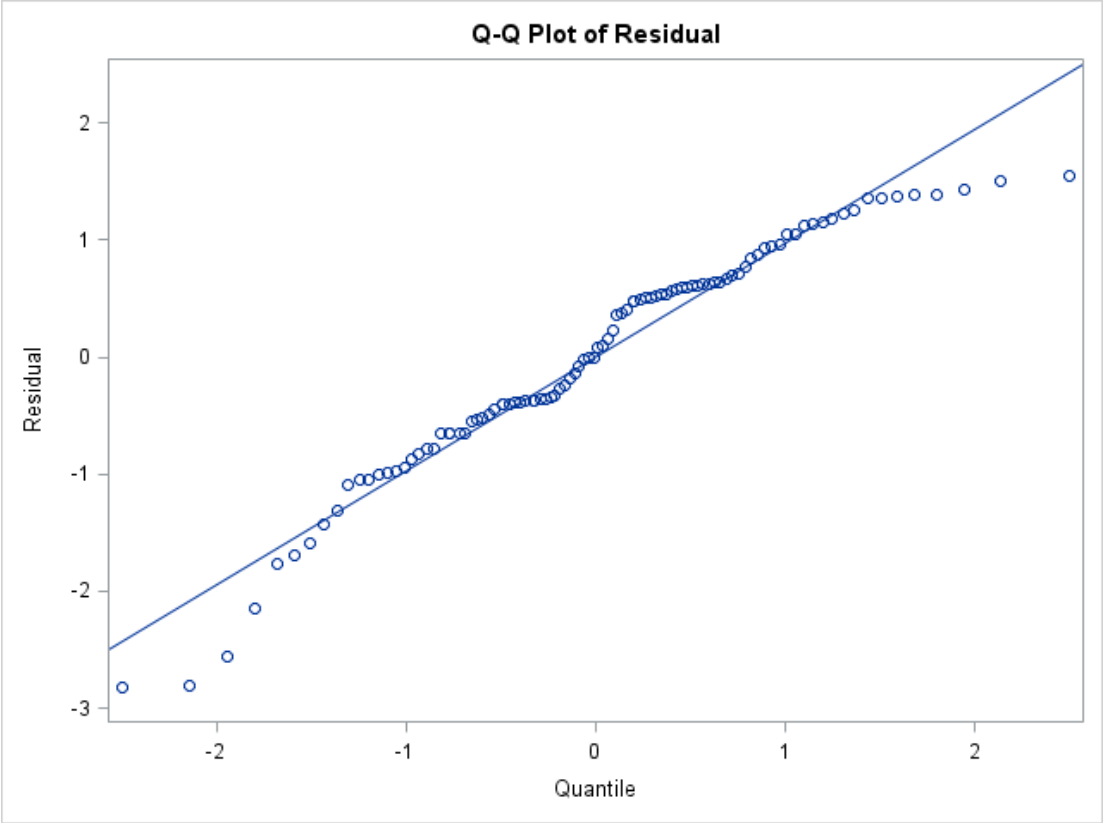

Figure 2c. "Q-Q Plot of Residual" Plot for Model 1c

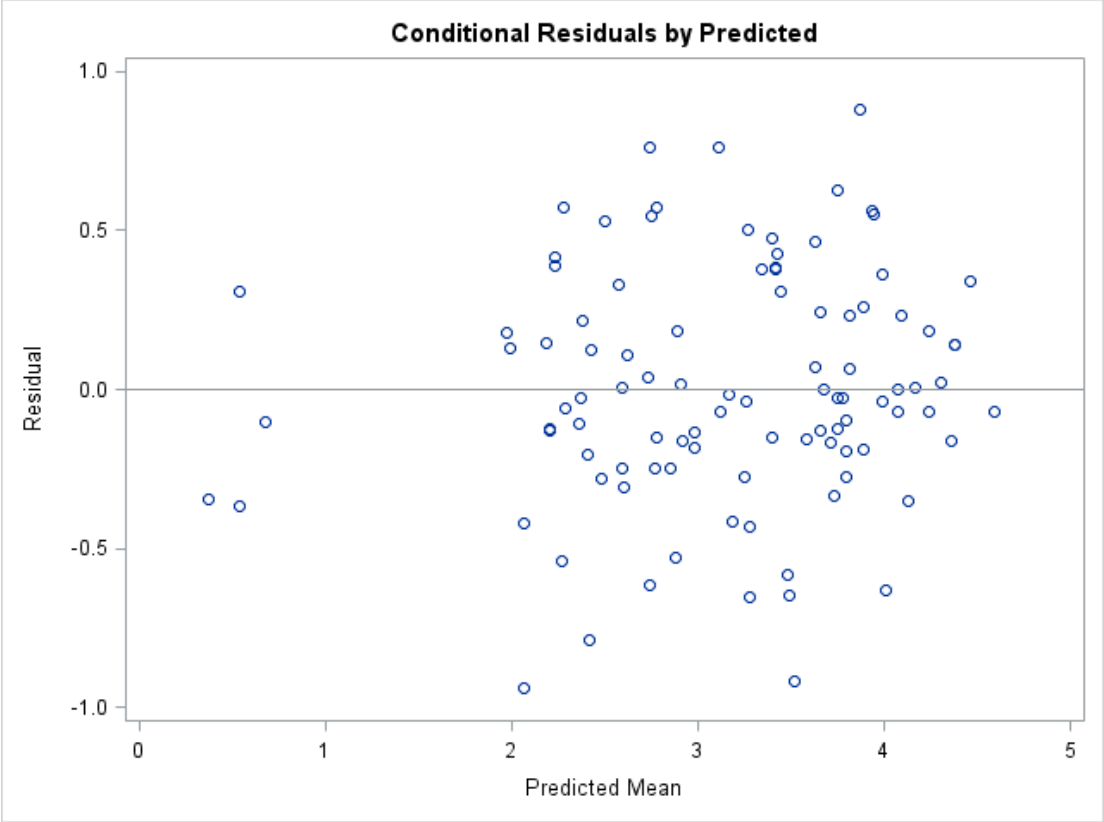

Figure 2d. "Conditional Residual by Predicted" Plot for Model 1c

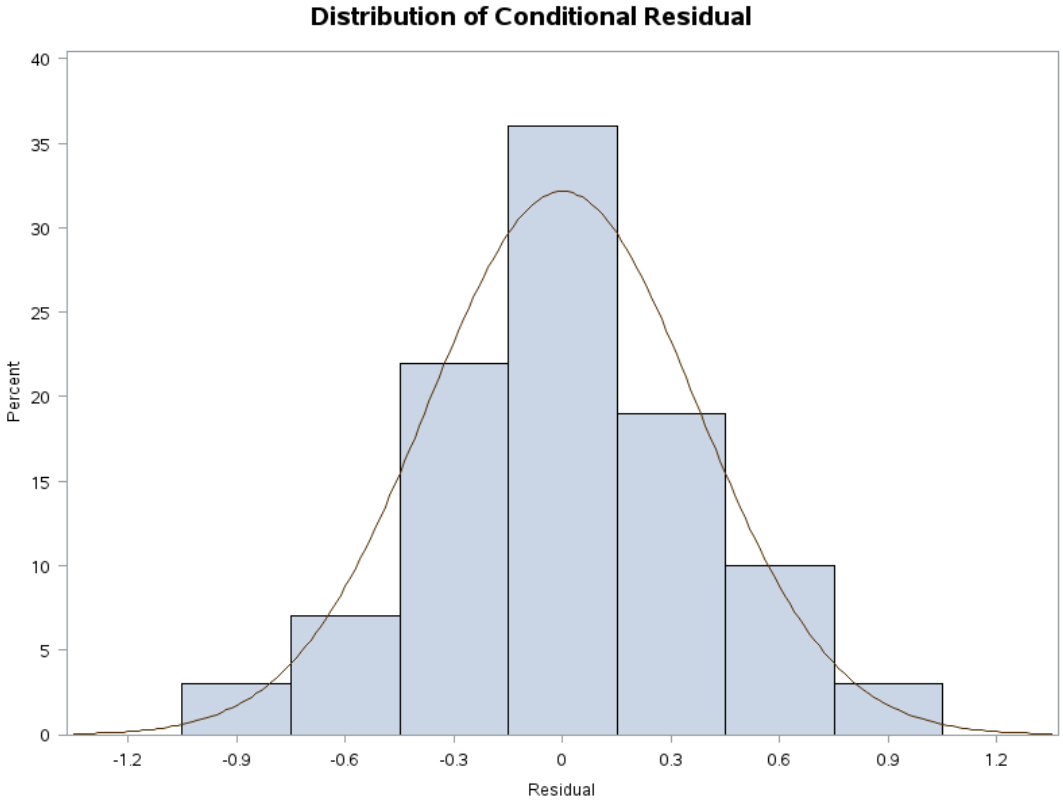

Figure 2e. "Distribution of Conditional Residual " Plot for Model 1c

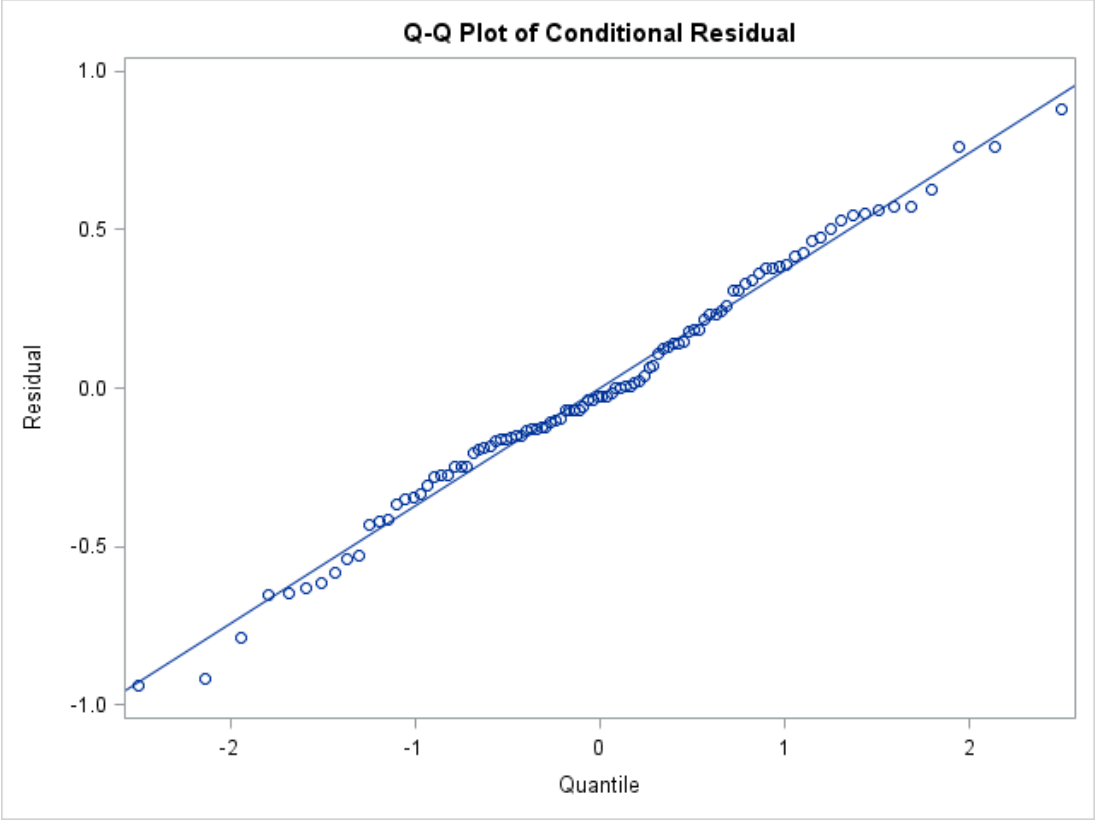

Figure 2f. "Q-Q Plot of Conditional Residual " Plot for Model 1c

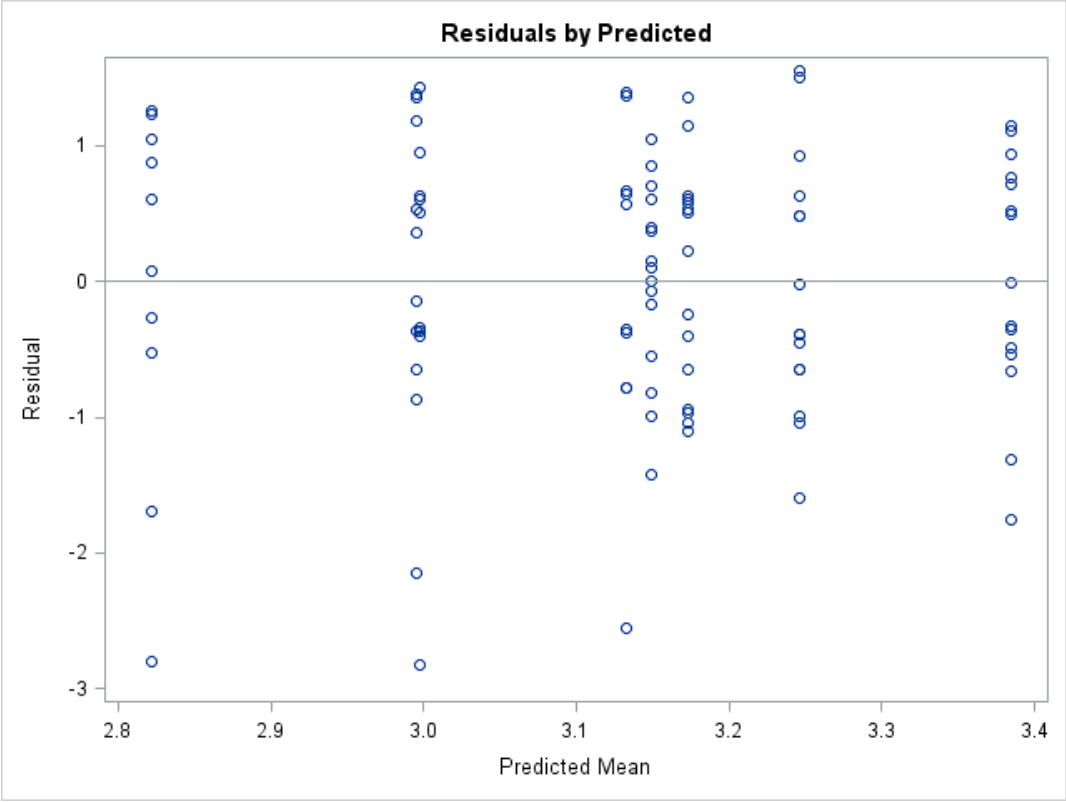

Figure 3a. "Residual by Predicted" Plot for Model 1d

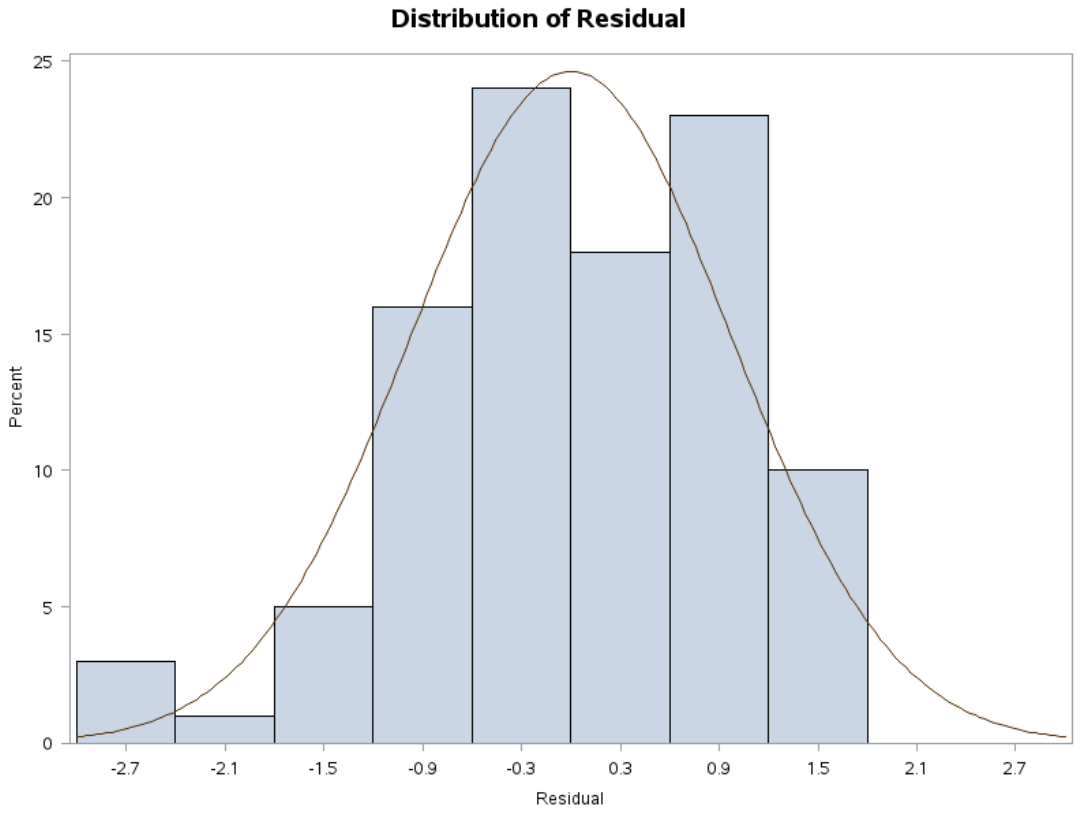

Figure 3b. "Distribution of Residual" Plot for Model 1d

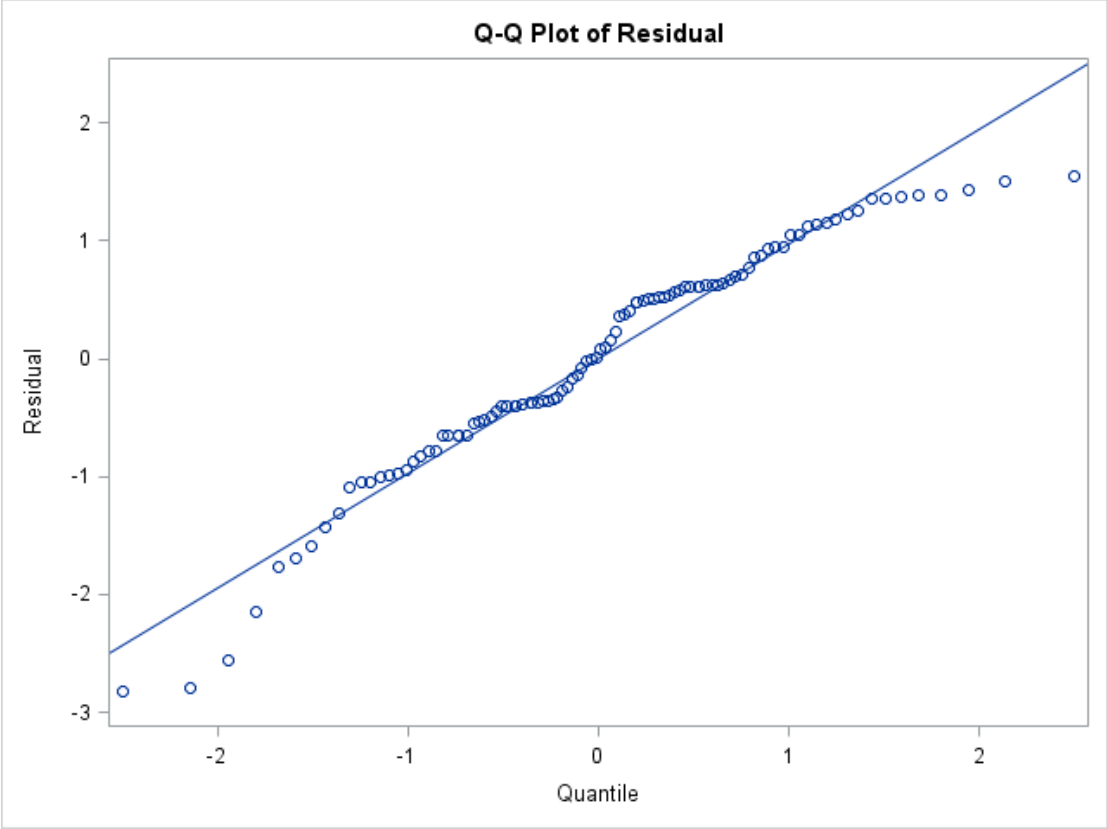

Figure 3c. "Q-Q Plot of Residual" Plot for Model 1d

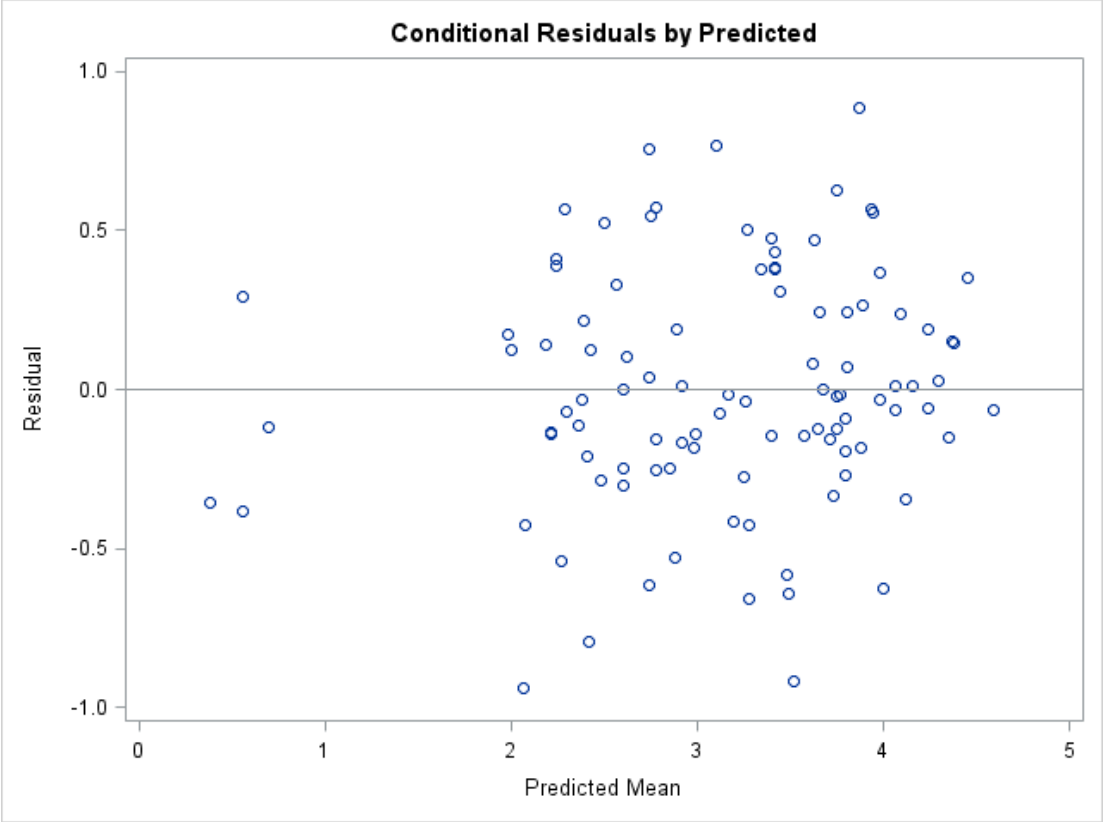

Figure 3d. "Conditional Residual by Predicted" Plot for Model 1d

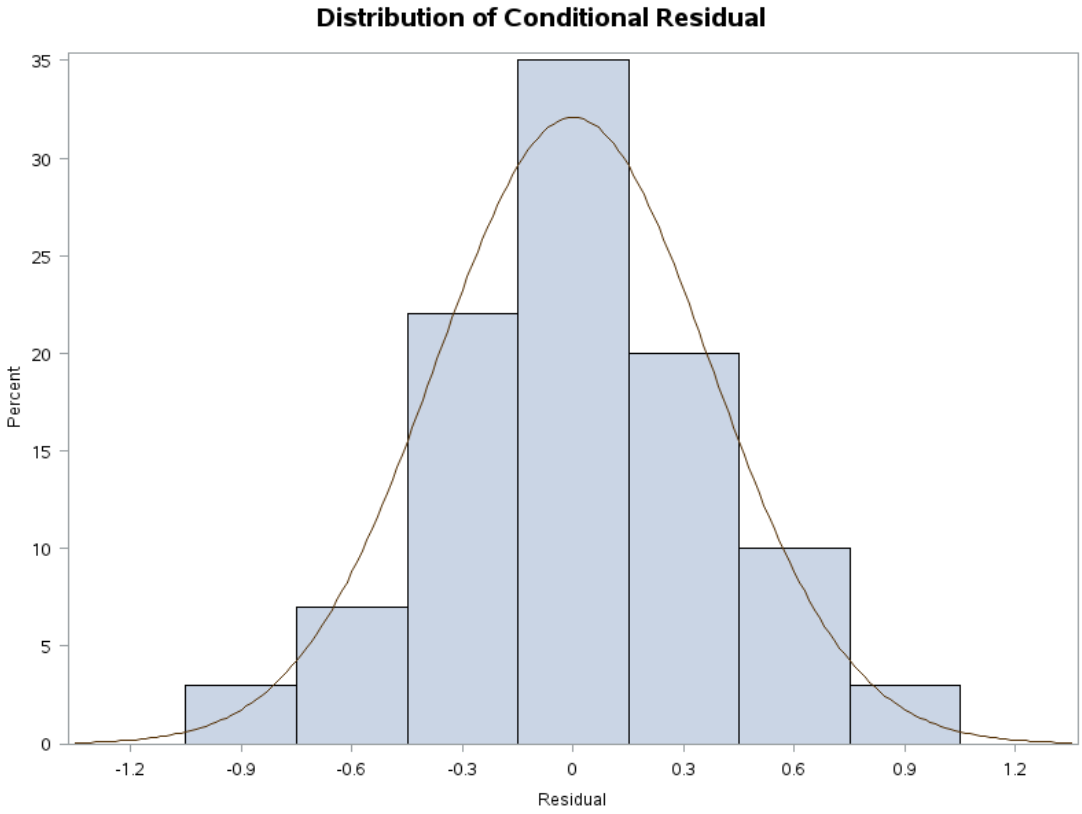

Figure 3e. "Distribution of Conditional Residual " Plot for Model 1d

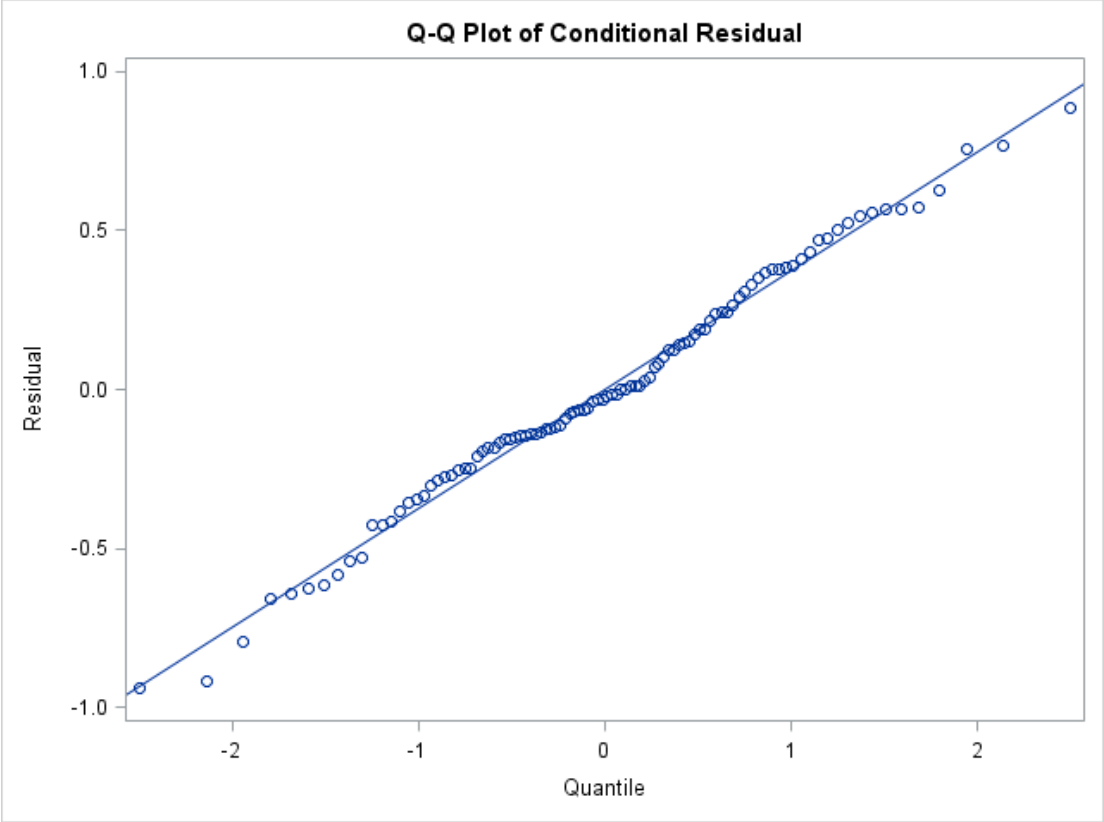

Figure 3f. "Q-Q Plot of Conditional Residual " Plot for Model 1d

