

A Validation Study for an Enzyme Analytical Method

Xuan Liu advised by Helle Rootzen and Peter Thyregod

November 10, 2007

Acknowledgements

I would like to thank my supervisors, Helle Rootzen and Peter Thyregod, for their incredible support, understanding and encouragement. They have given many great suggestions and advices and have been great help in structuring the thesis and interpreting the results. Furthermore, I also would like to thank my friend, Cheng Xu, for critical reading though statistics is a foreign language to him.

Preface

This thesis was written as a part of my studies for a Master of Science degree at the Department of Informatics and Mathematical Modeling (IMM) of the Technical University of Denmark (DTU), under the supervision of Professor Helle Rootzen in IMM and Peter Thyregod from Novozymes.

This thesis deals with inter-laboratory validation study with mixed linear model.

Lyngby, November 2007

Xuan Liu

Summary

In 2005 a new methodology has been developed to evaluate the activity of the phytase enzyme in liberating phytate bound phosphorus in animal feed. In this thesis validation study of this new method is performed. Two times of inter-laboratory studies with similar design were conducted. The larger inter-laboratory study involving more Labs and Materials is the main analysis object. The analysis of smaller study is also performed as a contrast with the larger one.

Different variance structures of Linear Mixed Models are presented, such as Homogeneous Variance Model and Heterogeneous Variance Model, to detect variability characteristics of measurement error. The main characteristics of performance precision are Repeatability and Reproducibility. Besides that, whether the Type (Solid/Liquid) of the Materials effects the evaluation is another topic of interest. A guess that the liquid materials may have more stable measurements and smaller variances will be investigated by Linear Mixed Models. The homogeneity of the Labs's capability in the evaluation is also investigated.

Usually mathematical models have to satisfy the assumptions. But in most cases inter-laboratory data could not satisfy the strict assumptions. The objectives of most validation studies are to reveal useful information and parameters of the data rather than finding a fitting model. Whereas this thesis also emphasizes particularly on modeling, which could supply more general characteristics of data.

From Homogeneous Variance models to Heterogeneous Variance Models, most of these models presented in this thesis were proven not to satisfy the assumptions. But the modeling and assumption checking process could supply great details of data, which could be the indication of modifying the covariance structures of models.

Key words: Inter-laboratory Validation Study, Linear Mixed Model, Homogeneous Variance Model, Heterogeneous Variance Model, Repeatability and Reproducibility, Assumptions Checking.

Reader Guide

The thesis consists of three big modules: Background introduction, Linear Mixed Model theory, and statistical implementation of the theory.

The introduction module supplies the information about design of the Study A and Study B and the experiment mechanism. The definition of the main characteristics: RSD_r and RSD_R are also presented.

Linear Mixed Model theory module include the general introduction of the LLM models. The model notation, covariance matrix specifications and assumptions accord with procedure mixed, the SAS 9.1 software. Hierarchical linear model is one kind of LLM. The HLM section focuses on describing the covariance matrices structures involved in this thesis. The readers who know mixed model and proc mixed well could skim the LLM section. But the HLM section is necessary to read for all because the covariance structure specifications are the cores of the HLM design.

In statistical implementation module, homogeneous variance models and heterogeneous models are processed both for study A and Study B. Because Study B is more complicated, it is set as the main object of analysis. While analysis of Study A is processed as the confirmation of the conclusion obtained from study B. The RSD_r and RSD_R obtained from different models are calculated. Besides, the approximation confidence intervals of RSD_r and RSD_R of are also discussed.

List of Abbreviations and Notations

REML-Restricted Maximum Likelihood

ML-Maximum Likelihood

RSD-Relative Standard Deviation

RSD_r -Relative Standard Deviation of repeatability

RSD_R -Relative Standard Deviation of reproducibility

WLS-Weighted Least Squares

GLS-General Least Squares

MS-Mean Square

ANOVA-Analysis of Variance

LMM-Linear Mixed Model

HLM-Hierarchical Linear Model

Contents

1	Introduction	11
1.1	Background Information	11
1.1.1	Design of the Inter-Lab Validation Study	12
1.2	Descriptive Statistics	14
1.2.1	Logarithmic Transformation of Raw Data and New Estimations of RSD	14
1.2.2	Data Description of Study A	15
1.2.3	Data Description of Study B	17
1.2.4	Explanatory Variables Relationships	19
1.2.5	Preliminary Analysis	21
2	Theoretical Methods and Modeling	29
2.1	Linear Mixed Model	29
2.1.1	Introduction	29
2.1.2	Notation, Assumptions and Algorithm of Linear Mixed Model	30
2.1.3	Homogeneous and Heterogeneous Variance Model	31
2.2	Hierarchical Linear Model	33
2.2.1	Notation and Covariance Structure for different layer models of Hierarchical Linear Model	34

3	Statistical Implementation of Theoretic Methods	41
3.1	Homogeneous Variance Model of Study B	41
3.1.1	Test of overall model reduction	42
3.1.2	Diagnostic Checking for Homogeneous Variance Model 3.4	51
3.1.3	Practical Implication of Model 3.4	60
3.1.4	RSD% of Homogeneous Variance Model 3.4	61
3.2	Heterogeneous Variance Model for Type of Study B	62
3.2.1	Motivation	62
3.2.2	Practical Issue of SAS statements	62
3.2.3	Statistical Implementation	63
3.2.4	RSD of Homogeneous Variance Model 3.12	66
3.2.5	Diagnostic Checking for Heterogeneous Variance Model 3.12	66
3.2.6	Brief Summary	71
3.3	Homogeneous Variance Model and Heterogeneous Variance Model for Type of Study A	72
3.3.1	Homogeneous Variance Model of Study A	72
3.3.2	Diagnostic Checking for Homogeneous Variance Model A3	72
3.3.3	RSD of Homogeneous Variance Model A3	76
3.3.4	Heterogeneous Variance Model for Type of Study A	76
3.3.5	RSD of Heterogeneous Variance Model AT6	78
3.4	Approximate Confidence Intervals on RSD	78
3.4.1	Homogeneous Variance Model 3.4: Approximate Confidence Intervals on RSD	80

<i>CONTENTS</i>	9
3.4.2 Homogeneous Variance Model A3 of Study A: Approximate Confidence Intervals on RSD	83
3.4.3 Comparison between Approximate Confidence Interval on RSD for Homogeneous Variance Model of Study A and Study B	86
3.5 Heterogeneous Variance Model for Material	87
3.5.1 Heterogeneous Variance Model for Material of Study B	87
3.5.2 Heterogeneous Variance Model for Material of Study A	89
3.5.3 Brief Summary	90
3.6 Heterogeneous Variance Model for Lab of Study B	91
3.6.1 Residual Analysis	94
3.6.2 Brief Summary	98
4 Conclusion	99
4.1 Result and Conclusion	99
4.1.1 Homogeneous Variance Model	99
4.1.2 Heterogeneous Variance Model	100
4.2 Further Study	101
5 Appendix	105
5.1 SAS Codes for Modeling	105
5.1.1 SAS Codes for Model3.4, homogeneous variance model	105
5.1.2 SAS Codes for Model3.11, between-type heteroscedasticity Model	106
5.1.3 SAS for RSD Approximation interval	106
5.1.4 SAS Codes for Model MB6, between-material heteroscedasticity Model	107

5.1.5	SAS Codes for Model LB3, between-Lab and between-Material heteroscedasticity Model	107
5.1.6	SAS Codes for Model LB5, between-Lab and between-Material heteroscedasticity Model	108
5.1.7	Log Ratio Likelihood Test code: LB3- $\hat{\mu}$ LB4	108
5.1.8	Normal distribution analysis of conditional residual of LB5	109
5.2	One way homogeneity test by Lab of conditional residual of LB5 for different materials	110
5.2.1	SAS output of LB5	110
5.3	SAS output of LB3 conditional residual homogeneity one way test by lab for 8 materials	112
5.4	Normality Checking for Random Effect Lab in LB3:	114

Chapter 1

Introduction

1.1 Background Information

The Background Information is generally from a validation study conducted by FEFANA (European Association of Feed Additive Manufacturers).[3]

Phytase is an enzyme that can be added to feed for monogastric animals. This enzyme can liberate phytate bound phosphorus in the digestive tract of animals thereby improving the nutritive value of feed by increasing the amount of available phosphorus for animal. In addition, it has a positive impact on the environment by reducing phosphorus excretion in animal manure.

A new evaluation method:

In 2005 a new methodology has been developed to evaluate the activity of the phytase enzyme in liberating phytate bound phosphorus in animal feed. Compared with other official analytical methods applied to this purpose ,for instance, AOAC method[4], the main advantage of this new method is its capability of measuring the phytase activity of all phytase products that currently exist on the market.

Principle and Mechanism of the new method:

The principle of the method is that the inorganic phosphate released by the enzyme in the presence of an acidic molybdate/vandate reagent forms a yellow complex. The yellow complex could be measured with a spectrophotometer at a wavelength of 415 nm and the released inorganic phosphate is expressed in optical density(OD₄₁₅). Then the optical density is

quantified through a phosphate standard curve. Finally the activity characteristic is expressed by in phytate units (U)/kg feed sample. One phytase unit(U) is the amount of enzyme that releases 1 μ mol of inorganic phosphate from phytate per minute under the above mentioned reaction conditions. The activity is calculated as following formula:

$$Activity = \frac{\Delta OD * D}{m * W * t} \quad (1.1)$$

$$\Delta OD = OD_{415sample} - OD_{415blank}$$

m=slope of the standard curve[$OD_{415}/(\mu mol * ml^{-1})$]

D =dilution factor (extraction volume* dilution of the extract)[ml]

W=weight of the sample [kg]

t= 30[min]

OD_{415} is the result of the measurement **od** the feed sample subjected to the whole analytical procedure, whereas $OD_{415blank}$ is the result of the determination of inorganic phosphate that is already present in the same sample before enzymatic reaction.

The Logarithm Transformation of *Activity* (U)/kg is the response values in all models in the thesis. The details will be discussed in section of **Descriptive Statistics**.

1.1.1 Design of the Inter-Lab Validation Study

Two statistical validations are organized. Inter-laboratory validations were performed as following structure: Some samples of different Materials were collected and distributed to different Laboratories.

Training period

Before formal validations some Labs were selected and participated the training exercises of applying new method protocol on some known samples of materials. After they have finished a period of test in their own labs 14 of them were selected to be involved in the final validation study.

Then two time of inter-laboratory validations were performed and their organizations are quite similar. Study B has one more layer of block design than Study A. For clarifying the design structures, we start from the simpler one: Study A.

Organization of the Validation Study A

Study A was carried out according to ISO guide 5725-2, thereby allowing for the determination of the standard deviation for repeatability and reproducibility. In detail, the repeatability standard deviation SD_r describes the within-laboratory variation obtained when applying the same method of analysis on the same sample under repeatability conditions (i.e. the same laboratory, the same operator, the same apparatus and short interval of time), whereas the reproducibility standard deviation SD_R describes the between-laboratory variation and is obtained when performing the same method of analysis under reproducibility conditions (i.e. on identical material obtained by operators in different laboratories). Extreme values reported by the participating laboratories were identified by sequential application of the Cochran and Grubbs outlier tests (at 2.5% probability level, 1 tail for Cochran and 2 tails for Grubbs) and were not included in the assessment of the method performance characteristics. The sequential application of these outlier tests stops when more than 22.2% (= 2/9) of the participating laboratories are identified as outliers and therefore excluded from the data set. However, the maximum number of outliers resulting for a single material was 2/14 laboratories.

Organization of the study: Fourteen participants from 9 different countries representing a cross-section of official feed control and laboratories with industry affiliations took part in this collaborative trial. Prior to the validation experiments the participating laboratories attended a training to get familiar with the method.

For second study each participant received:

- (1) 10 coded samples comprised of 5 feed samples separately fortified with 5 different phytase products and sent out in blind duplicates
- (2) The description of the method
- (3) A report template in Excel format in which the participants had to fill the results of their analysis

Organization of the Validation Study B

Study B also included the assessment of intermediate precision as suggested in ISO guide 5725-3, since the laboratories were requested to carry out duplicate analyses on the same day and to repeat these duplicate analyses on three different days. Therefore this study allowed to estimate the precision of the method under different circumstances regarding the execution of the method, namely (a) repeatability conditions (i.e. the same laboratory and

the same day), (b) intermediate conditions (i.e. the same laboratory, but different days) and (c) reproducibility conditions (different laboratories). The data set was also screened for the presence of outliers applying the same procedure as described for study A. Since in none of the cases more than 3 out of 14 laboratories, corresponding to 21 %, were identified as outliers, the criterion of the maximum number of 2 out of 9 laboratories (=22%) was respected throughout the whole study.

(1) 8 coded feed samples separately fortified with 8 different phytase products. The laboratories had to take sub-samples from the sample glasses to carry out the analyses on three days in duplicates

(2) The description of the method

(3) A report template in Excel format in which the participants had to fill the results of their analysis

1.2 Descriptive Statistics

1.2.1 Logarithmic Transformation of Raw Data and New Estimations of RSD

As introduced in Background Information Section, Chapter 1, the evaluations are the Activity (Unit/kg). The variance of this response value is not stable. Figure 2.1 shows the standard deviation tend to be proportional to the mean (that is, the relative standard deviation (RSD) is constant).

The log transform could stabilize the variance:

$$\sqrt{Var[Y_i]} = C * E([Y_i])$$

, where C is a constant value. The standard deviation of observations tend to be proportional to the mean.

$$Var[\ln(Y_i)] = \frac{Var[Y_i]}{E(Y_i)^2} \approx C^2$$

Here $C = \frac{\sqrt{Var[Y_i]}}{E([Y_i])}$, is the RSD (relative standard deviation) % of the raw data. Then the important inference is proven that the standard variance of the Log-transformation data

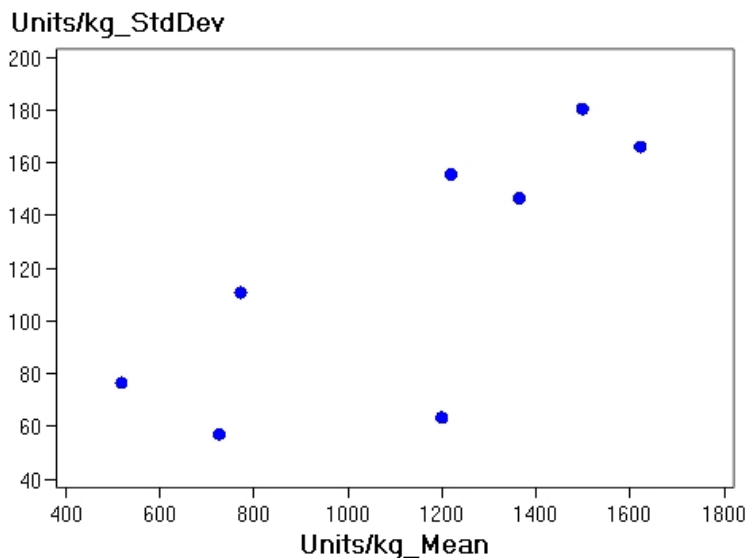


Figure 1.1: Standard deviation of Observations of Activity(Unit/kg) vs Means for different Materials, standard deviation tend to be proportional to the mean.

$(\sqrt{Var[\ln(Y_i)]})$ is almost the $RSD_R\%$ of the raw data.[?]

In the later analysis and modeling instead of the raw observations Activity(Unit/kg), the log-transformation of Activity($\ln(\text{Unit/kg})$) is set as response variable. Hence the standard variances estimated by following models are corresponding to the $RSD\%$ of raw data.

1.2.2 Data Description of Study A

The first several lines of DATA A are shown below:

Table 1.1: Data Structure of Study B

Lab	Material	Weighing	Type	$\ln(\text{Units/kg})$
1	1	1	Solid	6.753437919
1	1	2	Solid	6.633318433
2	1	1	Solid	6.683360946
2	1	2	Solid	6.452048954
3	1	1	Solid	6.841615476
3	1	2	Solid	6.650279049

Besides the response variables "Unit/kg" and $\ln(\text{Unit/kg})$ all the rest columns are corre-

sponding to 4 explanatory variables, **Material**, **Lab**, **Type** and **Weighing** as explanatory variables.

Fourteen participants(**Lab**) from 9 different countries representing a cross-section of official feed control and laboratories with industry affiliations took part in this collaborative trial. Those 14 **Labs** received 10 coded samples comprised of 5 feed samples separately fortified with 5 different phytase products(**Material**) and sent out in blind duplicates (**Weighing**). The Material is either solid or liquid(**Type**). Among those 5 kinds of materials two are solid type and the rest three are liquid type.

DATA A is hierarchical structure which could be shown below:

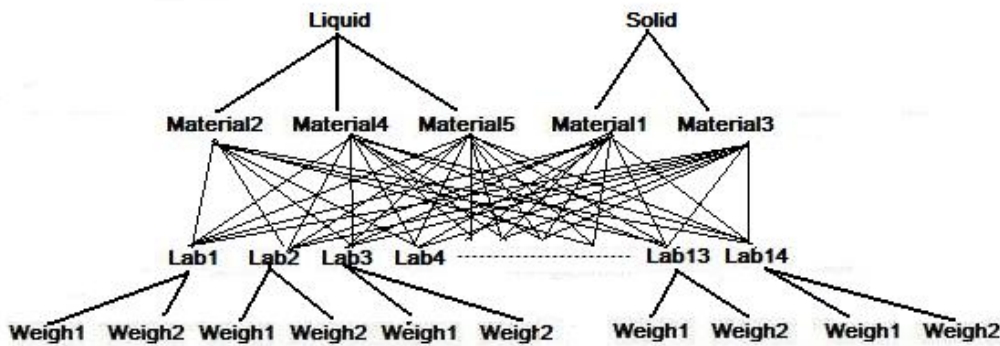


Figure 1.2: Data A Hierarchical structure: 5 Materials are nested within Type(Solid/Liquid). 5 Materials are cross with 14 labs. Weighs are nested within Days

The design of the experiment is complete and balanced. Because the outliers have been deleted already the structure of DATA A is incomplete. The principle and method of detecting outliers are introduced in section 1.1.1, Design of Inter-laboratory Validation Study.

After deleting the outliers the distribution of Materials versus Labs in DATA A is as below:

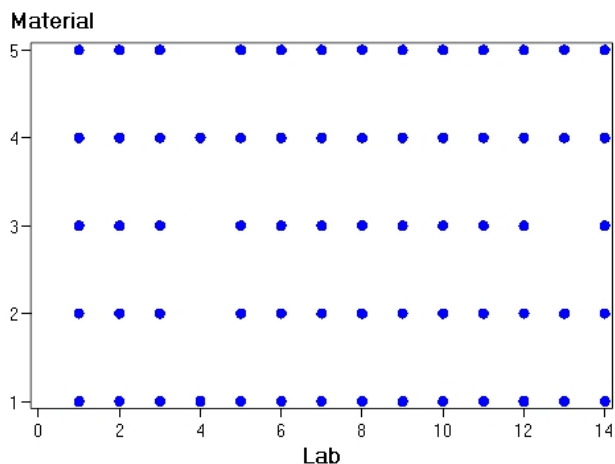


Figure 1.3: Distribution of Materials vs Different Labs in DATA A:For any position $X(i,j)$ ($i=1,2,3,\dots,14, j=1,2,3,4,5$) in the graph blue dot means No i lab have observations of No j Material in DATA A ,and the blank means No i Lab do not have observations in DATA A

1.2.3 Data Description of Study B

The first several lines of DATA B are shown below:

Table 1.2: Data Structure of DATA B

Lab	Day	Material	Weighing	Units/kg	$\ln(\text{Units/kg})$	Type
1	1	1	1	1600	7.377758908	Solid
1	1	1	2	1644	7.404887576	Solid
1	2	1	1	1524	7.329093736	Solid
1	2	1	2	1376	7.226936018	Solid
1	3	1	1	1688	7.431299675	Solid

The structure of DATA B is quite similar to the one of DATA A except for another Factor Day. Besides the response variables "Unit/kg" and $\ln(\text{Unit/kg})$ all the rest columns are corresponding to 5 explanatory variables and classification factor. 14 **Labs** received 8 samples composed of 8 different **Materials** and took 6 sub-samples from each bottles. Then in 2 duplicates(**Weigh**) within 3 days(**Day**). The Material is either solid or liquid(**Type**). Among those 8 kinds of materials four are solid type and the rest four are liquid type.

DATA B is of hierarchical structure which could be shown below:

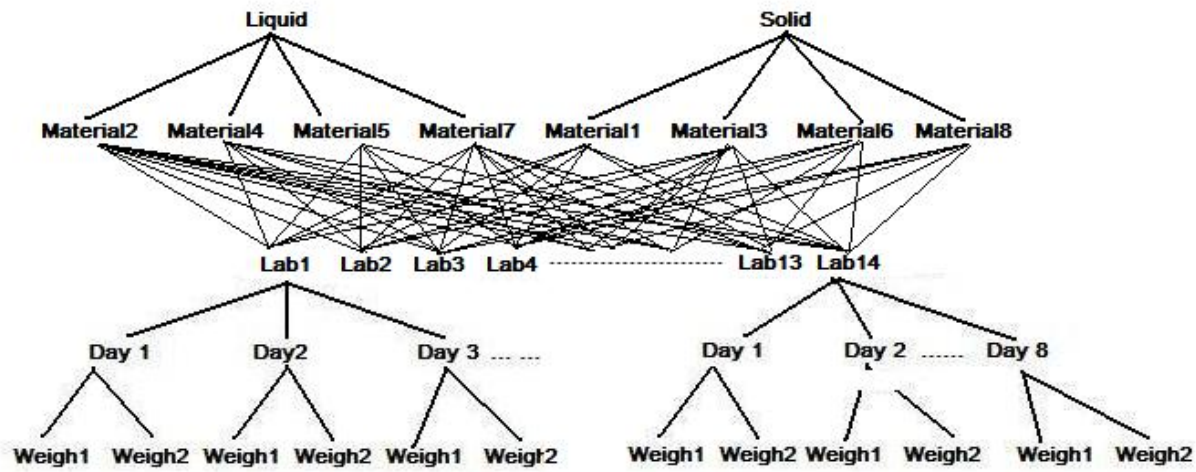


Figure 1.4: DATA B Hierarchical structure: 8 Materials are nested within Type(solid/Liquid). 8 Materials are cross with 14 labs. Days are nested within Labs. Weighs are nested within Days

The design of the experiment is complete and balanced. Because the outliers have been deleted already the structure of DATA B is incomplete. The principle and method of detecting outliers are introduced in section 1.1.1, Design of Validation Study.