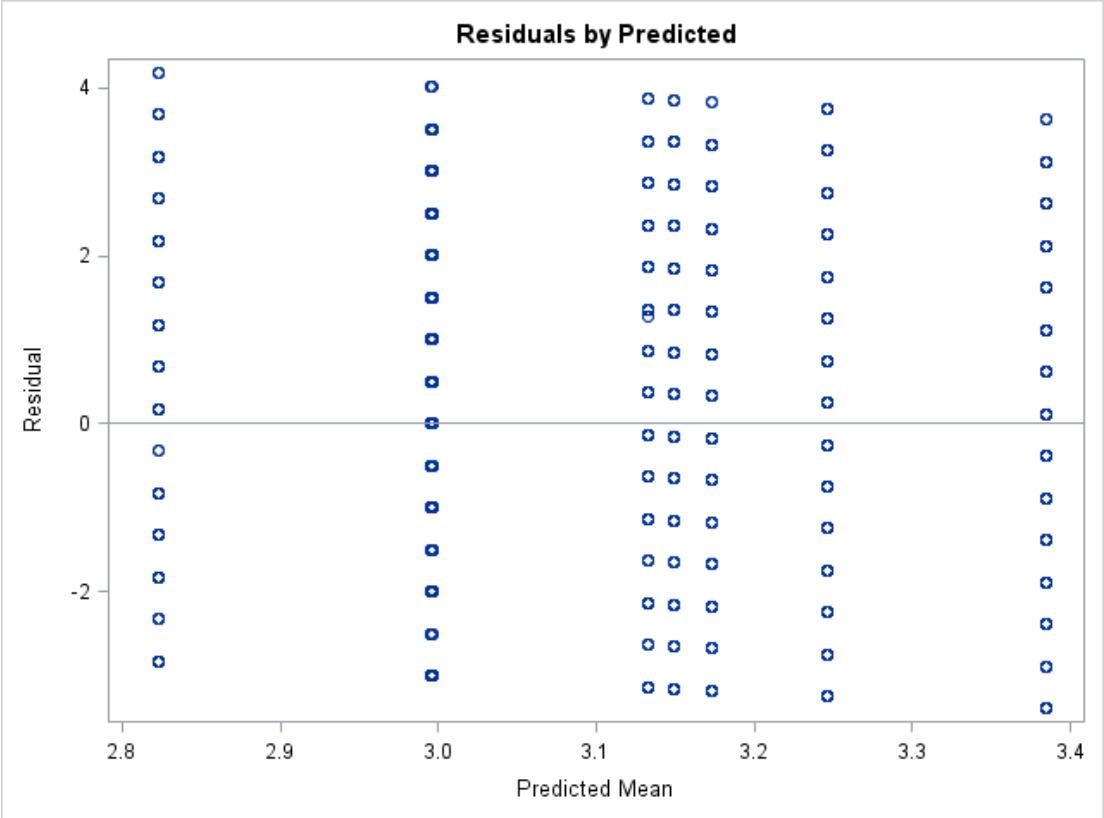

Figure 4a. "Residual by Predicted" Plot for Model 1f

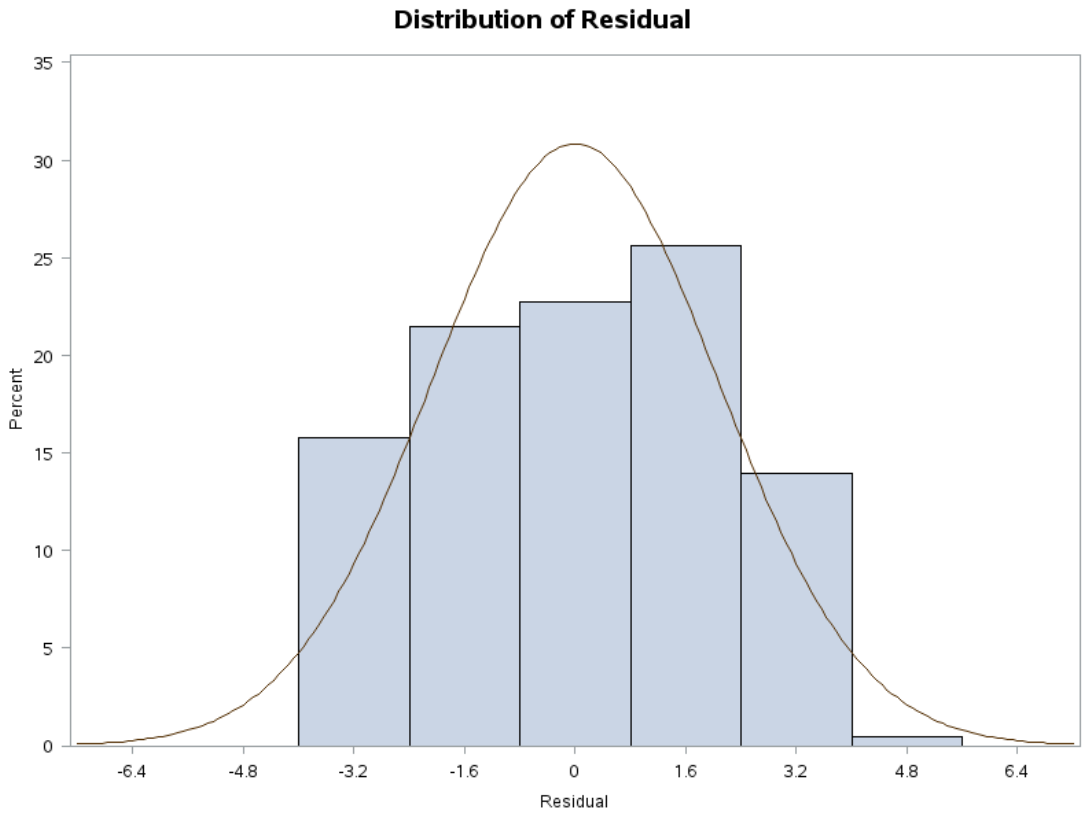

Figure 4b. "Distribution of Residual" Plot for Model 1f

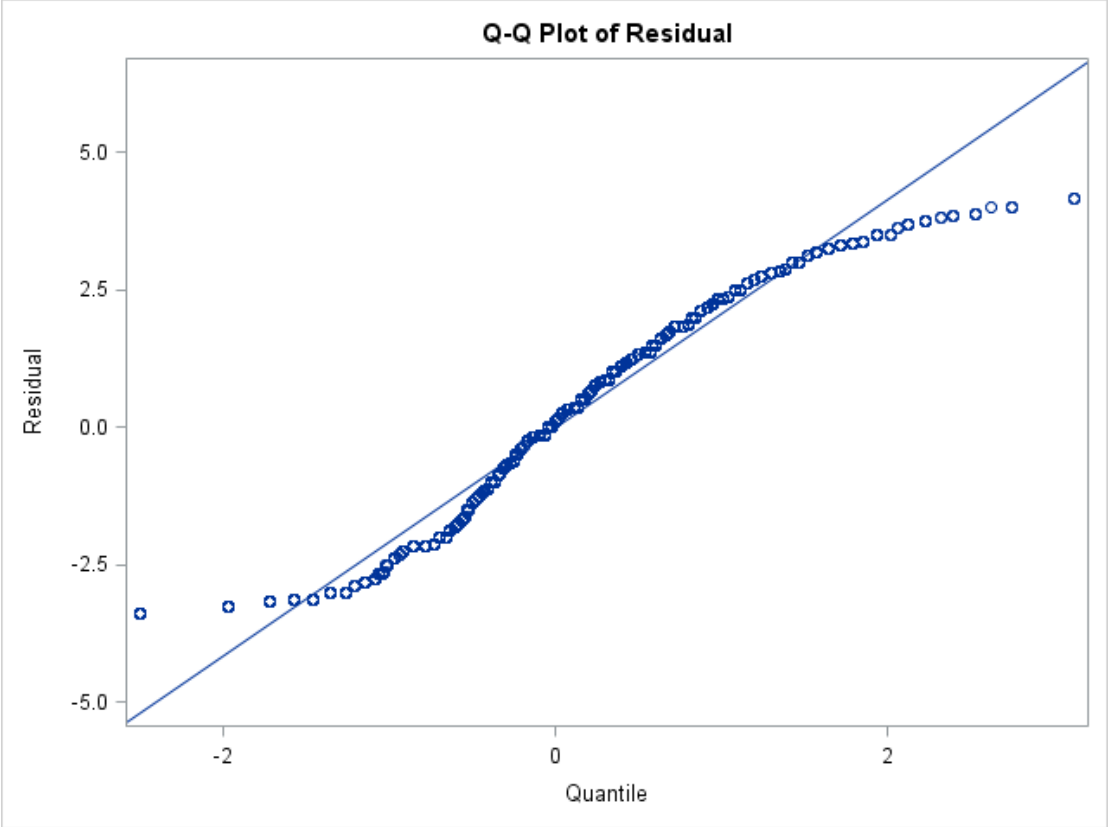

Figure 4c. "Q-Q Plot of Residual" Plot for Model 1f

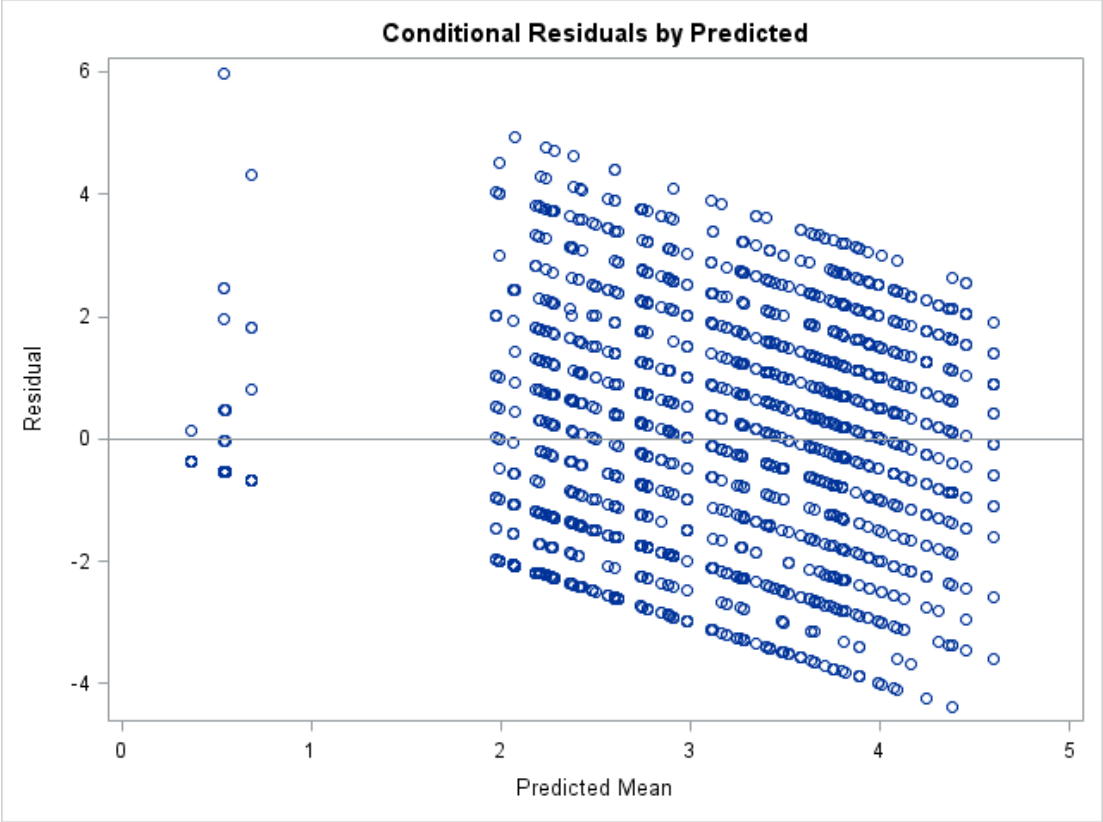

Figure 4d. " Conditional Residual by Predicted" Plot for Model 1f

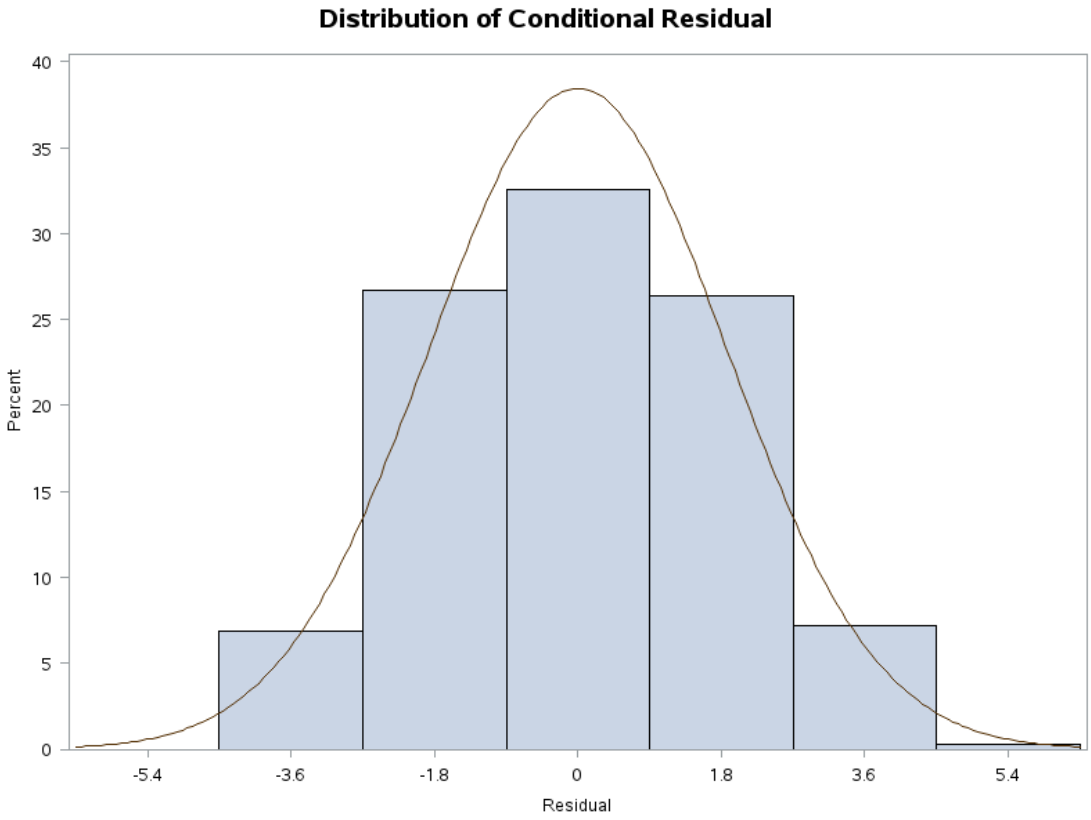

Figure 4e. "Distribution of Conditional Residual" Plot for Model 1f

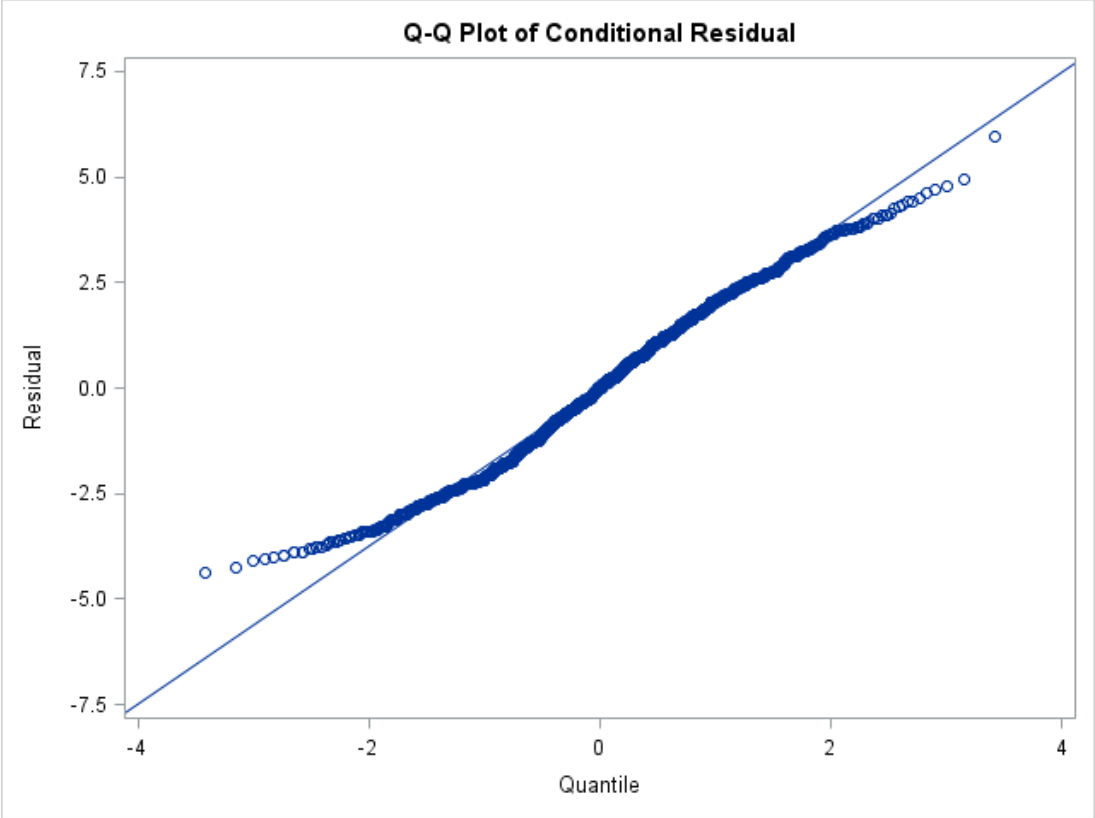

Figure 4f. "Q-Q Plot of Conditional Residual" Plot for Model 1f

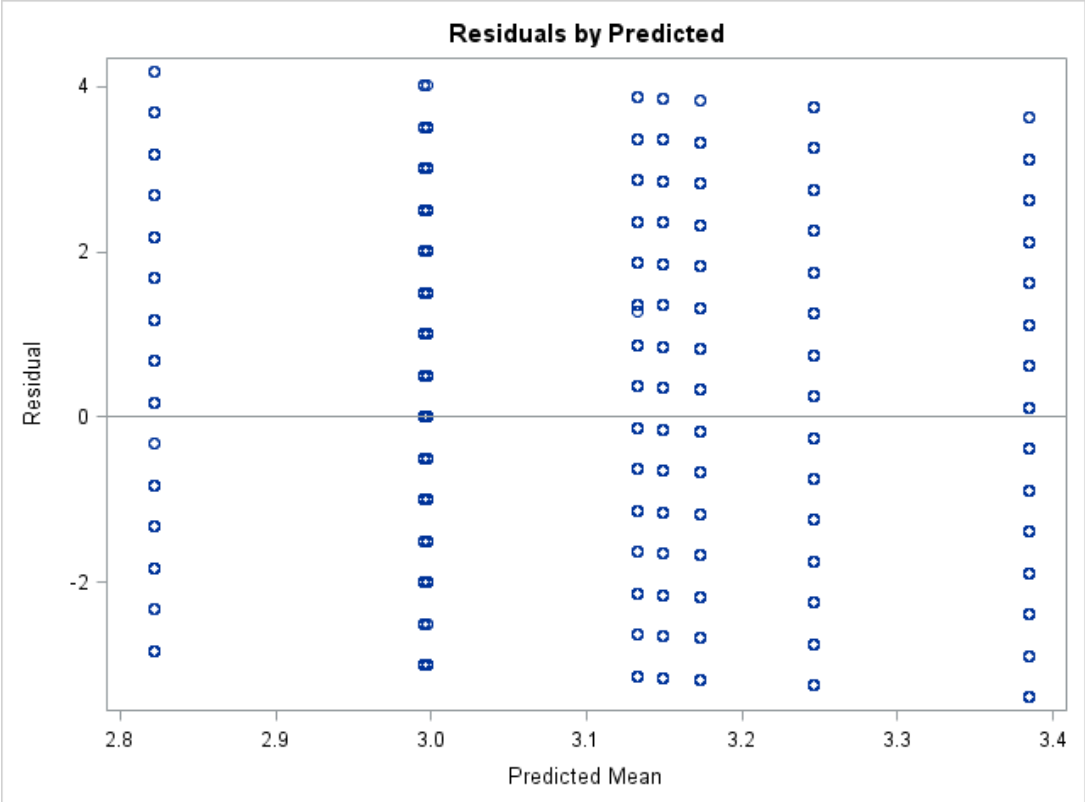

Figure 5a. "Residual by Predicted" Plot for Model 1g

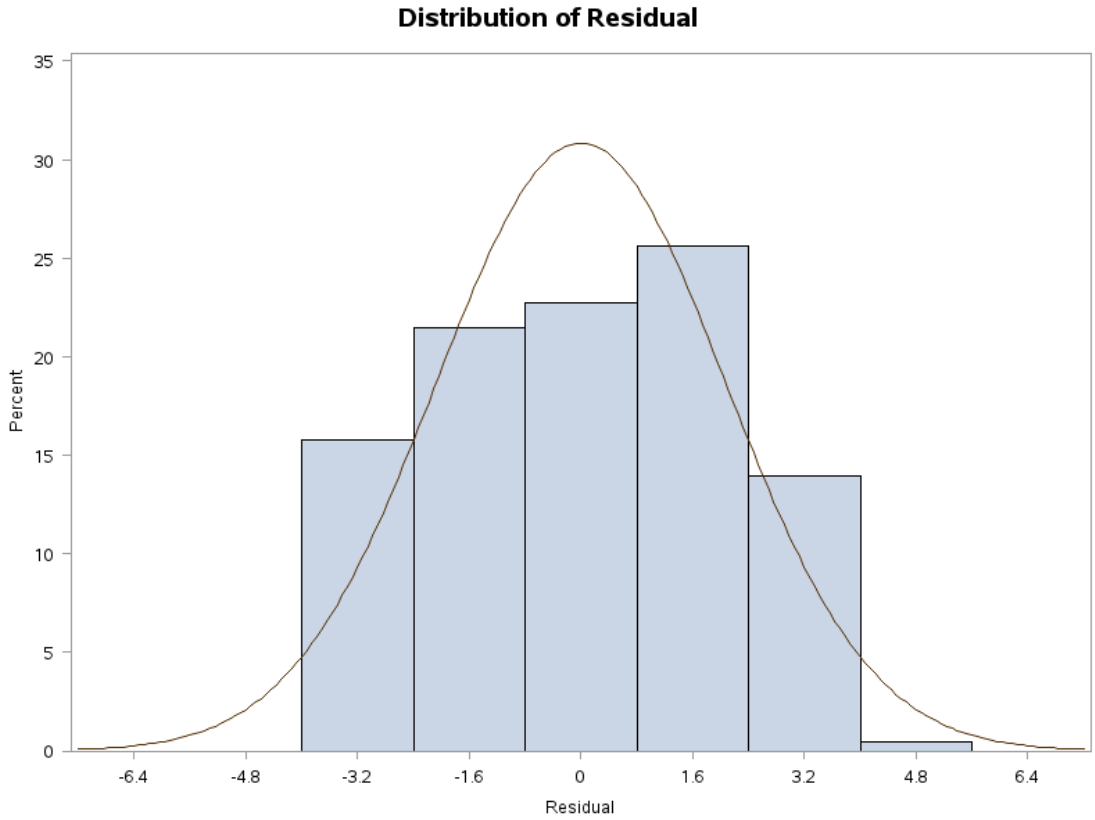

Figure 5b. "Distribution of Residual" Plot for Model 1g

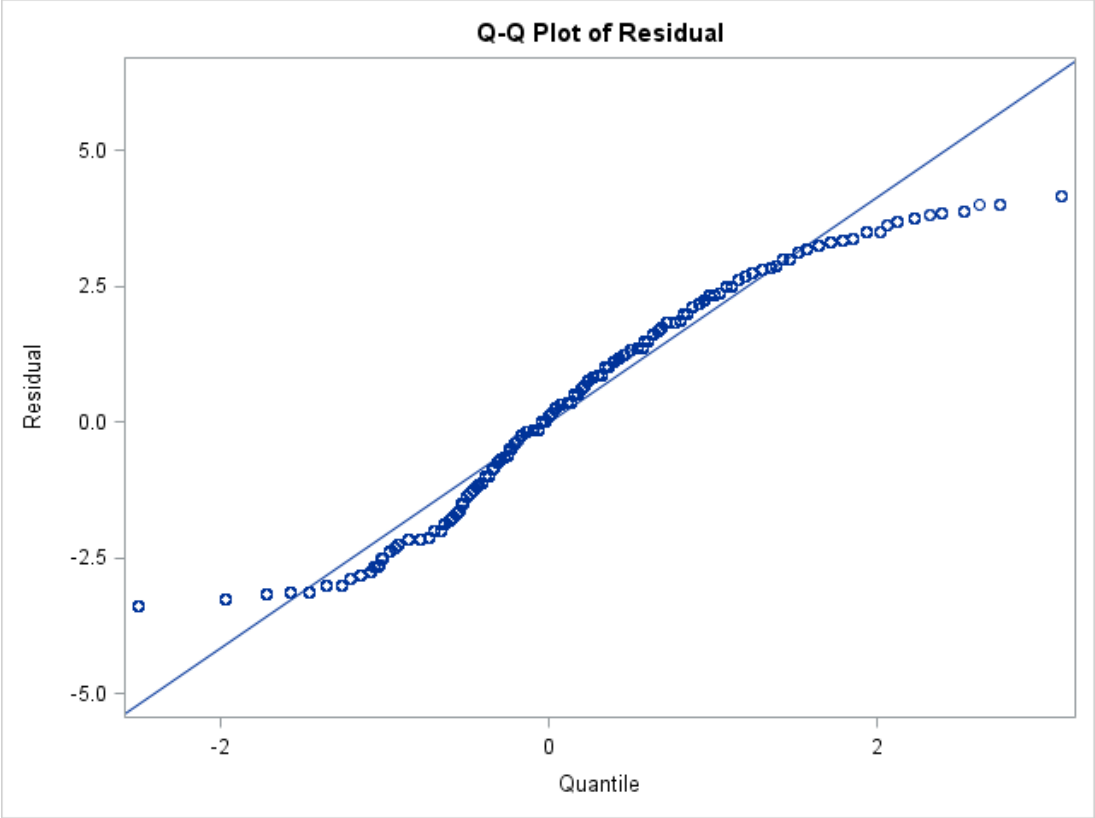

Figure 5c. "Q-Q Plot of Residual" Plot for Model 1g

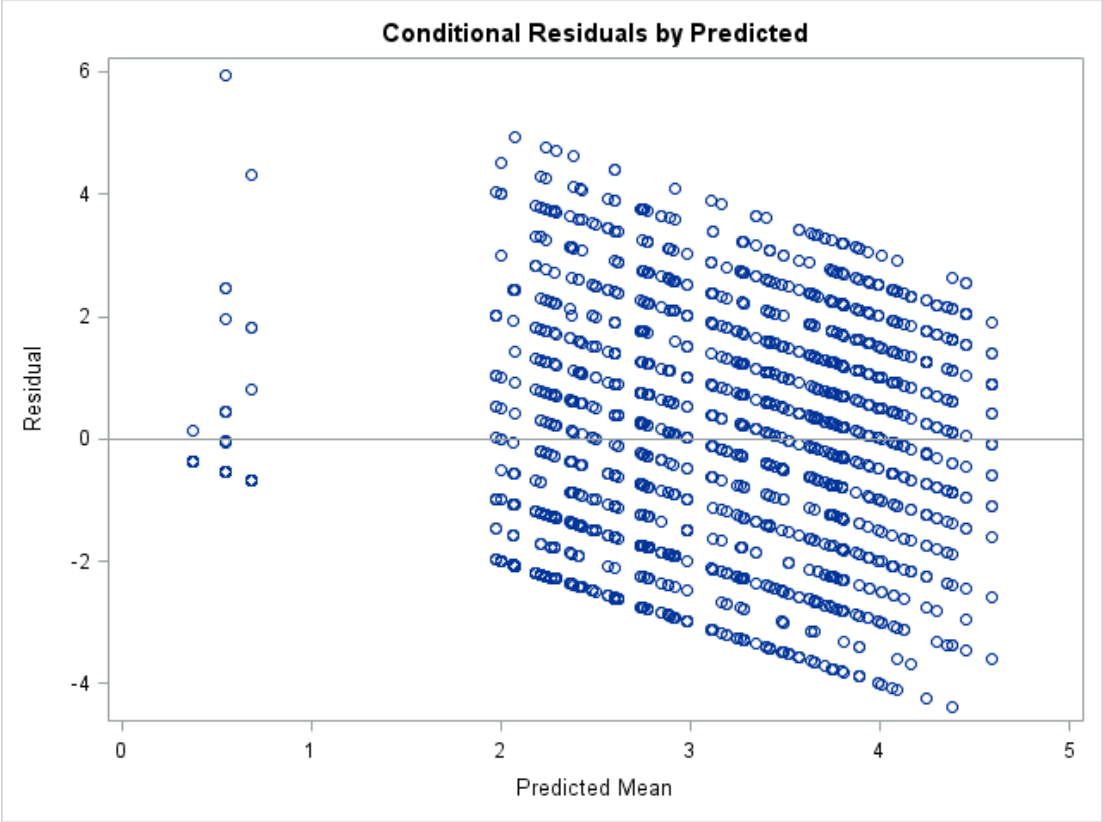

Figure 5d. "Conditional Residual by Predicted" Plot for Model 1g

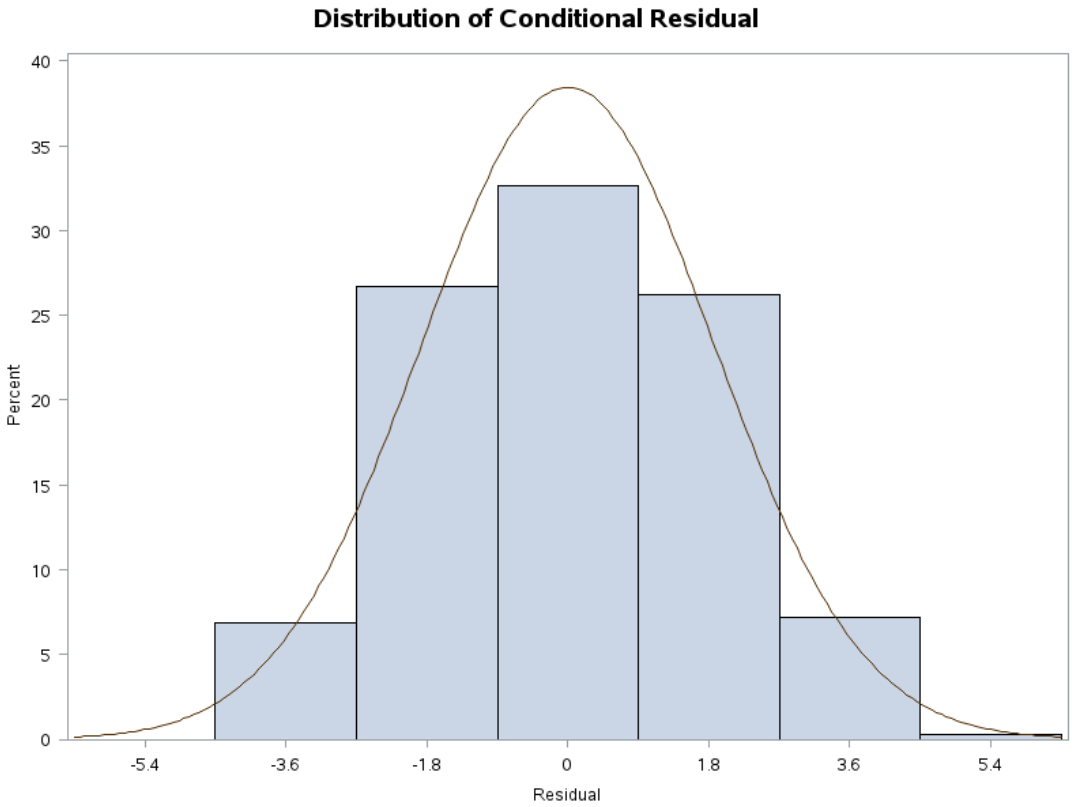

Figure 5e. "Distribution of Conditional Residual" Plot for Model 1g

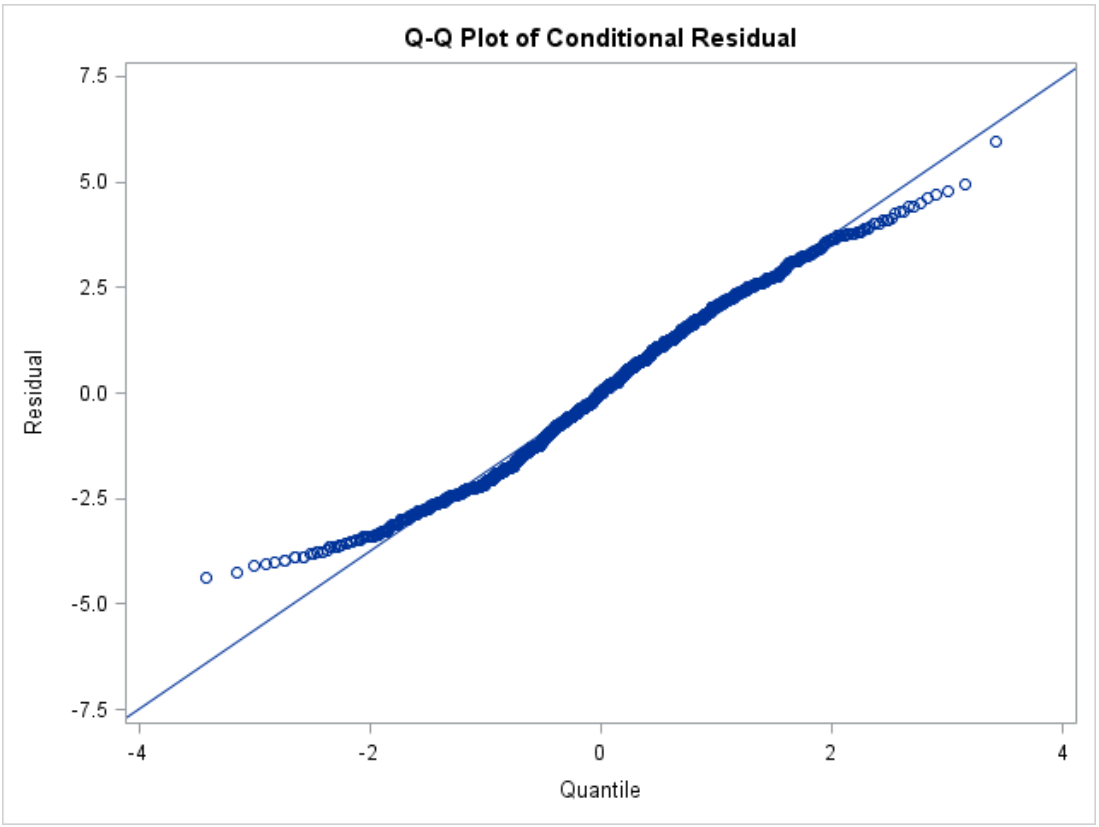

Figure 5f. "Q-Q Plot of Conditional Residual" Plot for Model 1g

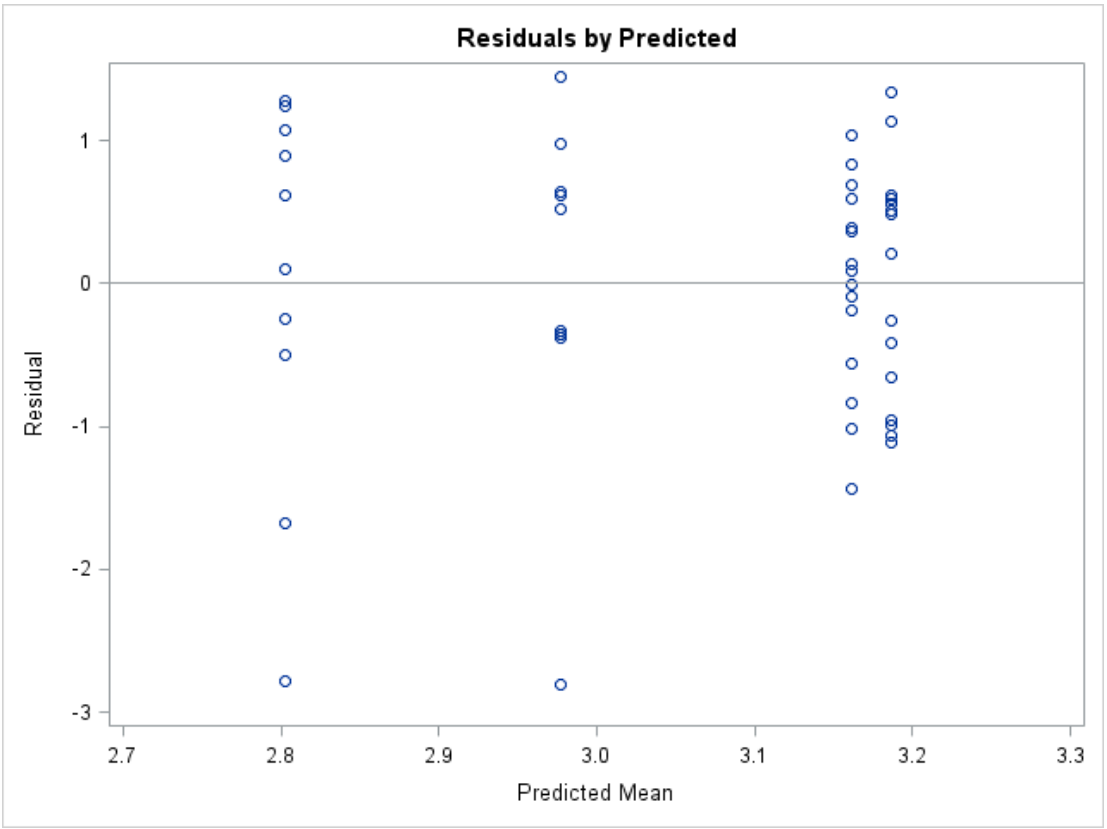

Figure 6a. "Residual by Predicted" Plot for Model 2a

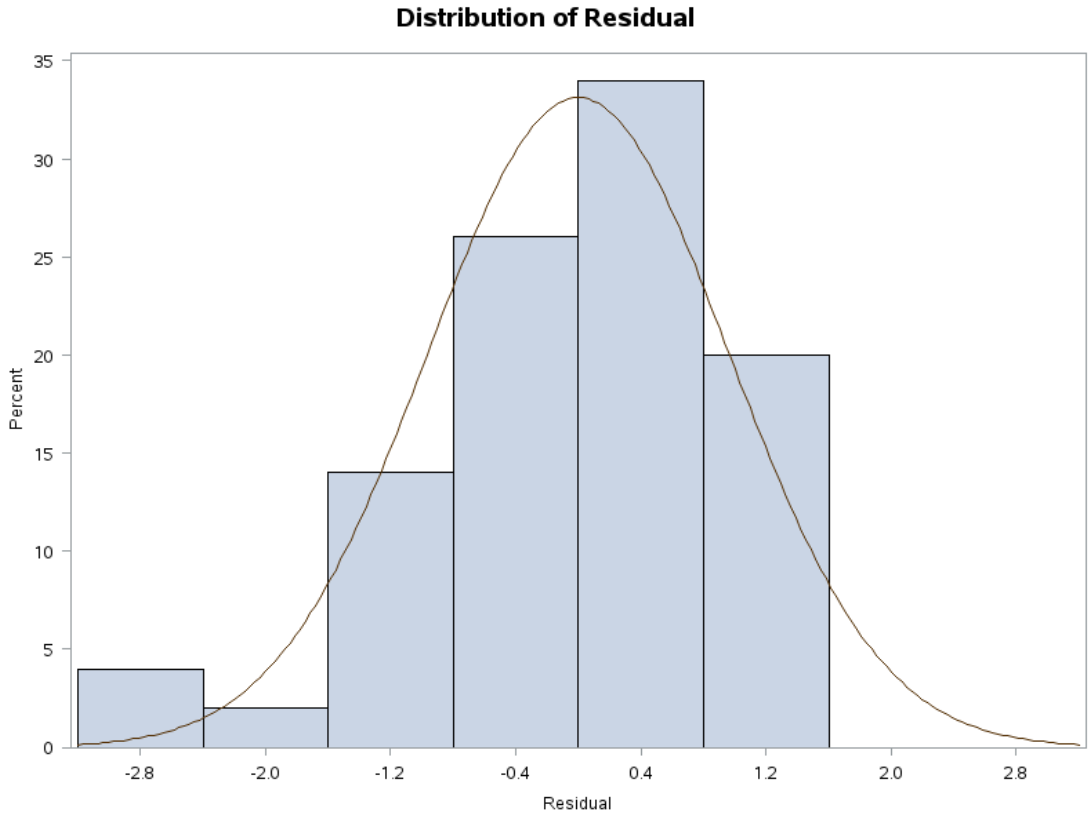

Figure 6b. "Distribution of Residual" Plot for Model 2a

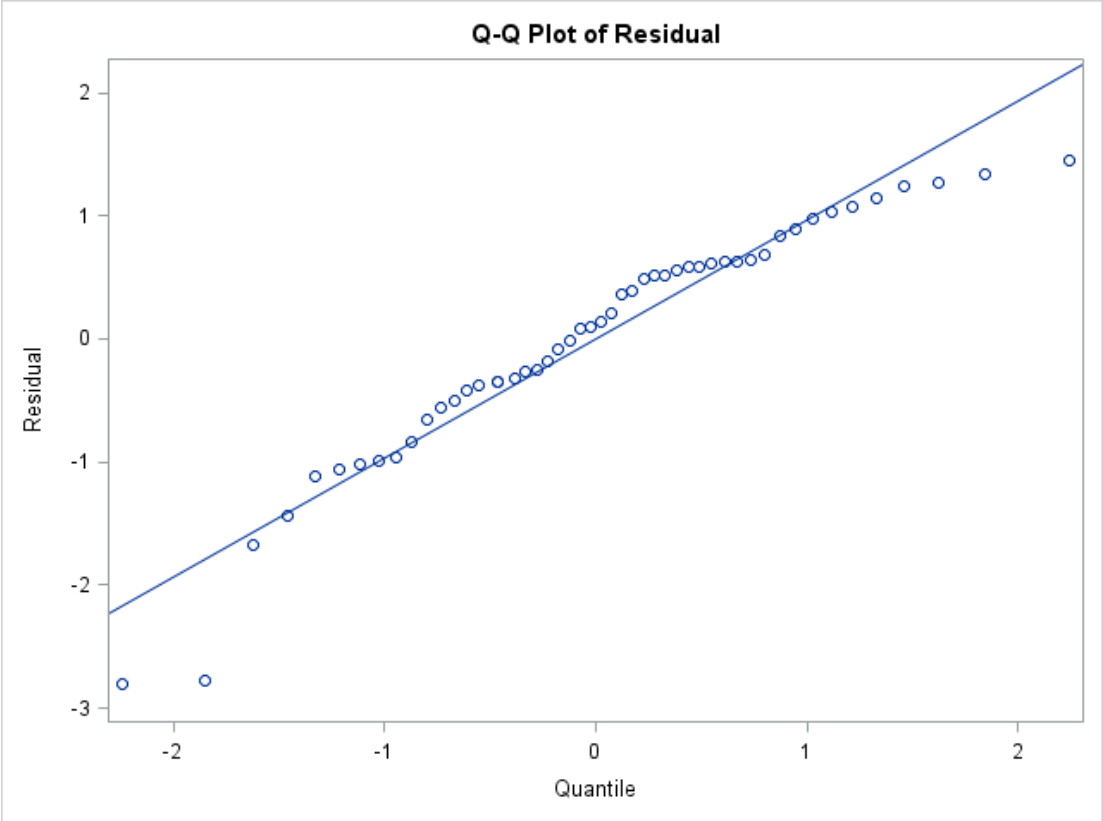

Figure 6c. "Q-Q Plot of Residual" Plot for Model 2a

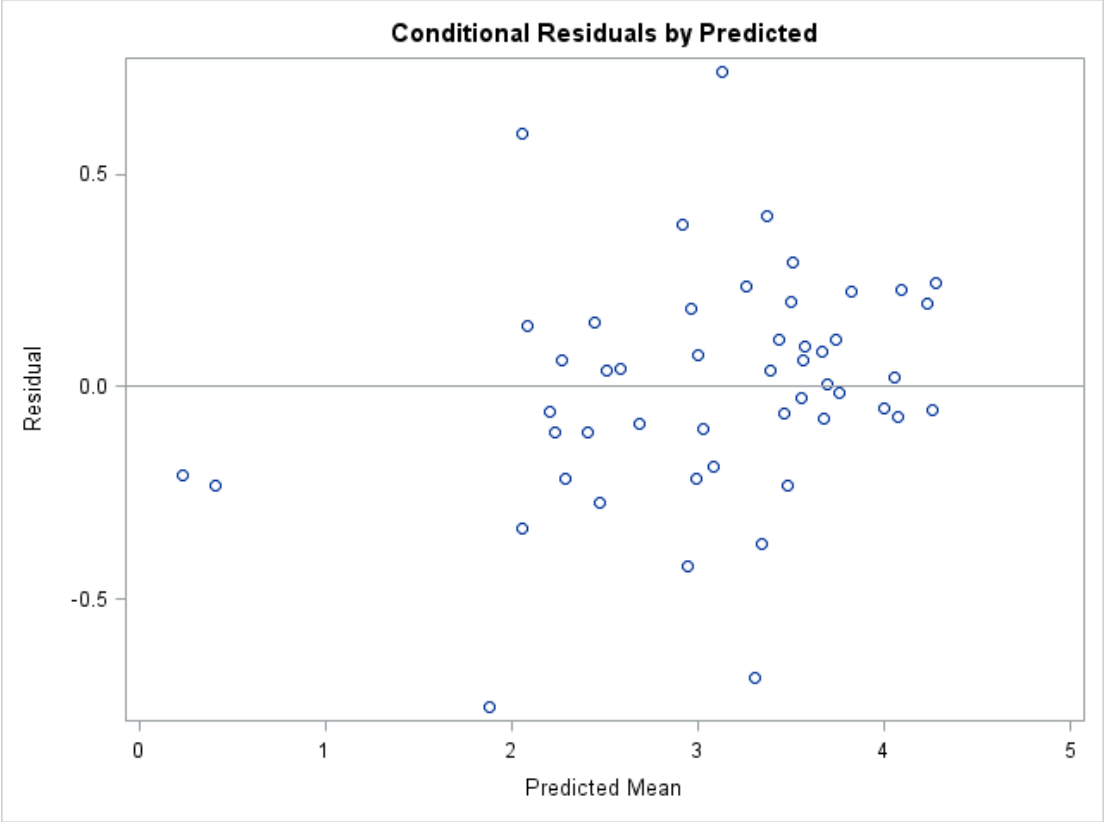

Figure 6d. "Conditional Residual by Predicted" Plot for Model 2a

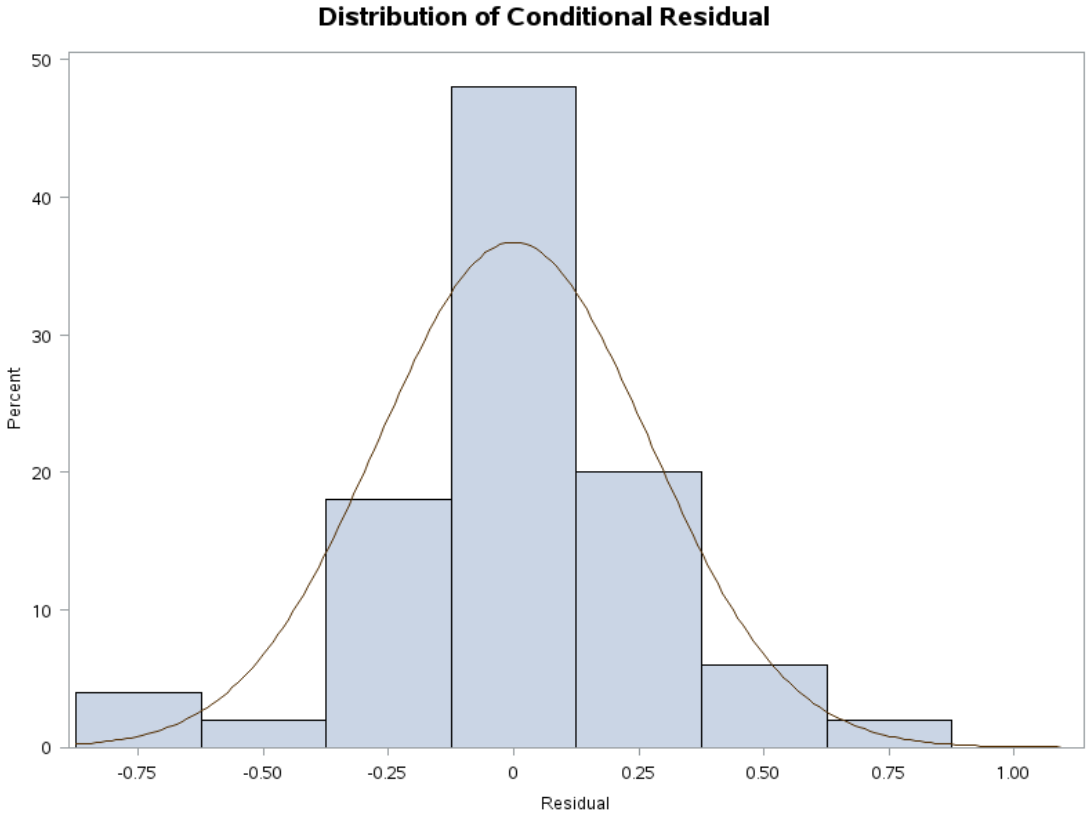

Figure 6e. "Distribution of Conditional Residual" Plot for Model 2a

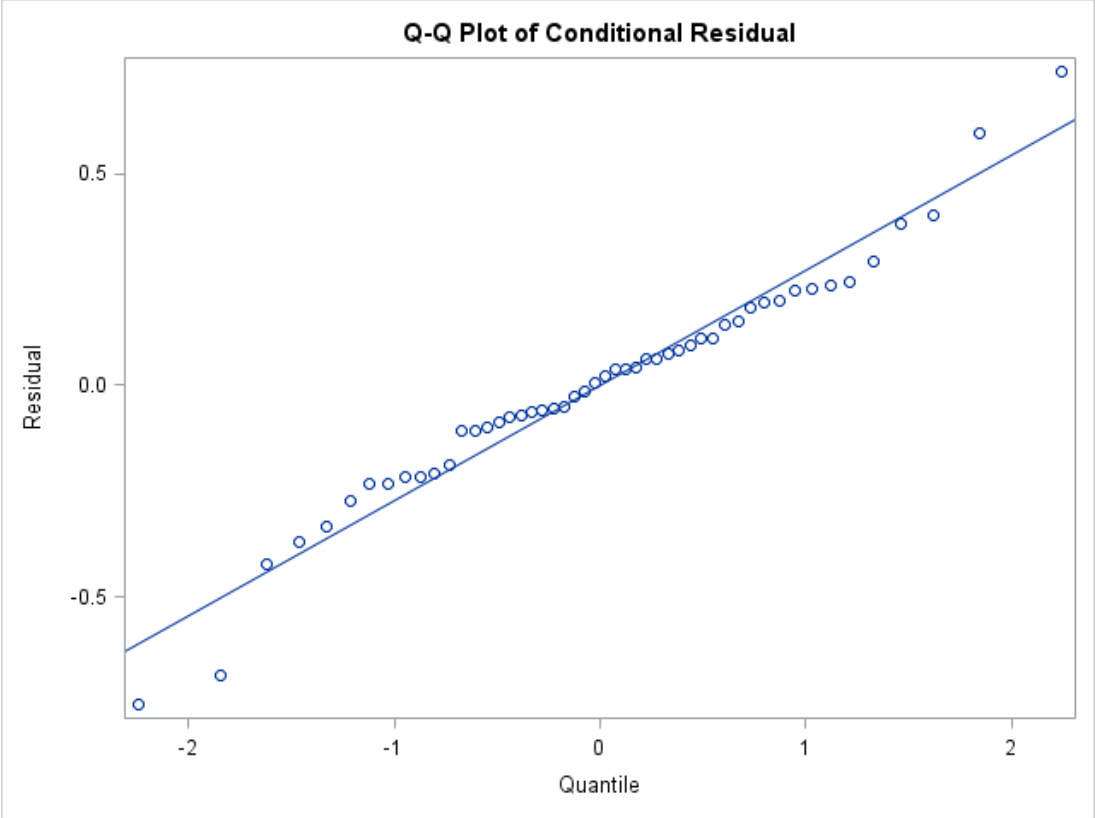

Figure 6f. "Q-Q Plot of Conditional Residual" Plot for Model 2a

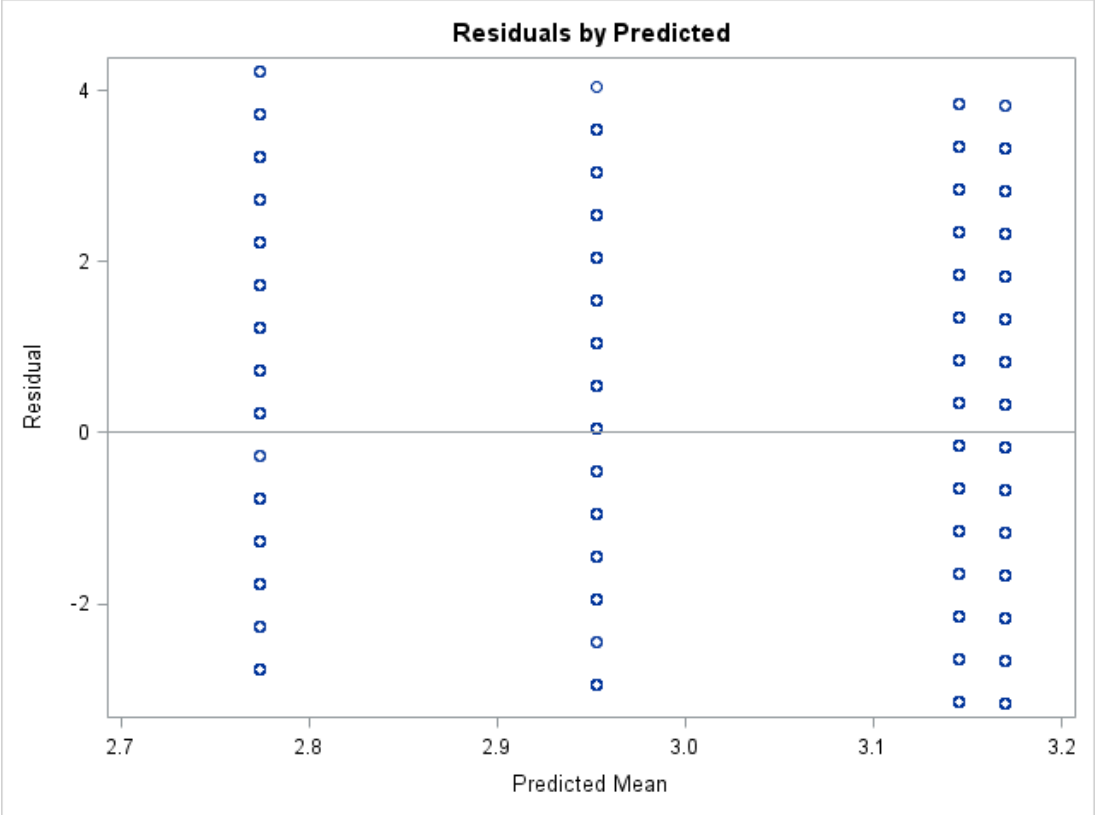

Figure 7a. "Residual by Predicted" Plot for Model 2b

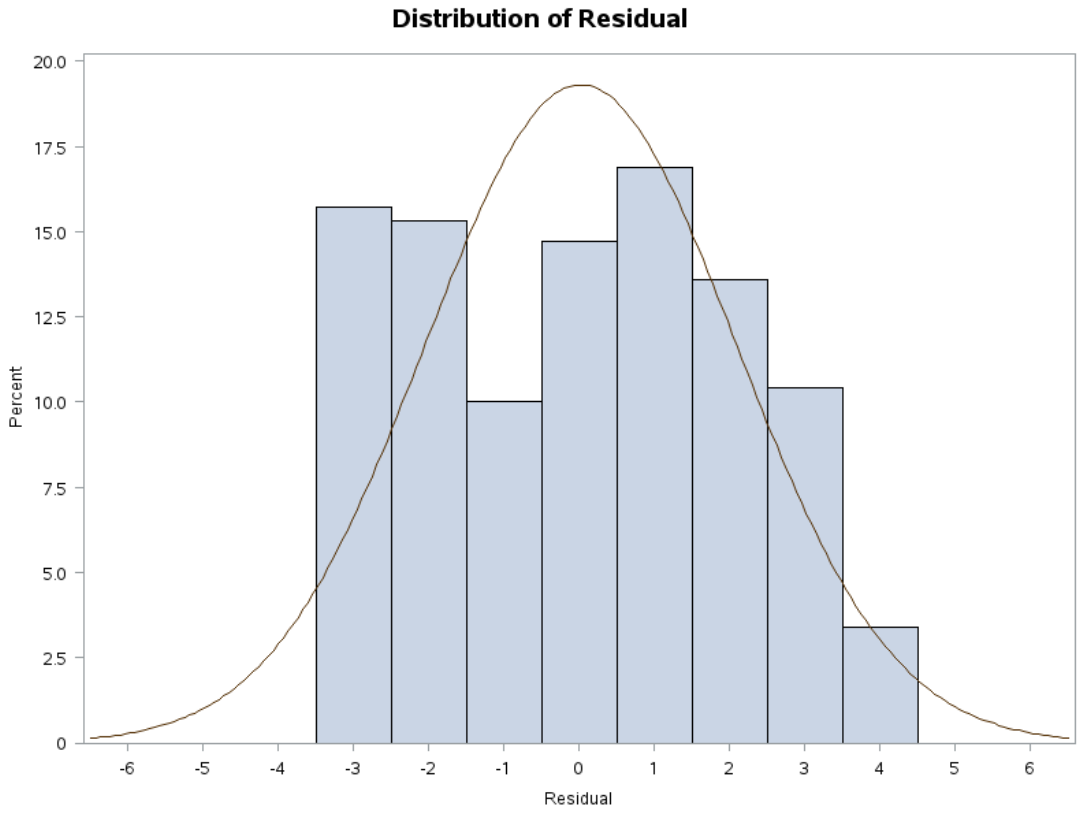

Figure 7b. "Distribution of Residual" Plot for Model 2b

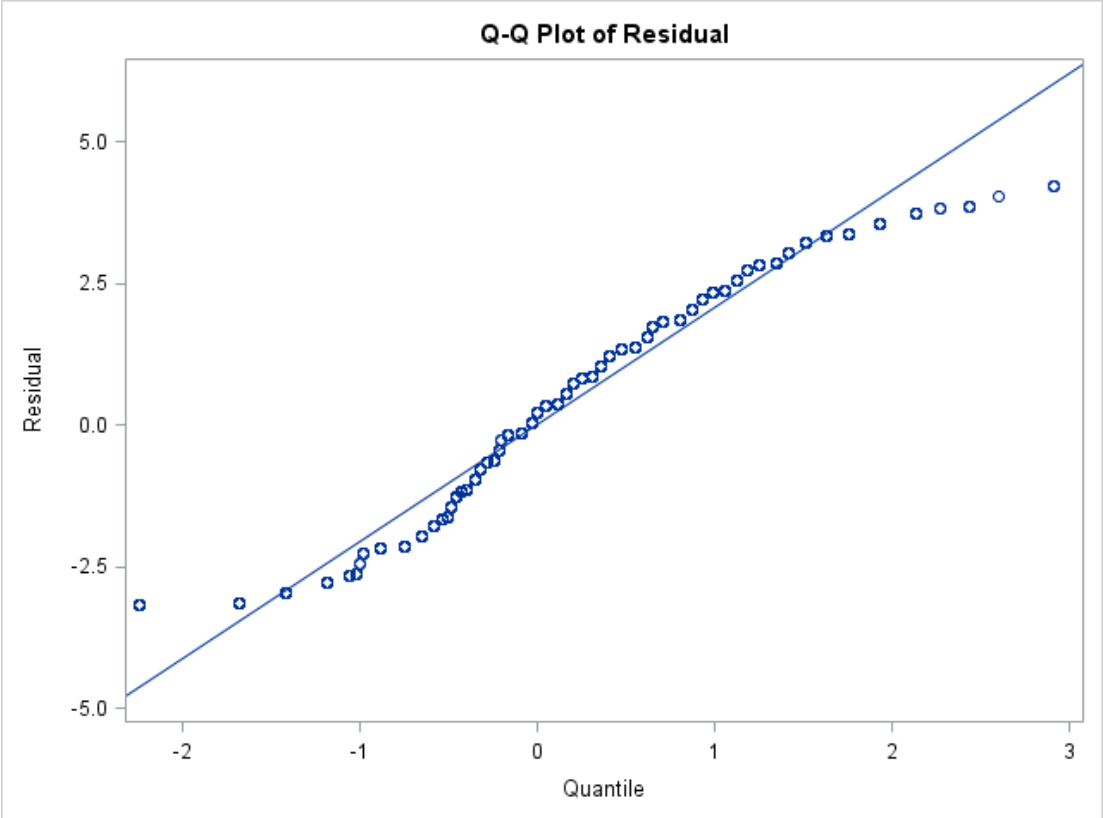

Figure 7c. "Q-Q Plot of Residual" Plot for Model 2b