After deleting the outliers the distribution of Materials versus Labs in DATA B is as below:

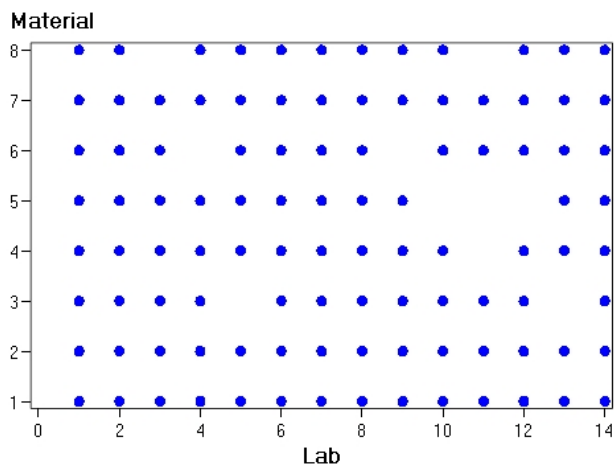


Figure 1.5: Distribution of Materials vs Different Labs in DATA B: For any position $X(i,j)(i=1,2,3,\dots,14,j=1,2,3,\dots,8)$ in the graph blue dot means No i lab have observations of No j Material in DATA A ,and the blank means No i Lab do not have observations in DATA A

1.2.4 Explanatory Variables Relationships

Figure 2.1 and 2.3 is the general structure graph of DATA A and DATA B. It is necessary to clarify the relationships between explanatory variables two by two before Modeling, which results will supply important inference of model specification later. Because DATA A and DATA B structures are quite similar, here I only state the DATA B case, which is more complicated because of another layer "DAY".

Labs VS Material

In this all the Labs received all 8 kinds of materials. First assume the relationship is hierarchical,nest Material within Lab:

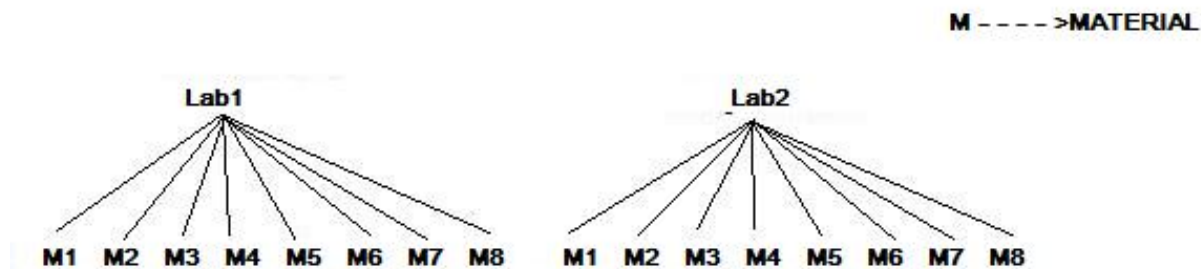


Figure 1.6: Assumption: If Material is nested within Lab, how does the relationship look like. Actually Material and Lab are cross with each other.

If the index of Material could be written as Material1, Material2,..Material8,Material9,Material10, then relationship could be Hierarchical. However, here the Material1 of Lab1 and Material1 of Lab2 are totally same samples. Therefore the relationship between Material and Lab is cross rather than Hierarchical.

Material VS Type



vs st.jpg

Figure 1.7: Material is nested within Type(Solid/Liquid).

The Materials are either solid or liquid. It is natural to nest **Material** within **Type**. There exists no contrast that one material have both liquid and solid types. The effects of Type and Material are totally mixed.

Labs VS. Day

Day is assumed to be nested within Labs. Each lab evaluated each kind of material during three days.

1.2.5 Preliminary Analysis

Here we use $\ln(\text{Units/kg})$ (the log transformation) instead of the original observations(Units/kg) as response value. For simplicity, later in this thesis the observations and response values just mean the log-transformation of original observations.

Observations VS. Material

Because one of this thesis aims is to detect whether the Type(Solid/Liquid) of Materials have different variances , observations and means of different materials of solid and liquid are plot separately. we start to plot and analysis from DATA B which is is more complicated and has more observation(618)..

In the 2 by 2 figure matrix of DATA B below the first line is of scatter plots and second line is of box plots with usual 4 interquartile range. The variances of the response value, $\ln(\text{Units/kg})$, are different from material to material. $\ln(\text{Units/kg})$ of Material No.5 and No.6 have obviously smaller variances than the rest. If we compare the plot of 4 liquid materials on the left with the plot of 4 solid materials on the right side, we could see that the measurements of Liquid materials are more dispersive than the ones of solid. But not all the ranges of observations of the Liquid material are wider than the ones of solid, for instance No5 is more compact than No 8. It is hard to make explicit conclusion in DATA B case. The 2 by 2 figure matrix of DATA A indicates a contrary situation: measurements of Liquid materials are more compact than the ones of solid.

The dispersion of response values are also effected by the Labs and Days. The random effect Lab and Day are confounded with Residuals totally. After the final model has been applied

to Data B we will return to comparisons of errors of different Types and Materials.

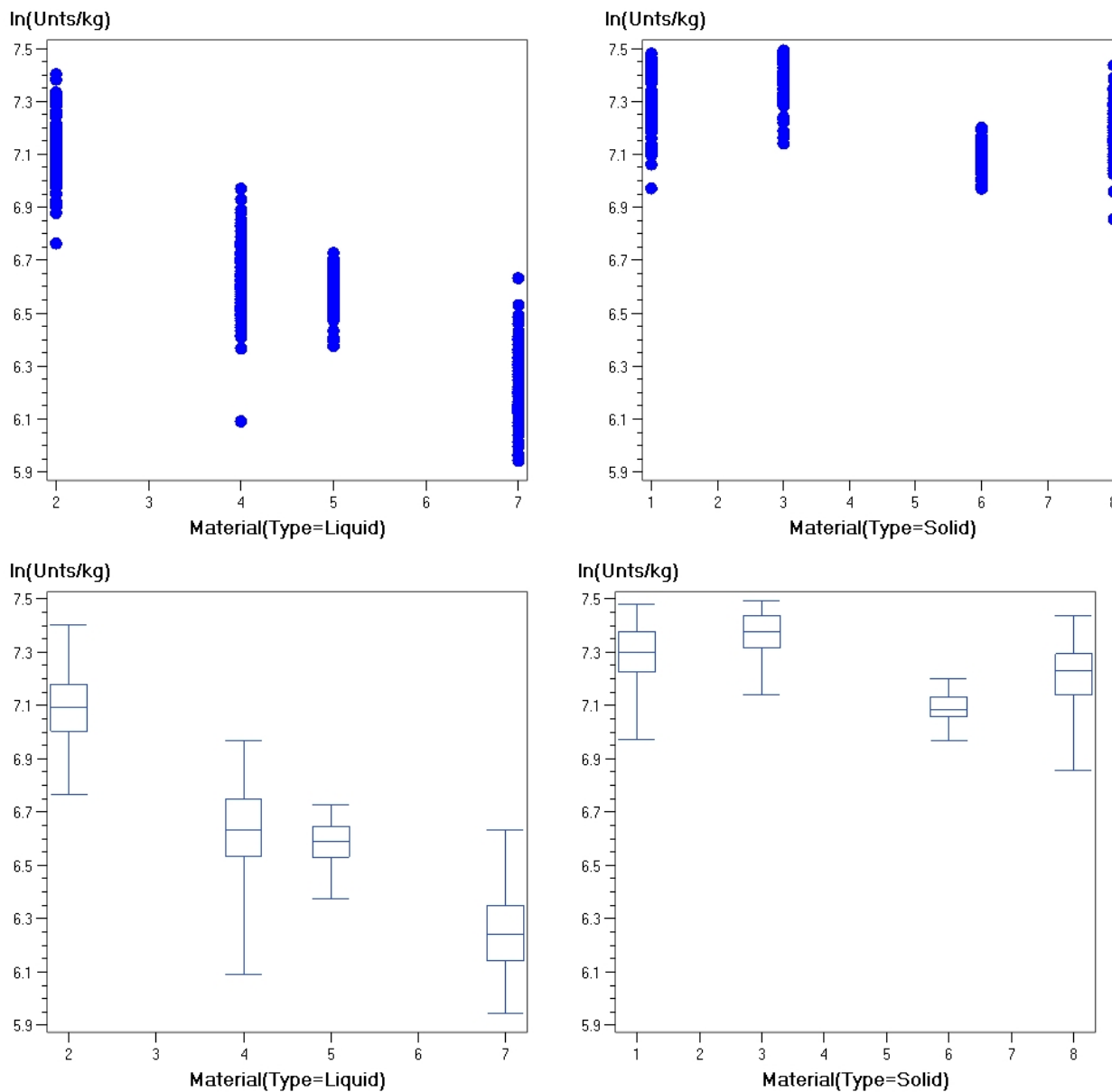


Figure 1.8: **DATA B**:The observations (Scatter/Box)plot vs. 8 kinds of Materials ,which are divided into two groups by Type(Liquid/Solid) The two figures on the left are of 4 Liquid Materials and two on the right are of Solid Materials. The Box Plots here indicate Interquartile ranges of response values($\ln(\text{Units/kg})$)

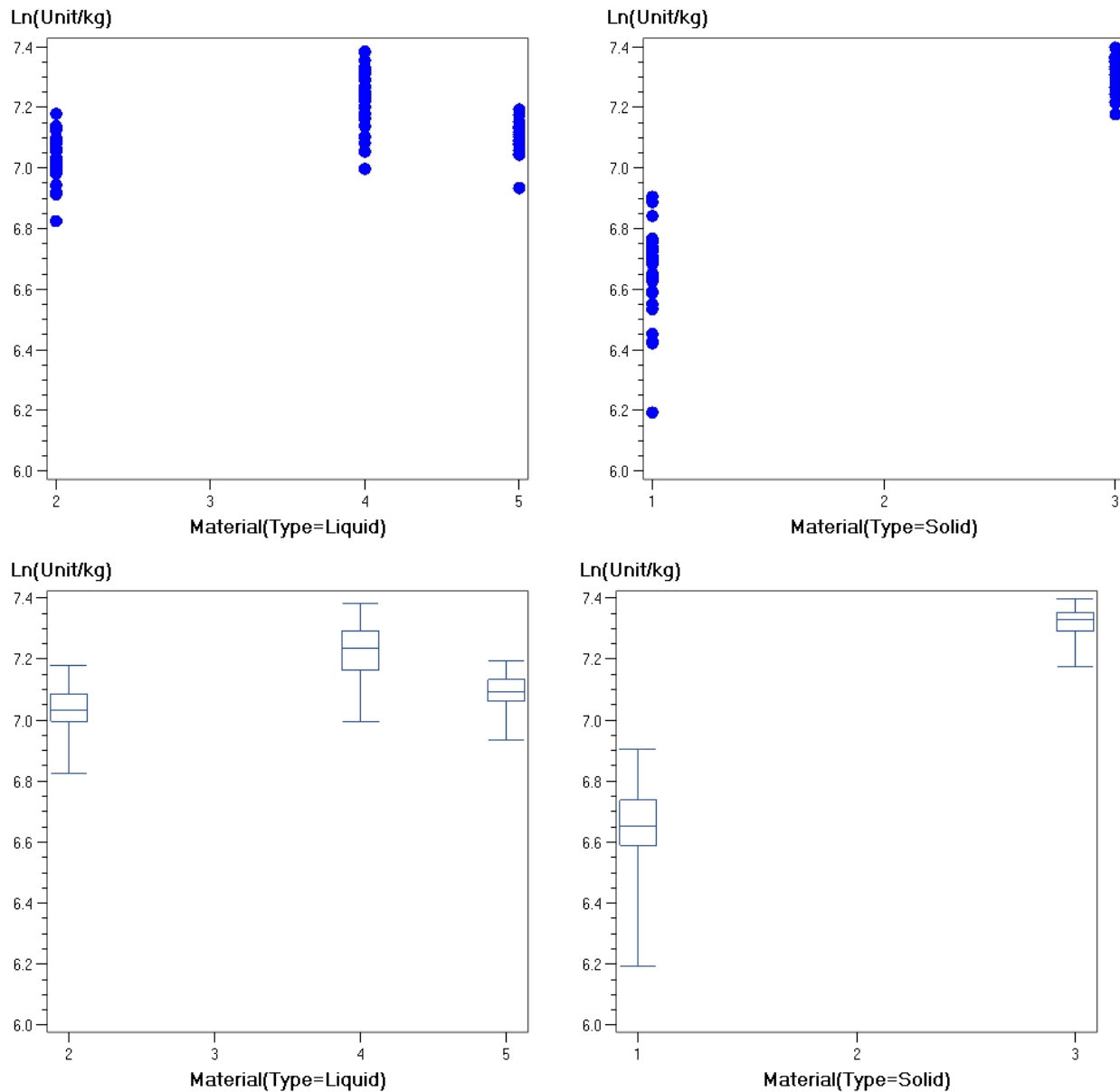


Figure 1.9: **DATA A:**The observations (Scatter/Box)plot vs. 5 kinds of Materials ,which are divided into two groups by Type(Liquid/Solid) The two figures on the left are of 4 Liquid Materials and two on the right are of Solid Materials. The Box Plots here indicate Interquartile ranges of response values($\ln(\text{Units}/\text{kg})$)

Comparing Activity($\ln(\text{unit}/\text{kg})$) of materials estimated by different Labs

In the 2 by 2 figures of DATA B in above section, for each Material there are more than 50 observations. Those 50 scatter points distribute on one line only 2 centimeters long-it is hard to see the clear distribution trends of the points. In the Box-plot the central line is marked and the acentric trends is clear.

How do different labs affects the observations distribution? Do some Labs always intend to supply higher measurements ? Do the labs give different estimates of the same material? A simple analysis and plot could give some hints.

Assuming we only have one lab, it processed the evaluations of one kind of material on three days and two duplicates one day. With those 6 observations this lab could give an estimate of this material.

Now labs could supply 14 estimates for each kind material. In the scatter plot there are only 14 points at most on one line. The Density of the scattng points is much smaller and we could see how they distribute clearly now.

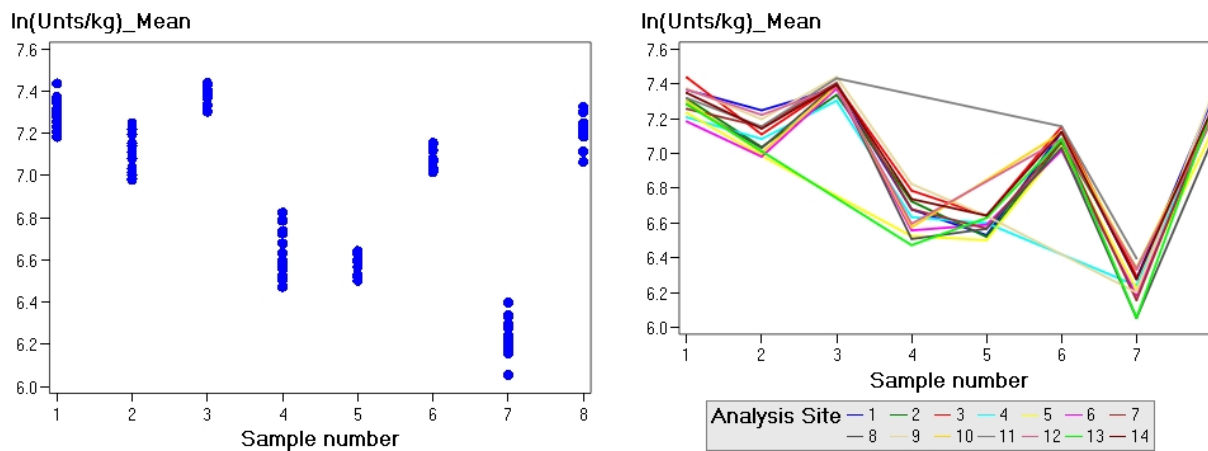


Figure 1.10: DATA B: Scatter and Line plots of Mean Values of Activity estimated by different Labs.

The left figure above is the scatter plot of estimates of 8 materials from 14 different Labs. The distributions of those estimates are still rather dispersive. The estimates of material activity vary from lab to lab. In addition the dispersive trends are kind of similar with the original observations plots in figure 1.8: Obviously Material No4 and No7 are most dispersive in both those 1.8 and 1.10 cases. It seems that the block effect ,Lab, is a big source of the variability of the measurements.

The line plot on the right is the same as scatter plot on the left except the estimates points from same lab are connected with colorful lines. Because the data structure is not complete and some labs have no estimations of some Materials, the 14 colorful lines could not form perfect cluster: Some lines "jump" away from some Material positions because of being lack of measurements of those kinds of materials. But there still exist some information valuable: Some lines tend to have trends to have higher or lower measurements. The most clear case is that: For all the materials Lab No 11(Grey line) always give higher estimations than Lab No.13(Green Line).

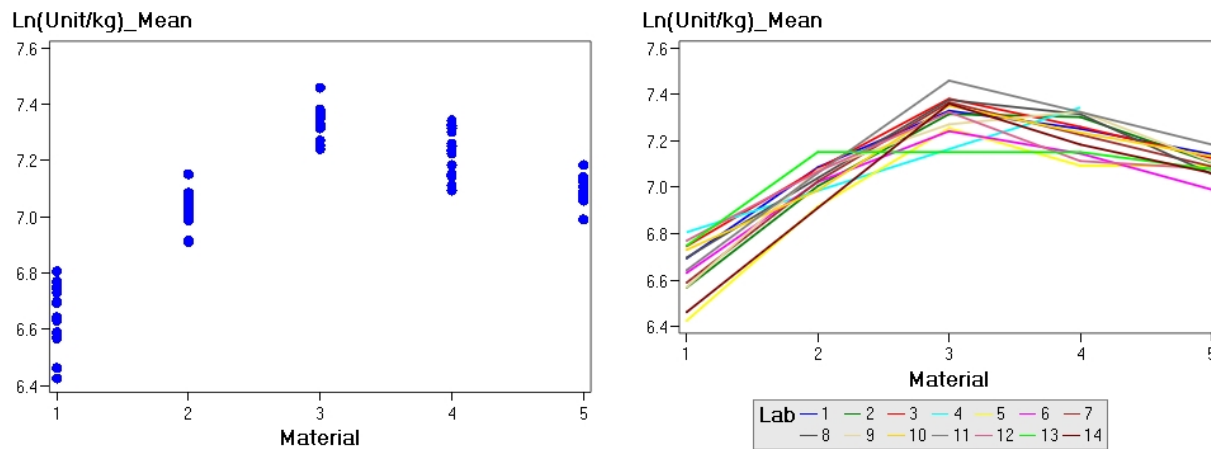


Figure 1.11: DATA A: Scatter and Line plots of Mean Values of Activity estimated by different Labs.

Because in Figure 1.10 the lines are too closed with each other, it is a little hard to distinguish the lines of different labs. New plots are processed: In figure 1.12 estimates activity of different materials is plot versus Labs. 8 different colorful lines indicate the 8 materials. In ideal case the 14 labs should have same capability and give almost same measurements of same material. Then those lines should almost parallel X-axis with some random oscillations.

Obviously this is not our case. In Figure 1.12 the lines have some obvious pattern for Labs. For instance, Lab 8 have smaller estimates of Materials 4, 7, 6 and 8. From the Material aspect, curves of Material 3,5 and 6 are more stable which means estimates of those materials from different labs are more closed to each other. And this point is also proven in scatter plot of Figure 1.8, in which dots of Material no 3,5 and 6 are obviously more compact.

In Figure 2.10 the curve of Material 1 oscillates more wildly than the rest, which inosculates the information in Figure 2.6 and 2.8 where the dots of Material 1 is most disperse.

Similar plots of DATA A is also presented as below. It also show some patterns for Labs. For instance Lab No 8 give lower estimates than the rest labs. And Material No 1 oscillated most wildly ,which is accord with Figure 1.9, Material No 1 has largest variance.

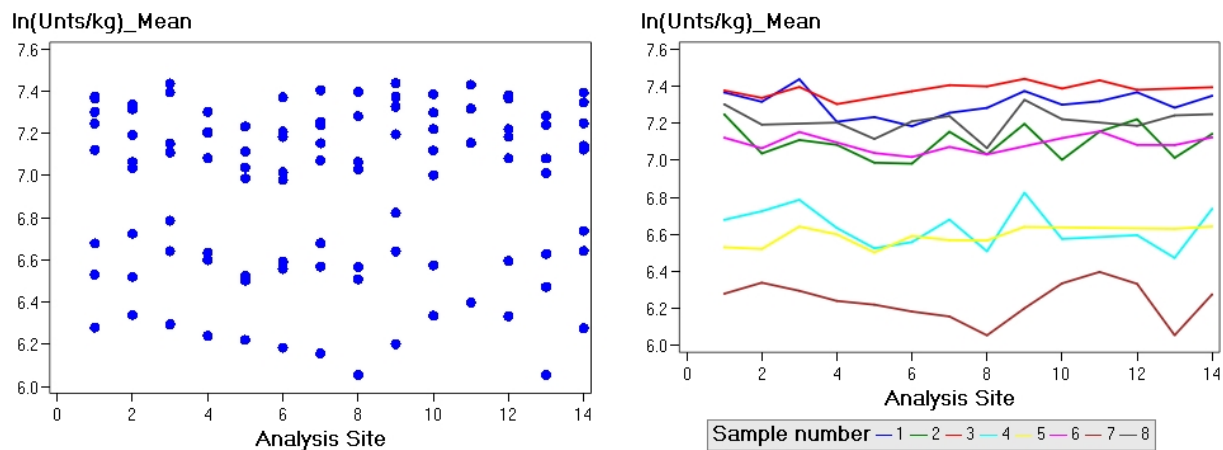


Figure 1.12: DATA B: Scatter and Line plots of Mean Values of Activity estimated by different Labs.

Further study of those supposes will be examined after modeling and discussed in the later Chapters.

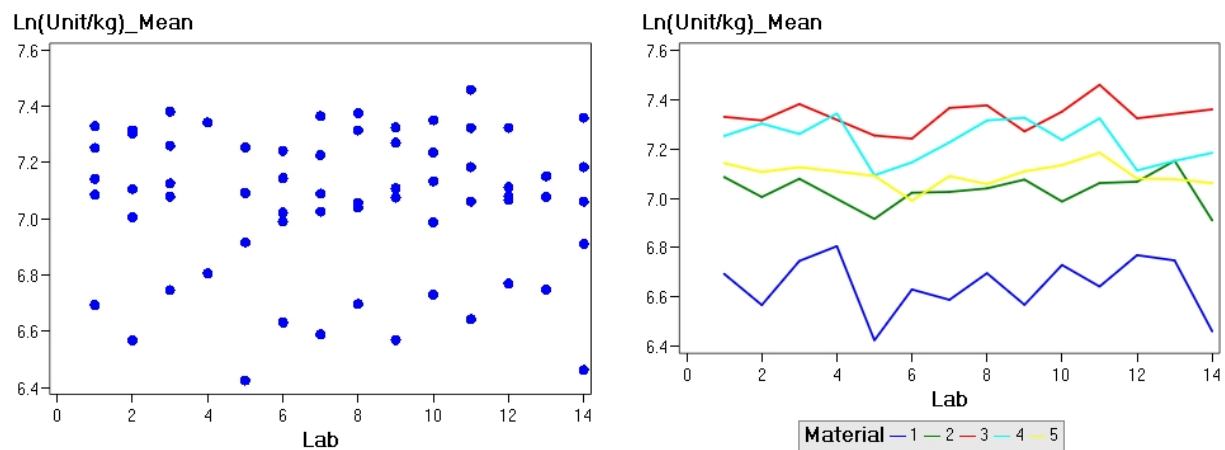


Figure 1.13: DATA A: Scatter and Line plots of Mean Values of Activity estimated by different Labs.

Chapter 2

Theoretical Methods and Modeling

2.1 Linear Mixed Model

2.1.1 Introduction

The classical variance analysis and regression analysis are based on rather strict assumptions about the data : The structure must be described by a linear model, observations, or rather the residual or error terms, must follow a normal distribution, they must be independent and the variability should be homogeneous.

The linear mixed models extends the general linear model by allowing a more flexible specification of the covariance matrix of errors. In other words, it allows for both correlation and heterogeneous variances. SAS **Proc Mixed** has recently explored the the great versatility of the mixed linear models. And still today, many statistical packages will only offer a limited version of the possibilities with mixed linear models. In this thesis I use **SAS Proc Mixed** as primary analysis tool. In the theory Chapter The definition , notation and assumption adapted here are on the basis of **Proc Mixed** Help document. Thereby the theory description could keep consistency with implementation and calculation.

The primary assumptions underlying the analyses performed by PROC MIXED are as follows: The data are normally distributed.

The means (expected values) of the data are linear in terms of a certain set of parameters.

The variances and covariances of the data are in terms of a different set of parameters, and they exhibit a structure matching one of those available in PROC MIXED.

2.1.2 Notation, Assumptions and Algorithm of Linear Mixed Model

Formulation of Linear Mixed Model

The formulation of Mixed Model could be written as:

$$y = X\beta + Z\gamma + \epsilon \quad (2.1)$$

Formula 3.1 is the general notation of mixed linear model which contain both random and fixed parameters.

Y is a $(n \times 1)$ vector of observed data, X is an $(n \times p)$ fixed-effects design or regressor matrix of rank k , Z is a $(n \times g)$ random-effects design or regressor matrix, β is a $(p \times 1)$ vector of fixed-effects, γ is a $(g \times 1)$ vector of random effects, and ϵ is an $(n \times 1)$ vector of model errors (also random effects). The **distributional assumptions** made by the MIXED procedure are as follows:

$$\mathbf{E} \begin{pmatrix} \gamma \\ \epsilon \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.2)$$

$$\mathbf{Var} \begin{pmatrix} \gamma \\ \epsilon \end{pmatrix} = \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{pmatrix} \quad (2.3)$$

1. γ is normal with mean 0 and variance \mathbf{G} ;
2. ϵ is normal with mean 0 and variance \mathbf{R} ;

The random components γ and ϵ are independent.

Estimating Variance Components (\mathbf{G} and \mathbf{R}) in the Mixed Model

The covariance structure of LMM is much more complicated than the one of GLM. Therefore the estimation of parameters is more complicated. The Least square method could not work because the violation of the assumption of GLM. SAS proc mixed implements Maximum Likelihood estimation which admits the random missing data and complicated variance structure.

restricted maximum likelihood (REML). REML estimates of the covariance components are based on residuals which are computed after estimating the fixed effects by WLS or by GLS and are estimates based on maximizing a marginal likelihood. REML estimates take into account the degrees of freedom used in estimating the fixed effects when estimating the covariance components. Besides, REML could constrain the variance component estimates to be non-negative. REML estimates of G and R are found by maximizing the following log-likelihood function

$$\mathbf{REML} : -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} r'V^{-1}r - \frac{n-p}{2} \log(2\pi) \quad (2.4)$$

where $V = ZGZ' + R$, the residuals vector: $r = y - X(X'V^{-1}X)^{-1}X'V^{-1}y$ and p is the rank of X .

PROC MIXED actually minimizes -2 times this function using a ridge-stabilized Newton-Raphson algorithm. The output of goodness value is -2 Res log Likelihood instead of Likelihood Value. Besides REML, there also exist other estimation methods such as ML, MIVQUE0, or Type1 -Type3 (Moment Estimate). REML and ML is the favorable method which is proven by the simulations evidence presented by Swallow and Monahan (1984). In all programming of this thesis the estimation option is set as REML.

Estimating β and γ in the Mixed Model

REML provides the estimates of G and R , \hat{G} and \hat{R} . According to the solution of mixed model equation the estimates of β and γ could be denoted as :

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y \quad (2.5)$$

$$\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(y - X\hat{\beta}) \quad (2.6)$$

2.1.3 Homogeneous and Heterogeneous Variance Model

The observations vector y has a multivariate normal distribution with an expected value of

$$E(y) = X\beta$$

and variance

$$V(y) = ZGZ + R$$

Parameters of this model are the fixed-effects β and all unknown variance components in the variance matrices G and R. The unknown **variance components** are referred to as the covariance elements in covariance matrix V. R is the within-subjects variance component. ZGZ' is the between-subjects variance component.

Proc MIXED could also allows the user to specify, separately and jointly, covariance structures that assume within-subjects and/or between-subjects heteroscedasticity.[?] Within-subjects heteroscedasticity occurs when the variances across repeated measures are unequal. Between-subjects heteroscedasticity occurs when covariance matrices differ across groups. Naturally the discrimination of the two kinds of Linear mixed model is:

Homogeneous variance model: All the observations have same covariance matrix.

Heterogeneous variance model: Within or between subjects heteroscedasticity occurs. In this thesis the heterogeneous variance structure for Type Lab and Material will be crucial topic discussed later.

Covariance Structure(R and G) and Relative Practical SAS Issue

The simple random model only contains random effect, $X=0$ and $R = \sigma^2 I_n$, where I_n is n by n identity matrix. In contrast general linear model only contain fixed effects and $Z=0$. And $R = \sigma^2 I_n$ still hold.

A real LMM(Linear Mixed Model) would have more complicated structure of R matrix. The **Repeated Statement** of Proc MIXED of SAS models the intra-individual variation and includes the structure of $R_i = V(\epsilon_1)$, where R_i is a block diagonal matrix for subject i. The **Group Option** of **Repeated Statement** could define a effect specifying heterogeneity in the covariance structure of R. All the observations having the same level of the group effect have same ϵ . Without specification by Repeated Statement R is just $\sigma^2 I_n$ and R_i are the same for all the subjects.[7]

ZGZ' is the between-subjects variance component. Random statement defines the random effects, $Z\gamma$, and the structure of G. The **Group option** defines an effect specifying heterogeneity in the covariance structure of G, ,the heteroscedasticity between-subjects. All observations having the same level of the group effect have the same covariance parameters. Each new level of the group effect produces a new set of covariance parameters with the same structure as

the original group.

Instead of denoting G and R individually the combination of them, covariance matrix V will be the primary tool used in the following Modeling Analysis section to show the covariance between two observations.

2.2 Hierarchical Linear Model

Hierarchical linear model (HLM) is a type of mixed model with hierarchical data - that is, where data exist at more than one layers. HLM models focus on differences between groups in explaining a dependent variable, which is just the study case of this thesis. By introducing HLM the variance structure of multi-layer models involved in this thesis could be illuminated.

The original definition of HLM was stated By Raudenbush and Bryk 1986: Hierarchical linear model a particular regression technique that is designed to take into account the hierarchical structure. Historically HLM has been used in educational research such as students nested with classes, classes nested within schools . Recently the new statistical computing ability made HLM used widely in many disciplines, including Biostatistics.

With traditional regression approaches, such as multiple regression and logistic regression, an underlying assumption is that the observations are independent. This means that the observations of any one individual are not in any way systematically related to the observations of any other individual. The assumption is violated, however, when some of observations sampled are from the same Laboratory, or the same equipment. When the assumption of independence is violated, the regression coefficients can be biased, and the estimates of standard errors are smaller than they should be.

Multi-level variance decomposition techniques such as HLM offer a number of advantages over traditional analysis techniques such as ANOVA and regression. First, because HLM separates out the criterion variance into within-and between-crew components, error terms are not systematically biased. This leads to more accurate effect size estimates and standard errors. Second, because HLM uses all available information, meaningful variance is not wasted. Finally, HLM allows for testing cross-level effects.[5]

2.2.1 Notation and Covariance Structure for different layer models of Hierarchical Linear Model

As introduced before HLM is a special case of mixed model. Hence the algorithm described in Mixed Model is totally accommodated to HLM. In this section the variance structure is the primary part of study. I have adapted the notation to comfort it to the next modeling step on our two data files.

The experiment design from DATA B is simply described as following:

1. The objective of Samples are of 8 Materials. Those samples are either solid ,or liquid.
2. 14 Laboratories have been chosen as participants. Each Labs received 8 samples comprised by 8 materials
3. Those labs implemented the validation measures on sub-samples in different days and 2 duplicates each day.
4. The outliers have been tested and deleted from the validation DATA. In the end the data file on which modeling is implemented is unbalanced and incomplete.

One-Layer Model

The simplest hierarchical model is one-layer model. The common notation is :

$$y_i = X\beta + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \quad (2.7)$$

Here are the observations are independent and follow normal distribution. $X\beta$ denotes the fixed effects part of the model because the fixed part is the predictors of mean value. This simple model is denoted here to clarify the structure of a more complicated hierarchical linear model step by step.

In this simple model mean value could be estimated by Maximum likelihood and denoted as:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.8)$$

Where β is the coefficient of fixed effects, X is the design matrix of fixed effects and y is observation values.

The un-bias estimate of the single variance in the model could be noted as:

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2 \quad (2.9)$$

N is number of observations, and p is number of means. $\hat{\mu}_i$ is the un-bias estimated mean value, which is calculated by $\hat{\mu}_i = X\hat{\beta}_i$

We start to analyze the Phytase validation DATA B with one layer model , thereby the random effects, **Labs** and **Day**, are ignored. The remaining fixed effect is **Material**, which has 8 levels. The one layer model for Study B becomes:

$$y_i = intercept + Material_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \quad (2.10)$$

Two-Layer Model

A two-layer Hierarchical model could be denoted as:

$$y_i = \mu + a(Subject_i) + \varepsilon_i, a(Subject_i) \sim N(0, \sigma^2), \varepsilon_i \sim N(0, \sigma^2) \quad (2.11)$$

Where both $a(Subject_i)$ and ε_i are independent. **The notation like $a(Subject_i)$, which brackets a effect, means this effect is set as random.** In this model the observations from the same subject are positively correlated with each other. For convenience and simplicity here V is used to denote covariance matrix instead of G and R.

The covariance between two observations is:

$$cov(y_i, y_j) = \begin{cases} 0 & subject_i \neq subject_j \text{ and } i \neq j; \\ \sigma_a^2 & subject_i = subject_j \text{ and } i \neq j; \\ \sigma_a^2 + \sigma^2 & i = j. \end{cases} \quad (2.12)$$

σ_a^2 is the variance between subjects, and σ^2 is variance between observations within the same object. The variance and fixed parameters are estimated by restrict maximum likelihood(REML), which result in a un-bias estimate. This estimate of fixed effects $\hat{\beta}$ is denoted

as :

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.13)$$

where X is design matrix of fixed effects, y is observations and V is covariance matrix of y which structure is as below,

$$V = \begin{pmatrix} \sigma_a^2 + \sigma^2 & \sigma^2 & \sigma^2 & 0 & 0 & 0 \\ \sigma^2 & \sigma_a^2 + \sigma^2 & \sigma^2 & 0 & 0 & 0 \\ \sigma^2 & \sigma^2 & \sigma_a^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_a^2 + \sigma^2 & \sigma^2 & \sigma^2 \\ 0 & 0 & 0 & \sigma^2 & \sigma_a^2 + \sigma^2 & \sigma^2 \\ 0 & 0 & 0 & \sigma^2 & \sigma^2 & \sigma_a^2 + \sigma^2 \end{pmatrix} \quad (2.14)$$

The variance matrix above is a block diagonal pattern which indicates the case that there are 6 observations within two subjects.

On the base of illumination of two-layer model structure above, we continue to modeling with DATA B. In this step the random effect **DAY** is set as subject.

$$y_i = intercept + Material_i + a(DAY_i) + \varepsilon_i, \text{ where } a(DAY_i) \sim N(0, \sigma_a^2), \varepsilon_i \sim N(0, \sigma^2) \quad (2.15)$$

Here i= 1,2,3,,618. 618 is number of the observations.

Three-Layer Model

A three-layer Hierarchical model could be denoted as:

$$y_i = \mu + a(Subject_i)(block_i) + b(block_i) + \varepsilon_i, b(block_i) \sim N(0, \sigma_b^2), a(Subject_i) \sim N(0, \sigma_a^2), \varepsilon_i \sim N(0, \sigma^2) \quad (2.16)$$

Where $b(block_i)$, $a(Subject_i)$ and ε_i are independent variances. Besides observations within same subject are correlated with each other, the subjects within the same block are also correlated. **The notation $a(Subject_i)(block_i)$ means subject is nested within block.** To

illuminate the covariance structure of three-layer model, the matrix is denoted as below:

$$\mathbf{V} = \begin{pmatrix} \sigma_b^2 + \sigma_a^2 + \sigma^2 & \sigma_b^2 + \sigma_a^2 & \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 + \sigma_a^2 & \sigma_b^2 + \sigma_a^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_a^2 + \sigma^2 & \sigma_b^2 + \sigma_a^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_a^2 & \sigma_b^2 + \sigma_a^2 + \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{pmatrix} \quad (2.17)$$

σ_b^2 is the variance between blocks, σ_a^2 is the variance between subjects, and σ^2 is variance between observations within the same subject. To simplify the structure we adopted 8 observations and two random effects variance matrix, which is also a block diagonal pattern. In the matrix the "*" symbol parts is the symmetrical block of the top-left 4 by 4 sub-matrix.

On the basis of the three-layer model structure above I implemented the model with the DATA B. The random effect **Lab** is set as the third layer.

$$y_i = intercept + Material_i + a(DAY_i)(Lab_i) + b(Lab_i) + \varepsilon_i \quad (2.18)$$

where $b(Lab_i) \sim N(0, \sigma_b^2)$, $a(DAY_i) \sim N(0, \sigma_a^2)$, $\varepsilon_i \sim N(0, \sigma^2)$. Here $i = 1, 2, 3, \dots, 618$. 618 is number of the observations in DATA B.

Three-Layer Model with Interaction Term

If there exists the interaction term between Lab and Material, the model is:

$$y_i = intercept + Material_i + a(DAY_i)(Lab_i) + b(Lab_i) + c(Material_i : Lab_i) + \varepsilon_i \quad (2.19)$$

where $b(Lab_i) \sim N(0, \sigma_b^2)$, $a(DAY_i) \sim N(0, \sigma_a^2)$, $c(Material_i : Lab) \sim N(0, \sigma_c^2)$, $\varepsilon_i \sim N(0, \sigma^2)$. The variance of any observation is

$$Var(y_i) = \sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma^2$$

and the covariance between two observations $i \neq j$ are:

$$\text{cov}(y_i, y_j) = \begin{cases} 0 & \text{Lab}_i \neq \text{Lab}_j ; \\ \sigma_b^2 & \text{Material}_i \neq \text{Material}_j \text{ and } \text{Lab}_i = \text{Lab}_j \text{ and } \text{Day}_i \neq \text{Day}_j ; \\ \sigma_a + \sigma_b^2 & \text{Material}_i \neq \text{Material}_j \text{ and } \text{Lab}_i = \text{Lab}_j \text{ and } \text{Day}_i = \text{Day}_j ; \\ \sigma_b^2 + \sigma_c^2 & \text{Material}_i = \text{Material}_j \text{ and } \text{Lab}_i = \text{Lab}_j \text{ and } \text{Day}_i \neq \text{Day}_j ; \\ \sigma_a^2 + \sigma_b^2 + \sigma_c^2 & \text{Material}_i = \text{Material}_j \text{ and } \text{Lab}_i = \text{Lab}_j \text{ and } \text{Day}_i = \text{Day}_j ; \end{cases} \quad (2.20)$$

It is convenient to think of the observations as an 16×1 vector. There are 2 labs conduct measuring on 2 kinds of material. Each lab takes measurements on 2 days, and 2 duplicates for each day. ($2 \times 2 \times 2 \times 2 = 16$) The first 8 observations are of first material. And within the material, first four of observations are of first lab. Within the lab first 2 observations are of the first day. The 16×16 covariance matrix structure is denoted as below:

$$V = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (2.21)$$

where $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ are 8×8 matrices as follows:

$$\Sigma_{11} = \Sigma_{22} = \begin{pmatrix} \sigma_y^2 & \sigma_a^2 + \sigma_b^2 + \sigma_c^2 & \sigma_b^2 + \sigma_c^2 & \sigma_b^2 + \sigma_c^2 & 0 & 0 & 0 & 0 \\ \sigma_a^2 + \sigma_b^2 + \sigma_c^2 & \sigma_y^2 & \sigma_b^2 + \sigma_c^2 & \sigma_b^2 + \sigma_c^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 + \sigma_c^2 & \sigma_b^2 + \sigma_c^2 & \sigma_y^2 & \sigma_a^2 + \sigma_b^2 + \sigma_c^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 + \sigma_c^2 & \sigma_b^2 + \sigma_c^2 & \sigma_a^2 + \sigma_b^2 + \sigma_c^2 & \sigma_y^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{pmatrix} \quad (2.22)$$

Σ_{11} and Σ_{22} are the two covariance matrices of two kinds of materials.

$$\Sigma_{12} = \Sigma'_{21} = \begin{pmatrix} \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & 0 & 0 & 0 & 0 \\ \sigma_b^2 & \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \end{pmatrix} \quad (2.23)$$

The matrix notation is inspired by Douglas C. Montgomery 2001.[6]

Chapter 3

Statistical Implementation of Theoretic Methods

3.1 Homogeneous Variance Model of Study B

DATA B is of hierarchical structure which could be shown below:

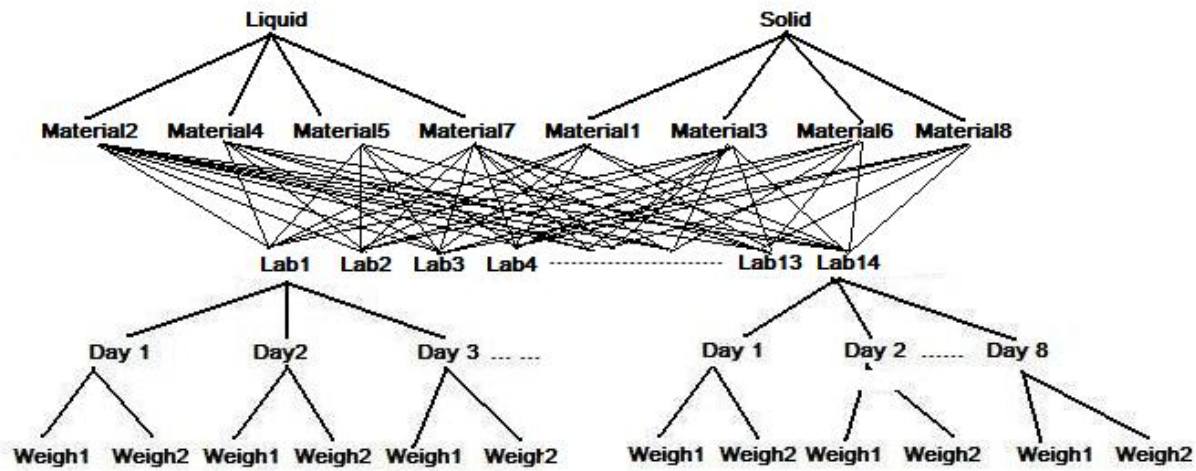


Figure 3.1: Data B Hierarchical structure: 8 Materials are nested within Type (solid/Liquid). 8 Materials are cross with 14 labs and the cross relationship is complete. Days are nested within Labs. Weighs are nested within Days

In this section Homogeneous Variance models are implemented, which means all the observations have same variance components.

3.1.1 Test of overall model reduction

I start with a overall model in which possible effects set as fixed effects. In the ANOVA table the p-values of those terms could be inferences of the following model selection and reduction.

The formula could be denoted as:

$$y_i = \mu + Type_i + Material_i(Type_i) + DAY_i(Lab_i) + Lab_i + Material_i : Lab_i + \varepsilon_i \quad (3.1)$$

where $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

All the concerning effects are described in Chapter One: Data description. Type is the state of material (Solid/liquid). There are 8 kinds of materials in total. 14 labs participated the Inter-Laboratory Study. Each Lab took sub-samples and evaluates those sub-samples in three Days, two duplicates for each Day. After the deletion of outliers, there are 618 observations left for modeling.

Part of the SAS analysis output is :

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Intercept	1	476	3471721	<.0001
Type	0	.	.	.
Material(Type)	0	.	.	.
LAB	13	476	16.18	<.0001
Day(LAB)	40	476	3.01	<.0001
LAB*Material	80	476	2.78	<.0001

Figure 3.2: SAS output of model 3.1

The ANOVA table shows that SAS could not return the F-value and P-value because of being lack of degree of freedom. Considering the situation that the model analysis machine could not hold so many terms as fixed effects, I remove **Type** from the the fixed model, and a sub-model of 3.1 is as follow:

$$y_i = \mu + Material_i + DAY_i(Lab_i) + Lab_i + Material_i : Lab_i + \varepsilon_i \quad (3.2)$$

where $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

Part of Result returned by SAS:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Intercept	1	476	3471497	<.0001
Material	7	476	1350.62	<.0001
LAB	13	476	16.18	<.0001
Day(LAB)	40	476	3.01	<.0001
LAB*Material	80	476	2.78	<.0001

Figure 3.3: SAS output of model 3.2

ANOVA table above shows that all the terms are significant at 0.01% level. Therefore there is no evidence to delete any term from the model.

From Model 3.1 to Model 3.2, **Type** was deleted to simplify the model by which the F-value and P-value could be calculated. But I still have interest to see in a real mixed model, which contains both random and fixed effects, how does **Type** affect the modeling. In model 3.3 all the terms in 3.1 are included, but **Day,Lab** and **Material:Lab** are set as random effects:

$$y_i = \mu + Type_i + Material_i(Type_i) + a(DAY_i)(b(Lab_i)) + b(Lab_i) + c(Material_i : Lab_i) + \varepsilon_i \quad (3.3)$$

where $b(Lab_i) \sim N(0, \sigma_b^2)$, $a(DAY_i) \sim N(0, \sigma_a^2)$, $c(Material_i : Lab_i) \sim N(0, \sigma_c^2)$, $\varepsilon_i \sim N(0, \sigma^2)$.