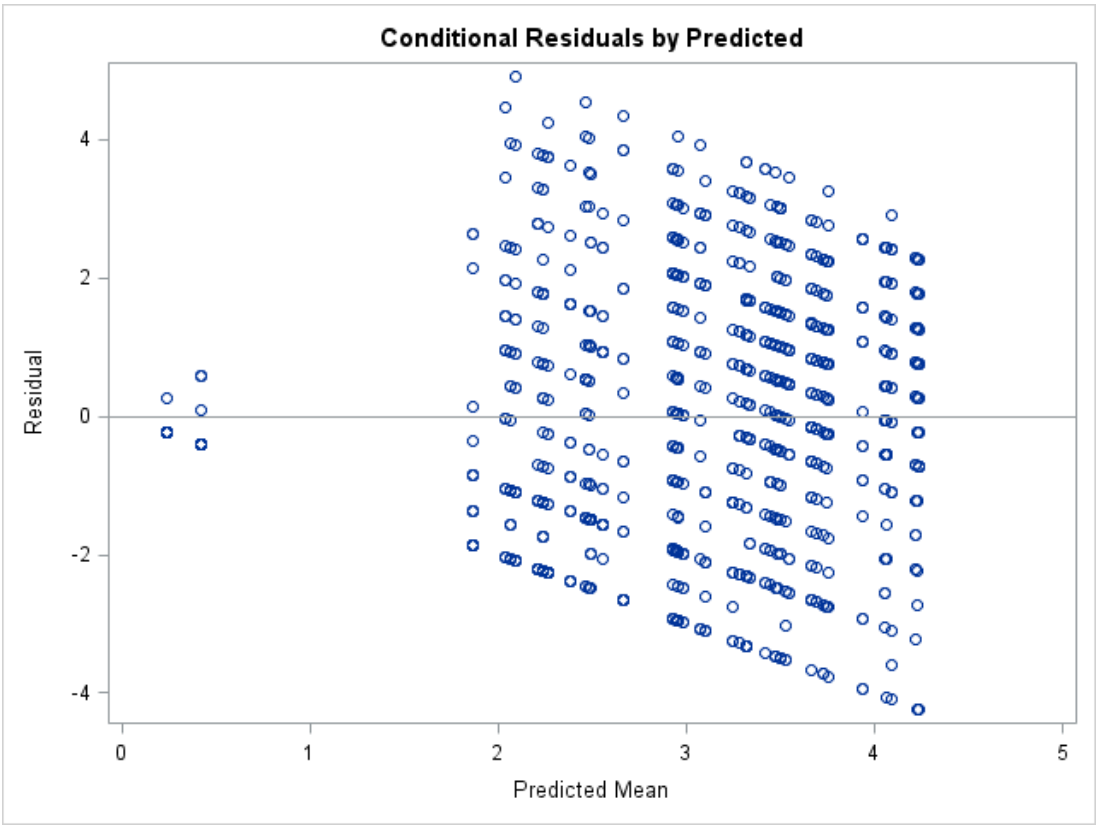

Figure 7d. "Conditional Residual by Predicted" Plot for Model 2b

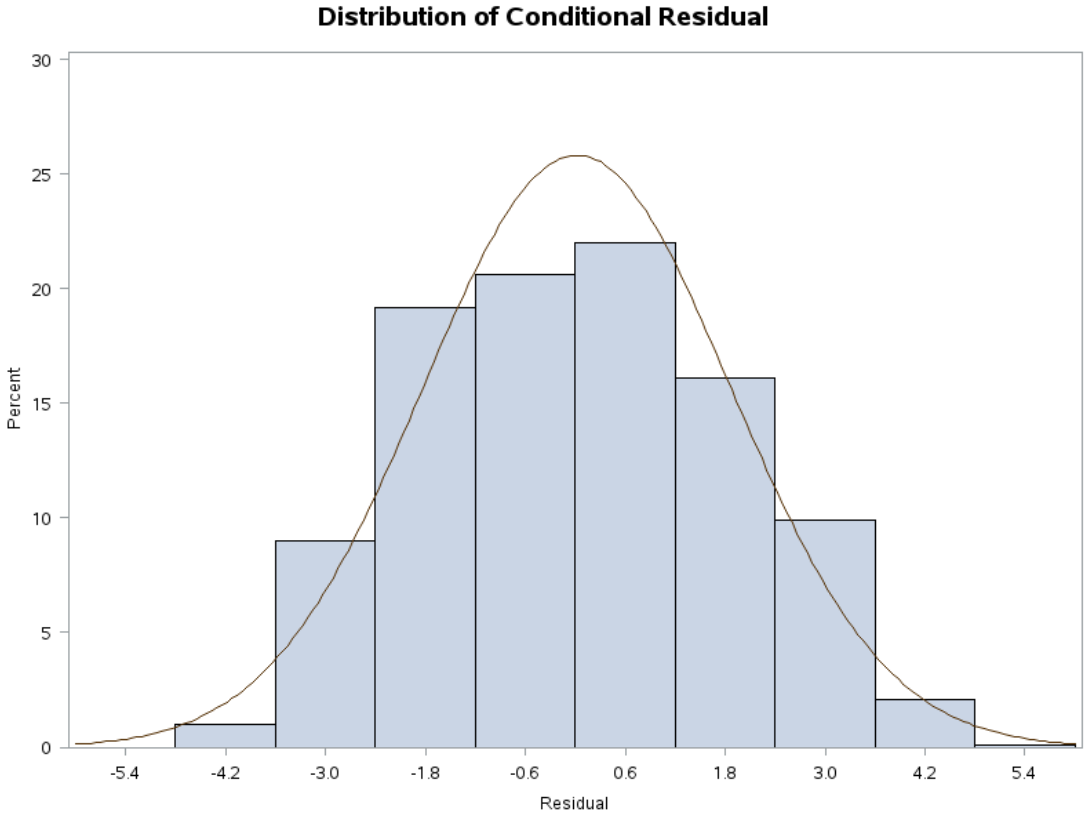

Figure 7e. "Distribution of Conditional Residual" Plot for Model 2b

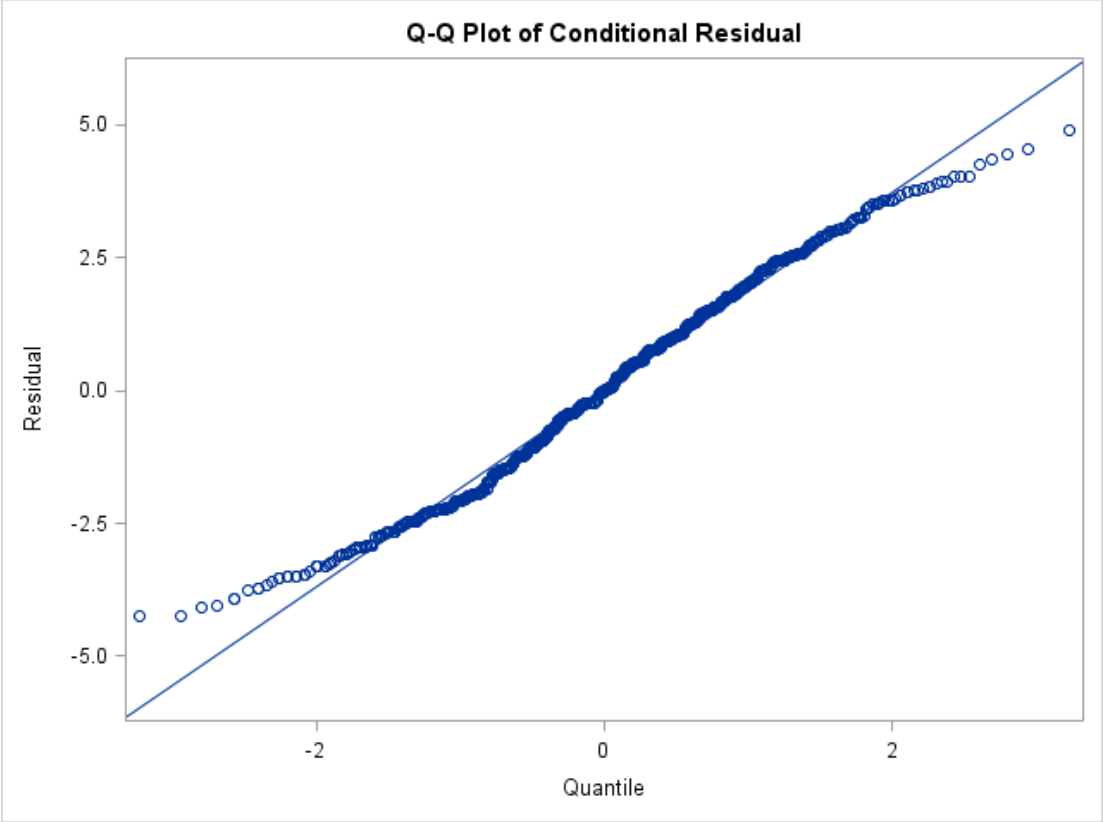

Figure 7f. "Q-Q Plot of Conditional Residual" Plot for Model 2b

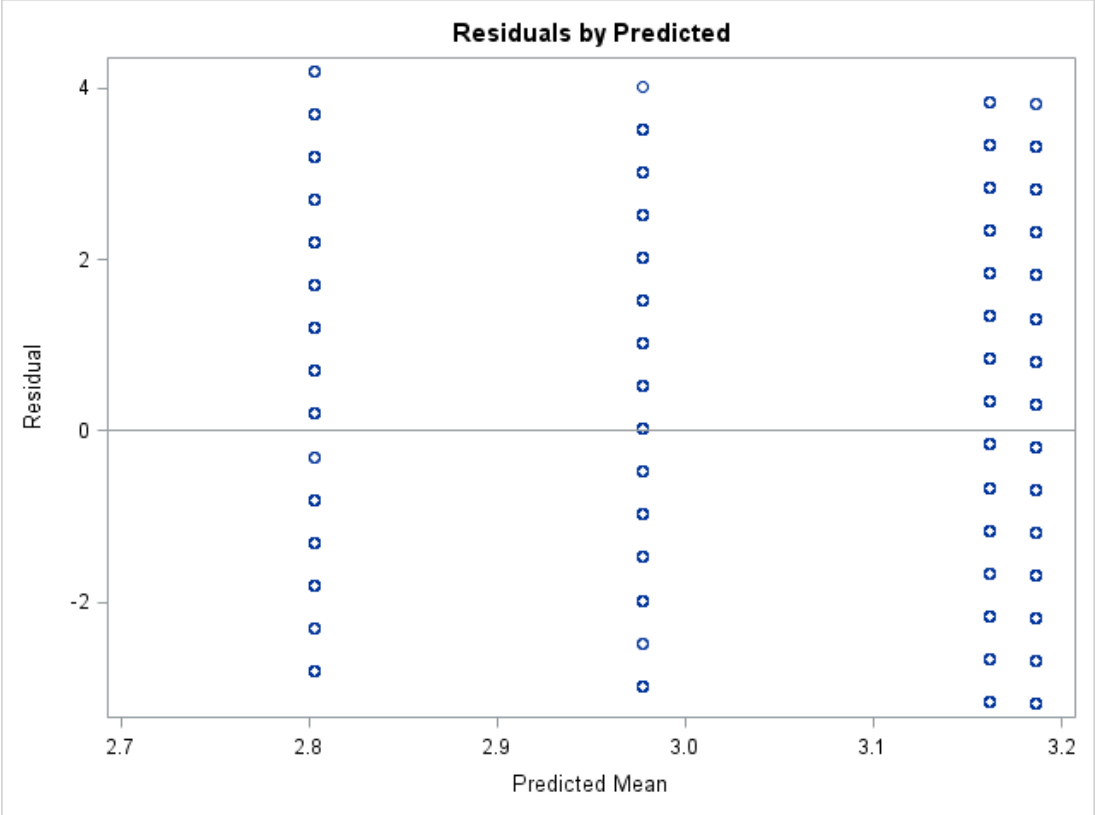

Figure 8a. "Residual by Predicted" Plot for Model 2c

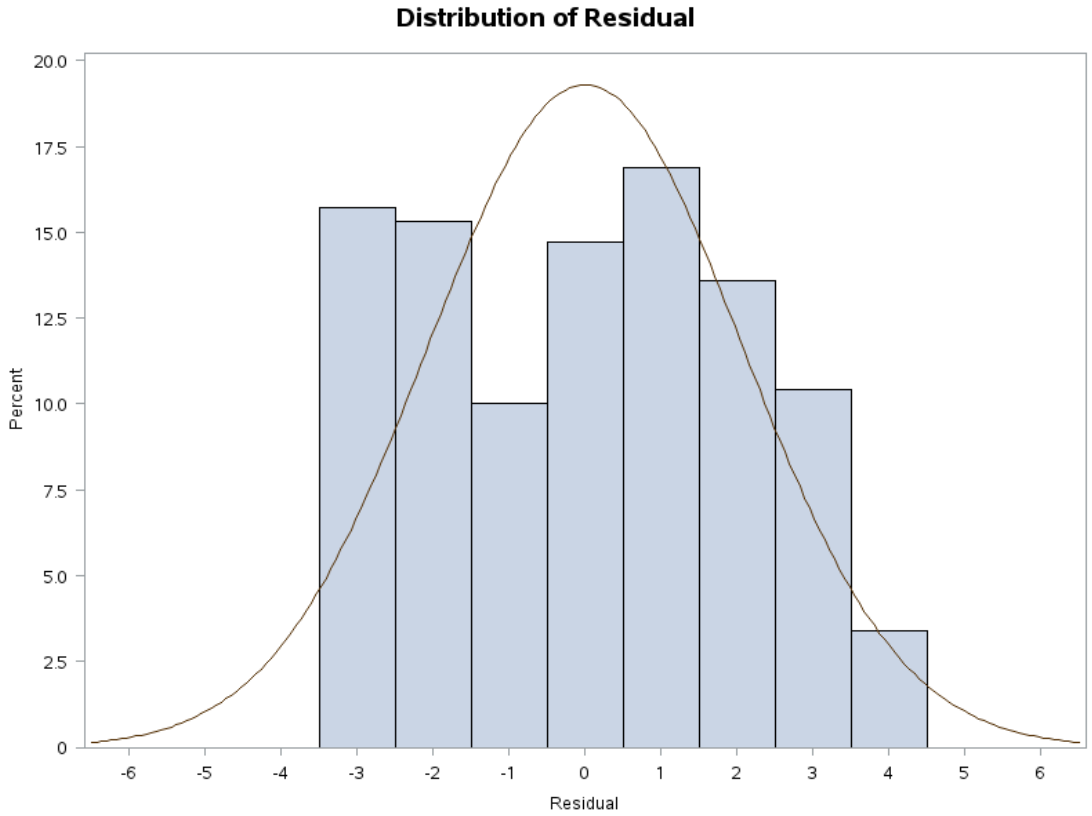

Figure 8b. "Distribution of Residual" Plot for Model 2c

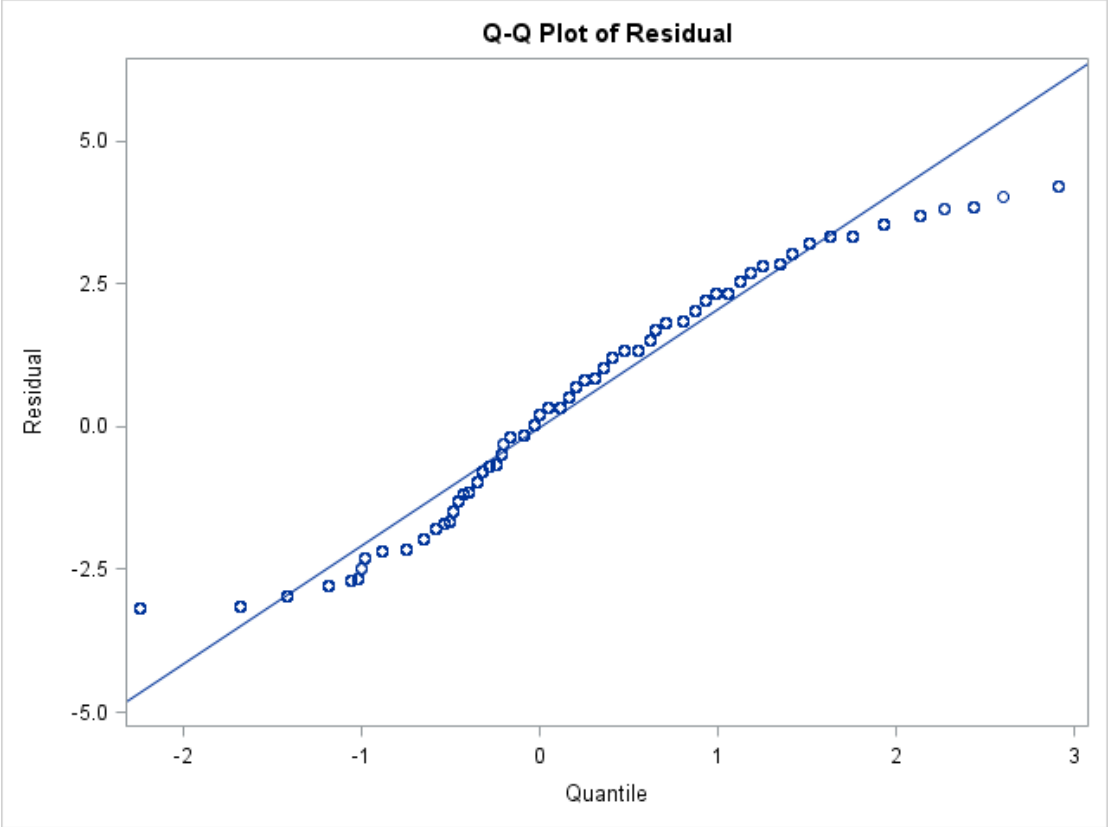

Figure 8c. "Q-Q Plot of Residual" Plot for Model 2c

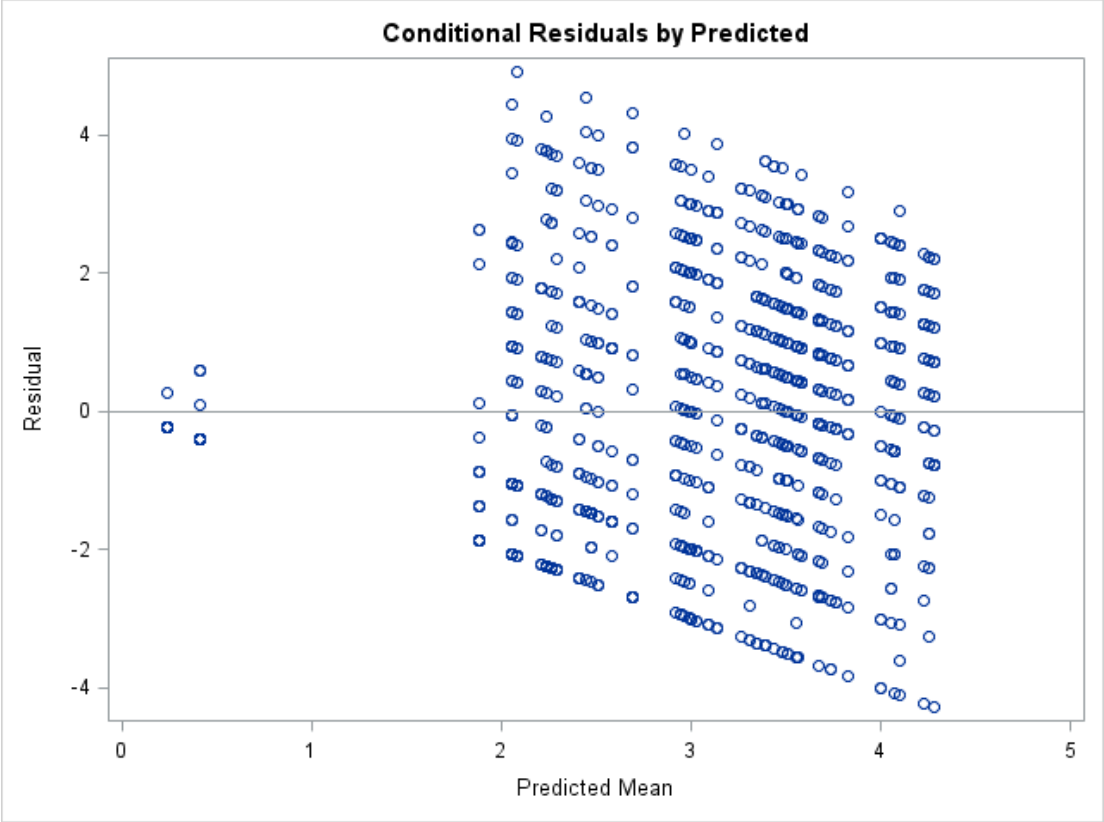

Figure 8d. "Conditional Residual by Predicted" Plot for Model 2c

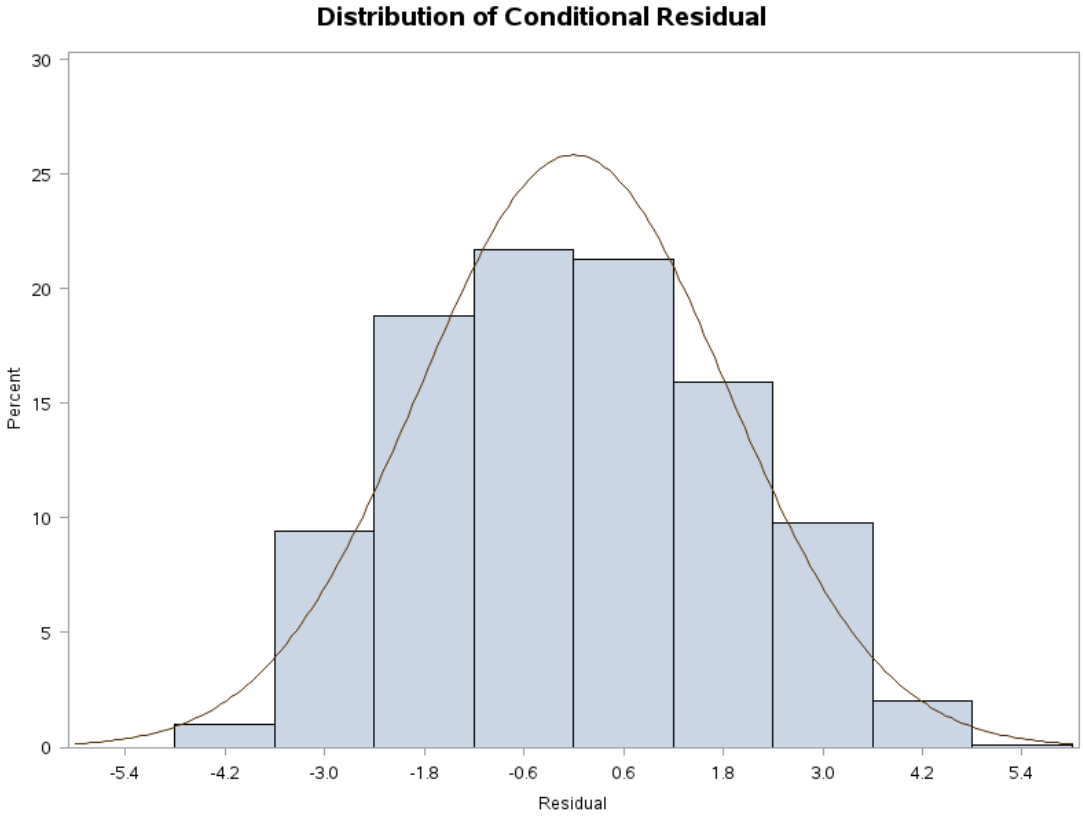

Figure 8e. "Distribution of Conditional Residual" Plot for Model 2c

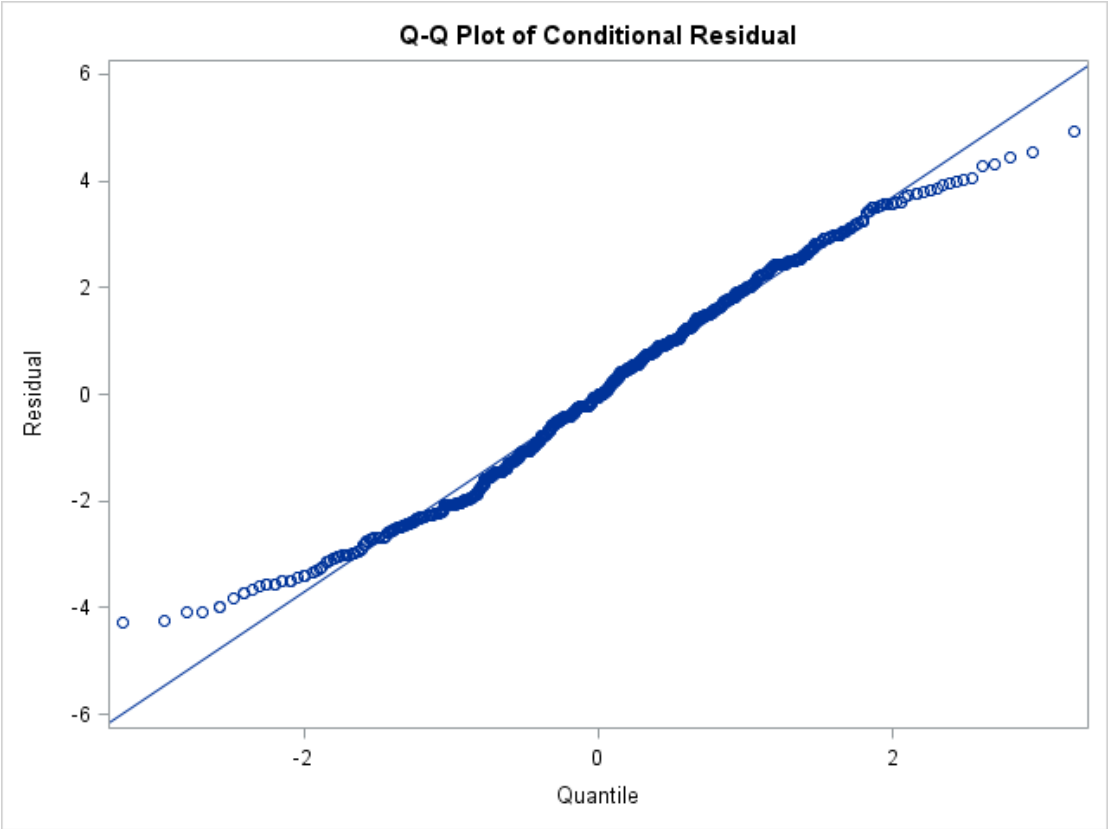

Figure 8d. “Q-Q Plot of Conditional Residual” Plot for Model 2c

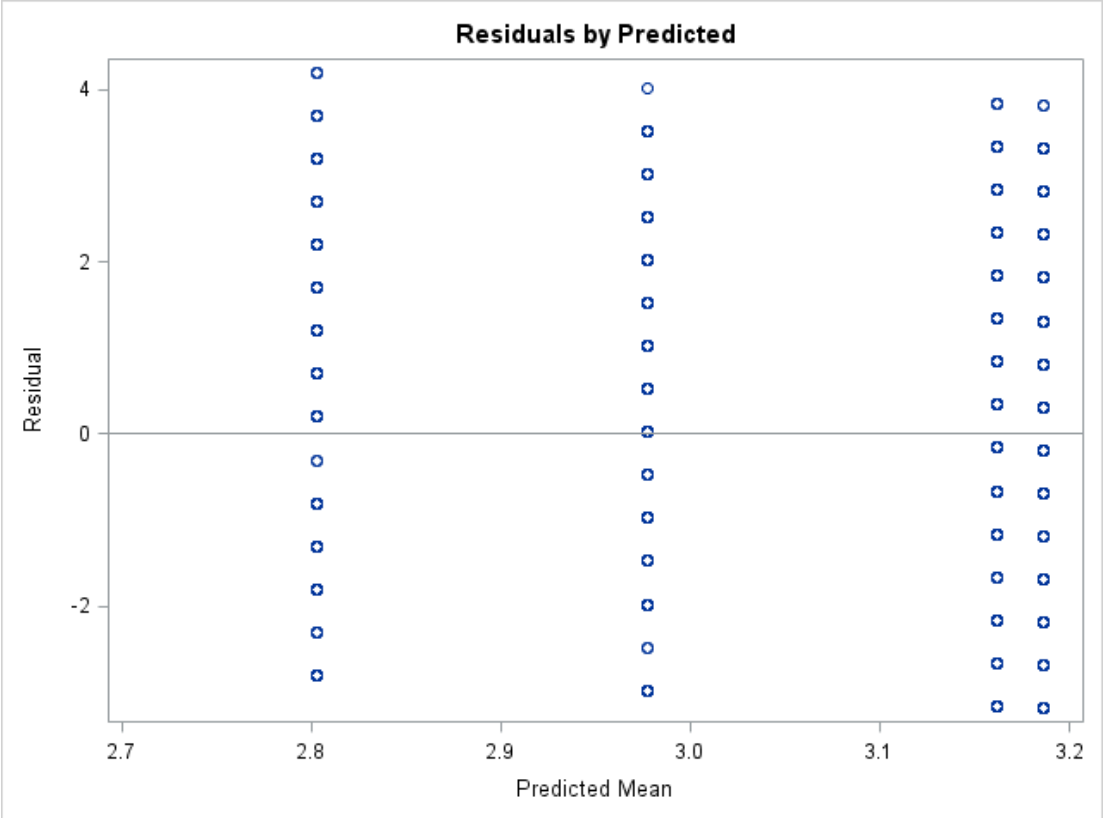

Figure 9a. “Residual by Predicted” Plot for Model 2d