Part of the SAS analysis output is :

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate
LAB	0.002102
LAB*Material	0.002348
Day(LAB)	0.001426
Residual	0.008135

Fit Statistics

-2 Res Log Likelihood	-1011.4
AIC (smaller is better)	-1003.4
AICC (smaller is better)	-1003.4
BIC (smaller is better)	-1000.9

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Intercept	1	610	222613	<.0001
Material(Type)	6	610	250.12	<.0001
Type	1	610	2465.78	<.0001

Least Squares Means

Effect	Sample Type	Sample number	Estimate	Standard Error	DF	t Value	Pr > t
Material(Type)	Liquid	2	7.0962	0.02109	610	336.47	<.0001
Material(Type)	Liquid	4	6.6434	0.02169	610	306.32	<.0001
Material(Type)	Liquid	5	6.5923	0.02300	610	286.57	<.0001
Material(Type)	Liquid	7	6.2419	0.02114	610	295.26	<.0001
Material(Type)	Solid	1	7.3039	0.02109	610	346.31	<.0001
Material(Type)	Solid	3	7.3741	0.02223	610	331.69	<.0001
Material(Type)	Solid	6	7.0928	0.02229	610	318.20	<.0001
Material(Type)	Solid	8	7.2226	0.02230	610	323.89	<.0001
Type	Liquid		6.6435	0.01591	610	417.66	<.0001
Type	Solid		7.2483	0.01596	610	454.24	<.0001

Figure 3.4: SAS output of model 3.3

Figure 3.4 shows in model 3.3 SAS could estimate all the fixed effects and P-values. This model could be a candidate of final model. In the Fit Statistics Table, four kinds of goodness-of-fit values are supplied. In this thesis -2Restricted log Likelihood values was chosen as the main criterion to select the optimum model.

Model 3.3's -2Res log Likelihood is 1011.4.

On the basis of this full model, several sub-models were constructed by deleting the fixed effects and random effects, **Type**, **Day**, **Lab** , and random cross term **Material:Lab** one by one. All the goodness of fit values of those models were compared in the end.

In model 3.3 all the levels of **Material** and **Type** are significant at 0.01% level. In practical **Type** and **Material** effects are mixed because one kind of material is either solid or liquid. From this aspect a sub-model 3.4 of 3.3 is implemented which drops the **Type** effect:

$$y_i = \mu + Material_i + a(DAY_i)(b(Lab_i)) + b(Lab_i) + c(Material_i : Lab_i) + \varepsilon_i \quad (3.4)$$

where $b(Lab_i) \sim N(0, \sigma_b^2)$, $a(DAY_i) \sim N(0, \sigma_a^2)$, $c(Material_i : Lab) \sim N(0, \sigma_c^2)$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

Part of the SAS output:

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
Material	0
LAB	0.002102
Day(LAB)	0.001426
LAB*Material	0.002348
Residual	0.008135

Fit Statistics

-2 Res Log Likelihood	-1011.4
AIC (smaller is better)	-1003.4
AICC (smaller is better)	-1003.4
BIC (smaller is better)	-1003.1

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Intercept	1	610	222613	<.0001
Material	7	610	564.98	<.0001

Figure 3.5: SAS output of model 3.4

Figure 3.5 shows $-2\text{Res log Likelihood}$ value of model is 1011.4, as same as the value of 3.3, which means the fixed effect **Type** in model 3.3 could not improve the model fitting.

Model 3.5 is a sub-model of model 3.4 rejecting the random cross effect **Material:Lab** .

$$y_i = \mu + \text{Material}_i + a(\text{DAY}_i)(b(\text{Lab}_i)) + b(\text{Lab}_i) + \varepsilon_i, \quad (3.5)$$

where $b(\text{Lab}_i) \sim N(0, \sigma_b^2)$, $a(\text{DAY}_i) \sim N(0, \sigma_a^2)$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

The SAS returned the -2 (Restricted log-likelihood) value of model 3.5: -968.4

Model 3.6 is a sub-model of model 3.5 rejecting the random block effect **Lab**.

$$y_i = \mu + \text{Material}_i + a(\text{DAY}_i) + \varepsilon_i, \quad (3.6)$$

where $a(\text{DAY}_i) \sim N(0, \sigma_a^2)$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

The SAS returned the -2 (Restricted log-likelihood) value of model 3.6: -855.2

Model 3.7 is a sub-model of model 3.5 rejecting the random block effect **Day**.

$$y_i = \mu + \text{Material}_i + b(\text{Lab}_i) + \varepsilon_i, \quad (3.7)$$

where $b(\text{Lab}_i) \sim N(0, \sigma_b^2)$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 618$

The SAS returned the -2 (Restricted log-likelihood) value of model 3.7: -947.9

Model 3.8 is a sub-model of model 3.7 and only contains the fixed effect **Material** and the interception μ .

$$y_i = \mu + \text{Material}_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \quad (3.8)$$

The SAS returned the -2 (Restricted log-likelihood) value of model 3.8: -854.6

Model 3.9 is another possible reduction way. If the cross effect **Material:Lab** is proved to be significant and could not be left out the model we could not drop the basic effect **Material** and **Lab** neither. Hence the new investigation is whether **Day** is a necessary effect.

$$y_i = \mu + Material_i + b(Lab_i) + c(Material_i : Lab_i) + \varepsilon_i \quad (3.9)$$

here $b(Lab_i) \sim N(0, \sigma_b^2)$, $c(Material_i : Lab) \sim N(0, \sigma_c^2)$, $\varepsilon_i \sim N(0, \sigma^2)$

The SAS returned the -2 (Restricted log-likelihood) value of model 3.9: -978.9

Model Comparison

Restricted Likelihood Ratio Test is used here to compare different Models. For instance Model B is a sub-model of model A. Model A has r more parameters than Model B. Twice the change in log likelihood is referred to a chi_r^2 distribution. The additional parameters could be either fixed parameters or random(variance components).[1] The test statistic could be written as:

$$G_{A \rightarrow B} = 2le^A - 2le^B = -[(-2le^A) - (-2le^B)] \quad (3.10)$$

Where $2le$ indicates the 2 log-likelihood values of models. The restricted likelihood ratio test G statistic should follow chi_{df}^2 -distribution, where df denotes the difference of number of variance components between Model A and B.

For the model 3.3 to 3.9 seven different models with different variance structures were compared:

Table 3.1: Model Comparison Table

Model	-2le	G-value	df	P-value
3.3: Material,Type,Day,Lab,Lab:Material included	-1011.4	$G_{3.3 \rightarrow 3.4} = 0$	1	1
3.4: Material,Day,Lab,Lab:Material included	-1011.4	$G_{3.4 \rightarrow 3.5} = 32.5$	1	1.1919E-8
3.5: Material,Day,Lab included	-968.4	$G_{3.5 \rightarrow 3.6} = 113.2$	1	0
3.5: Material,Day,Lab included	-968.4	$G_{3.5 \rightarrow 3.7} = 20.5$	1	.000005963
3.6: Material,Day included	-855.2	$G_{3.6 \rightarrow 3.8} = 0.6$	1	0.43858
3.7: Material,Lab included	-947.9	$G_{3.7 \rightarrow 3.8} = 93.3$	1	0
3.8: Only Material included	-854.6			
Another Reduction Way: start with dropping Day effect: 3.4: Material,Day,Lab,Lab:Material included	-1011.4	$G_{3.4 \rightarrow 3.9} = 43$	1	5.474E-11
3.9: Material,Lab,Lab:Material included	-978.9			

Table 3.1 gives all the possible models which may appear in the reducing process.

We start from the full model 3.3. The statistic of test From 3.3 to 3.4 is 0 with P-value zero,which means **Type** as fixed effect can not improving Model fitting . The statistic of test from 3.4 to 3.5 shows the interaction effect **Material:Lab** is significant 0.0001% level with tiny p-value 1.1919E-8. Therefore the reduction steps of basic effects **Material** and **Lab** are unnecessary to be considered.

Then following the reducing of Day from model 3.4. The statistic of test from 3.4 to 3.9 shows the random effect **Day** is also significant at 0.0001% level with small p-value 5.474E-11.

Finally Model 3.4 is testified to be necessary and fit better than the reduced models:

$$y_i = \mu + Material_i + a(DAY_i)(b(Lab_i)) + b(Lab_i) + c(Material_i : Lab_i) + \varepsilon_i$$

,where $b(Lab_i) \sim N(0, \sigma_b^2), a(DAY_i) \sim N(0, \sigma_a^2), c(Material_i : Lab) \sim N(0, \sigma_c^2), \varepsilon_i \sim N(0, \sigma^2)$

The analysis of variance of 3.4 is summarized in Figure 3.5. Materials are fixed effect. Lab ,Day and Material:Lab are random.

3.1.2 Diagnostic Checking for Homogeneous Variance Model 3.4

The major tool used in diagnostic checking is residual analysis. There are two different residuals.

Two Types of Raw Residuals and Random effects

The marginal residuals are the difference between observations and estimated (marginal mean) mean values:

$$e_{mi} = y_i - \hat{y}_i = y_i - x\hat{\beta}$$

, where $i=1:618$

The conditional residuals are the difference between observations and predicted values of observations.

$$e_{ci} = y_i - \hat{y}_i = y_i - x\hat{\beta} - z\hat{\gamma}$$

, where $i=1:618$

The conditional residuals are more close to the real residuals because the random effects have been considered in the prediction. Compared with the conditional residuals, the marginal residuals reflect the information of the total variability from variance components: $z\hat{\gamma}$ and the real errors ϵ .

The residual vectors of these two raw residuals are denoted as ϵ_c and ϵ_m . The variance of ϵ_m is V , the covariance matrix of the observations. From this aspect marginal residuals could be used to investigate the whole variability of observations.[8]

According to the assumption of LLM, the normality of the random effects γ is also a topic. This will be processed in the end of the thesis for the optimal model. In this section only Residuals Diagnostic Checking is presented.

Normality and Homogeneity Checking of Residuals

Besides normality and homogeneity testing of residuals some plots are also made to reveal the structure and suggest appropriate further modeling.

1. Residuals (conditional residuals ϵ_c) against predicted values:

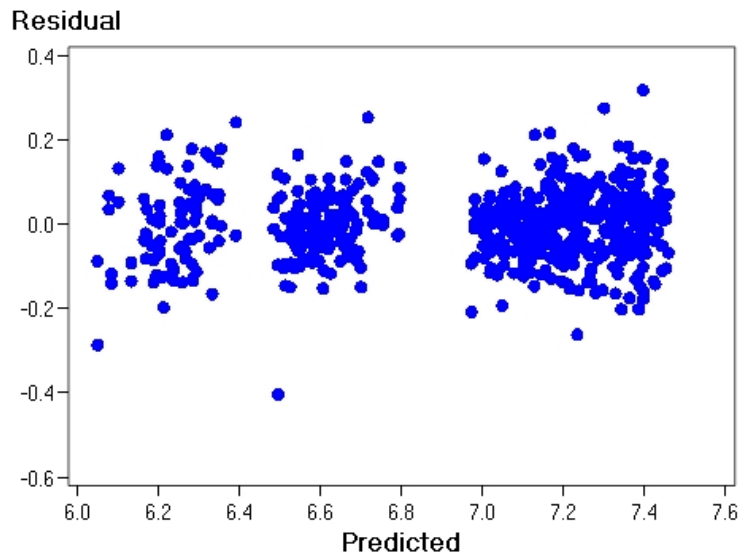


Figure 3.6: Residuals ϵ_c vs predicted values

The ϵ_c seems quite constant and not increasing with the predicted values.

2. Histogram of conditional residuals ϵ_c

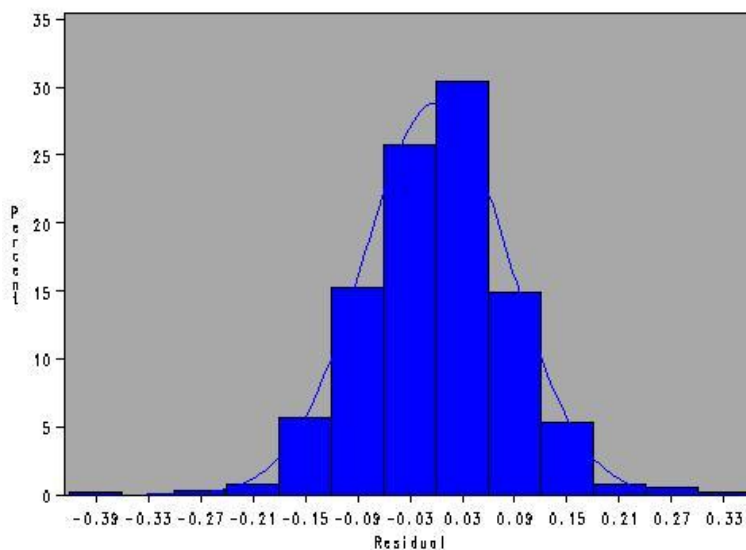


Figure 3.7: Hist-gram of ϵ_c vs fitted values

The distribution of residuals looks only a little asymmetrical.

3. The normal Q-Q plot for conditional residuals ϵ_c

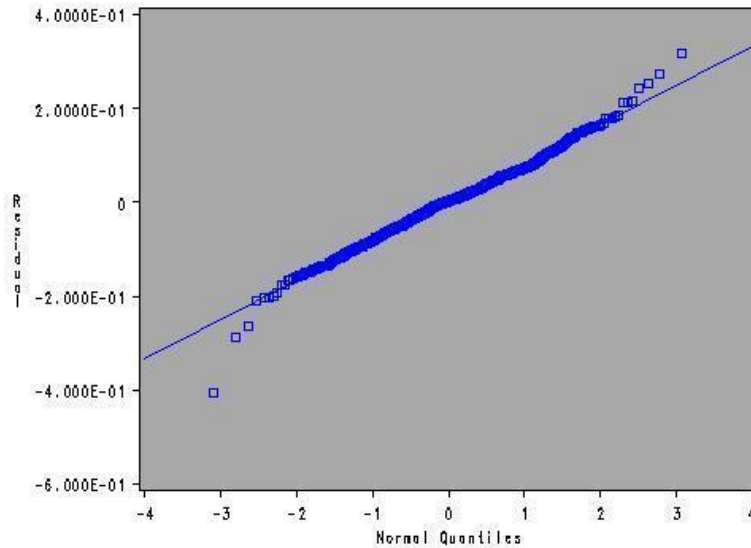


Figure 3.8: A normal Q-Q plot of conditional residuals ϵ_c

There are light departure trends at two tails. It does not fit normal distribution perfectly.

Histogram and Q-Q plot are the comparisons by eye to the normal distribution. To get quantitative statistic for normality checking, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling tests are performed. Those tests have same null hypothesis:

H_0 : The data given for the test follow normal distribution.

Goodness-of-Fit Tests for Normal Distribution

Test	Statistic	p Value
Kolmogorov-Smirnov	D 0.03598477	Pr > D 0.050
Cramer-von Mises	W-Sq 0.15895375	Pr > W-Sq 0.019
Anderson-Darling	A-Sq 0.92863435	Pr > A-Sq 0.020

Figure 3.9: Goodness-of-Fit Tests for Normal Distribution (conditional residuals ϵ_c)

The P-values returned by last three tests are significant at 5% level. Those kinds of tests are not very reliable, we can not get conclusion that normality assumption is violated just because

these tests.

4. Residuals against **Type**

As introduced before **Type** (Solid/Liquid) is an important effect of interest. In this part both Marginal and Conditional Residuals are test to reveal more information.

marginal residuals ϵ_m against Type:

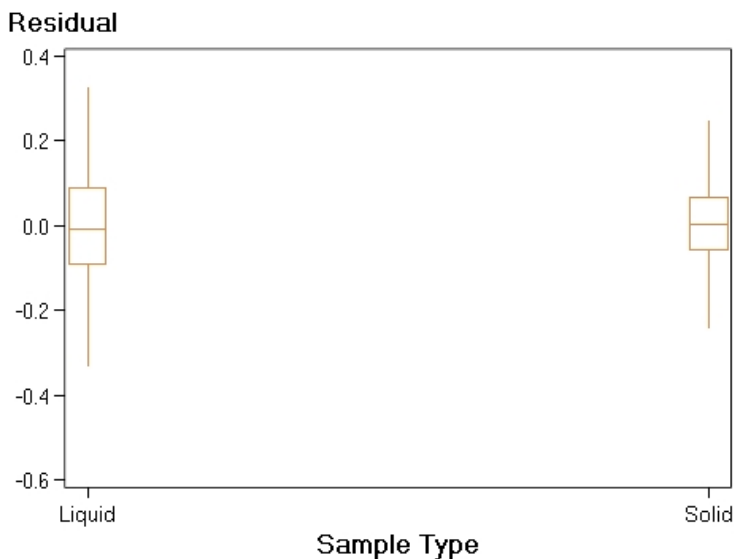


Figure 3.10: marginal residuals ϵ_m against Type, ± 1.5 times interquartile range

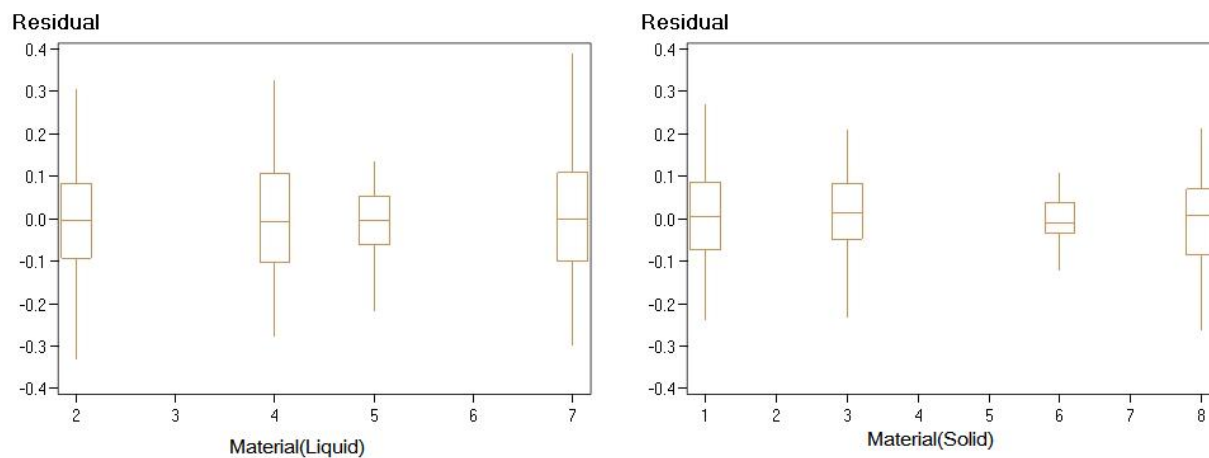


Figure 3.11: DATA B: marginal residuals ϵ_m against Materials, divided by Type group, ± 1.5 times interquartile range

It seems the ϵ_m of the Liquid type are more dispersive than the ones of Solid. For further quantitative information the one way homogeneity tests for Type of ϵ_m are performed.

Levene's Test for Homogeneity of Resid Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Type	1	0.00673	0.00673	11.69	0.0007
Error	616	0.3548	0.000576		

Bartlett's Test for Homogeneity
of Resid Variance

Source	DF	Chi-Square	Pr > ChiSq
Type	1	19.0825	<.0001

Figure 3.12: Levene's and Bartlett's one way tests for Homogeneity of marginal residuals ϵ_m against Type

For both Levene's and Bartlett's one way tests for ϵ_m the statistics are significant at 0.1% level. The standard deviations for the Liquid and Solid are 0.13 and 0.1. Those could be strong evidences of heterogenous variance structure for different Type.

conditional residuals ϵ_c against Type: For Levene's and Bartlett's one way tests for ϵ_c the statistics are not significant with p value 0.25 and 0.15.

There still exists the possibility that the between-Type heteroscedasticity is actually caused by between-Material heteroscedasticity. Because we only have four materials for one Type, the sample size of Materials within Type is only 4, which is quite small and could not present the situation of the population of all kinds of materials. Therefore the similar analysis will be implemented on DATA A to compare with the analysis here.

5.conditional residuals ϵ_c against **Material**

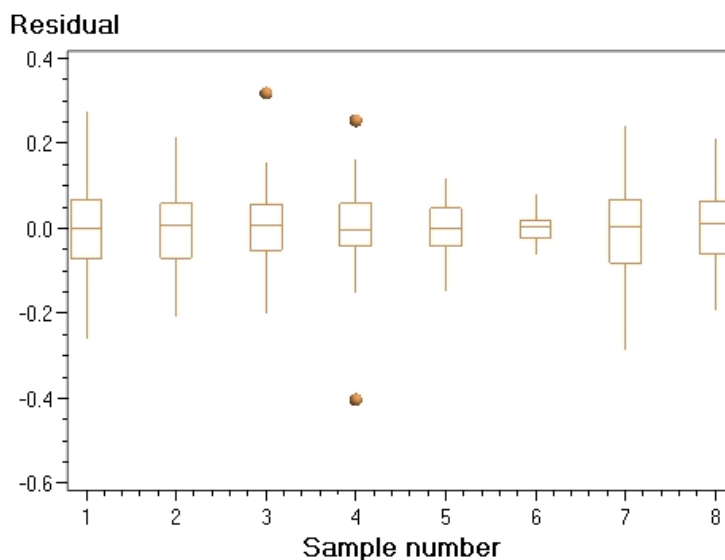


Figure 3.13: conditional residuals ϵ_c against Material, ± 1.5 times interquartile range

Levene's Test for Homogeneity of Resid Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Material	7	0.00528	0.000754	5.13	<.0001
Error	610	0.0897	0.000147		

Bartlett's Test for Homogeneity of Resid Variance			
Source	DF	Chi-Square	Pr > ChiSq
Material	7	117.0	<.0001

Figure 3.14: Levene's and Bartlett's one way tests for Homogeneity of conditional residuals ϵ_c against Lab

For both Levene's and Bartlett's one way tests for ϵ_c the statistics are significant at 0.01% level. Those give us strong evidences of heterogenous variance structure for different Materials. Besides, the residual box plots vs Material seem quite nonuniform.

At present Material is already in model 3.4 as a fixed effect and the main predictor to affect the mean value. The variability within different Martials looks different now. What is the source of the difference? Is it because of different Type(Liquid/Solid)?

An interest phenomena is the dispersive extents for different Material in Figure 3.13 are not totally same with the extents in Figure 1.8 of Chapter 1, Preliminary Analysis section. Although Material No.7 and No.8 still tend to have large ϵ_c , but Material No.4 tends to have smaller variance than in Figure 1.8. The difference of the between ϵ_c of materials seem smaller than the difference between variance of materials in Figure 1.8. Although model 3.4 violated the homogeneity assumption of errors obviously, it still explain rather large part of the variability of the DATA B.

6. conditional residuals ϵ_c against **Lab**

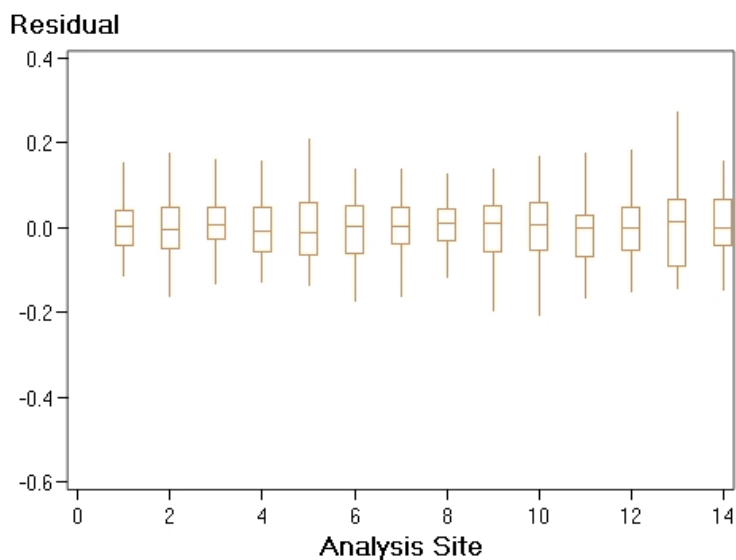


Figure 3.15: conditional residuals ϵ_c against Lab, ± 1.5 times interquartile range. The dispersions of ϵ_c for different Labs looks obvious nonuniform.