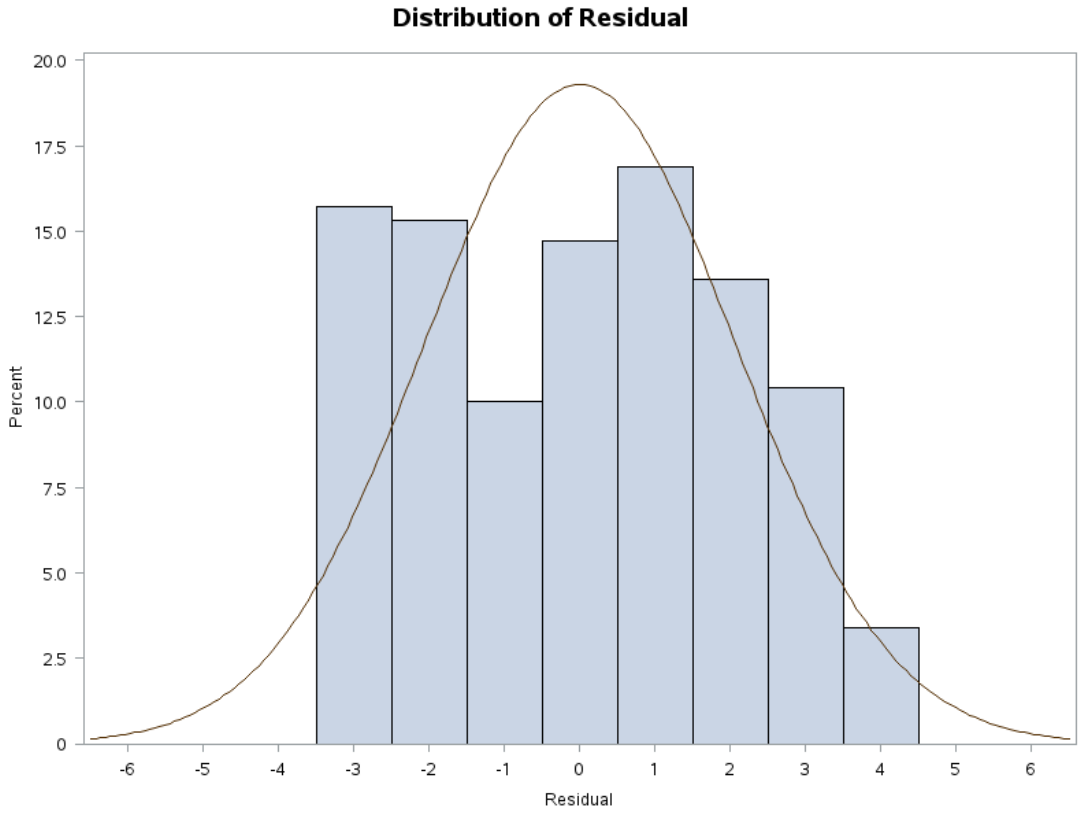

Figure 9b. "Distribution of Residual" Plot for Model 2d

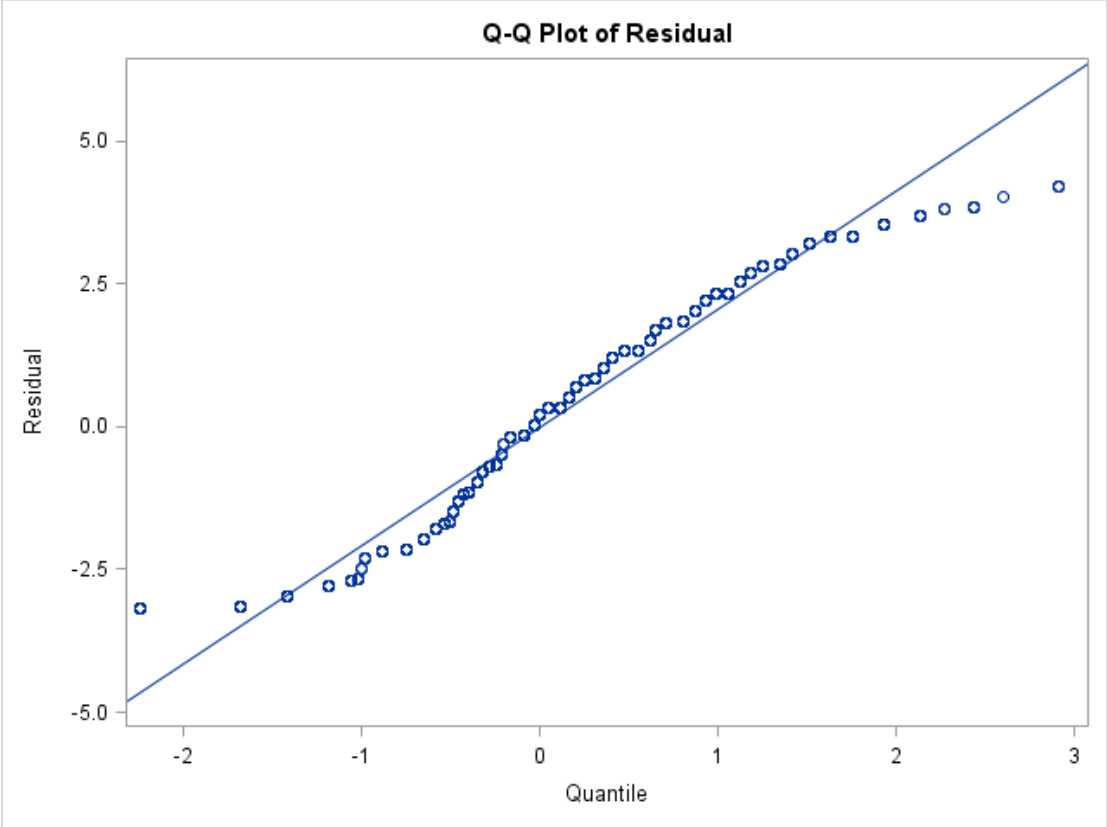

Figure 9c. "Q-Q Plot of Residual" Plot for Model 2d

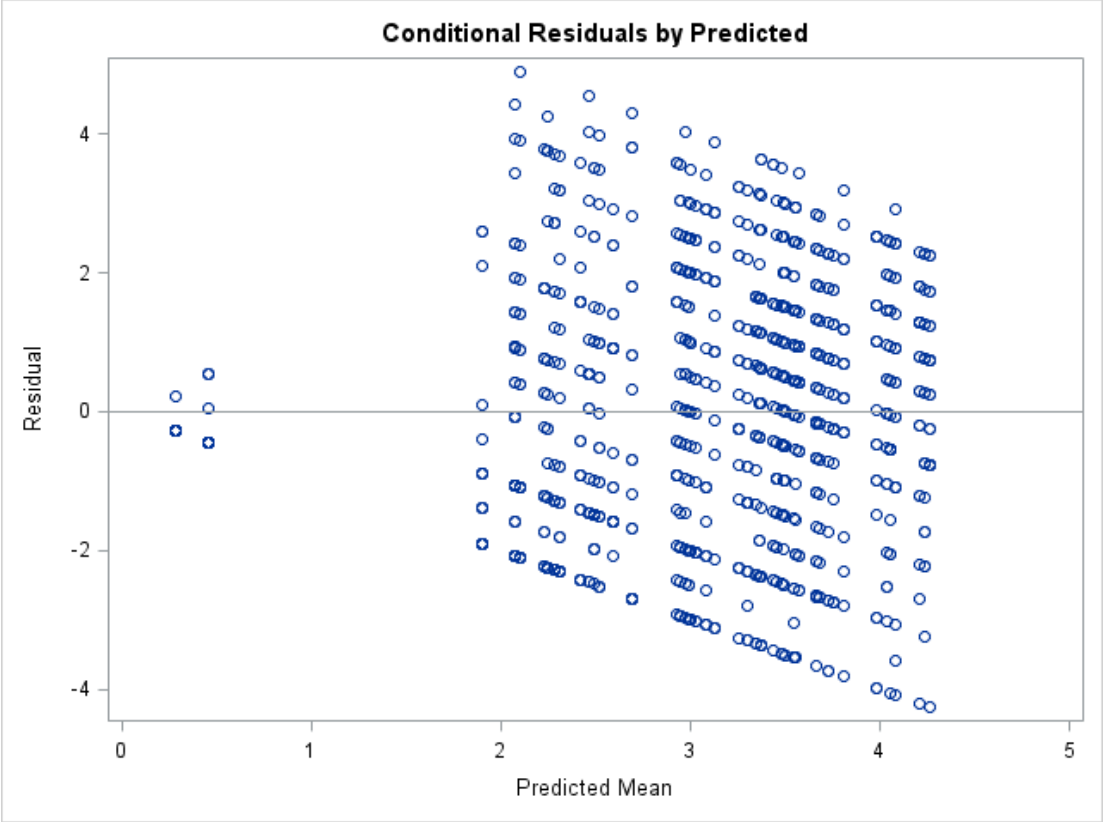

Figure 9d. "Conditional Residual by Predicted" Plot for Model 2d

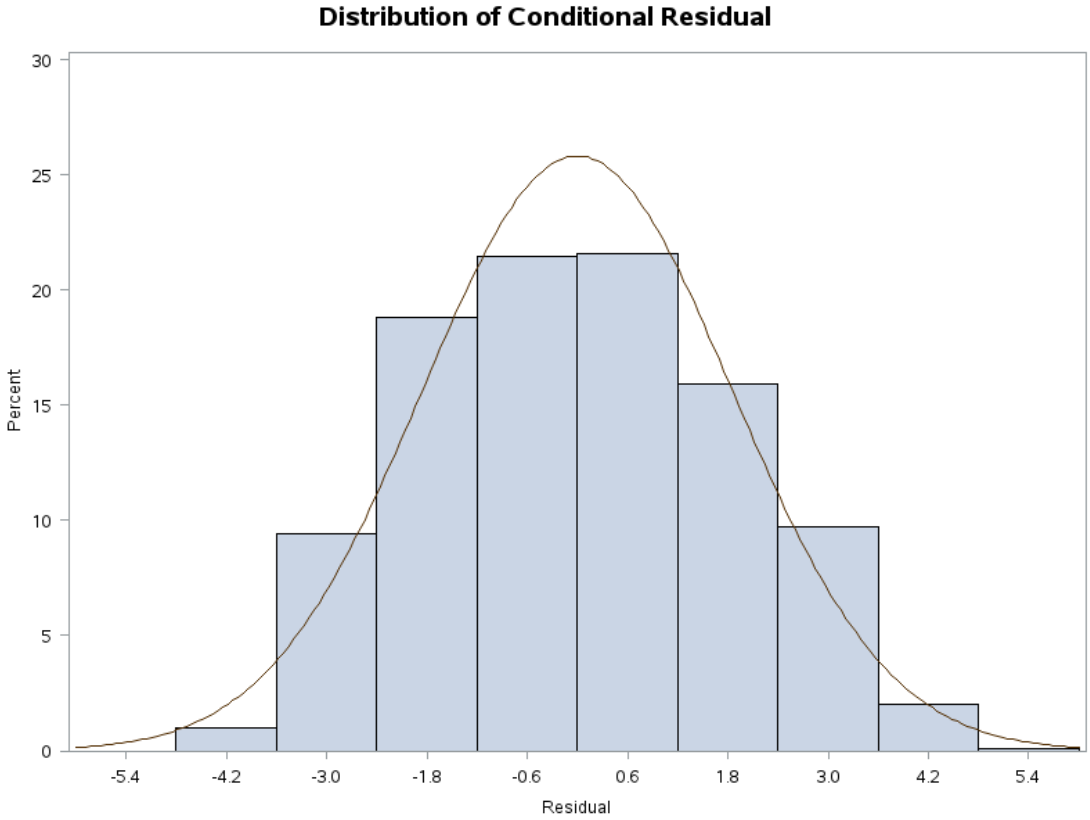

Figure 9e. "Distribution of Conditional Residual" Plot for Model 2d

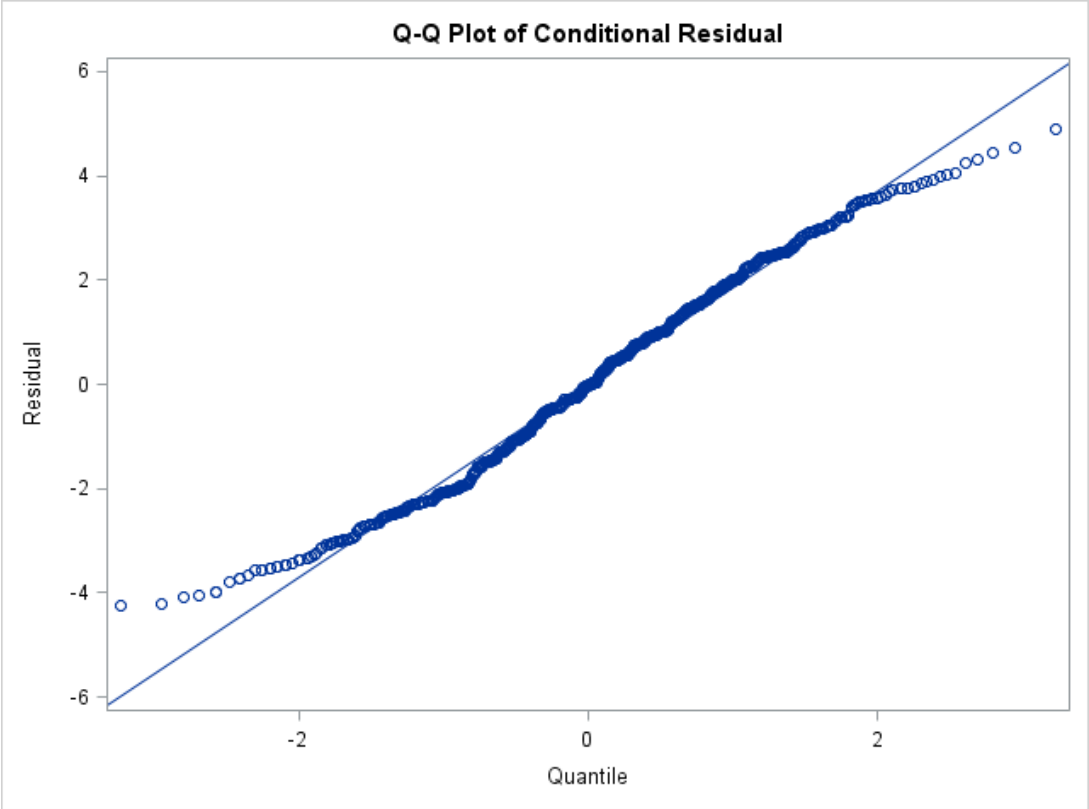

Figure 9f. "Q-Q Plot of Conditional Residual" Plot for Model 2d

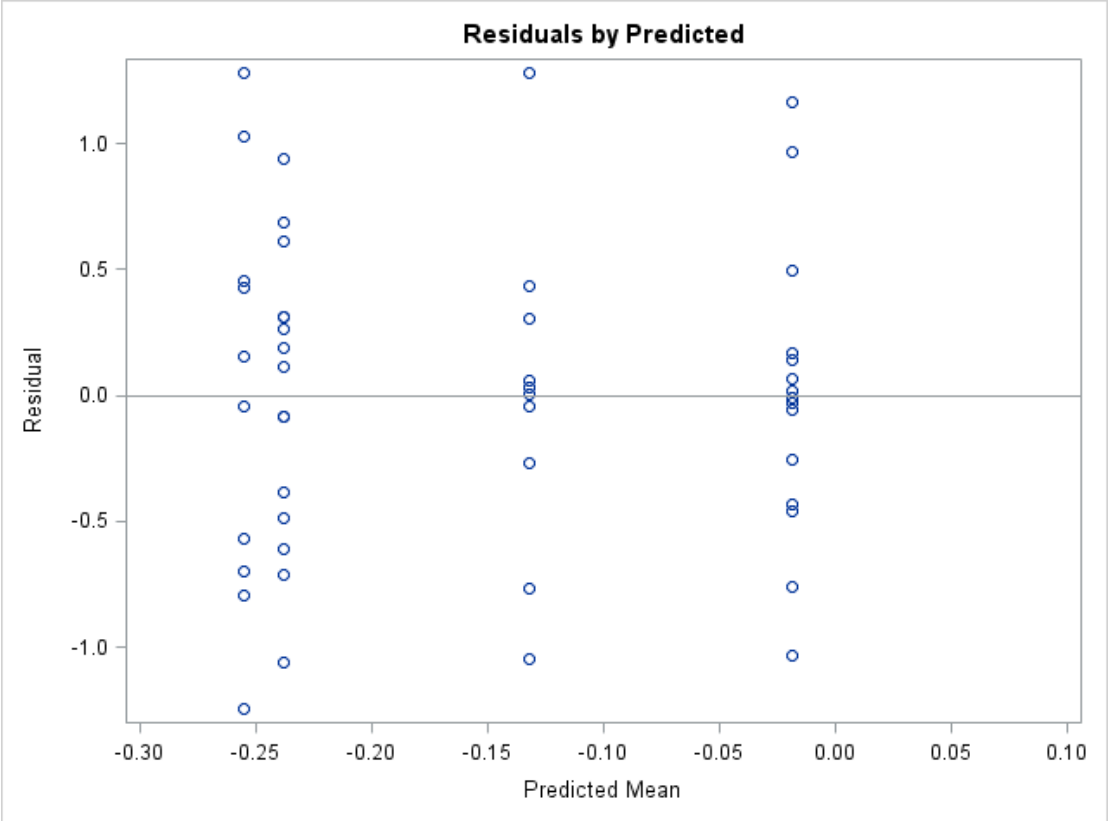

Figure 10a. "Residual by Predicted" Plot for Model 3b

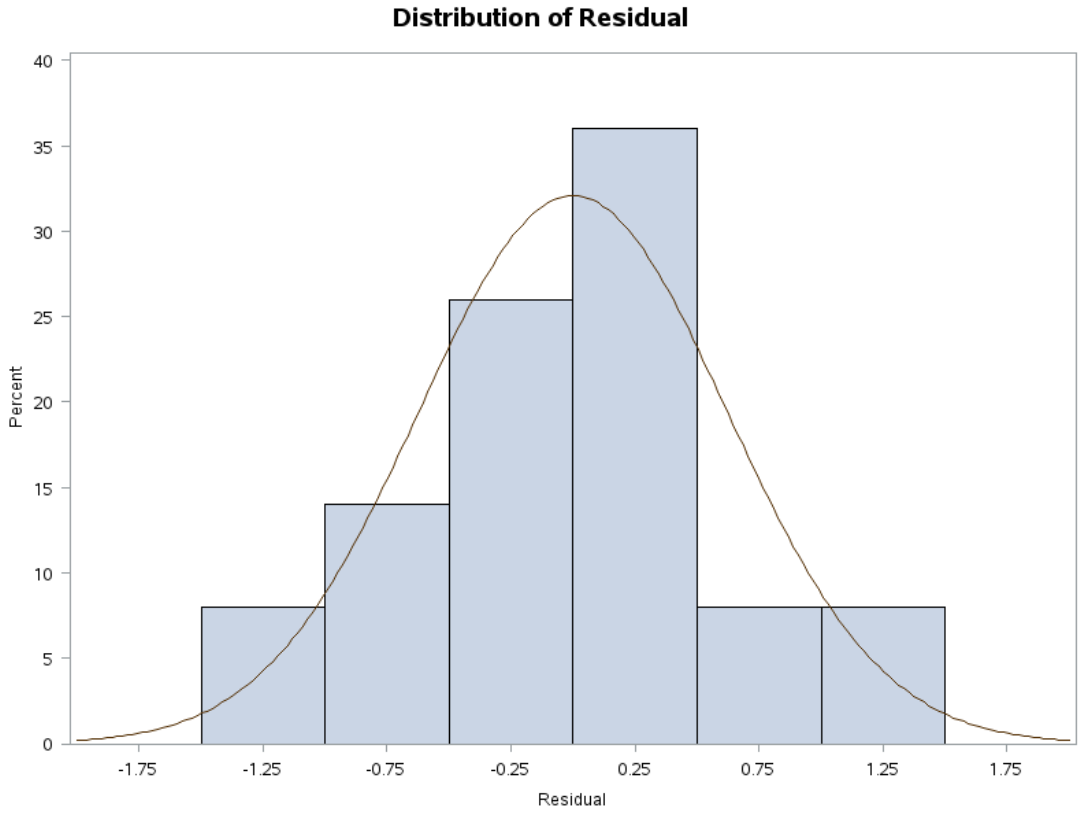

Figure 10b. "Distribution of Residual" Plot for Model 3b

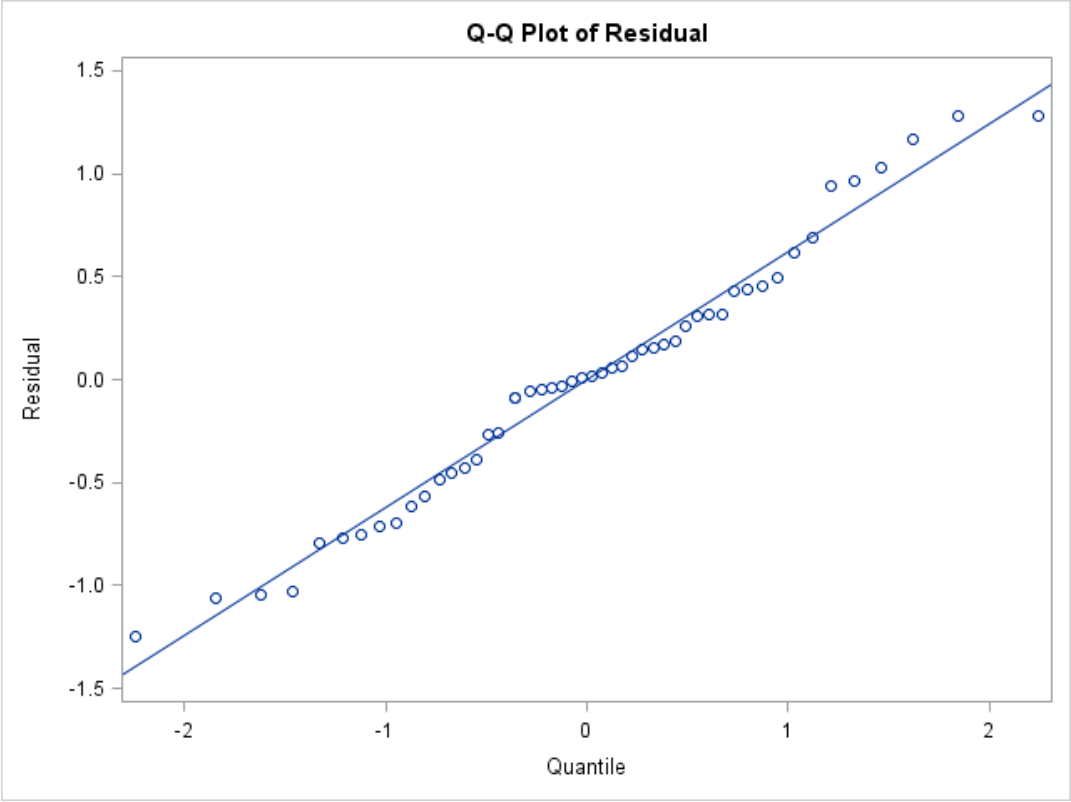

Figure 10c. "Distribution of Residual" Plot for Model 3b

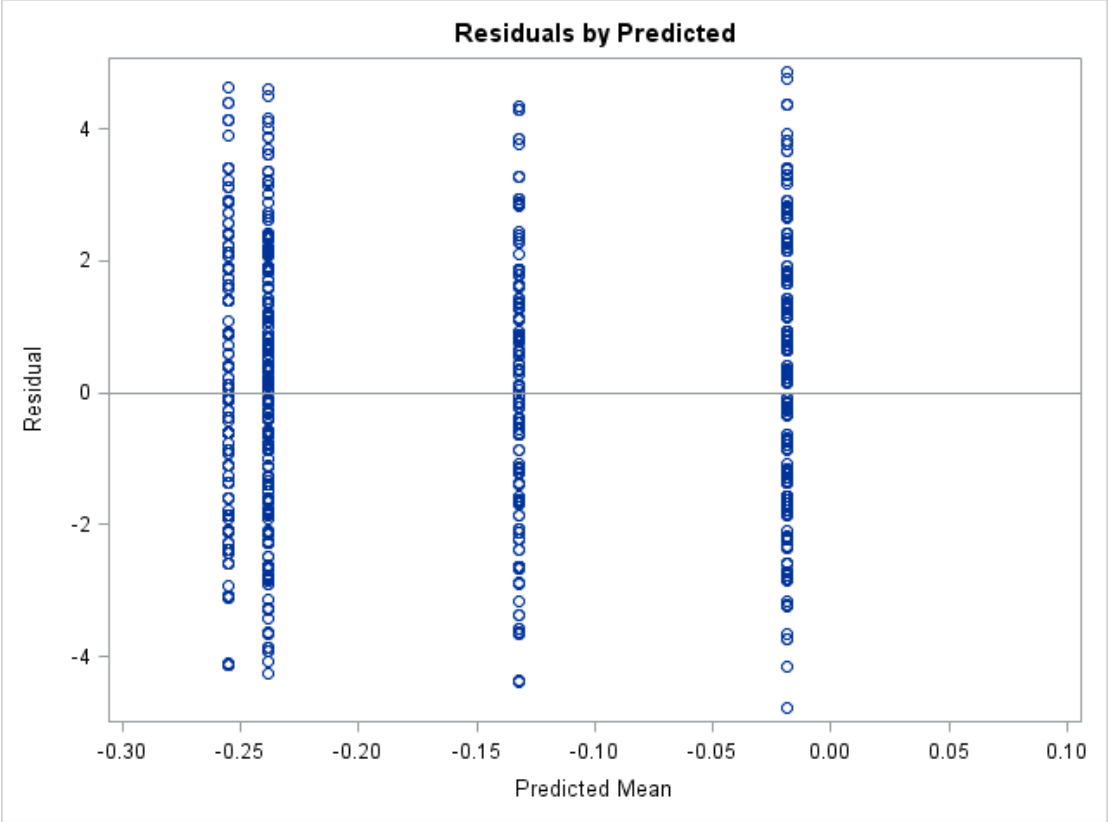

Figure 11a. "Residual by Predicted" Plot for Model 3e

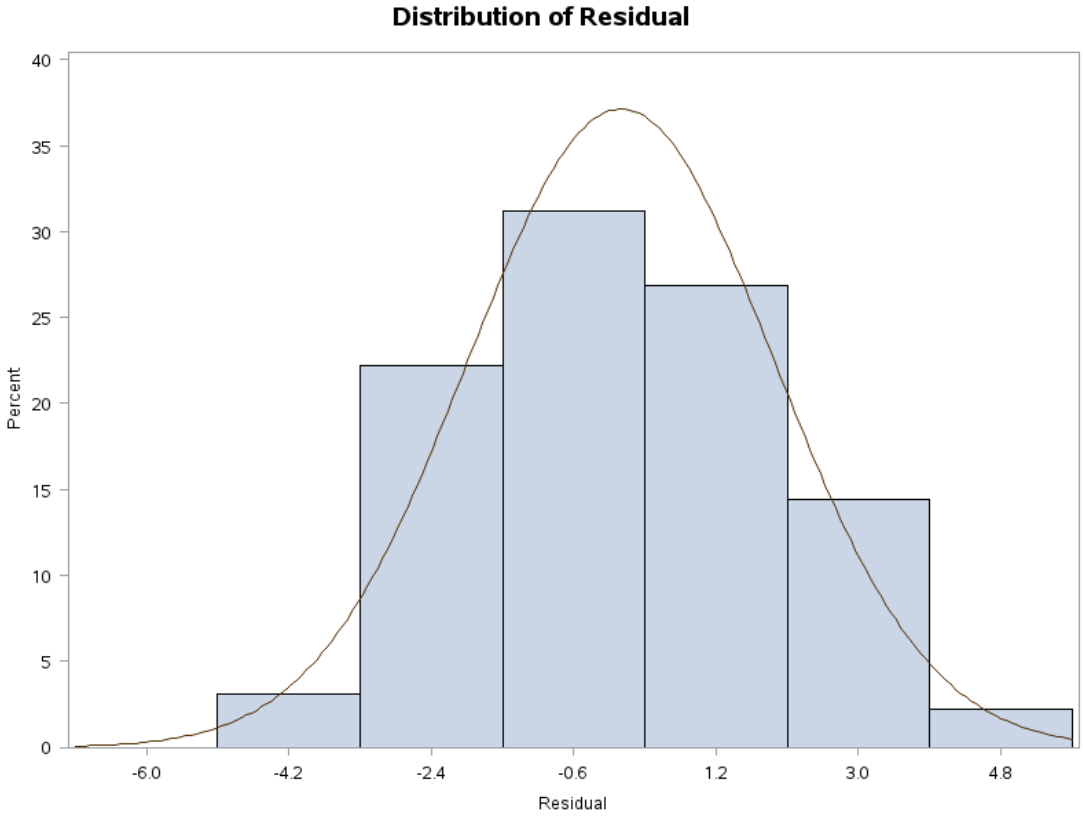

Figure 11b. "Distribution of Residual" Plot for Model 3e

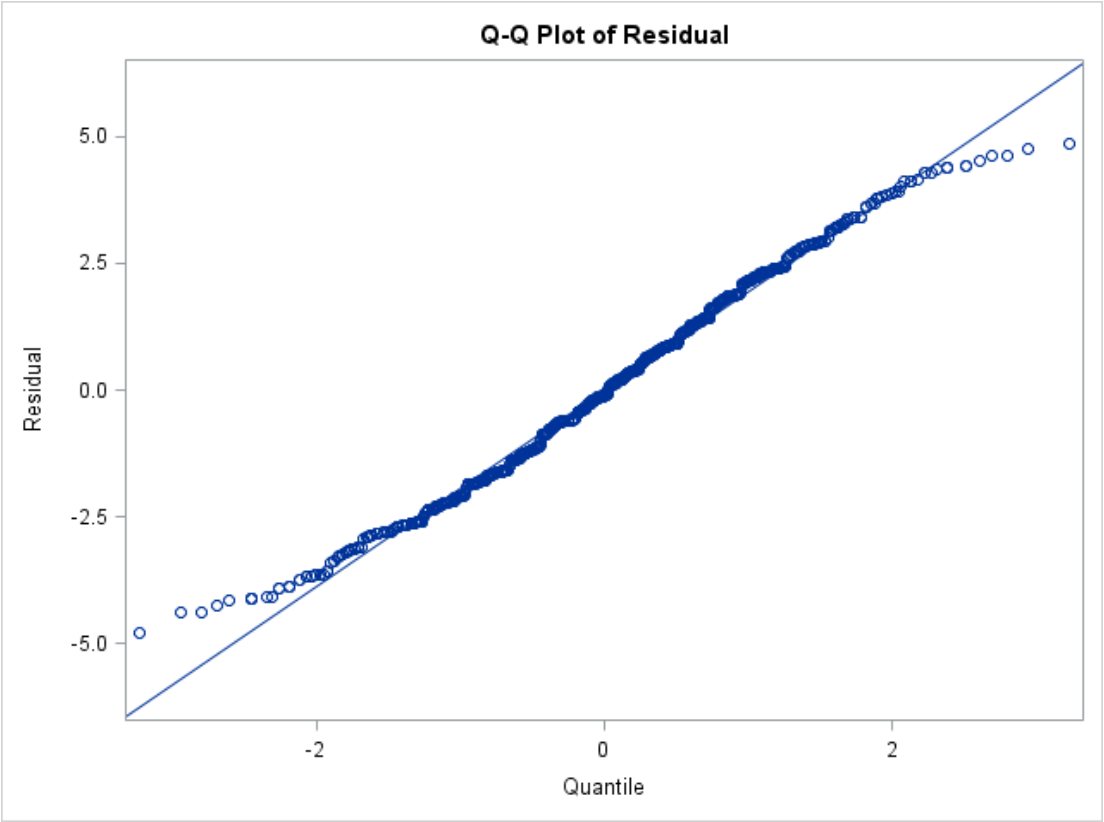

Figure 11c. "Q-Q Plot of Residual" Plot for Model 3e

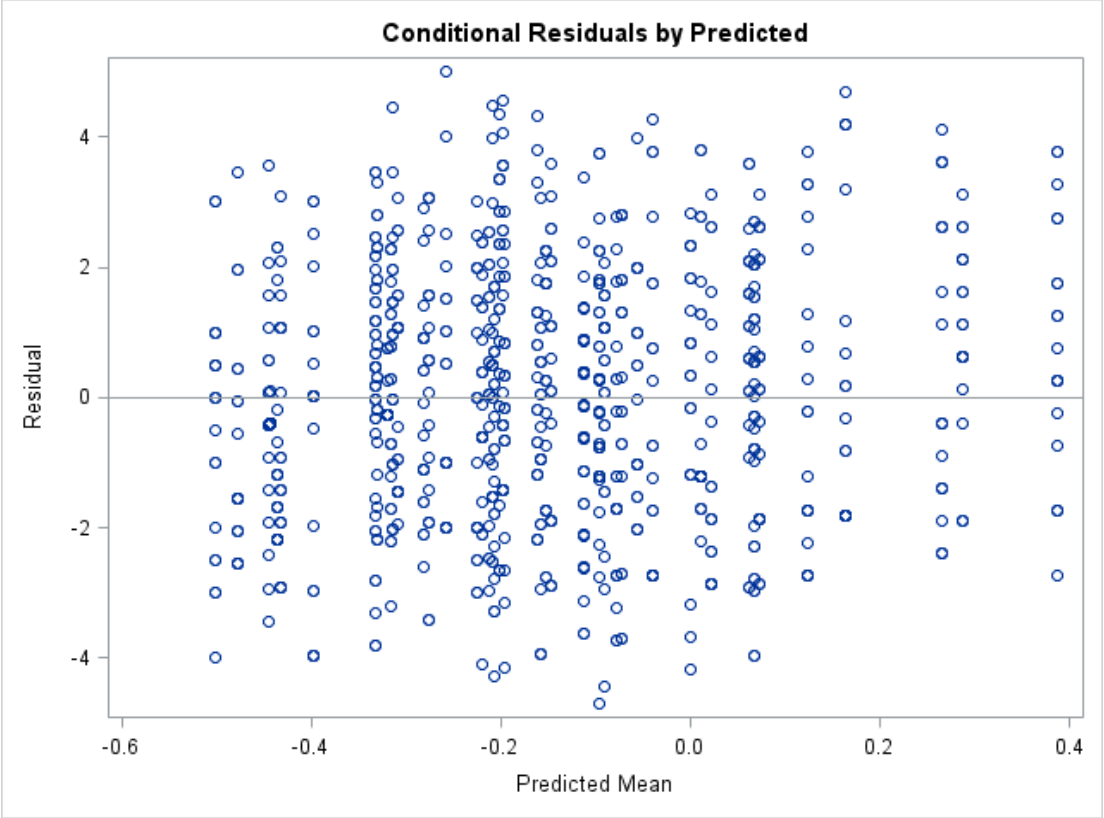

Figure 11d. "Conditional Residual by Predicted" Plot for Model 3e

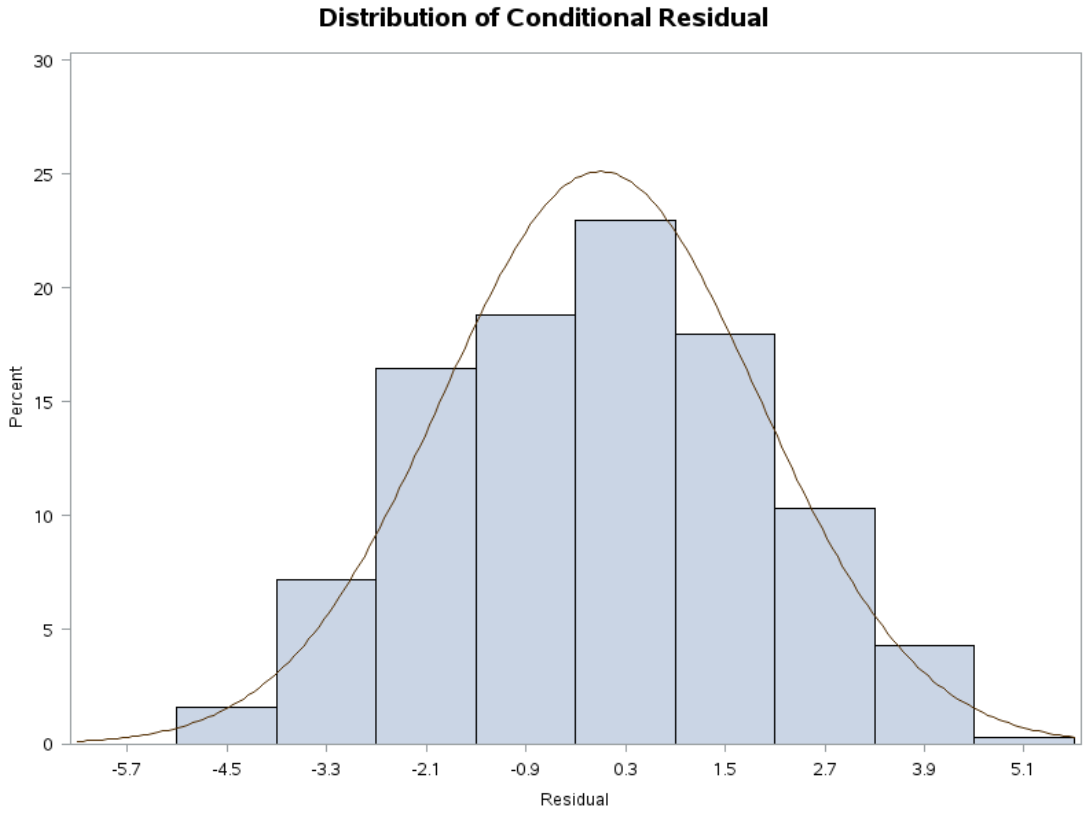

Figure 11e. "Distribution of Conditional Residual" Plot for Model 3e

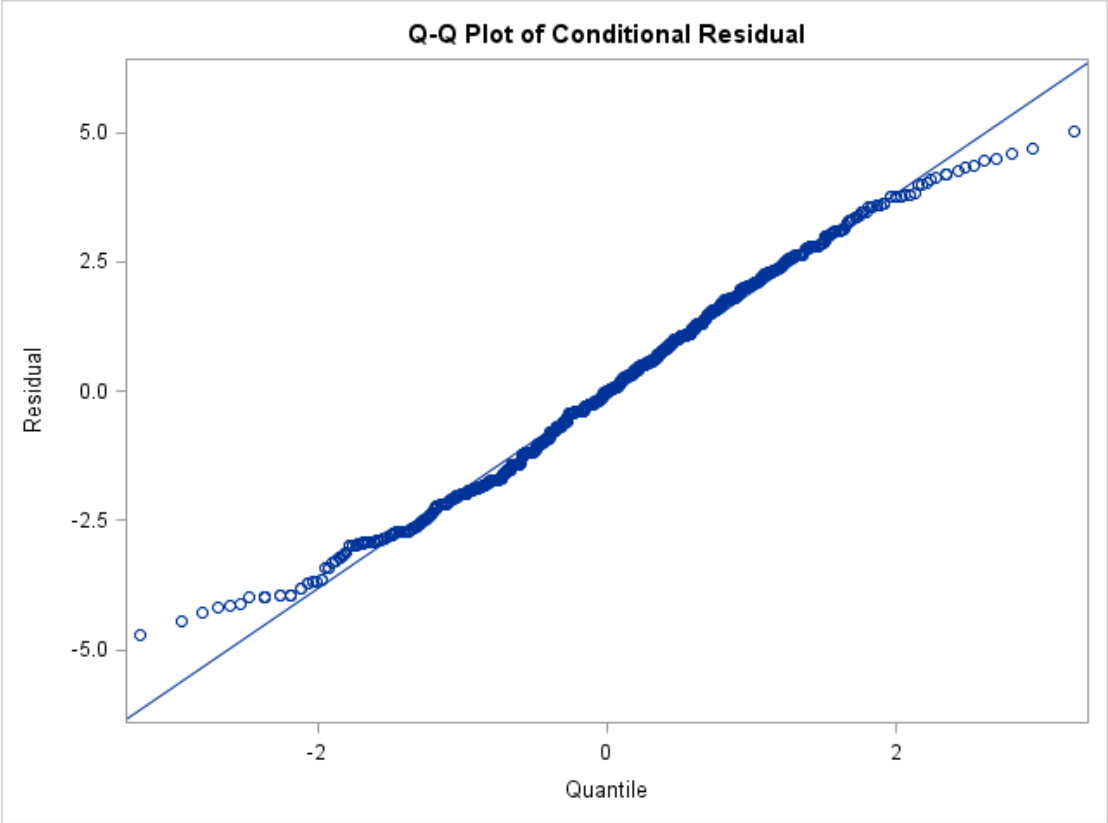

Figure 11f. "Q-Q Plot of Conditional Residual" Plot for Model 3e

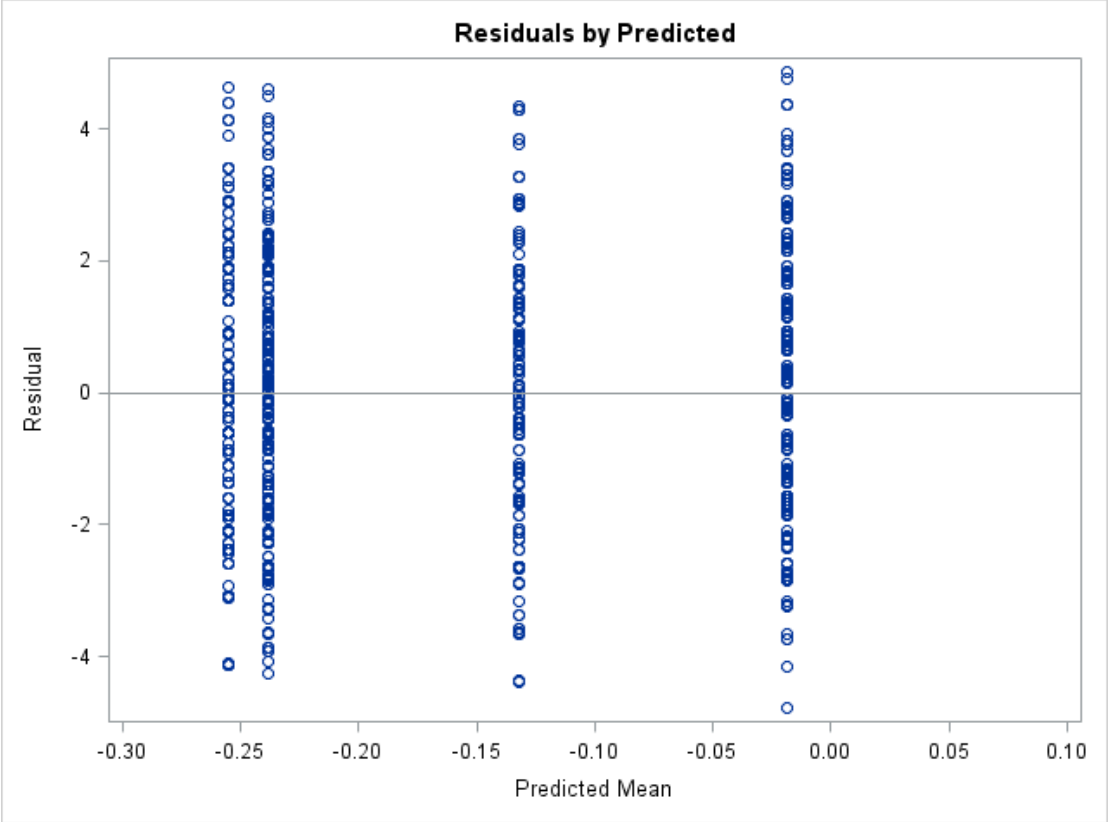

Figure 12a. “Residual by Predicted” Plot for Model 3f

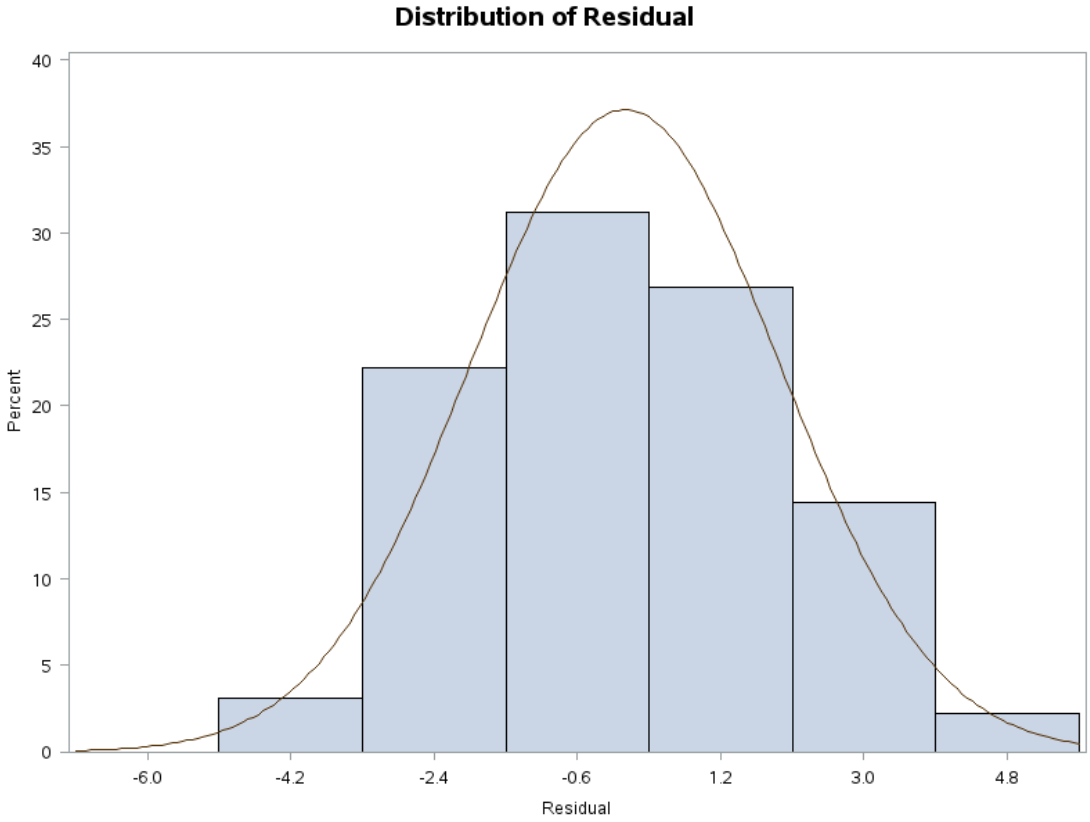

Figure 12b. “Distribution of Residual” Plot for Model 3f

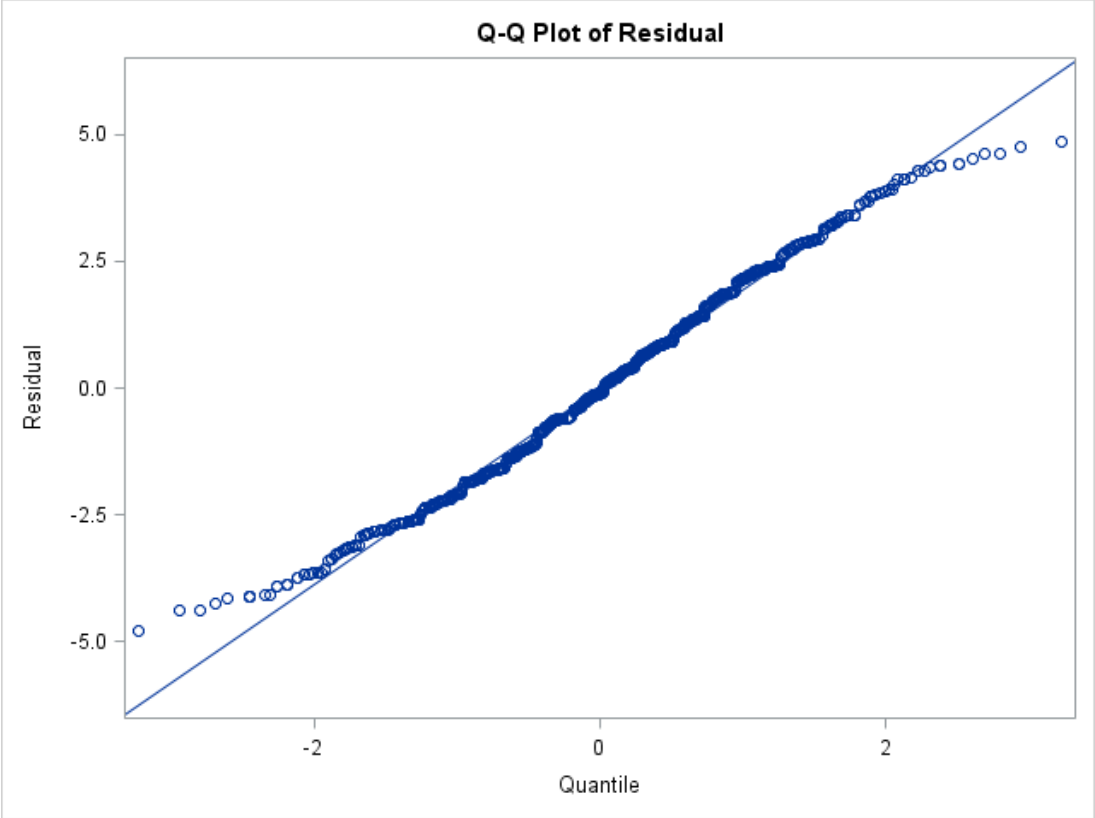

Figure 12b. "Q-Q Plot of Residual" Plot for Model 3f

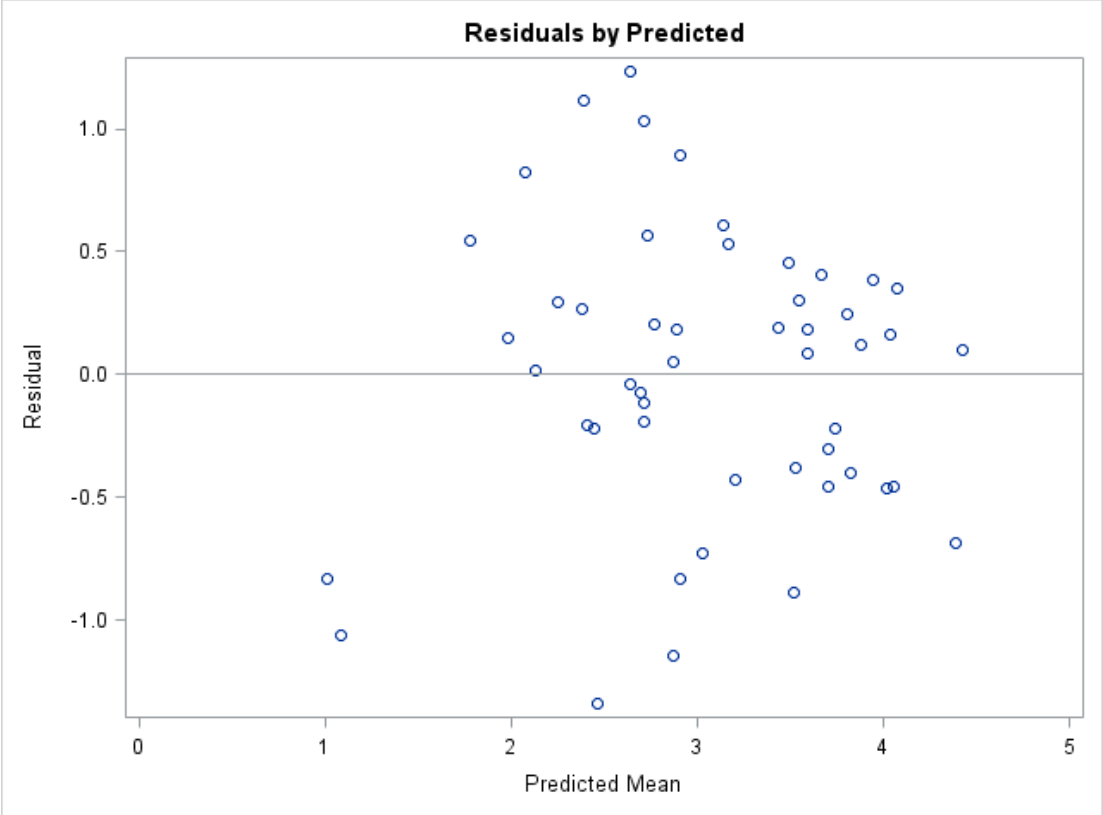

Figure 13a. "Residual by Predicted" Plot for Model 4a

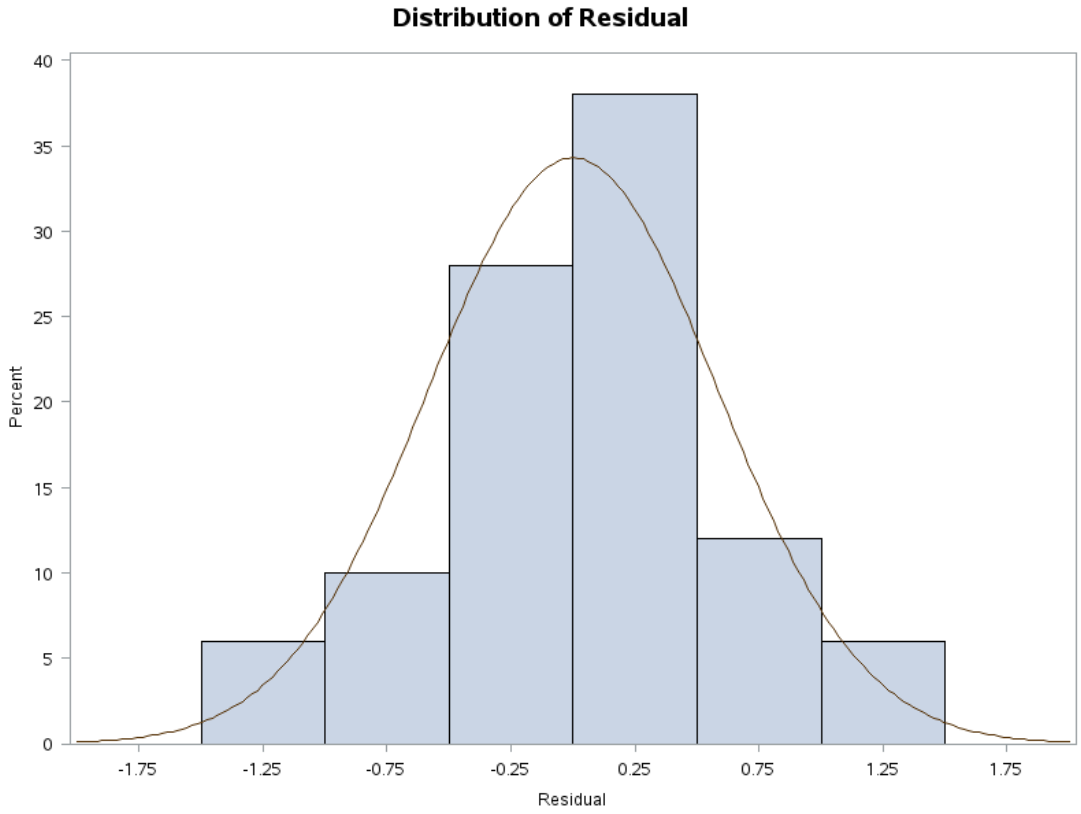

Figure 13b. "Distribution of Residual" Plot for Model 4a

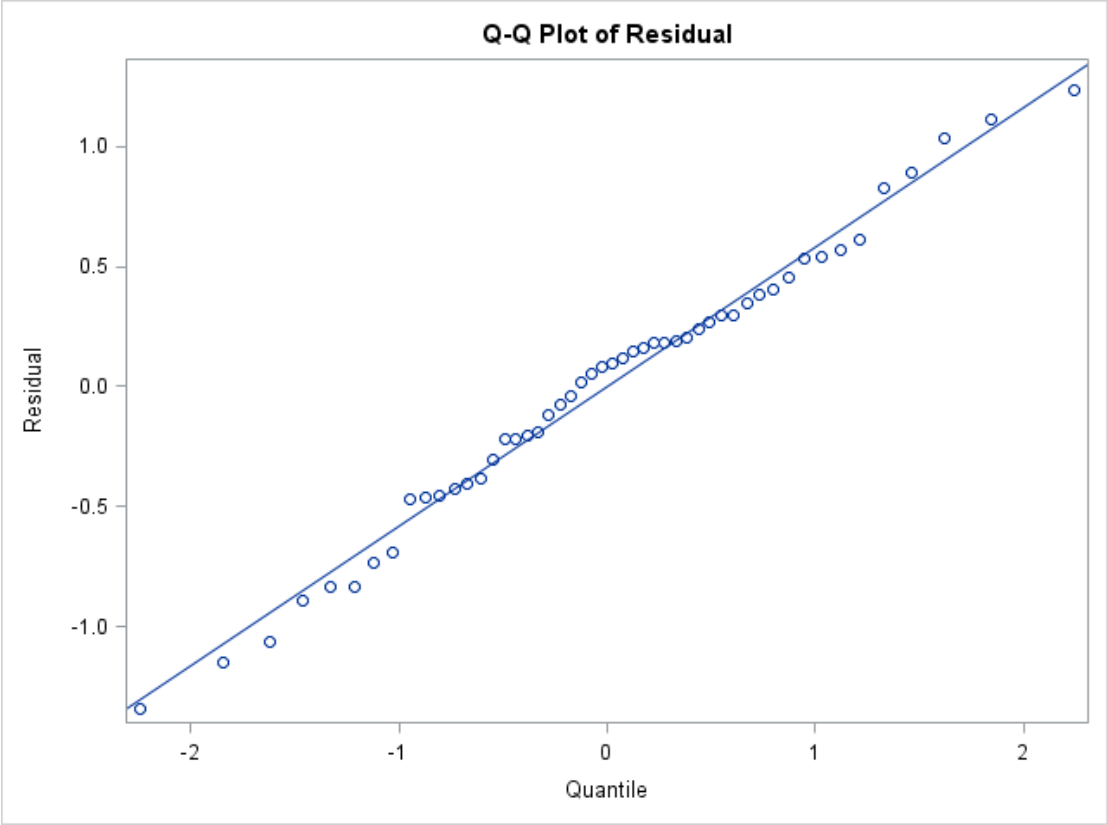

Figure 13c. "Q-Q Plot of Residual" Plot for Model 4a

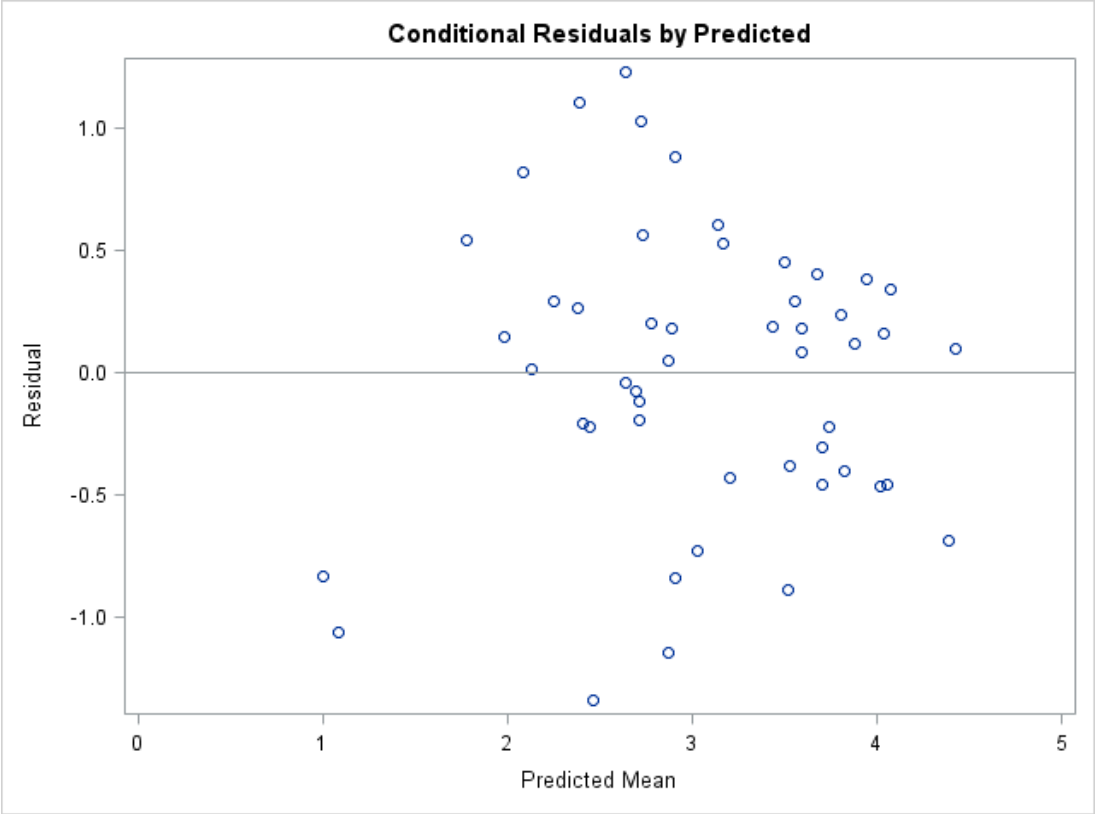

Figure 13d. "Conditional Residual by Predicted" Plot for Model 4a