The one way homogeneity tests for Lab were performed:

Levene's Test for Homogeneity of Resid Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
LAB	13	0.00421	0.000323	2.20	0.0084
Error	604	0.0887	0.000147		

Bartlett's Test for Homogeneity
of Resid Variance

Source	DF	Chi-Square	Pr > ChiSq
LAB	13	35.3057	0.0008

Figure 3.16: Levene's and Bartlett's one way tests for Homogeneity of conditional residuals ϵ_c against Lab

For both Levene's and Bartlett's one way tests of ϵ_c the statistics are significant at 1% level. Those give us evidences of between-lab heteroscedasticity. Do some lab supply more precise evaluation? We already know there exists significant Day-to-Day variability, which is the variance within Lab. But is the variability between Days the same for all the Labs? It will be studied later.

New models of heterogeneous variance structure will be implemented on the basis of hints above.

3.1.3 Practical Implication of Model 3.4

The practical implication of this study is quite important. We are interested in the source of the variability of measurements. In Model 3.4 Lab, Day and Material:Lab are all significant. It means the variability results from Lab, Day and Material:Lab. After the residuals analysis, some indications are listed:

1. For the homogeneity tests the residuals are heterogeneous cross Type, Lab and Materials. Type have been tested in homogeneous model and tend to be not significant as fixed effect. But at present it seems that it affects the measurement in another way: Different Types of Materials may have different variance structures. This will be an individual topic later.

Although Lab and Material effects are already in the model 3.4 as random effects, it seems the specification of them are not proper. Heterogeneous variance structure for Lab and Material would be implemented later. It is natural to guess maybe some labs have smaller within variances and could supply more precise results.

2. The Day to Day variation within Labs is significant source of variability. Therefore may it is necessary to modify the work process of method and labs internal quality assurance system.

3. It is hard to give a practical conclusive interpretation of the Material:Lab interaction exactly before a well fit model is conducted. So far there are several possibilities.

In model 3.4 one assumption is that variance components are homogeneous for all observations. But in practical maybe some materials have larger variances than the rest, for instance, Material No 7. (See figure 3.14) The variance within Material No7 may tend to be larger than the ones of Material 6. Model 3.4 uses the single Lab, Day and Residual random effect to estimate the variances of different Materials. Hence the random interaction term could a "cushion" for the between-material heteroscedasticity.

Some Labs may could supply more precise evaluations. That means even for same material, the observations from "better" have smaller variance than the ones from the "worse" labs. Even if a "worse" Lab sometimes may be good at measure certain kind of material and give more precise evaluations with smaller variance than other labs. So far we do not know whether the all the Labs work stable from material to material.

In those cases the interaction term could be a "cushion" for this extra "variability".

3.1.4 RSD% of Homogeneous Variance Model 3.4

According to the SAS output of Model 3.4 the estimated values of Variance Components are:

LAB 0.002102

Day(LAB) 0.001426

LAB*Material 0.002348

Residual 0.008135

As introduced before in section Log-transformation of Raw Data, Chapter 2,

$$Var[\ln(Y_i)] = \frac{Var[Y_i]}{E(Y_i)^2} \approx C^2 = RSD_R^2$$

Here the variance of log-transformed observations :

$$RSD_R^2 = Var[\ln(Y_i)] = \sigma_{Lab} + \sigma_{Day} + \sigma_{Lab*Material} + \sigma_{Residual} = 0.002102 + 0.001426 + 0.002348 + 0.008135 = 0.014011$$

Hence the RSD_r and RSD_R of model 3.4 :

Table 3.2: RSD for model 3.4

$RSD_r\% = RSD_{Residual}\%$	9%
$RSD_R\% = \sqrt{0.014011}$	11.84%

3.2 Heterogeneous Variance Model for Type of Study B

3.2.1 Motivation

When Type is set as fixed effect in the model the fixed effect Material is totally mixed with it. Because all kinds of Materials are either liquid or solid and there is no direct contrast to compare the Type effects on the same material. Unsurprisingly Type(Solid/liquid) as a fixed effect in the Homogeneous modeling is not significant and proven not to improve the model fitting by the Log ratio likelihood test from model 3.3 to 3.4.

An important part of this thesis aim is to detect whether the different Type (solid/liquid) of the Materials have different variance components. According to this idea the SAS statement **Repeated** is conducted, by which the variance components are estimated separately for different Type. The homogeneity assumption of error is also reset: The variance should be homogeneous of within different Types.

The goodness-of-fit values of the new models will be compared with the one of the general model 3.4 presented in last section by Likelihood Ratio Test to detect whether Materials of different Types have different covariance structure.

If the difference of variance does exist, which variance component leads to it? Is it lead by the effects DAY, LAB or residuals within Types? A series sub-models will be conducted. Same as the model selection mechanism in homogeneous model section, an optimal heterogeneous variance model for Types will be chosen by Likelihood Ratio Test.

3.2.2 Practical Issue of SAS statements

The relative SAS statements have already been introduced in Section 2.1.3. Here I just list the main points briefly.

In SAS Proc Mixed **Repeated** statement is to define the structure of residuals matrix **R**. In default output $R = \sigma^2 I_n$. Here R only contain one variance component.

When Repeated is set as: repeated/ group= subject1, then Proc mixed will calculated individual residuals for each level of the subject. If this subject contain p levels then R contains

p different variance components on the diagonal line.

Another similar function is the group option following the **Random** statement, which could set the covariance structure of **G**. For instance, in SAS it is written as:

Random *Day/group = Lab*; Then for each Lab SAS will return an individual estimated variance component of DAY.

In other words, Group option following **Repeated** statement defines an effect specifying heterogeneity in the covariance structure of R. While group options following Random statement specifies the heterogeneity in the covariance structure of G.

3.2.3 Statistical Implementation

For comparing with the homogeneous variance model 3.4 in last section I choose the same Model Formula. But in new model different types corresponds to two different system of covariance components.

$$y_i = \mu + Material_i + a(DAY_i)(b(Lab_i)) + b(Lab_i) + c(Material_i : Lab) + \varepsilon_i \quad (3.11)$$

if $i \in Group(Solid)$: $b(Lab_i) \sim N(0, \sigma_{bs}^2)$, $a(DAY_i) \sim N(0, \sigma_{as}^2)$, $c(Material_i : Lab) \sim N(0, \sigma_{cs}^2)$, $\varepsilon_i \sim N(0, \sigma_s^2)$;

if $i \in Group(Liquid)$: $b(Lab_i) \sim N(0, \sigma_{bl}^2)$, $a(DAY_i) \sim N(0, \sigma_{al}^2)$, $c(Material_i : Lab) \sim N(0, \sigma_{cl}^2)$, $\varepsilon_i \sim N(0, \sigma_l^2)$;

i=1:618.

In model 3.11 there are different estimates of LAB ,DAY Lab*Material and Residuals for different Type(Liquid/solid).

Part of the SAS output is as below:

The model 3.11 has goodness value -2lld: -1022.2.

Comparing Heterogeneous Model 3.11 with Homogeneous Model 3.4:

Model 3.11 have 4 more parameters than model 3.4, twice the change in log likelihood is referred to a χ_4^2 distribution. The Null hypothesis of the test is :

$$H_0 : \sigma_{bs}^2 = \sigma_{bl}^2 \cap \sigma_{as}^2 = \sigma_{al}^2 \cap \sigma_{cs}^2 = \sigma_{cl}^2 \cap \sigma_s^2 = \sigma_l^2$$

Covariance Parameter Estimates		
Cov Parm	Group	Estimate
LAB	Type Liquid	0.001839
LAB	Type Solid	0.001615
Day(LAB)	Type Liquid	0.003766
Day(LAB)	Type Solid	0.000776
LAB*Material	Type Liquid	0.004468
LAB*Material	Type Solid	0.000472
Residual	Type Liquid	0.007687
Residual	Type Solid	0.007589
Fit Statistics		
-2 Res Log Likelihood		-1022.2

Figure 3.17: SAS output of model 3.11

The Likelihood Ratio Test between those two models with degree of freedom 4:

$$G_{3.11 \rightarrow 3.4} = 2le^{3.11} - 2le^{3.4l} = 1022.2 - 1011.4 = 10.8 \rightarrow P_{\chi^4} = 0.028906$$

The test is significant at 5% level. Model 3.11 to fit better than model 3.4.

Heterogeneous Variance Model Reduction

The conclusion obtained in last section is that Model 3.11 fits better than the general model 3.4. And different Types of Materials are proven to have different covariance structure. But which block effect leads this difference? Do all of the variance components of Solid/Liquid are different or only some terms different? For simplicity, only the goodness of fit values of Models with different covariance structure by Type are compared and tested in following table. The

SAS output of model 3.4 shows that the residual of different types are similar. Therefore we start with hypothesis test is : $H_0 : \sigma_s^2 = \sigma_l^2; (3.11 - > 3.12)$

Table 3.3: Heterogeneous Variance Model Comparison Table

Model	-2le	G-value	df	P-value
3.11: Day,Lab,Lab:Material,Residual are different for Type	-1022.2	$G_{3.11 \rightarrow 3.12} = 0$	1	1
3.12: Day,Lab,Lab:Material are different for Type	-1022.2	$G_{3.12 \rightarrow 3.13} = 4.9$	1	0.0268
3.13: Lab,Lab:Material are different for Type	-1017.3			
3.12: Day,Lab,Lab:Material are different for Type	-1022.2	$G_{3.12 \rightarrow 3.14} = 11.1$	1	0.0039
3.14: Day,Lab are different for Type	-1011.1			
3.12: Day,Lab,Lab:Material are different for Type	-1022.2	$G_{3.12 \rightarrow 3.15} = 28.9$	1	7.6213E-8
3.15: Day,Lab:Material are different for Type	993.3			

The table above shows the procedure of model reduction. From model 3.11 to 3.12, the statistic is not significant with p value 1. The goodness value of fit of 3.11 and 3.12 are even the same. Hence the null hypothesis is accepted.

But from model 3.12 to 3.13, 3.14 and 3.15, all three tests are significant at 5% level. Model 3.12 could not be reduced further.

According to the log likelihood ratio test, the only term could be combined is residual. Part of SAS output of 3.12:

```
Covariance Parameter Estimates
Cov Parm Group Estimate
LAB Type Liquid 0.001835
LAB Type Solid 0.001617
Day(LAB) Type Liquid 0.003780
Day(LAB) Type Solid 0.000769
LAB*Material Type Liquid 0.004477
LAB*Material Type Solid 0.000463
Residual 0.007638
```

The assumption of residual of model 3.12 is residual should be homogeneous.

3.2.4 RSD of Homogeneous Variance Model 3.12

The RSD value estimated in model 3.12: Table 3.4 shows that the RSD_R of Liquid is much

Table 3.4: Data Structure

item	Liquid	Solid
$RSD_r\% = RSD_{Residual}\%$	8.74%	8.74%
$RSD_R\%$	13.3%	10.2%

larger than $tRSD_R$ of Solid.

RSD_r is 8.74 which is a little smaller than RSD_r (9.0) homogeneous variance model. While RSD_R of homogeneous variance model 3.4 is 11.84, which is between the RSD_R of Liquid and Solid (13.3, 10.2).

3.2.5 Diagnostic Checking for Heterogeneous Variance Model 3.12

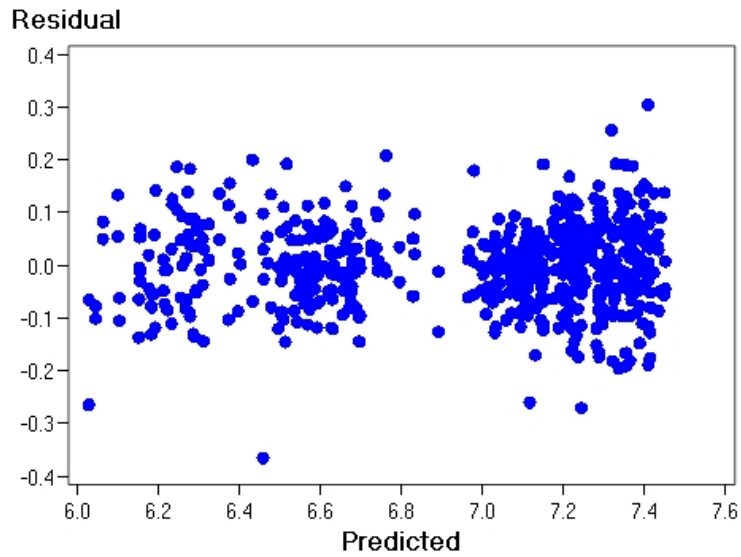
The major tool used in diagnostic checking of this section is still residual analysis.

Model 3.12 have assumption that: Variance components for Lab, Day and Lab:Material are heterogeneous for Type, but residual is homogeneous.

The homogeneity of residual was checked as in the homogeneous model last section.

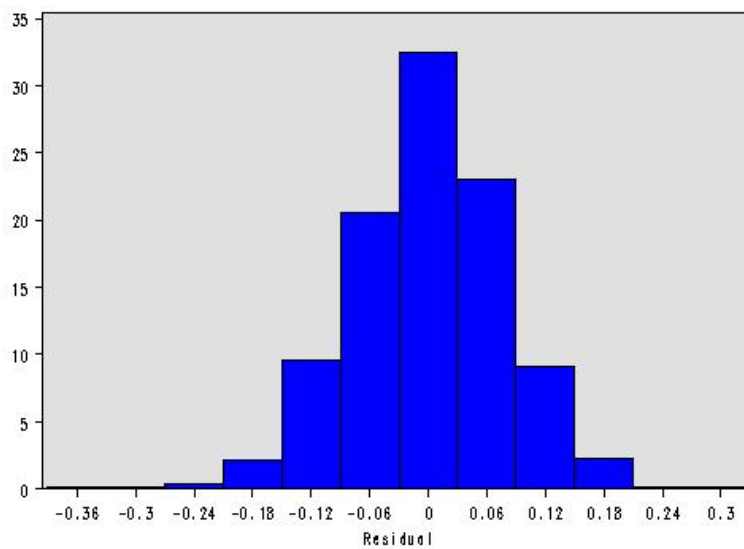
I still use conditional residuals as the analysis object.

1. conditional residuals ϵ_c) against fitted values:

Figure 3.18: ϵ_c) vs fitted values

The ϵ_c seem quite constant and having no pattern for the predicted values.

2. Histogram of conditional residuals ϵ_c

Figure 3.19: Hist-gram of ϵ_c vs fitted values

The distribution of residuals looks a little asymmetrical. But the bias extent is not large.

3. The normal Q-Q plot for conditional residuals ϵ_c

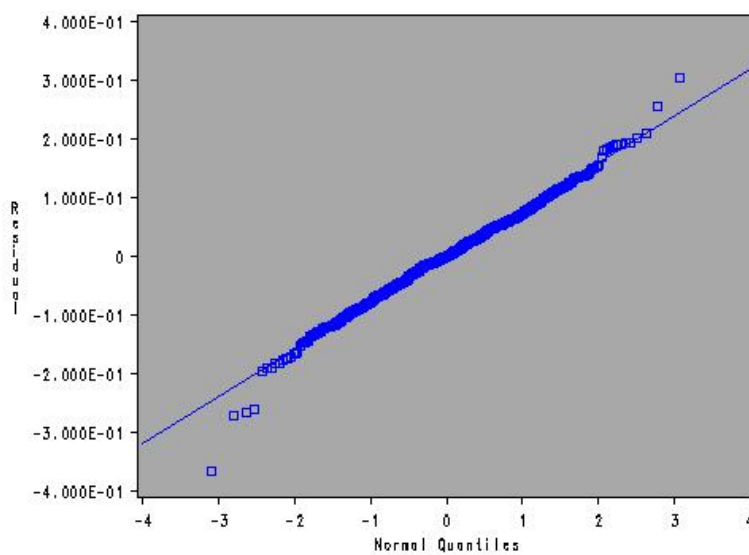


Figure 3.20: A normal Q-Q plot of conditional residuals ϵ_c

There are several wild points and light departure trend at two tails. Generally it fits normal distribution.

Those Normality tests have same null hypothesis:

H_0 : The data given for test follow normal distribution.

Goodness-of-Fit Tests for Normal Distribution

Test Statistic p Value

Kolmogorov-Smirnov D 0.0365655 Pr > D 0.044

Cramer-von Mises W-Sq 0.11036836 Pr > W-Sq 0.085

Anderson-Darling A-Sq 0.69274680 Pr > A-Sq 0.074

The last two tests are not significant at 5% level. Only Kolmogorov-Smirnov test is significant at 5% level. Those kinds of tests are not very reliable, we can not draw the conclusion that normality assumption is violated.

4. conditional residuals ϵ_c against Type, plots and homogeneity test.

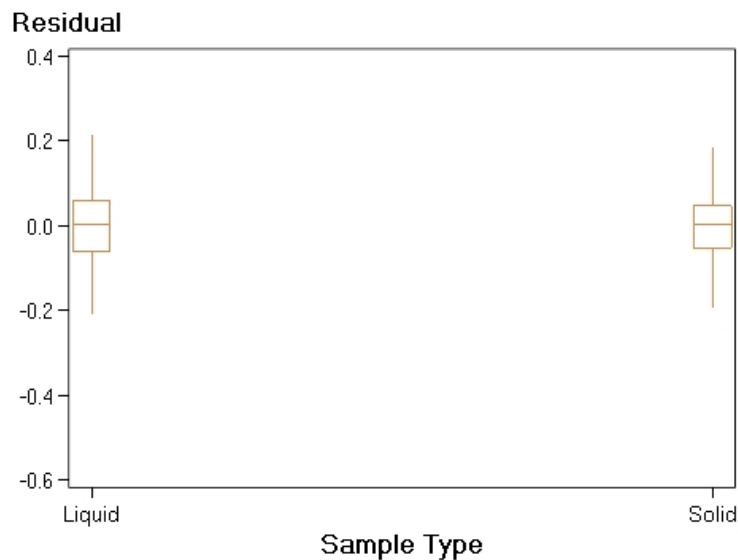


Figure 3.21: Conditional residuals ϵ_c against Type, ± 1.5 times interquartile range

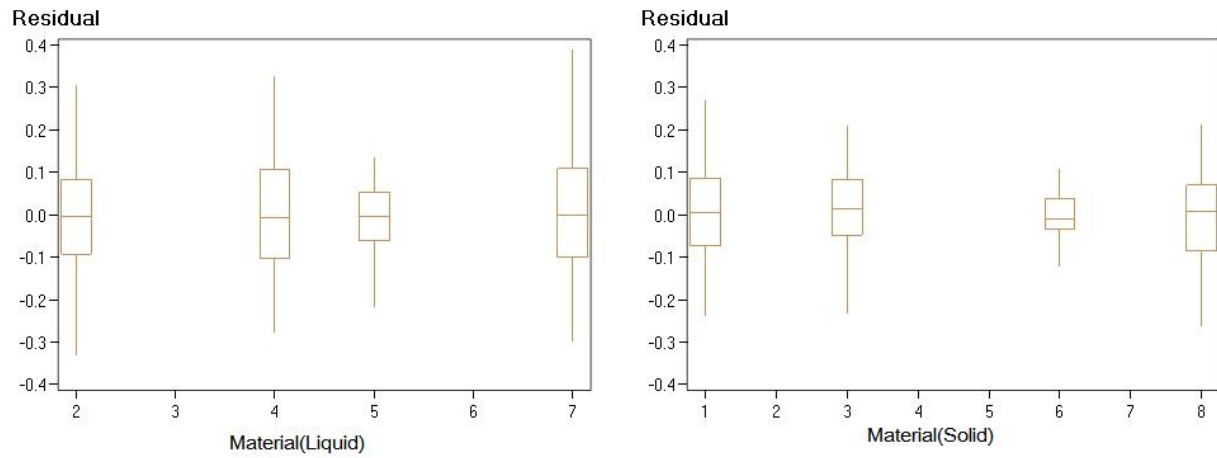


Figure 3.22: DATA B: residuals ϵ_c against Materials, divided by Type group, ± 1.5 times interquartile range

It seems ϵ_c for the Liquid type are more dispersive than the ones of Solid. For further quantitative information the one way homogeneity tests for Type are performed.

Bartlett's Test for Homogeneity of Residual Variance:

Source DF Chi-Square Pr >ChiSq

Type 1 2.0662 0.1506

For Bartlett's one way test for ϵ_c the statistic is not significant at 5% level. The ϵ_c is not heterogeneous for Type.

5.conditional residuals ϵ_c against **Material**

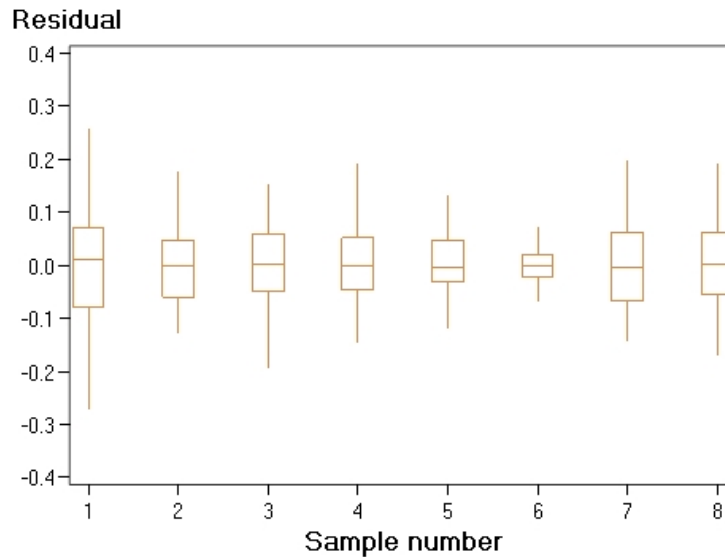


Figure 3.23: conditional residuals ϵ_c against Material, ± 1.5 times interquartile range

Bartlett's Test for Homogeneity of Residual Variance

Source DF Chi-Square Pr > ChiSq

Material 7 107.5 <0.0001

For Bartlett's one way test for ϵ_c the statistic is significant at 0.01% level. The residual is heterogeneous for Material. The homogeneity assumption of residual of model 3.12 is violated. In model 3.12 the Residual distribution has obvious pattern for Materials.

3.2.6 Brief Summary

In Model 3.12 the assumptions of its' covariance structure specification are violated: the residual is not homogeneous. The variance estimates in 3.12 could not describe the heteroscedasticity between Material. Although the RSD_R of Liquid is larger than RSD_R of Solid in 3.12, but it could not lead to conclusion that the Liquid Material have larger variance than the Solid Material. Furthermore, the materials involved in Study B are only 8 samples of the whole population of materials. Similar models and tests on Study A were processed to check whether variance of another 5 kinds of materials have similar characteristics or not.