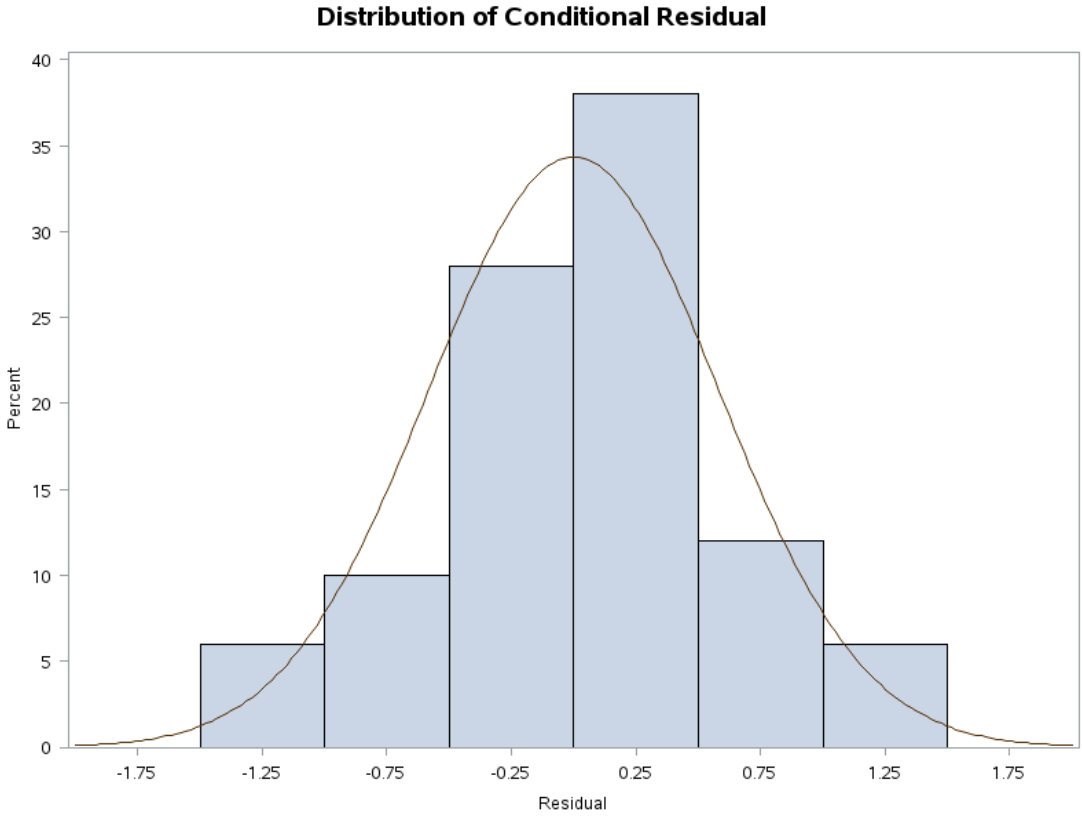

Figure 13e. "Distribution of Conditional Residual" Plot for Model 4a

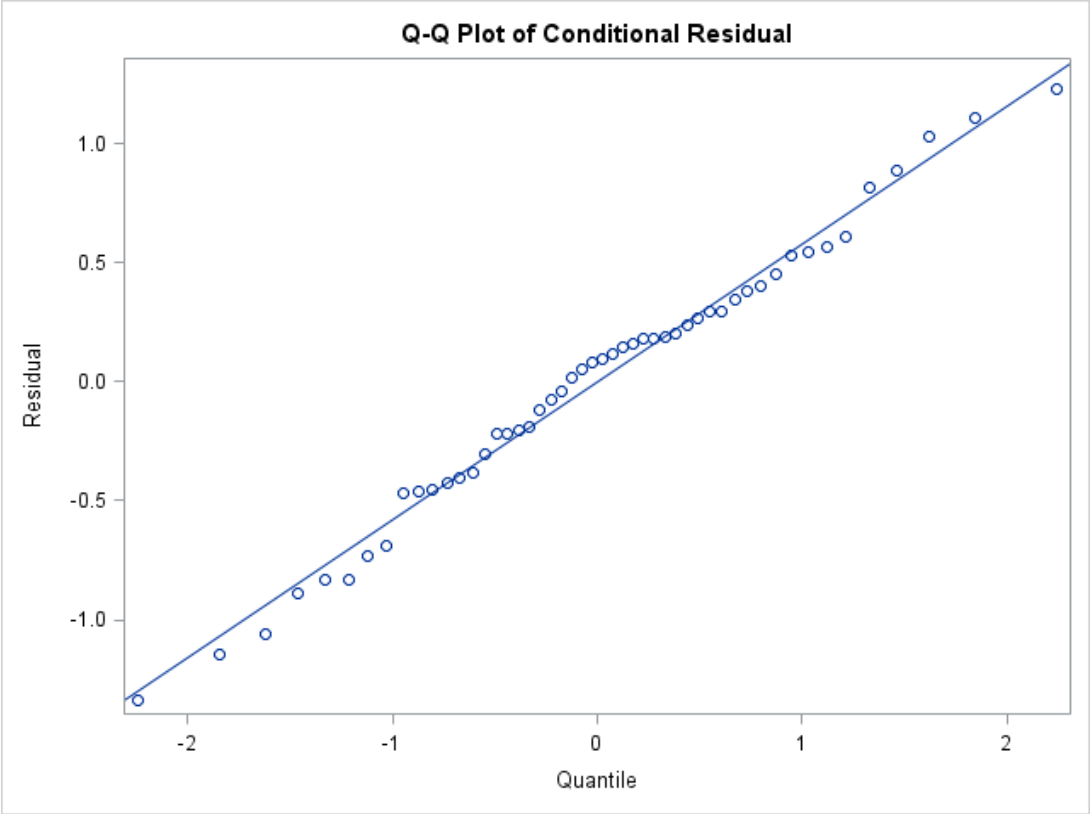

Figure 13f. "Q-Q Plot of Conditional Residual" Plot for Model 4a

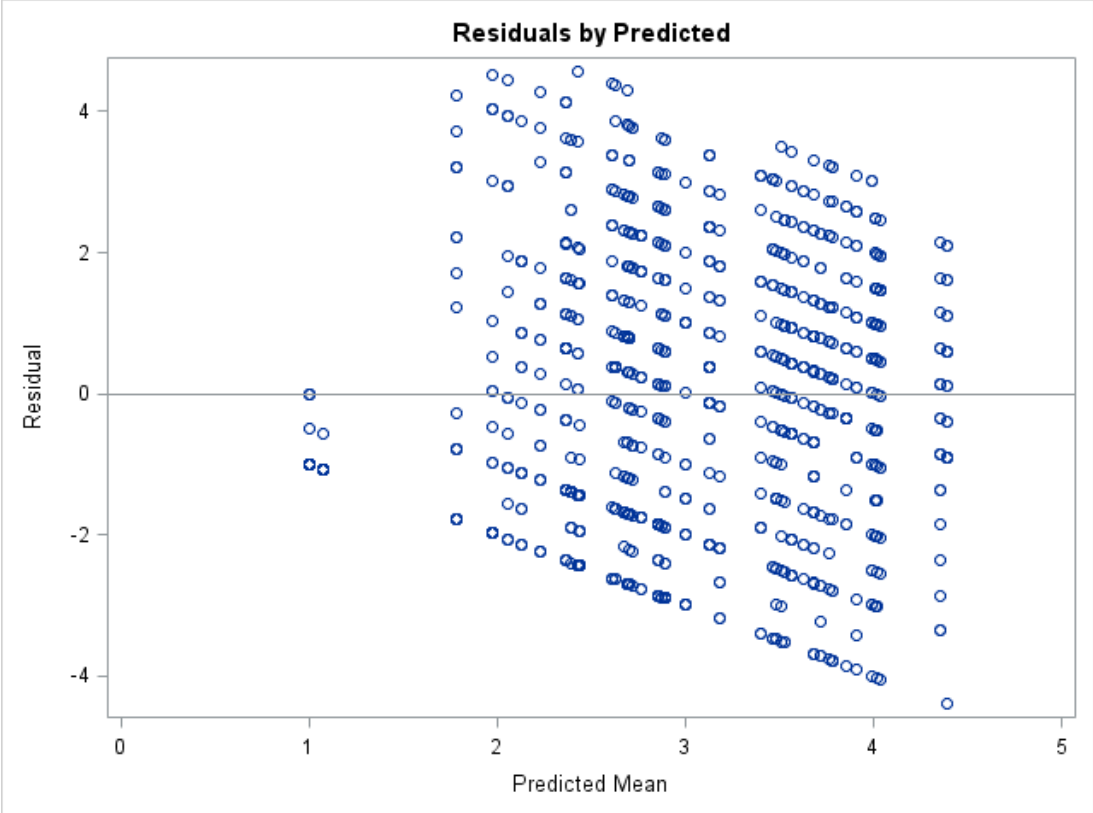

Figure 14a. "Residual by Predicted" Plot for Model 4b

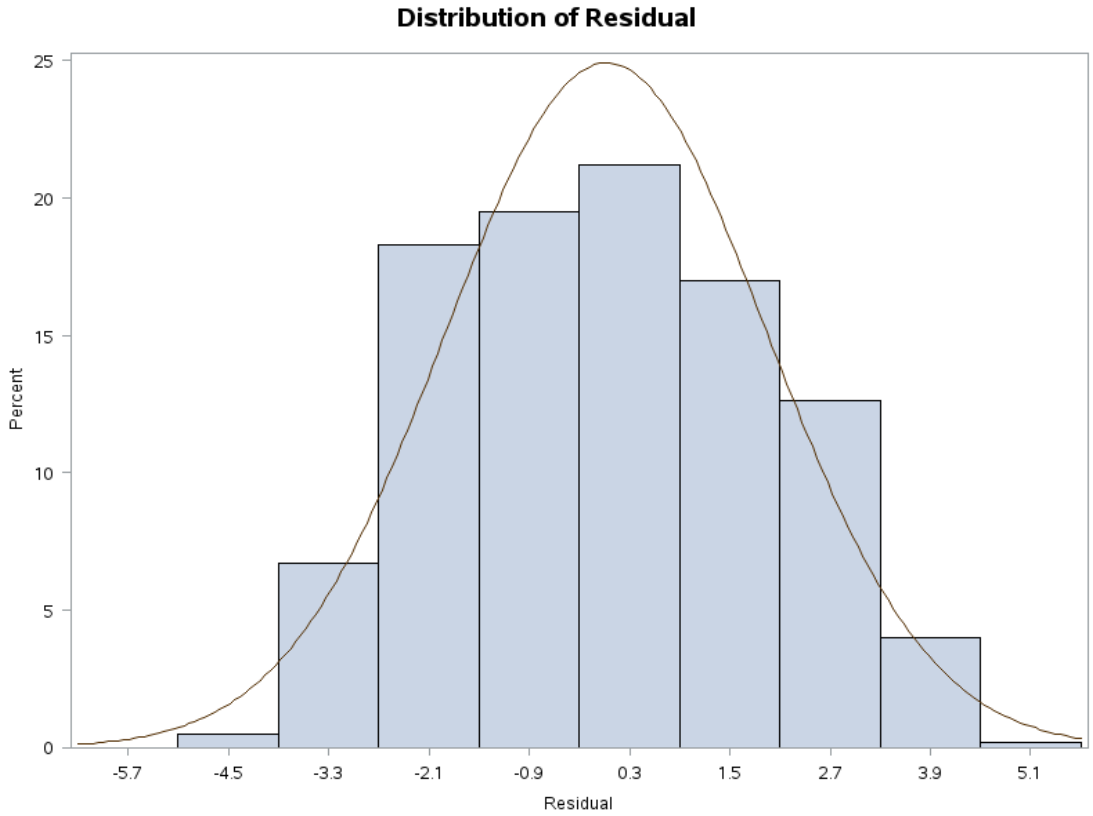

Figure 14b. "Distribution of Residual" Plot for Model 4b

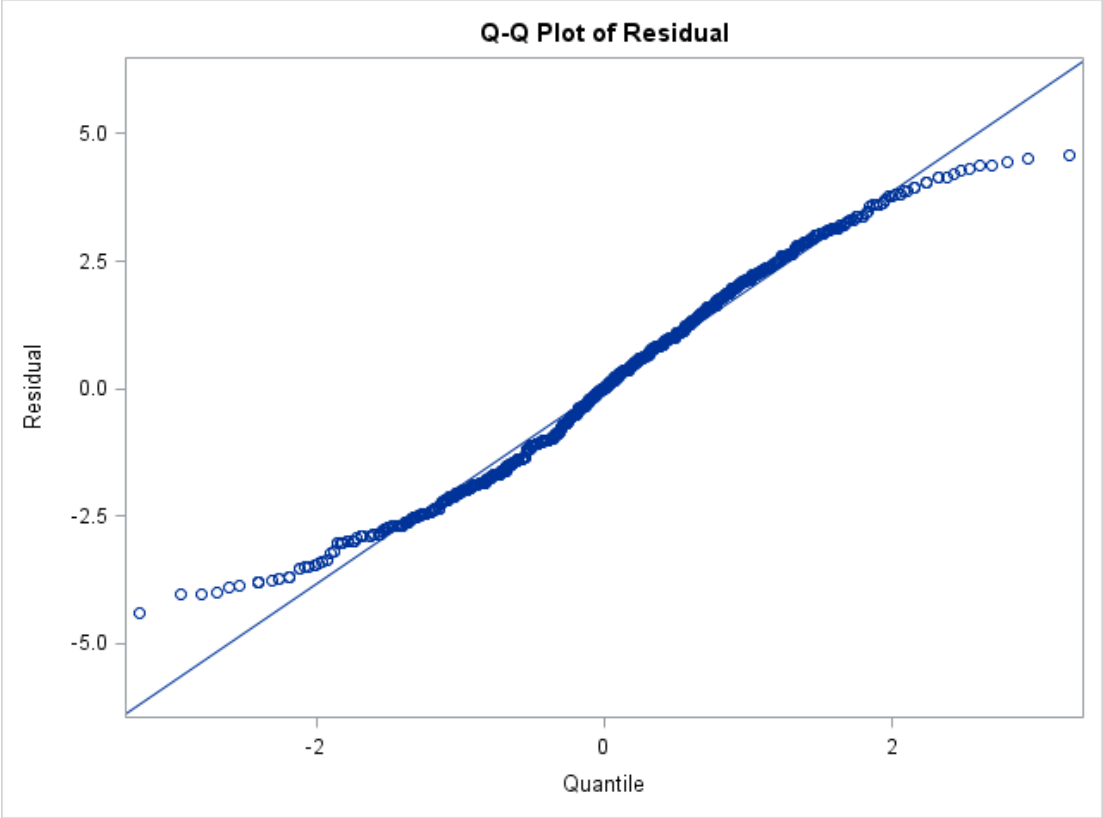

Figure 14c. "Q-Q Plot of Residual" Plot for Model 4b

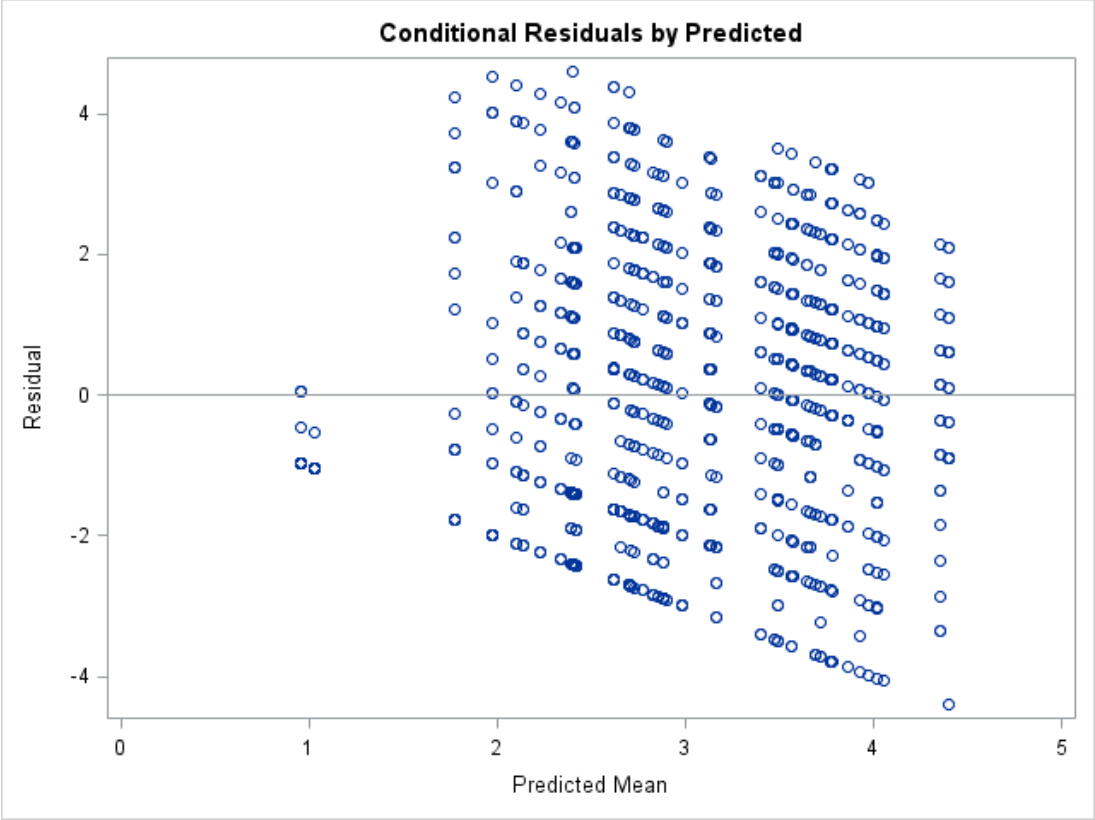

Figure 14d. “Conditional Residual by Predicted” Plot for Model 4b

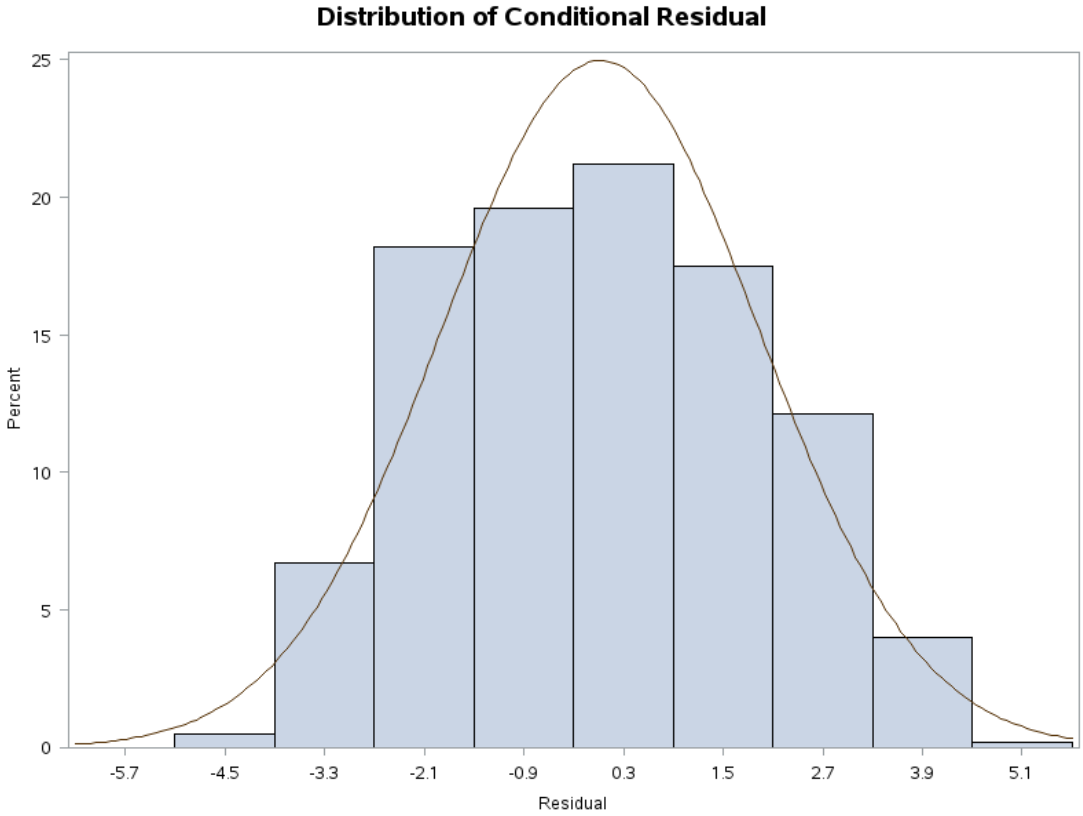

Figure 14e. “Distribution of Conditional Residual” Plot for Model 4b

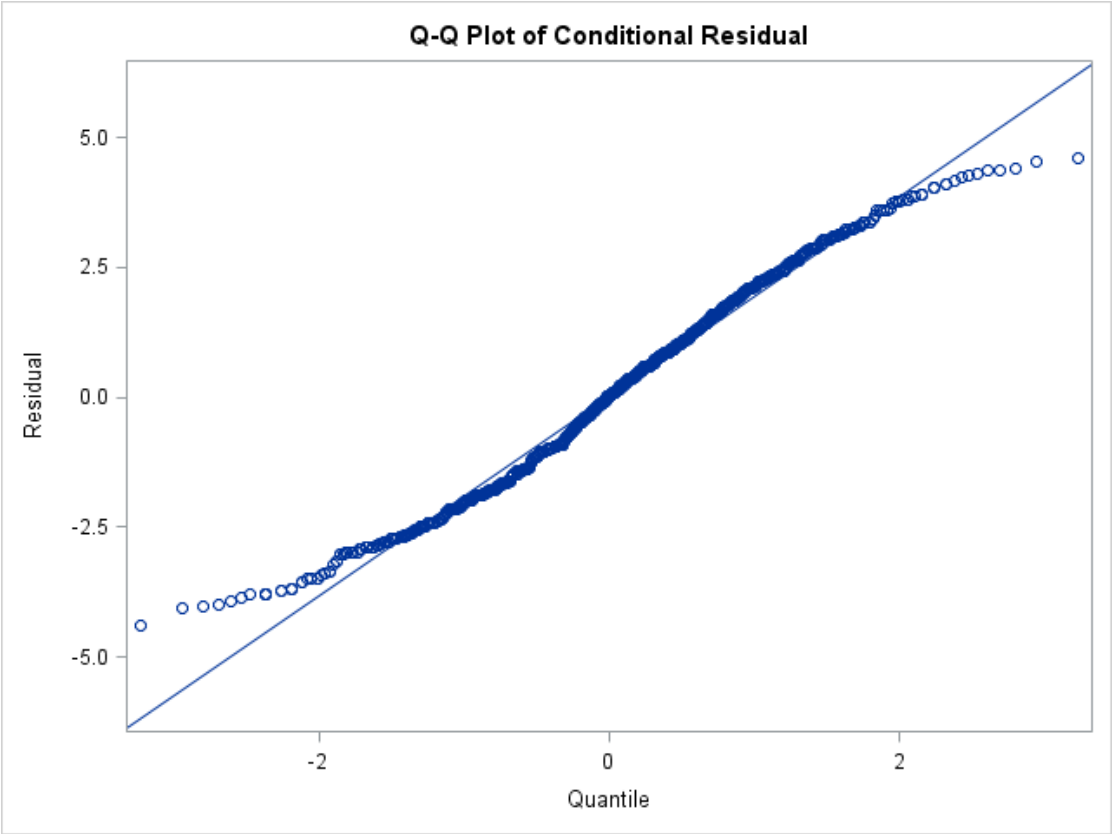

Figure 14f. “Q-Q Plot of Conditional Residual” Plot for Model 4b

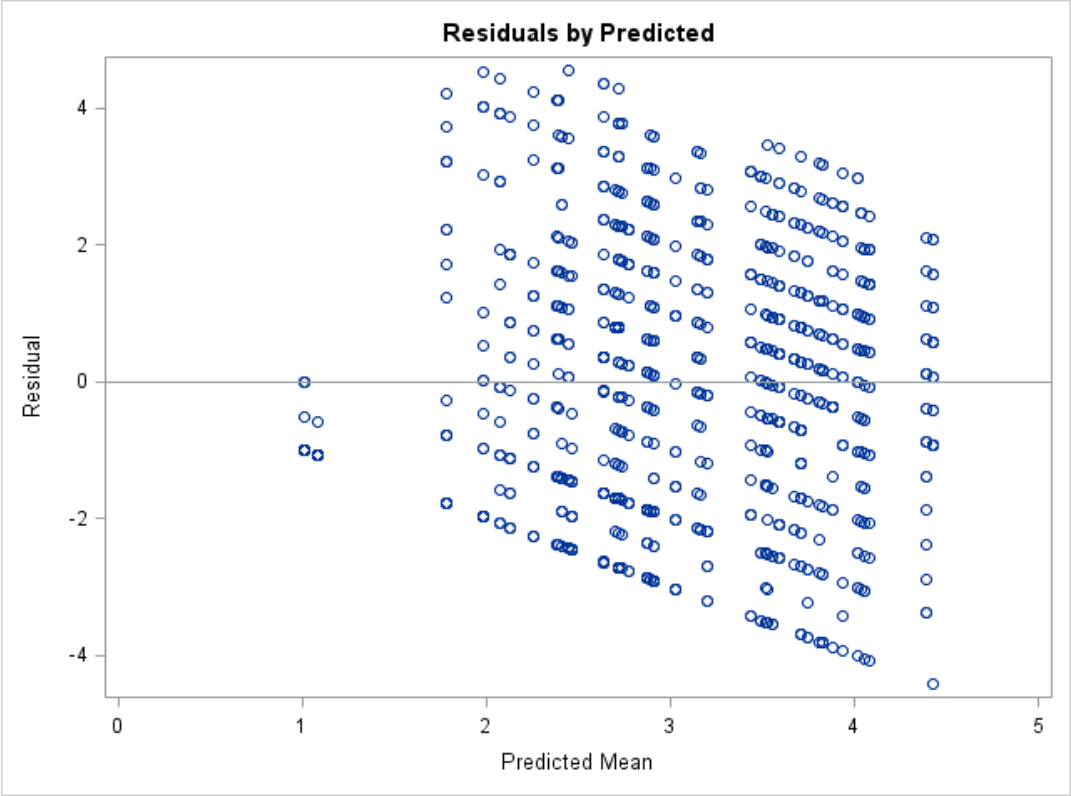

Figure 15a. "Residual by Predicted" Plot for Model 4c

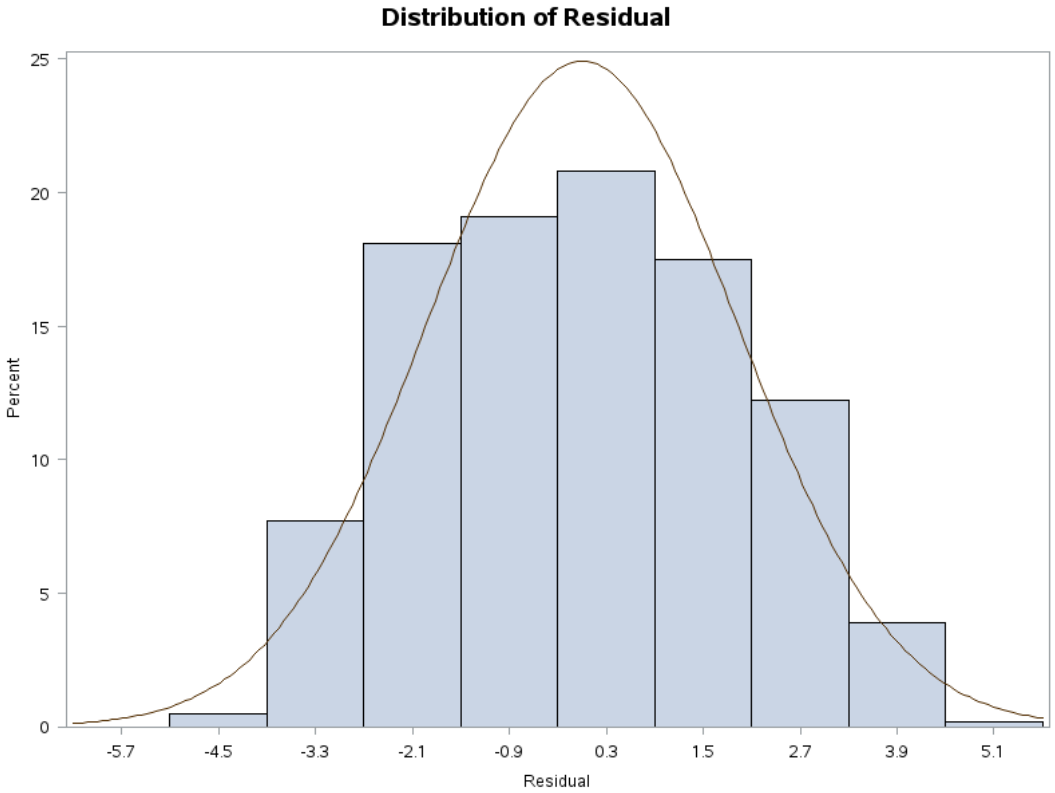

Figure 15b. "Distribution of Residual" Plot for Model 4c