3.3 Homogeneous Variance Model and Heterogeneous Variance Model for Type of Study A

3.3.1 Homogeneous Variance Model of Study A

For simplicity, the model reduction steps are presented in below below. All the models specifications and tests are similar to Study B, except there is no block effect "Day". Material(5 levels) is fixed effect nest within another fixed effect, Type(Solid/liquid). Lab(14) and Interaction term, Lab:Material, are random effects. 4 different variance structures were compared from model A1 to model A4:

Table 3.5: Model Comparison Table of Homogeneous Variance Model for Type of Study A

Model	-2le	G-value	df	P-value
A1: Material(Type),Type,Lab,Lab:Material included	-226.9	$G_{A1 \rightarrow A2} = 0$	1	1
A2: Material,Lab,Lab:Material included	-226.9	$G_{A2 \rightarrow A3} = 0.2$	1	0.65
A3: Material,Lab included	-226.7	$G_{A3 \rightarrow A4} = 15.3$	1	.00009
A4: Material included	-211.4			

The series of Log ratio Likelihood tests of model reduction show only the statistic of step " $G_{A3 \rightarrow A4}$ " is significant. The optimal homogeneous variance model is A3:

$$y_i = \mu + Material_i + b(Lab_i) + \varepsilon_i,$$

where $b(Lab_i) \sim N(0, \sigma_b^2)$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1 : 132$

3.3.2 Diagnostic Checking for Homogeneous Variance Model A3

1. Residuals(conditional residuals ϵ_c) against predicted values:

3.3 Homogeneous Variance Model and Heterogeneous Variance Model for Type of Study A73

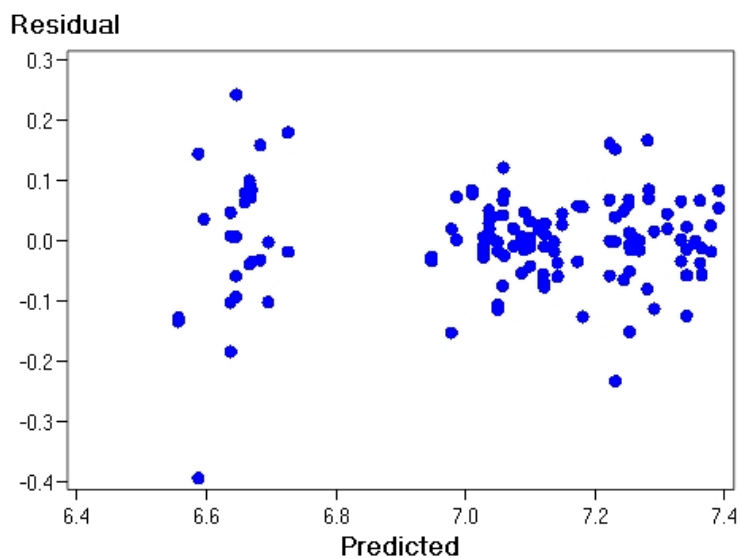


Figure 3.24: Residuals ϵ_c vs fitted values

The ϵ_c is not uniform. The variance of the residual for prediction value around 6.6 looks larger than the rest. But it seems having no increasing pattern with the predicted values.

2. Histogram of conditional residuals ϵ_c

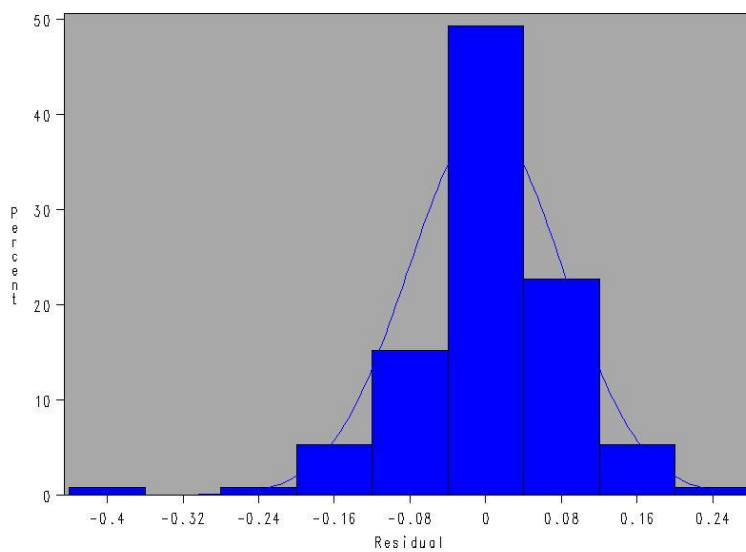
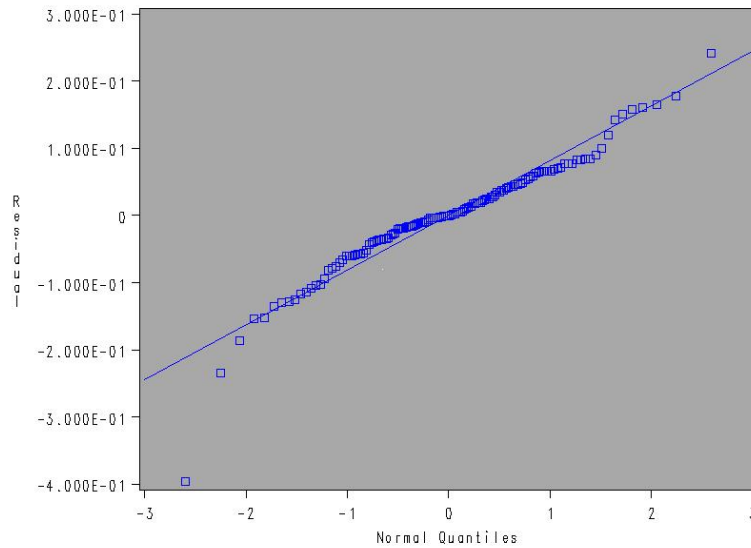


Figure 3.25: Hist-gram of ϵ_c vs fitted values

The distribution of residuals seems asymmetrical.

3. The normal Q-Q plot for conditional residuals ϵ_c Figure 3.26: A normal Q-Q plot of conditional residuals ϵ_c

There are clear departure trends in Q-Q plot. To obtain quantitative statistics for normality checking, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling tests are performed. Those tests have same null hypothesis:

H_0 : The data given for test follow normal distribution.

Goodness-of-Fit Tests for Normal Distribution

Test Statistic p Value

Kolmogorov-Smirnov D 0.10193254 Pr > D <0.010

Cramer-von Mises W-Sq 0.27921116 Pr > W-Sq <0.005

Anderson-Darling A-Sq 1.60299622 Pr > A-Sq <0.005

The P-values returned by last three tests are significant at 1% level. Combining the departure situation in Q-Q plot, we could come to the conclusion that normality assumption of residuals is violated.

4. Residuals(conditional residuals ϵ_c) against Type

3.3 Homogeneous Variance Model and Heterogeneous Variance Model for Type of Study A75

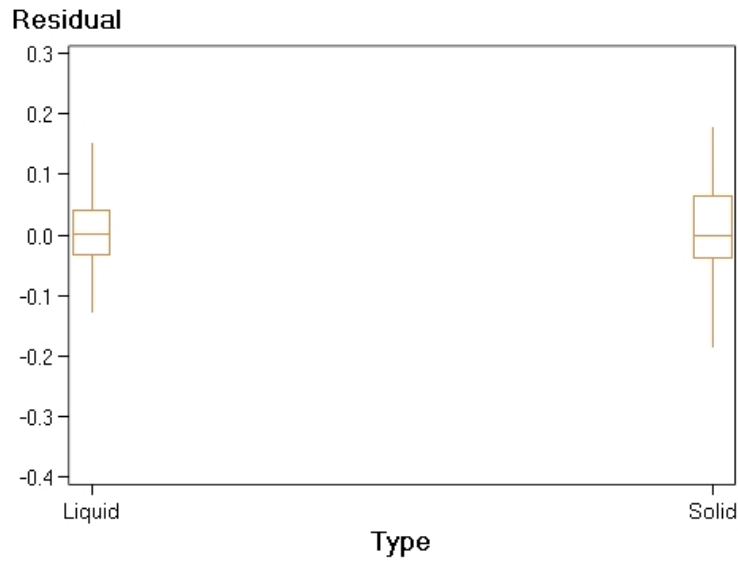


Figure 3.27: conditional residuals ϵ_c against Type, ± 1.5 times interquartile range

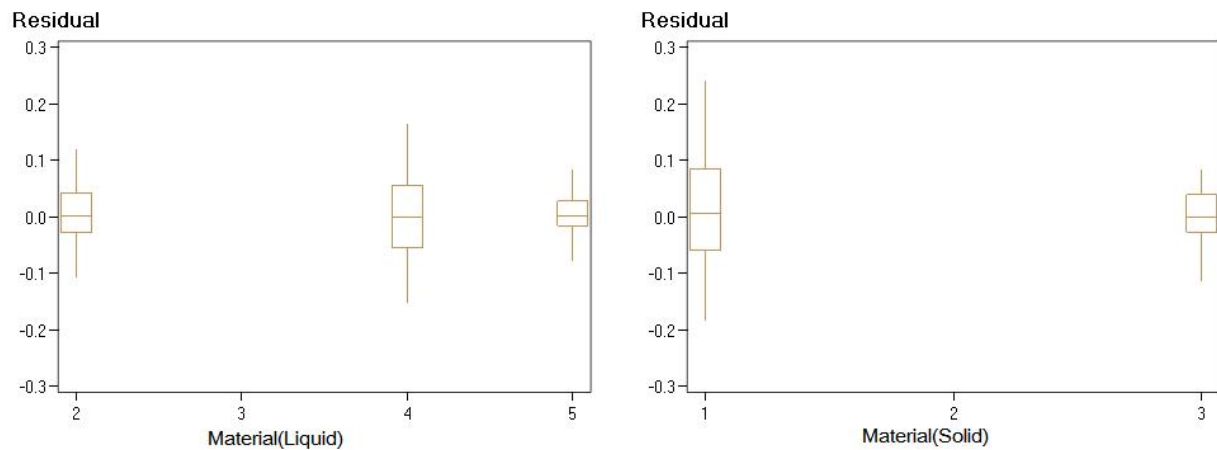


Figure 3.28: DATA B: conditional residuals ϵ_c against Materials, divided by Type group, ± 1.5 times interquartile range

There is indeed interesting thing in those plots. In model A3 Solid Materials seem more dispersive variance than the Liquid Material. It is the opposite phenomena with Study B. But not all the solid materials variances are larger than the variances of liquid, for instance, Material No 3(silod) have smaller variance than Material No 4(liquid). And Study B also have similar analysis results. For further quantitative information the one way homogeneity tests for Type are performed.

Bartlett's Test for Homogeneity of Residual Variance

Source DF Chi-Square Pr > ChiSq

Type 1 10.1829 0.0014

The Bartlett's one way tests for ϵ_c is significant at 0.1% level. The standard deviations for the Liquid and Solid are 0.067 and 0.1. Those give us some evidences of heterogenous variance structure for different Type.

5.Homogeneity Test of conditional residuals ϵ_c against **Material**

Bartlett's Test for Homogeneity of Residual Variance

Source DF Chi-Square Pr > ChiSq

Material 4 36.8695 <.0001 Bartlett's one way test for ϵ_c is significant at 1% level. Those give us strong evidences of heterogenous variance structure for different Materials. This point is consistent to analysis in DATA A.

3.3.3 RSD of Homogeneous Variance Model A3

Covariance Parameter Estimates by A3:

Parameter Estimate

Lab 0.002616

Residual 0.007450

Table 3.6: RSD for Homogeneous Model A3

$RSD_r\% = RSD_{Residual}\%$	8.6%
$RSD_R\% = RSD_{Lab}\% + RSD_{Residual}\%$	10.03%

The RSD_r of A3 is almost the same with RSD_r of Homogeneous Model 3.4(9) and Heterogeneous Model 3.12(8.7). Because of no "Day" effect the variance RSD_R of A3(10.03) in Study A is smaller than RSD_R of models in Study B(11.83).

3.3.4 Heterogeneous Variance Model for Type of Study A

Heterogeneous Variance Components Specification and Test

The reduction table above is rather special. The full model for heterogeneous variance for type is model AT1 which has 6 variance components. The test $G_{AT1 \rightarrow AT2} = 0.1$ is not significant.

3.3 Homogeneous Variance Model and Heterogeneous Variance Model for Type of Study A77

Table 3.7: Heterogeneous Variance Model for Type Comparison Table of Study A

Model	-2le	G-value	df	P-value
AT1: Lab(2),Res(2),Lab:Mat(2) are diff for Type	-231.6	$G_{AT1 \rightarrow AT2} = 0.1$	1	0.75183
AT2: Lab(2),Res(2) are diff for Type,Lab:Mat(1) is single	-231.5		0	0
AT3: Res(2),Lab:Mat(2) is diff for Type,Lab(1) is single	-237.7	$G_{AT3 \rightarrow AT6} = 1.2$	2	0.54881
AT4: Lab:Mat(2) Lab(2) is diff for Type,Res(1) is single	223.7			
AT5: Res(2)l is diff for Type,Lab:Mat(1) Lab(1) is single	237.3	$G_{AT5 \rightarrow AT6} = 0.4$	1	0.52709
AT6: Res(2) is diff for Type,Lab(1) is single	236.5			
AT7: Res(2) ,Lab(2) is diff for Type	230.2			

The special point starts from AT3, which has 5 variance components, but highest goodness-of-fit value 237.7. The covariance matrix of AT3 was printed out. However, everything seems right. There exist same problems for AT5 and AT6, which have fewer variance components but higher goodness-of-fit values than AT1.

Several methods were adopted to solve the problems: 1. change the estimate method from REML to ML. After all those 7 models were calculated again, the goodness-of-fit values changed a little, but AT3, AT5 and AT6 were still have higher values. 2. The covariance matrix is printed out for each model to check the structures and calculate the variance number. So far there was no error of the covariance matrix V.

So far as I know this is the first time this kind of problems appear when proc mixed is running. At present we just keep this problem and go on analysis with AT6, which is the most simplified model with high goodness-of-fit values. The reduction step is in the table. In model AT6, different types corresponds to two different error variances. Model AT6:

$$y_i = \mu + Material_i + b(Lab_i) + \varepsilon_i \quad (3.12)$$

$b(Lab_i) \sim N(0, \sigma_b^2)$, if $i \in Group(Solid)$: $\varepsilon_i \sim N(0, \sigma_s^2)$;

if $i \in Group(Liquid)$: $\varepsilon_i \sim N(0, \sigma_l^2)$;

i=1:132.

Comparing Heterogeneous Model AT6 with Homogeneous Model A3:

Model AT6 have 1 more parameters than model A3, twice the change in log likelihood is

referred to a χ_1^2 distribution. The Null hypothesis of the test is :

$$H_0 : \sigma_s^2 = \sigma_l^2$$

The Likelihood Ratio Test between those two models with degree of freedom 1:

$$G_{AT6 \rightarrow A3} = 2le^{AT1} - 2le^{AT1} = 236.5 - 226.7 = 9.8 \rightarrow P_{\chi_1} = 0.001745119$$

The test is significant at 0.5% level. Model AT6 does fit better than model A3.

3.3.5 RSD of Heterogeneous Variance Model AT6

Covariance Parameter Estimates by AT6:

Parameter Group Estimate

Lab 0.002323

Residual Type Liquid 0.004921

Residual Type Solid 0.01149

Table 3.8: RSD of Heterogeneous Variance Model AT6

RSD	Solid	Liquid
$RSD_r\% = RSD_{Residual}\%$	10.7	7%
$RSD_R\%$	11.75%	8.5%

In this RSD table, the RSD values of Liquid materials are much smaller than the Solid materials.

3.4 Approximate Confidence Intervals on RSD

As denoted before, all the RSD_r and RSD_R of different kinds of models have been calculated by combination of the standard squares of variance components. RSD_r and RSD_R are important precision characteristics in inter-laboratory valid study. The confidence Intervals on the RSD_r and RSD_R could also be calculated by variance components inference.

The existing method for estimating the confidence intervals on variance components, "Satterthwaites", is on the base of the mean squares in the analysis of variance, which could be supplied by EM(Expected Mean Squares Estimation) method. However, in the beginning of the thesis REML method is adopted as estimation method, which could not return the mean squares. For calculate the confidence intervals I just adopted Expected Mean Squares Estimation to process the models again, which estimation values are same with the results returned by REML.

It is easy to find the exact confidence interval on the function of variance components which are the expected values of one mean squares, for instance, the error mean square.[9] In our case $RSD_r^2 = \sigma^2 = E(MS_E)$. The quantity

$$f_E MS_E / \sigma^2 = f_E \hat{\sigma}^2 / \sigma^2$$

has a chi-square distribution. The $100(1 - \alpha)$ percent confidence interval of σ^2 :

$$\frac{f_E MS_E}{\chi_{\alpha/2, f_E}^2} \leq \sigma^2 \leq \frac{f_E MS_E}{\chi_{1-\alpha/2, f_E}^2} \quad (3.13)$$

, in which $\sigma^2 = RSD_r^2$

For variance components which could be not expressed as expected value of a single mean square, it is impossible to find exact confidence intervals. But we can use Satterthwaites approximation to construct an approximate confidence interval for variance components that are estimated by a linear combination of mean squares: Satterthwaites approximation, says that if

$$\hat{\sigma}_0 = \sum_{i=1}^k c_i MS_i$$

then $f \times \frac{\hat{\sigma}_0}{\sigma_0}$ is approximately a chi-squared variable with $df = f$ where

$$f = \frac{(\sum_{i=1}^k c_i MS_i)^2}{\sum_{i=1}^k c_i MS_i^2 / f_i}$$

and

$$P\{\chi_{1-\alpha/2,f}^2 \leq f \times \frac{\hat{\sigma}_0}{\sigma_0} \leq \chi_{\alpha/2,f}^2\} = 1 - \alpha$$

$$P\left\{\frac{f \times \hat{\sigma}_0}{\chi_{\alpha/2,f}^2} \leq \sigma_0 \leq \frac{f \times \hat{\sigma}_0}{\chi_{1-\alpha/2,f}^2}\right\} = 1 - \alpha$$

Therefore an approximate $100(1-\alpha)$ percent confidence interval on σ_0 is

$$\frac{f \times \hat{\sigma}_0}{\chi_{\alpha/2,f}^2} \leq \sigma_0 \leq \frac{f \times \hat{\sigma}_0}{\chi_{1-\alpha/2,f}^2} \quad (3.14)$$

3.4.1 Homogeneous Variance Model 3.4: Approximate Confidence Intervals on RSD

The ANOVA table with mean squares returned by SAS :

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Intercept	1	28154	28154	Var(Residual) + 5.8657 Var(LAB*Material) + 41.827 Var(LAB) + 11.805 Var(Day(LAB)) + Q(Intercept, Material)	-0.0535 MS(Day(LAB)) + 0.982 MS(LAB) + 0.0099 MS(LAB*Material) + 0.0616 MS(Residual)	12.882	219424	<.0001
Material	7	76.674132	10.953447	Var(Residual) + 5.4577 Var(LAB*Material) + Q(Material)	0.9367 MS(LAB*Material) + 0.0633 MS(Residual)	83.926	506.32	<.0001
Day(LAB)	40	0.975473	0.024387	Var(Residual) + 10.35 Var(Day(LAB))	MS(Residual)	476	3.01	<.0001
LAB	13	1.706221	0.131248	Var(Residual) + 5.9143 Var(LAB*Material) + 42.592 Var(LAB) + 12.585 Var(Day(LAB))	1.216 MS(Day(LAB)) + 1.0151 MS(LAB*Material) - 1.2311 MS(Residual)	63.016	3.08	0.0014
LAB*Material	80	1.803773	0.022547	Var(Residual) + 5.8263 Var(LAB*Material)	MS(Residual)	476	2.78	<.0001
Residual	476	3.860341	0.008110	Var(Residual)				

Covariance Parameter Estimates

Cov Parm	Estimate
Day(LAB)	0.001573
LAB	0.002082
LAB*Material	0.002478

Figure 3.29: The ANOVA table of model 3.4 with mean squares estimations

As mentioned in Chapter One, after the outliers deleted, the design is unbalanced. Unsurprisingly in mean squares the coefficients of variance components are not integers. The variance components could be expressed as the linear combination of Means square:

$$\hat{\sigma}^2 = MS(Residual) = RSD_r^2$$

$$\hat{\sigma}_{Day}^2 = (MS(Day) - MS(Residual))/10.35$$

$$\hat{\sigma}_{Lab*Material}^2 = (MS(Lab * Material) - MS(Residual))/5.8263$$

$$\hat{\sigma}_{Lab}^2 = (MS(Lab) - (1.216MS(Day) + 1.0151MS(LAB*Material) - 1.2311MS(Residual)))/42.592$$

Then RSD_R^2 could be denoted as the linear combination of mean squares:

$$RSD_R^2\% = \hat{\sigma}_{Lab}^2 + \hat{\sigma}_{Day}^2 + \hat{\sigma}_{Lab*Material}^2 + \hat{\sigma}^2 = 0.02348MS(Lab) + 0.068MS(Day) + 0.1478MS(LAB*Material) + 0.76065MS(Residual)$$

According to formula 3.14, the degree of freedom f:

$$f = \frac{(\sum_{i=1}^k c_i MS_i)^2}{\sum_{i=1}^k c_i^2 MS_i^2 / f_i} = 199.22$$

The $100(1 - 0.05)$ percent confidence interval of $RSD_R^2\%$:

$$\frac{f * RSD_R^2\%}{\chi_{0.975, f}^2} \leq (RSD_R^2\%)^2 \leq \frac{f * RSD_R^2\%}{\chi_{0.025, f}^2}$$

$$\frac{199 * (11.83\%)^2}{\chi_{0.975, 199}^2} \leq (RSD_R^2\%)^2 \leq \frac{199 * (11.83\%)^2}{\chi_{0.025, 199}^2}$$

$$\frac{199 * 0.014011}{240} \leq (RSD_R^2\%)^2 \leq \frac{199 * 0.014011}{161.8}$$

$$0.011617 \leq (RSD_R^2\%)^2 \leq 0.017232$$

$$10.78 \leq RSD_R \leq 13.13$$

The $100(1 - 0.05)$ percent confidence interval of $RSD_r^{2\%}$:

$$\frac{f_E MS_E}{\chi_{1-\alpha/2, f_E}^2} \leq (RSD_r^{2\%})^2 \leq \frac{f_E MS_E}{\chi_{\alpha/2, f_E}^2}$$

$$\frac{476 * 0.00811}{\chi_{0.975, 476}^2} \leq (RSD_r^{2\%})^2 \leq \frac{476 * 0.00811}{\chi_{0.025, 476}^2}$$

$$\frac{476 * 0.00811}{538.4} \leq (RSD_r^{2\%})^2 \leq \frac{476 * 0.00811}{417.45}$$

$$0.0071700 \leq (RSD_r^{2\%})^2 \leq 0.00924748$$

$$8.4676 \leq RSD_r \leq 9.616$$

3.4.2 Homogeneous Variance Model A3 of Study A: Approximate Confidence Intervals on RSD

The ANOVA table with mean squares returned by SAS :

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Intercept	1	6160.279218	6160.279218	Var (Residual) + 8.7887 Var(Lab) + Q (Intercept, Material)	MS(Lab) + 0.0635 MS (Residual)	13.433	213396	<.0001
Material	4	7.541325	1.885331	Var (Residual) + Q (Material)	MS (Residual)	114	254.84	<.0001
Lab	13	0.394205	0.030323	Var (Residual) + 9.3846 Var(Lab)	MS (Residual)	114	4.10	<.0001
Residual	114	0.843394	0.007398	Var (Residual)

Covariance Parameter Estimates

Cov Parm	Estimate
Lab	0.002443
Residual	0.007398

Figure 3.30: The ANOVA table of model A3 with mean squares estimations

The variance components could be expressed as the linear combination of Means square:

$$\hat{\sigma}^2 = MS(Residual) = RSD_r^2$$

$\hat{\sigma}_{Lab}^2 = (MS(Lab) - MS(Residual))/9.3846$ The RSD value estimated by expected mean square type3 method are almost same with the RSD estimated by REML:

Table 3.9: RSD for Homogeneous Model A3

$RSD_r\% = RSD_{Residual}\%$	8.6%
$RSD_R\%$	9.92%

Then RSD_R^2 could be denoted as the linear combination of mean squares:

$$RSD_R^2\% = \hat{\sigma}_{Lab}^2 + \hat{\sigma}^2 = 0.10655755MS(Lab) + 0.89344245MS(Residual)$$

According to formula 3.14, the degree of freedom f:

$$f = \frac{(\sum_{i=1}^k c_i MS_i)^2}{\sum_{i=1}^k c_i^2 MS_i^2 / f_i} = 9.68419709277134e - 005 / 1.1863274083841e - 006 = 81.632$$

confidence interval of $RSD_R^2\%$:

$$\frac{f * RSD_R^2\%}{\chi_{0.975, f}^2} \leq (RSD_R^2\%)^2 \leq \frac{f * RSD_R^2\%}{\chi_{0.025, f}^2}$$

$$\frac{81.632 * 0.009841}{\chi_{0.975, 82}^2} \leq (RSD_R^2\%)^2 \leq \frac{81.632 * 0.009841}{\chi_{0.025, 82}^2}$$

$$\frac{0.803340512}{108.94} \leq (RSD_R^2\%)^2 \leq \frac{0.803340512}{58.85}$$

$$0.0073741556 \leq (RSD_R^2\%)^2 \leq 0.0136506459$$

$$8.59 \leq RSD_R \leq 11.68$$

The 100(1 - 0.05) percent confidence interval of $RSD_r^2\%$:

$$\frac{f_E MS_E}{\chi_{1-\alpha/2, f_E}^2} \leq (RSD_r^{2\%})^2 \leq \frac{f_E MS_E}{\chi_{\alpha/2, f_E}^2}$$

$$\frac{114 * 0.007398}{\chi_{0.975, 114}^2} \leq (RSD_r^{2\%})^2 \leq \frac{114 * 0.007398}{\chi_{0.025, 114}^2}$$

$$\frac{114 * 0.007398}{145.45} \leq (RSD_r^{2\%})^2 \leq \frac{114 * 0.007398}{86.35}$$

$$0.005798 \leq (RSD_r^{2\%})^2 \leq 0.0097669$$

$$7.61 \leq RSD_r \leq 9.88$$

3.4.3 Comparison between Approximate Confidence Interval on RSD for Homogeneous Variance Model of Study A and Study B

Table 3.10: Approximate Confidence Interval on RSD for Homogeneous Models

	Study A	Study B
RSD_r	8.6	9
95% Confidence Interval for RSD_r	$7.61 \leq RSD_r \leq 9.88$	$8.4676 \leq RSD_r \leq 9.616$
RSD_R	9.92	11.84
95% Approximate Confidence Interval for RSD_R	$8.59 \leq RSD_R \leq 11.68$	$10.78 \leq RSD_R \leq 13.13$

From the plots above it seems the confidence intervals of Study A are wider than the intervals of Study B for both RSD_R and RSD_r . It is because Study B have much more observations(618) than Study A observations(132). Therefore Study B have smaller variances of the estimated parameters, including the variance components, which lead to more narrow confidence intervals.

According to the original definition of the confidence interval, the RSD intervals here mean if many other studies with same design are conducted , 95% of RSD values from those studies

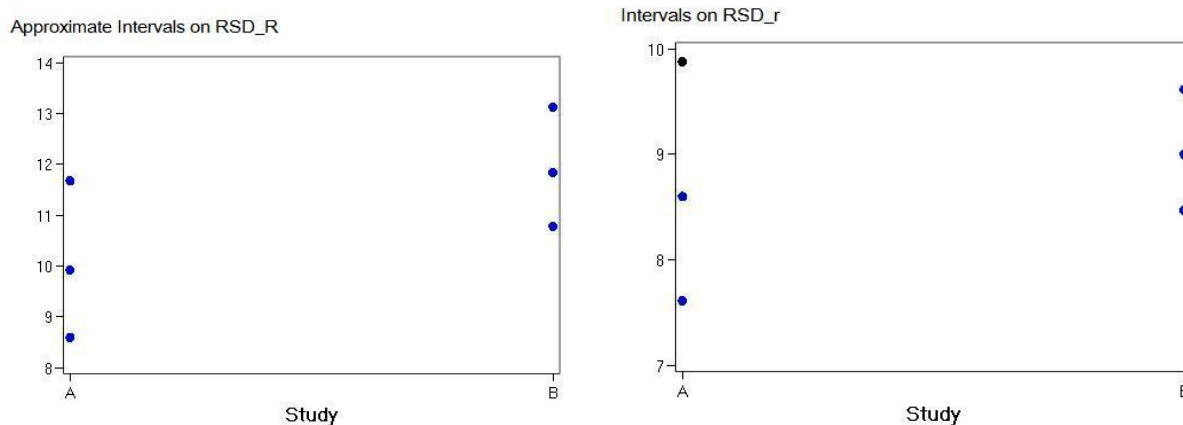


Figure 3.31: The 95% confidence interval of RSD_R and RSD_r vs study A and Study B. For each column in the plots, there are three points on the line: the upper bound point, lower bound point and the RSD value.

should be within the intervals.

Because $f \times \frac{\hat{\sigma}_0}{\sigma_0}$ has an approximately chi-squared distribution, all those confidence intervals are not symmetrical around the RSD values, and all those RSD values are closer to the lower bounds of the confidence intervals than the upper bounds.

3.5 Heterogeneous Variance Model for Material

3.5.1 Heterogeneous Variance Model for Material of Study B

In model 3.4 model residual checking, (see Residual vs.Material) we could see the evidence of heteroscedasticity between Materials. The one way homogeneity test for Material is also significant. In this section the distinct variance components for Material will be estimated.

Firstly Heterogeneous Variance Models for Material reduction steps are as follows:

Table 3.11: Model Comparison Table of Heterogeneous Variance Model for Material of Study B

Model	-2le	G-value	df	P-value
MB1: Lab,Day,Residual are different for Material	-1083.3	$G_{MB1 \rightarrow MB2} = 0$	7	1
MB2: Lab,Residual are diff for Material, Day is same	-1083.3	$G_{MB2 \rightarrow MB3} = 6$	7	0.53975
MB3: Residual is diff for Material,Lab and Day are same	-1077	$G_{MB3 \rightarrow MB4} = 2$	1	0.15730
MB4: Residual is diff for Material,Lab is same. No Day !	-1075	$G_{MB4 \rightarrow MB5} = 125$	1	0
MB5: Residual is diff for Material,No Lab , No Day!	-950			

Table above shows the series of models reduction steps from M1 to M5.

Formula of MB1:

$$y_i = \mu + Material_i + a(DAY_i)(b(Lab_i)) + b(Lab_i) + \varepsilon_i \quad (3.15)$$

if $y_i \in Group(Material_j)$: $b(Lab_j) \sim N(0, \sigma_{bj}^2)$, $a(DAY_j) \sim N(0, \sigma_{aj}^2)$, $\varepsilon_j \sim N(0, \sigma_j^2)$, $j=1:8; i=1:618$. There are 8 kinds of Materials. While in MB1 each kind of materials have its own 3 variance components: $\sigma_{lab}^2, \sigma_{day}^2, \sigma_{residual}^2$. Model MB1 involves 24 independent variance components.

As denoted in the Model Comparison Table, in MB2 all kinds material have same σ_{day}^2 , but different σ_{lab}^2 and $\sigma_{residual}^2$. Hence MB2 only have 17 independent variance components.

The log-likelihood ratio test : $G_{MB1 \rightarrow MB2}$ is with degree of freedom:7.

The reduction test from MB2 to MB3 is similar. Model MB3 have 10 independent variance components.

MB4 is sub-model of model MB3, without σ_{day}^2 . Hence the degree of freedom of $G_{MB3 \rightarrow MB4}$ is 1.

MB5 is sub-model of model MB4, without σ_{Lab}^2 . Similarly, the degree of freedom of $G_{MB4 \rightarrow MB5}$ is also 1.

The reduction tests are all not significant except for $G_{MB4 \rightarrow MB5}$. Therefore MB4 is the optimal model temporarily.

The interaction term (Lab:Material) is significant in Model 3.4. I have also tried to put the

interaction term (Lab:Material) in the model MB1. But the computation complexity is too high and computer could not return the result. Hence I just insert the interaction term in Model MB4. The new model with interaction term is denoted as MB6. MB4 is a sub-model of MB6. The log-likelihood ratio test : $G_{MB6 \rightarrow MB4} = 1092.7 - 1075 = 17.7$ is with degree of freedom:1, p value :0.000025863. The new MB6 is significantly better than MB4 at 0.001% level. Covariance Parameter Estimates of MB6:

Cov Parm Group Estimate
 LAB 0.002281
 LAB*Material 0.001485
 Residual Material 1 0.01149
 Residual Material 2 0.01009
 Residual Material 3 0.01014
 Residual Material 4 0.01292
 Residual Material 5 0.004728
 Residual Material 6 0.000943
 Residual Material 7 0.01600
 Residual Material 8 0.008775

Table 3.12: RSD for Heterogeneous Variance Model MB6: for Material of Study B

Material:	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8
$RSD_r = RSD_{Residual}$	10.72	10.04	10.07	11.37	6.88	3.07	12.65	9.37
RSD_R	12.35	11.77	11.79	12.92	9.22	6.86	14.06	11.20

3.5.2 Heterogeneous Variance Model for Material of Study A

Table 3.13: Model Comparison Table of Heterogeneous Variance Model for Material of Study A

Model	-2le	G-value	df	P-value
MA1: Lab,Residual are different for Material	-260	$G_{MA1 \rightarrow MA2} = 0$	4	1
MA2: Residual is diff for Material, Lab is same	-267	$G_{MA2 \rightarrow MA3} = 22$	1	.000002727
MA3: Residual is diff for Material,No Lab	-245			
MA2: Residual is diff for Material, Lab is same	-267	$G_{MA2 \rightarrow A3} = 40.7$	4	3.101E-8
A3: homogenous model :Residual and Lab are same	-226.7			

MA2 is temporarily the optimal model in the reduction table. On the basis of MA2, a new model **MA4** with interaction term (Lab:Material) is also processed. The log-likelihood ratio test : $G_{MA4 \rightarrow MA2} = 270.27 - 267.23 = 3.07$ is with degree of freedom:1, p value :0.08. The new MA4 is significantly better than MA2 at 10% level. By those two reduction processes, Study A and Study B resulted in similar final optimal model: Residual is heterogeneous for Material, Lab and interaction term (Lab:Material) are homogeneous between Material.

parameter estimated by Model MA4:

Covariance Parameter Estimates

Cov Parm Group Estimate

Lab 0.001605

Lab*Material 0.000878

Residual Material 1 0.01857

Residual Material 2 0.003447

Residual Material 3 0.002322

Residual Material 4 0.008119

Residual Material 5 0.001033

Table 3.14: RSD for Heterogeneous Variance Model MA4: for Material of Study A

Material:	M 1	M 2	M 3	M 4	M 5
$RSD_r = RSD_{Residual}$	13.63	5.87	4.82	9.01	3.21
RSD_R	14.51	7.7	6.93	10.3	5.93

3.5.3 Brief Summary

In this section the heterogeneous variance models for material for Study A and Study B are processed. The RSD_r vary from 3 to 13 in both studies. RSD_R , which are a little larger than the corresponding RSD_r , vary from 5.93 to 14.5. Apparently, most of the variability is attribute to difference between duplicates within Materials.

The goodness of fit values of model MB6 and MA4 are much larger than the values of model 3.4 and model A3, which are the homogeneous variance models of study A and study B.

Table 3.15: Model comparison of Study A and Study B

Model :	goodness of fit	variance components number
homo model of study A: model A3	-226.7	1(Lab)+1(Res)=2
hetero model of study A: model MA4	-270.27	1(Lab)+5(Res)+1(Lab:Material)=7
homo model of study B: model 3.4	-1011.4	1(Lab)+1(Day)+1(Lab:Material)+1(Res)=4
hetero model of study B: model MB6	-1092	1(Lab)+8(Res)+1(Lab:Material)=10
hetero model of study B: model MB7	-1094.0	1(Lab)+8(Res)+1(Day)+1(Lab:Material)=11

The table shows that A3 could be regarded as the sub-model of MA4. The log-likelihood ratio test : $G_{MA4 \rightarrow A3} = 44$ is significant at 0.0001% level with degree of freedom:5, P-value: 2.3162E-8. Model 3.4 is not the exact sub-model of MB6 which does not have the random effect Day. We could use model MB7 instead of MB6, which contains the random effect: Day. Model 3.4 is the sub-model of MB7. $G_{MA7 \rightarrow 3.4} = 83$ significant at 0.00001% level with p-value:3.3307E-15. So far the between-Material heteroscedasticity is conclusive. The heterogeneous variance model for Material fit much better than the models before.

The interaction term (Lab:Material) in both MB6 and MA4 are still significant, especially in MB4. It means for some Material, the Lab variances may be different. The test $G_{MB2 \rightarrow MB3}$ is not significant, it seems that setting distinct lab variances for each kind material could not improve model fitting a lot. But the difference between lab variances of different materials may still exist. Moreover, the assumptions of model MB7 are not checked. If the assumptions are violated, the model reduction steps are not that reliable. The heteroscedasticity between labs labs still need to be considered.

3.6 Heterogeneous Variance Model for Lab of Study B

Table 3.16: Model Comparison Table of Heterogeneous Variance Model for Lab of Study B

Model	-2le	G-value
LB1: Day,Residual are different for Lab	-1000	$G_{MB1 \rightarrow MB2} = 0$
LB2:Residual are diff for Material, Day is same , with interaction Lab:Material	-987	$G_{MB2 \rightarrow MB3} = 6$

Now we may ponder the thought of between-lab heteroscedasticity. Several heterogeneous variance models for lab are processed. The goodness-of-fit values are much smaller than heterogeneous variance models for Material. Although heteroscedasticity among labs may exist, it is not the main variability of the measurements.

For combining the between-lab and between-material heteroscedasticity together, new model LB3 is processed on the basis of model MB2: 8 distinct residuals and 8 distinct lab variance components are set for 8 kinds of materials, 14 distinct day variance components are set for 14 labs. Several reduced models of LB3: LB4, LB5, are also presented:

Table 3.17: Model Comparison Table of Heterogeneous Variance Model for Material and Lab of Study B

model variance components specification	-2le	G-value	df	P-value
LB3: Lab(8),Residual(8) are diff for Material,Day(14) are diff for Lab	-1155	$G_{LB3 \rightarrow LB4} = 31$	7	0.00006
LB4: Residual(8) are diff for Material,Day(14) are diff for Lab, Lab(1) is single	-1124			
LB5: Residual(8) are diff for Material,Day(14) are diff for Lab,Lab:Material(1) and Lab(1) are single	-1151	$G_{LB5 \rightarrow LB4} = 27$	1	0.0000002
LB3: Lab(8),Residual(8) are diff for Material,Day(14) are diff for Lab	-1155	$G_{LB3 \rightarrow MB2} = 71$	13	5.2508E-10
MB2: Lab(8),Residual(8) are diff for Material,Day(1) is single	-1083.3			

The table above shows that LB5 and LB3 are both significantly better than LB4. LB3 and LB5 are different at the random effect lab: LB3 has 8 distinct lab variance components for 8 different materials, while LB5 has one single lab variance component and interaction term (Lab:Material). The goodness of fit values of LB3 and LB5 are similar. LB5 has 6 fewer parameters than LB3 and approaches almost the same extent goodness of fit value. According to the usual model selection rules, LB5 is a better model. But LB3 contains the more information about the lab variance value within material. The test $G_{LB3 \rightarrow MB2} = 71$ is very significant with p-value 5.2508E-10, therefore the null hypothesis of this test:

$$\sigma_{Day1}^2 = \sigma_{Day2}^2 = \sigma_{Day3}^2 = \dots \sigma_{Day14}^2$$

, where 14 σ_{Day}^2 are corresponding to 14 labs, is rejected, which proves the between-lab heteroscedasticity.

Covariance Parameter Estimates of LB3:

Cov Parm Group Estimate

Day LAB 1 1.45E-36

Day LAB 2 0

Day LAB 3 0.003098

Day LAB 4 0.003217

Day LAB 5 0.004211

Day LAB 6 0.002352

Day LAB 7 0.000068

Day LAB 8 0.004671

Day LAB 9 0.003890

Day LAB 10 0.002399

Day LAB 11 0.001724

Day LAB 12 0.000272

Day LAB 13 0.000277

Day LAB 14 0.000517

LAB Material 1 0.000637

LAB Material 2 0.004462

LAB Material 3 0

LAB Material 4 0.006125

LAB Material 5 0.001290

LAB Material 6 0.000650

LAB Material 7 0.006979

LAB Material 8 0.001300

Residual Material 1 0.01120
 Residual Material 2 0.008409
 Residual Material 3 0.009952
 Residual Material 4 0.01022
 Residual Material 5 0.003851
 Residual Material 6 0.000472
 Residual Material 7 0.01258
 Residual Material 8 0.008056

Table 3.18: RSD for Heterogeneous Variance Model LB3 and MB6 of Study B

Material:	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8
MB6: $RSD_r = RSD_{Residual}$	10.72	10.04	10.07	11.37	6.88	3.07	12.65	9.37
LB3: $RSD_r = RSD_{Residual}$	10.58	9.17	9.97	10.11	6.02	2.17	11.21	8.96
LB3: RSD_{LAB}	2.52	6.68	0	7.83	3.59	2.55	8.35	3.6

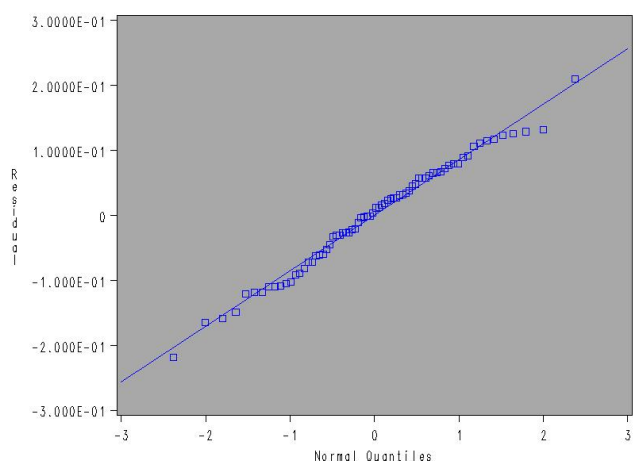
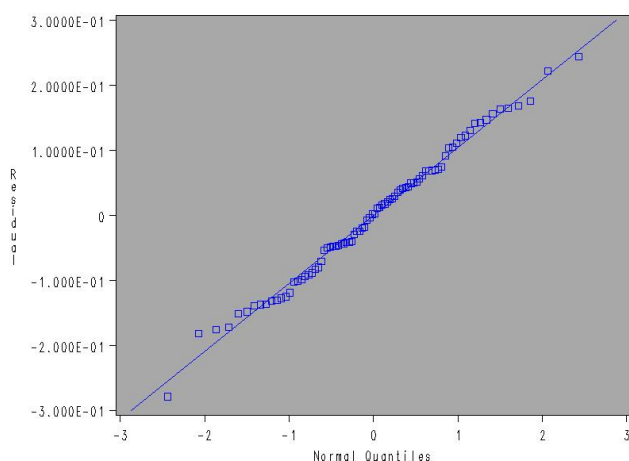
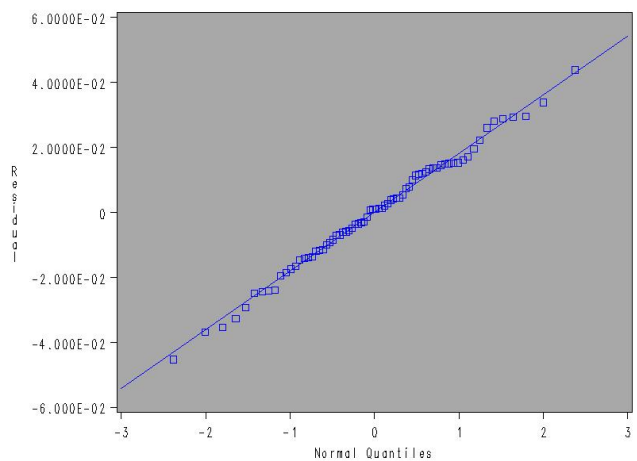
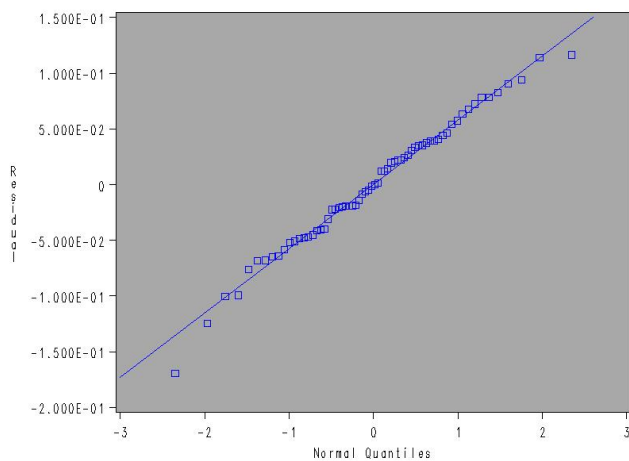
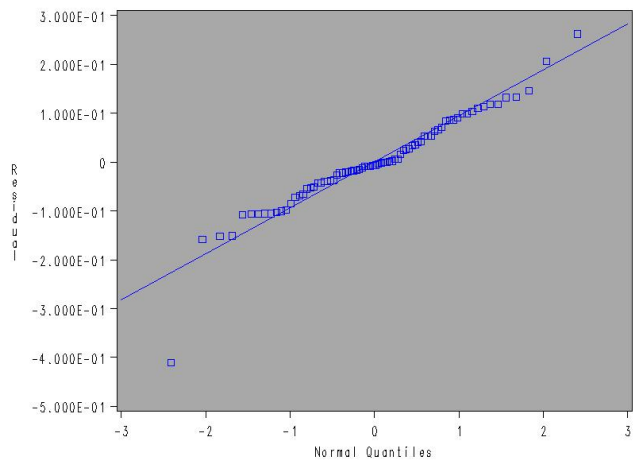
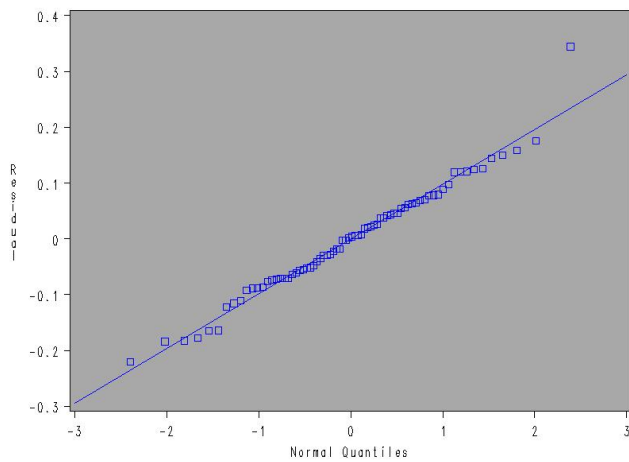
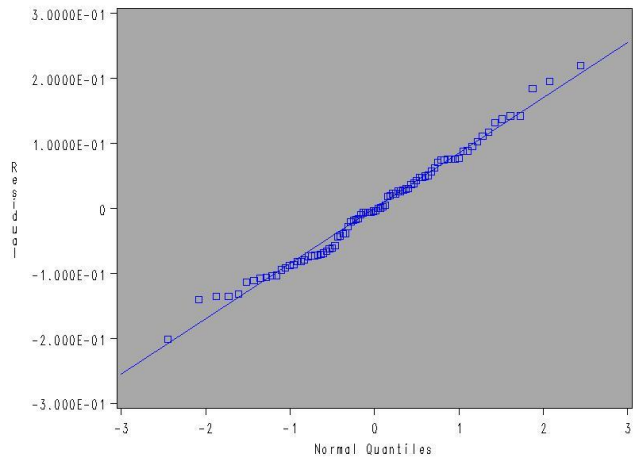
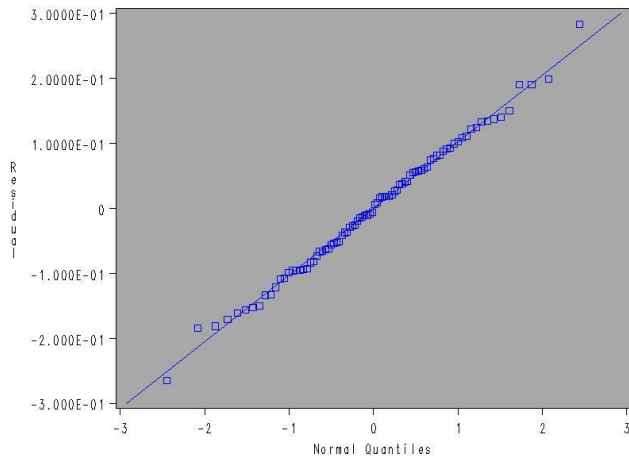
The RSD_r of LB3 and MB6 are listed in the table. All the RSD_r of LB3 are smaller than the RSD_r of MB6. In LB3 more viability of the data are explained by random effects, Day and Lab.