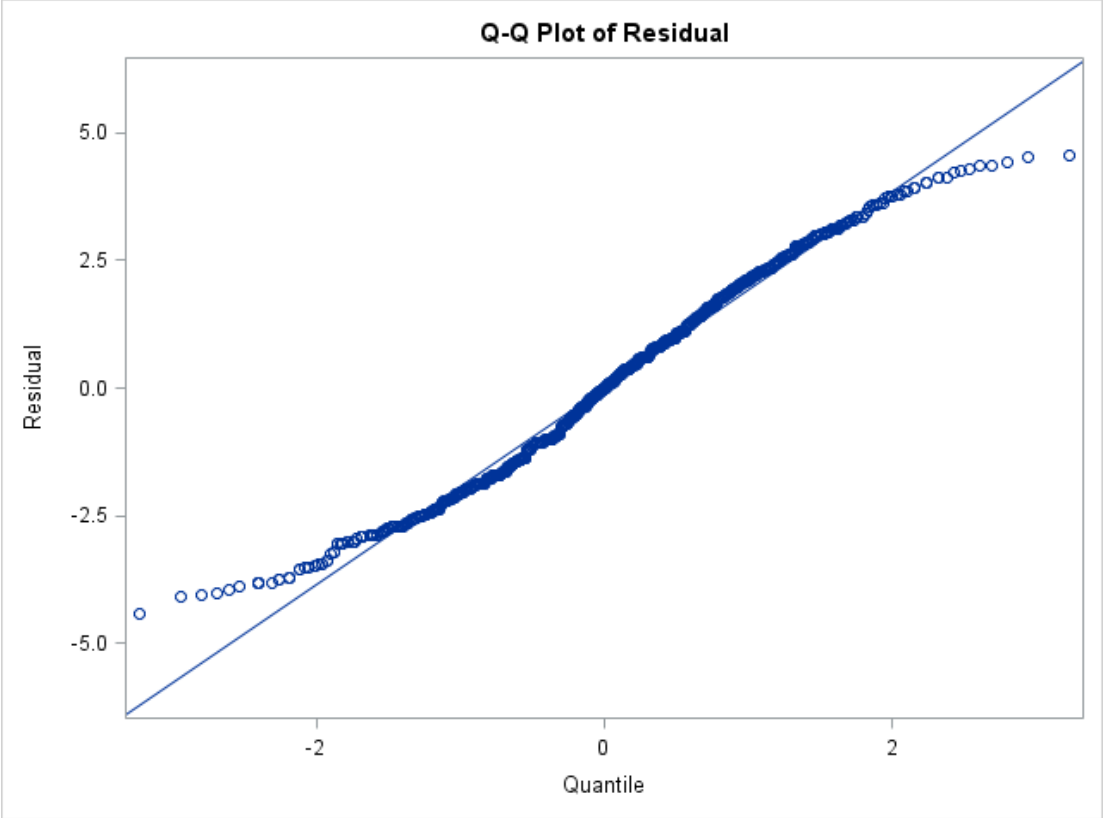

Figure 15c. "Q-Q Plot of Residual" Plot for Model 4c

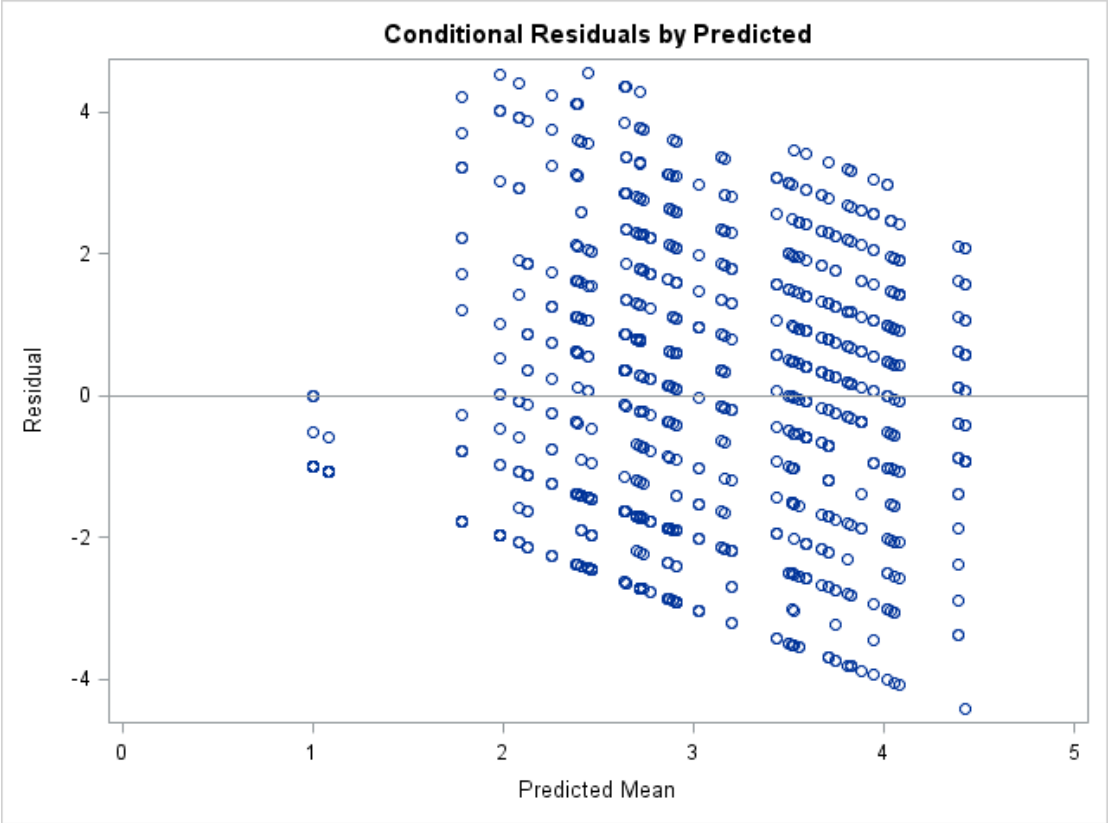

Figure 15d. "Conditional Residual by Predicted" Plot for Model 4c

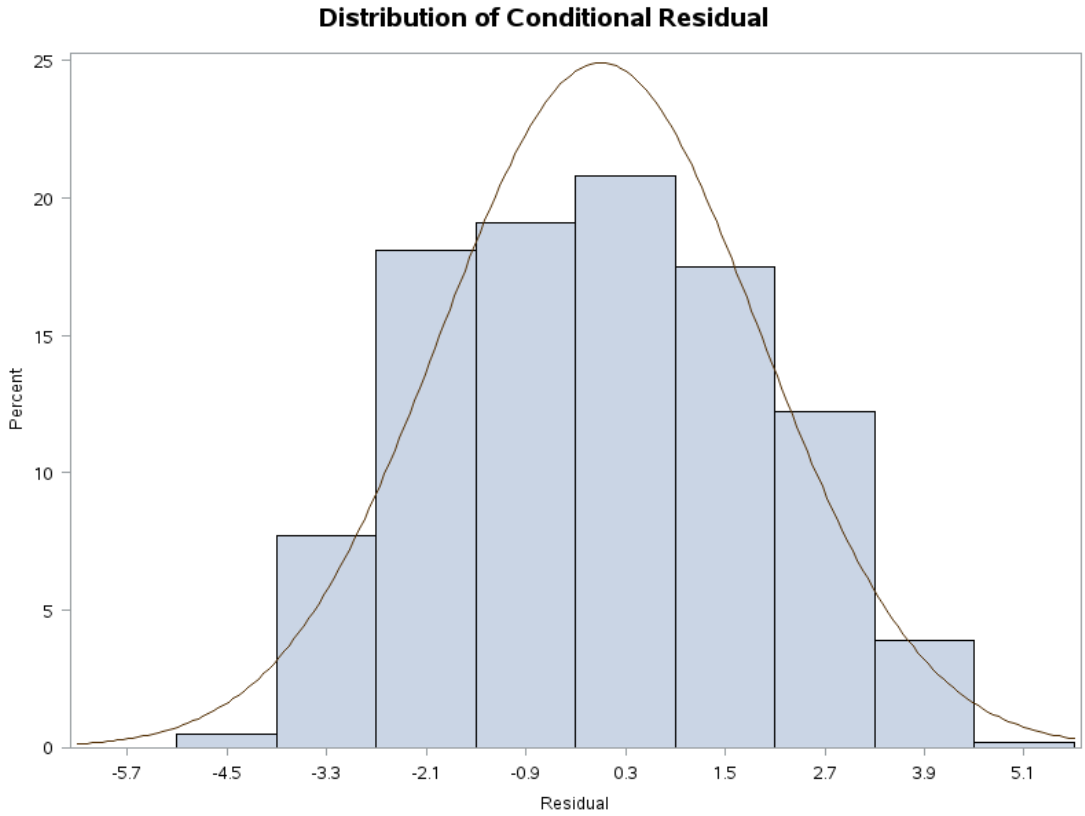

Figure 15e "Distribution of Conditional Residual" Plot for Model 4c

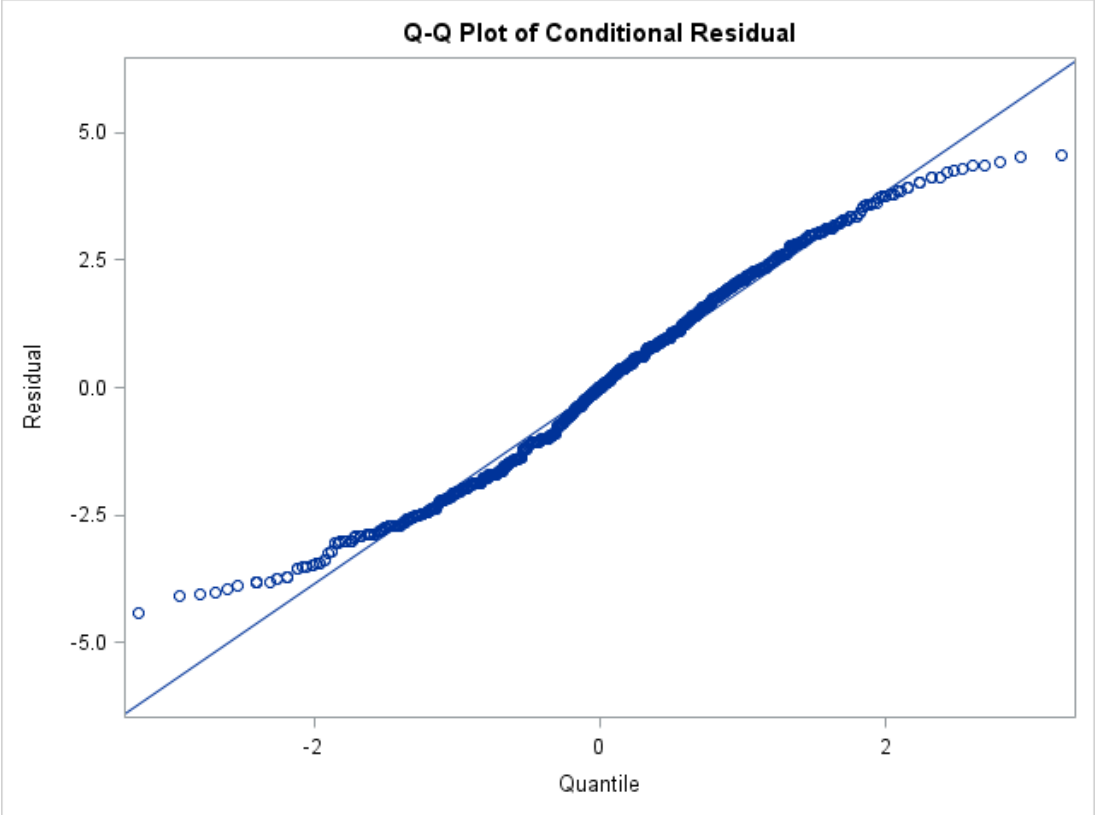

Figure 15f "Q-Q Plot of Conditional Residual" Plot for Model 4c

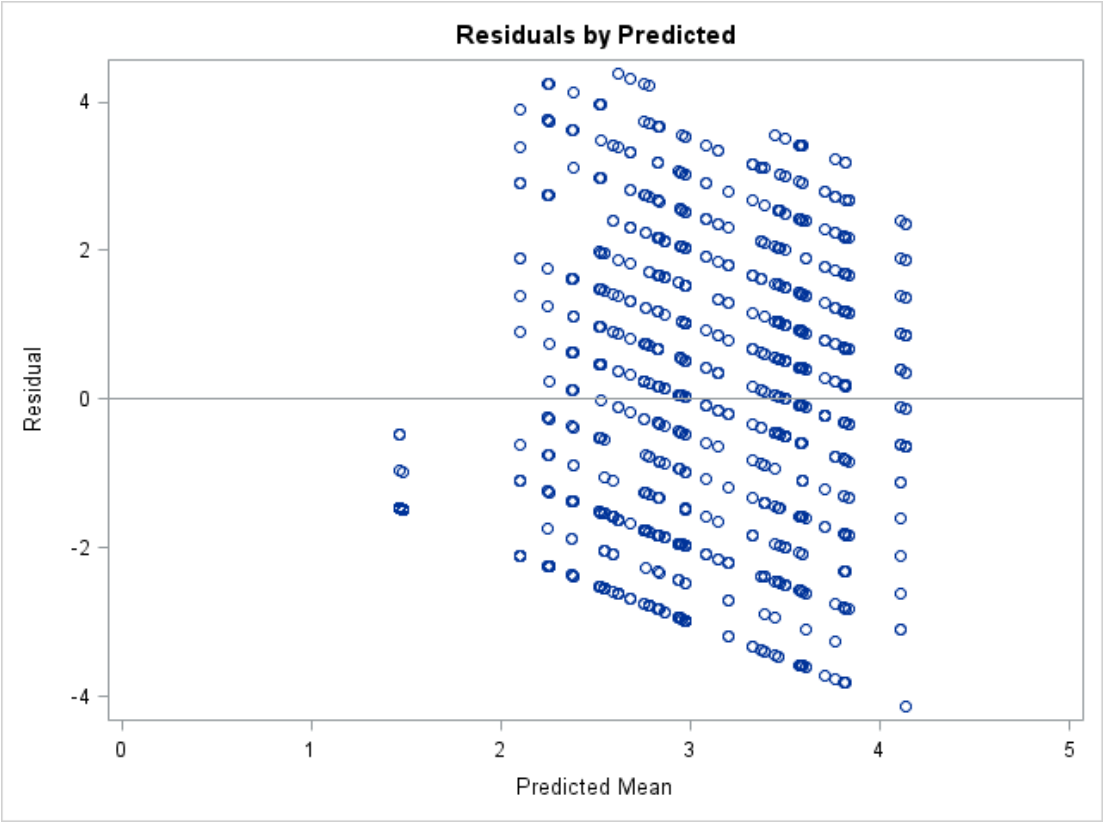

Figure 16a. "Residual by Predicted" Plot for Model 4d

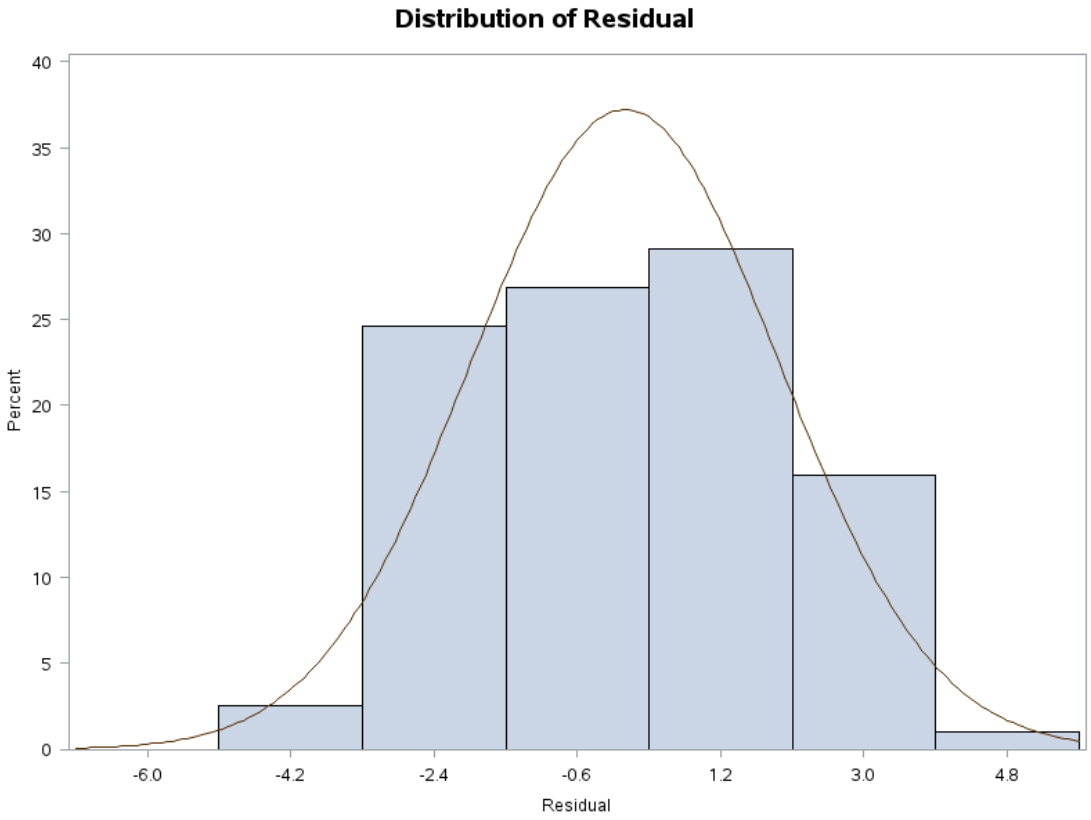

Figure 16b. "Distribution of Residual" Plot for Model 4d

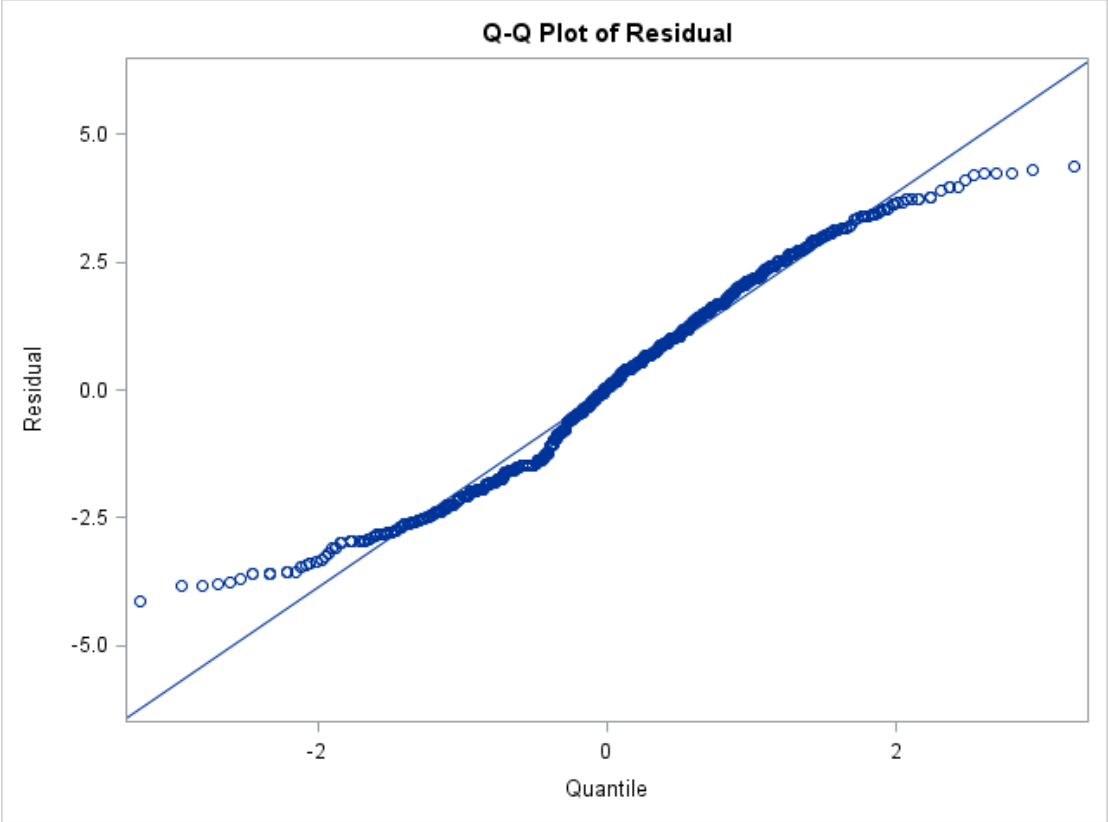

Figure 16c. “Q-Q Plot of Residual” Plot for Model 4d

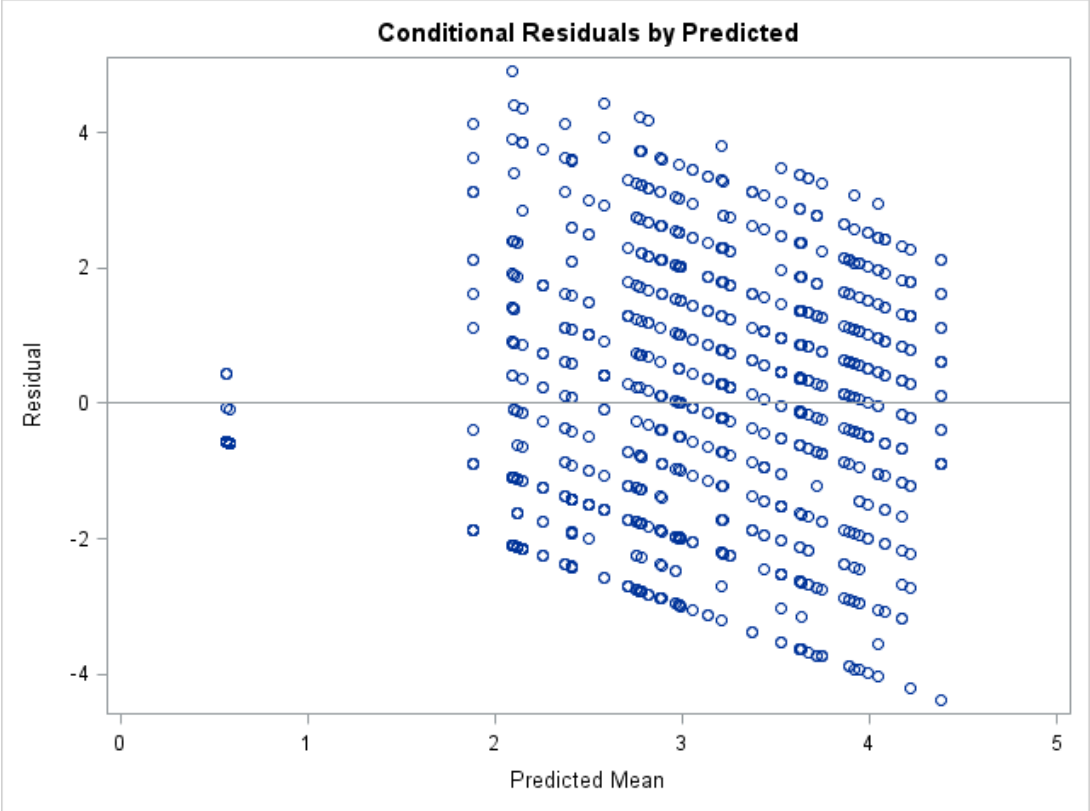

Figure 16d. “Conditional Residual by Predicted” Plot for Model 4d

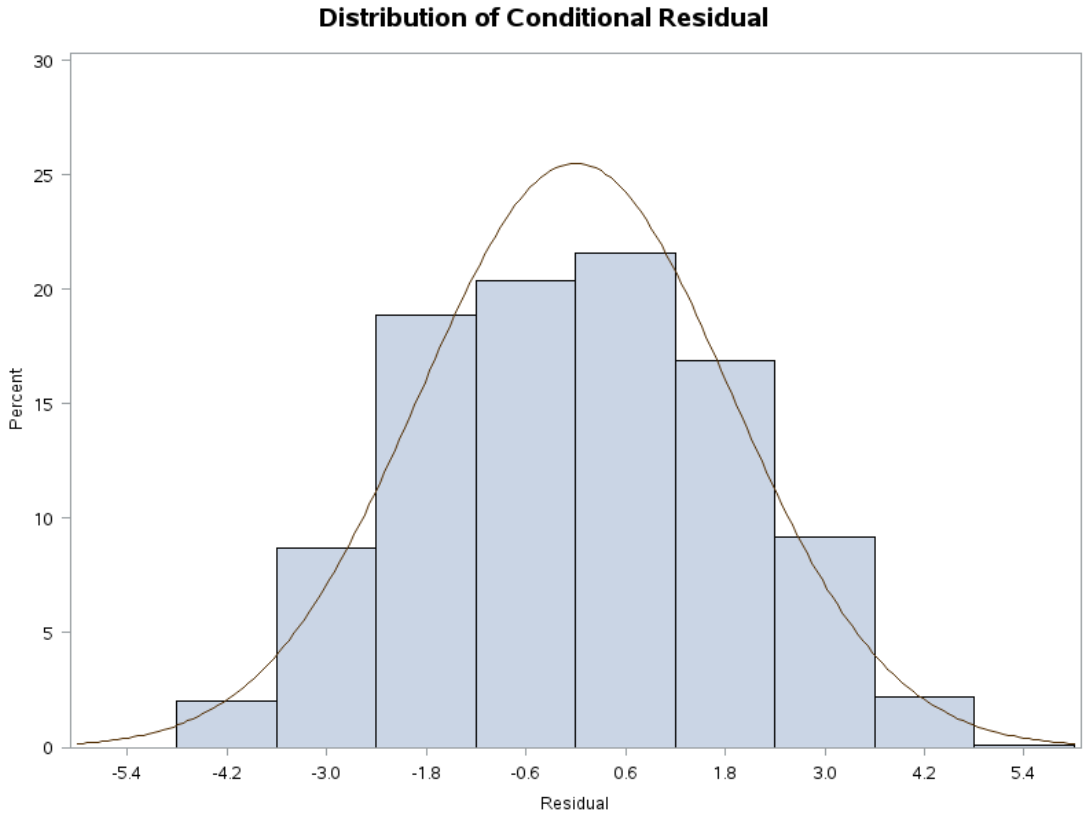

Figure 16e. "Distribution of Conditional Residual" Plot for Model 4d

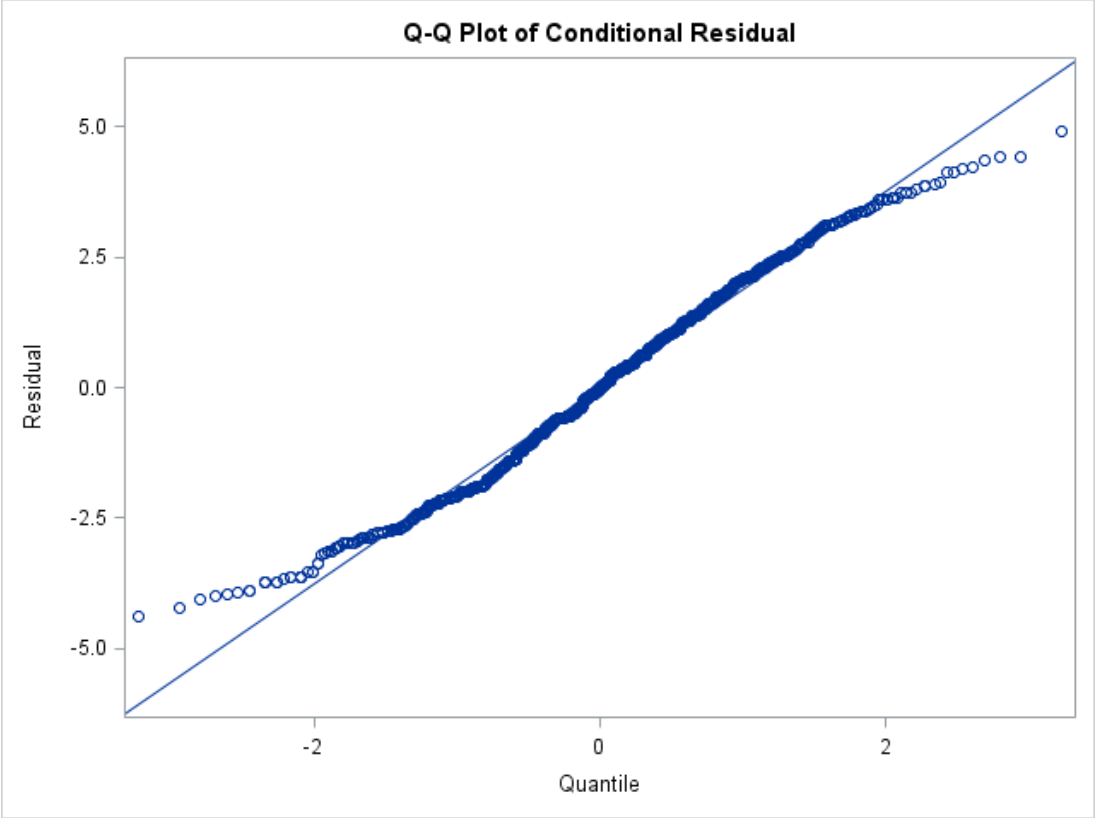

Figure 16f. "Q-Q Plot of Conditional Residual" Plot for Model 4d