3.6.1 Residual Analysis

One basic assumption of model LB3 is the homoscedasticity and normality of error within Material. The normality test and homogeneity one way test for residuals are processed in all material blocks.

normality test

The Q-Q plots of residuals for 8 materials are as blow, which seem fit normal distributions well. The normal tests, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling for those eight sets of residuals are all not significant.



homogeneity test

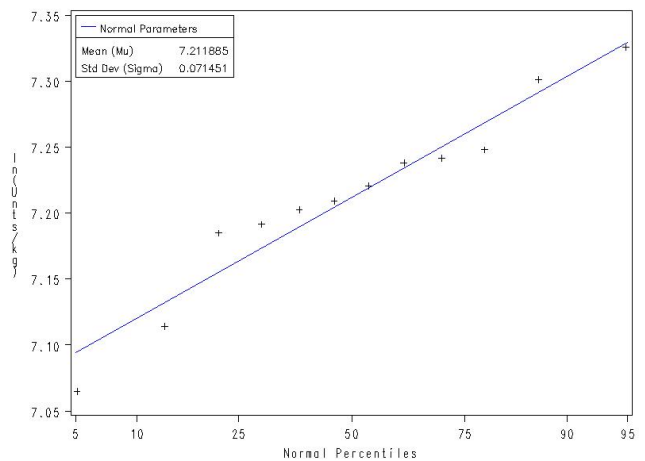
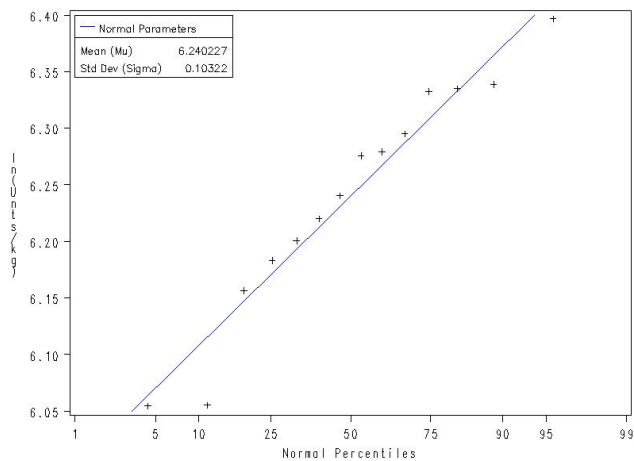
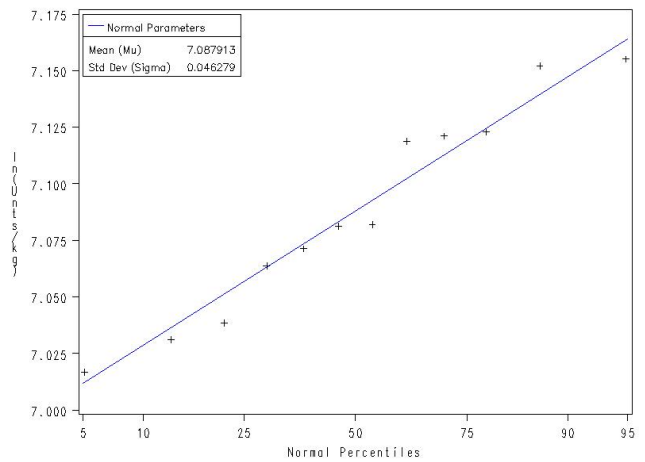
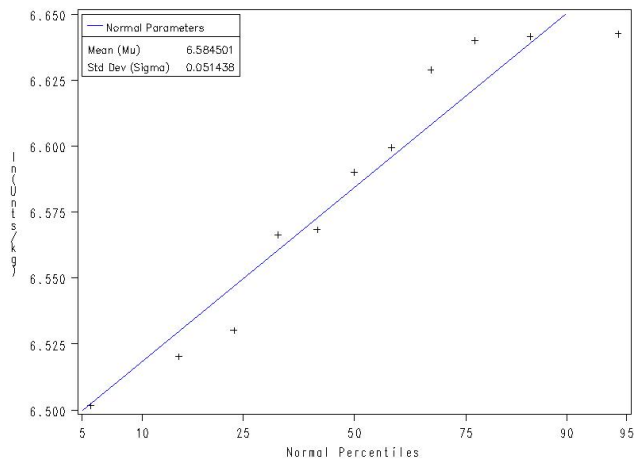
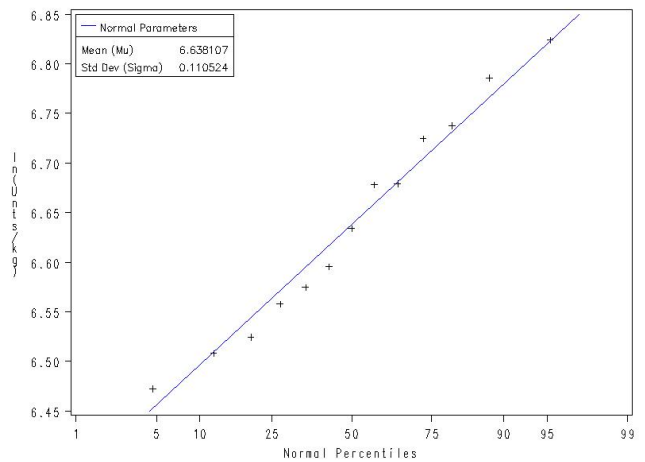
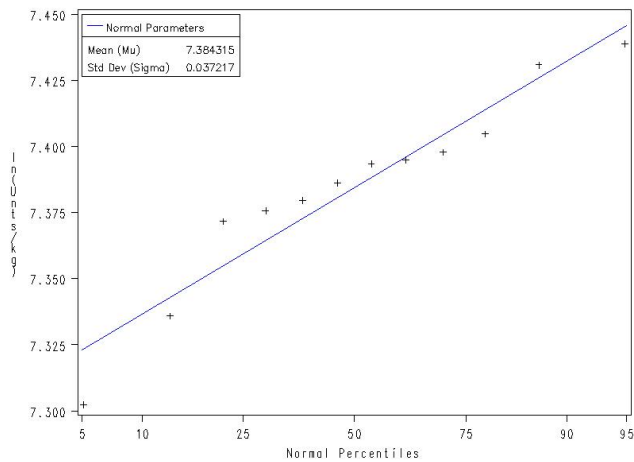
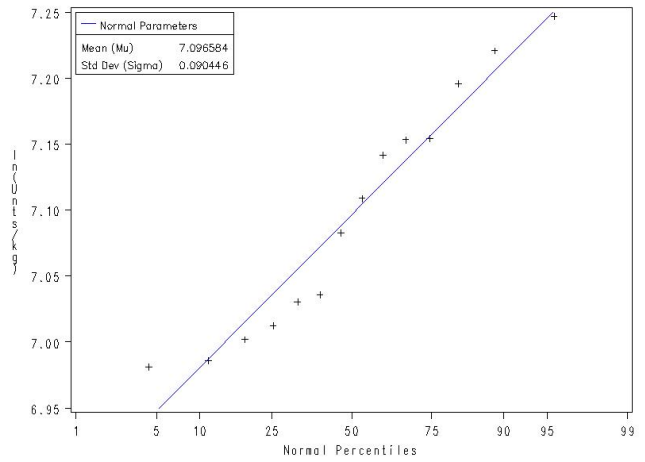
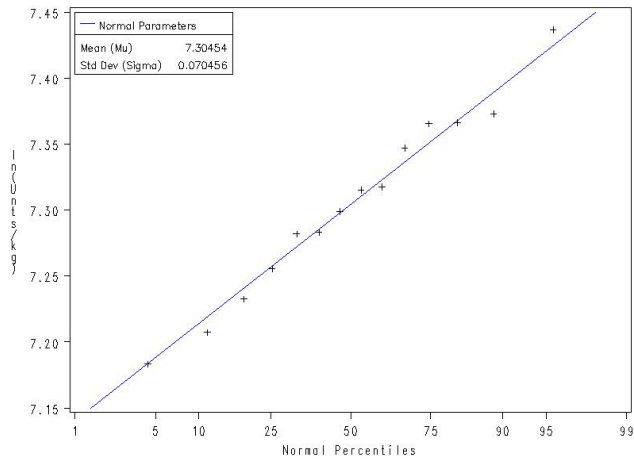
One way homogeneity tests for those 8 sets of residuals of materials are implemented, 3 tests of the are significant at 0.05% level. The test results returned by SAS are in the Appendix.

Check for random effects normality

The assumption of random effects of LB3 is : if $y_i \in Group(Material_j)$: $b(Lab_j) \sim N(0, \sigma_{b_j}^2)$ $j=1:8; i=1:618$ The random effects Labs follow normal distribution within Material. There are 8 Q-Q plots for 8 materials. While the random effect Day in this model is hard to print.[11] It is impossible to plot random effects as residuals. I use probability plots of averages of the data corresponding to the random factors to check whether it is severe lack of normality. However, this method has some limitations:

- 1.Firstly, "it is only a rough approximate approach as the averages consist of contributions from several random effects". [11]The normal effects confounded with each other . It is hard to distinguish which effect leads the non-normal.
- 2.The raw average will be effected by the fixed part of the model.
- 3.The approach only could apply to the factor which number of levels is not too small.
- 4.If non-normal happens,we have no method to deal with it!

In model LB3, Day only have three levels. Therefore it is impossible to investigate it by this raw average method. I only present the plots of raw average for eight random effect day within eight materials. There are no indication of severe lack of normality:



3.6.2 Brief Summary

In the the **homogeneous** variance model section, the residual analysis shows that there exist heteroscedasticity both among Labs and Materials. SAS Mixed Procedure have "repeated" statement with group option to specify the residual variance structure, which only allow for one factor to define the distinct residual components. Material is the main source of heteroscedasticity of residual. When the group option after "repeated" statement is set as Material, the goodness of fit value of the model is much higher than the values in the case of "Group=Lab". (MB6 1092 \gg LB2 987)

For the limitation of the SAS mixed model specification of covariance structure, only Material is set as group option factor after "repeated" statement. In Study B the random effect Day could be set as 14 distinct variances for labs to explain the difference of evaluation precision of different labs. LB3 is the case that in distinct residual and lab variances is set for different Materials, and distinct Day variance components are set for different labs. LB3 is the best model with goodness of fit value 1155 so far.

The residual checking of Model LB3 shows the residual assumption are satisfied by and large. The characteristics got from LB3 are reliable, although model LB3 is not perfect.

The estimate of all the day variance components of different labs are listed in this section, which vary from 0 to 0.004671, corresponding to RSD_{day} : 0 to 6.8.

SAS output of LB5 is in the attachment.

Chapter 4

Conclusion

4.1 Result and Conclusion

In the beginning of the analysis, log-transformation of the data is processed. Log-transformation reduces the variance increasing trend with mean values, but does not eliminate the heteroscedasticity.

4.1.1 Homogeneous Variance Model

The homogeneous variance model is on the basis of the assumption that all the observations have same variances. In the residual checking of homogeneous variance model the homogeneity assumption is obviously violated: Residual have obvious patterns with Material and Lab.

The confidence interval of RSD_R and RSD_r could be estimated by Satterthwaites approximation, which requires the Mean Squares of ANOVA table. To get the Mean Squares, method-of-moment variance component estimates are used instead of REML. [2] In SAS procedure Moment Estimate specifications apply only to variance component models with no REPEATED statement and subject option.

The confidence interval of RSD_R and RSD_r in the homogeneous variance model could not predict the future evaluation RSD because the assumptions of the models are violated obviously. However, in case of inter-laboratory of one kind of material with ANOVA table without any repeated statement, the RSD confidence interval Satterthwaites approximation could predict the percentage interval of RSD. The aim of RSD confidence interval section is to implement

the Satterthwaites approximation method.

4.1.2 Heterogeneous Variance Model

In the early part of this study we believe that solid materials may have larger variance than the Liquid Material. One basic assumption of the heterogeneous variance model for **Type** (Solid/Liquid) is the within Type homoscedasticity. The assumption checking for residual shows this assumption is violated: Bartlett's one way homogeneity test by Material of Residual within Type is significant.

The variance estimated from Study B shows solid material have smaller variance than liquid material. But in Study A has the contrary result with Study B. Therefore, according to Study A and Study B, the guess that solid material have larger variance than the liquid material is not true.

Heterogeneous variance model MB6 for Material are processed. And RSD_R and RSD_r for each material are presented. The RSD values vary from 3 to 13. Moreover, considering the between-lab heteroscedasticity, model LB3 is conducted on the basis of MB6 assisted with the distinct Day random effects for each Lab. LB3 is the best fit model obtained so far, which specify the heteroscedasticity between -Lab and heteroscedasticity between-Material at the same time.

Although MB6 does not fit as well as LB3, it is still a useful tool to achieve the general characteristics (RSD) of the study. Because LB3 sets 14 Day variance components for 14 labs, the RSD_R for different materials could not be calculated directly. RSD_R could only be calculate for certain material evaluated by certain lab.

LB5 contains the interaction term (Lab:Material), which means for different material, the lab variance vary from material to material. This point of view also accords with model LB3, which sets 8 lab variance components for different materials.

Both LB5 and LB3 are are optimal models with similar goodness-of-fit values. LB5 uses interaction term plus ones single lab variance instead of 8 distinct lab variances, and therefore is more simplified. While LB3 could supply more details of the Between-Material heteroscedasticity.

4.2 Further Study

1. The test $G_{LB3 \rightarrow MB2} = 71$ is very significant which proves the between-lab heteroscedasticity does exist and 14 distinct variance components of **Day** are significant. Since the precision of labs evaluation capability are different, it is reasonable to guess that the between-lab heteroscedasticity at residual level exists at the same time.

The inter-laboratory Study A and Study B could be regarded as series of distinct inter-laboratory studies for different Materials. In most conventional validation studies, characteristics of each material estimated by different labs are presented. For instance, Study B could be seemed as 8 times of inter-lab validation of 8 different materials. If 8 models are processed for eight kinds of materials separately, the group option following statement "Random" and "repeated" could be set as "group=Lab", the between-lab heteroscedasticity of residual and random effect Day could be revealed at the same time. After 8 times of distinct mixed modelings, a 14×8 RSD table could be constructed. The average RSD values for each labs and each material can be obtained by this table. It could be a comparison with the characters resulted from the models in this thesis.

Besides, If "repeated, group=Lab" is not used in sas mixed proc model for each material, the ANOVA table could be returned. With ANOVA table Satterthwaites approximation could be adapted to estimate the confidence intervals of RSD_R for each material. just as the RSD interval for homogeneous variance model section.

2. When SAS procedure mixed apply REML to estimate covariance matrix, it could return the asymptotic covariance matrix for variance components, and also could requests confidence limits for the covariance parameter estimates. A Satterthwaite approximation is used to construct limits for all parameters.

These limits take the form Refer to Milliken and Johnson (1992) and Burdick and Graybill (1992)[2]. In their thesis it does not talk about how to conduct the approximate confidence interval of linear function of the variance components. So far all the papers I read are based on the MS (mean square) to estimate the approximation confidence interval of linear function of variance components. It will be an very interesting and piratical topic to study how to do that with asymptotic covariance, and the reliability of this kind of approximation.

3. In Linear Mixed Model, the assumptions include linearity of effects, normality, homoscedasticity and independence of variance components. Three types of of disturbances for LLM: conditional error, random effects and marginal error are of interest. However, non-of these

disturbances could be observed directly, and in the estimating process all kinds of residual and variance are confounded.[10] The confounding of residuals reduces the analyst's power to detect violations of modeling assumptions. Some least confounded estimate methods are presented.[10]

The more precisely assumption checking is adopted, the more reliable of characteristics resulted by the model could be. In this thesis I mostly adopt normality and homogeneity checking for the conditional residual. In further study, more diagnostic checking on random effects and marginal residual could be implemented.

4. Group option of repeated and random statement could create large number of covariance parameters and strange covariance pattern. The problem in AT6 is quite singular. All the models converge successfully and supply estimates of all the parameters, but the sub-model have higher goodness-of-fit values. It was discussed with supervisors , but it was still unsolved. The covariance structure was also checked. This may be lead by the ML algorithm and need to be studied seriously later.

Bibliography

- [1] Mixed Effects Models for Longitudinal Data, Encyclopedia of Biostatistics, Peter Armitage & Theodore Colton, WILEY
- [2] SAS help document, The mixed procedure, Copyright 2003 by SAS Institute Inc., Cary, NC, USA. All rights reserved.
- [3] Validation study on a new method for the determination of phytase activity in feed: Results from an interlaboratory study conducted according to the IUPAC harmonised protocol, Gisele Gizzi and Christoph von Holst, European Commission Directorate General Joint Research Centre Institute for Reference Materials and Measurements Food Safety and Quality Unit 8 December 2005
- [4] AOAC Official Method 2000.12, Phytase in feed. Colorimetric enzymatic method first action 2000.
- [5] Data with constant coefficient of variation Gamma distributed data *Søren Højsgaard* and Ulrich Halekoh, Biometry Research Unit Sen and Danish Institute of Agricultural Sciences March 4, 2005
- [6] The application of hierarchical linear modeling (HLM) techniques in commercial aviation research. Beaubien, J. M., Hamman, W. R., Holt, R. W., and Boehm-Davis, D. A. (2001).
- [7] Design and Analysis of Experiments , page 548 ,Douglas C. Montgomery 2001
- [8] Using linear mixed models in longitudinal studies: Application of SAS PROC MIXED.
- [9] Multilevel Diagnostics for Mixed and Hierarchical Linear Models by James Andrew Hilden-Minton 1995. page 59.
- [10] Chapter 12-7 Design and Analysis of Experiments, Douglas C. Montgomery 2001.
- [11] Multilevel Diagnostics for Mixed and Hierarchical Linear Models by James Andrew Hilden-Minton 1995.

- [12] ST113: ST113/Mixed Linear Models, Prepared by The Statistics Groups at KVL and DTU. -

Chapter 5

Appendix

5.1 SAS Codes for Modeling

5.1.1 SAS Codes for Model3.4, homogeneous variance model

PS: the V option following Random effect will return the 618×618 covariance matrix for 618 observations.

```
proc mixed data=SASUSER.DATA1 ;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material
/
DDFM=CONTAIN
INTERCEPT
Residual
OUTP = sasuser._byMatResday
;
Random Day(Lab) Lab Lab*Material/ V=1 ;
LSMEANS Material
;
run;
```

5.1.2 SAS Codes for Model3.11, between-type heteroscedasticity Model

```

proc mixed data=SASUSER.DATA1 ;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material
/
DDFM=CONTAIN
INTERCEPT
Residual
OUTP = sasuser._ResdaykabinterbyType
;
Random Day(Lab) Lab Lab*Material/ group=Type ;
repeated/group=Type;
LSMEANS Material
;
run;

```

5.1.3 SAS for RSD Approximation interval

Codes for Model3.4 with Moment estimate , which could supply ANOVA table.

```

proc mixed data=SASUSER.DATA1 ;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material Method= Type3
/
DDFM=betwithin
INTERCEPT
Residual
OUTP = sasuser._byMatResday
;
Random Day(Lab) Lab Lab*Material ;
LSMEANS Material
;
run;

```

5.1.4 SAS Codes for Model MB6, between-material heteroscedasticity Model

```

proc mixed data=SASUSER.DATA1;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material
/
HTYPE =3
DDFM=betwithin
INTERCEPT

Residual
OUTP = sasuser._resbyMatdaylabinter
;
random Lab Lab*Material;
repeated / group=Material;
LSMEANS Material
;
run;

```

5.1.5 SAS Codes for Model LB3, between-Lab and between-Material heteroscedasticity Model

```

proc mixed data=SASUSER.DATA1;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material
/
HTYPE =3
DDFM=betwithin
INTERCEPT
Residual
OUTP = sasuser._reslabbyMatdaybylab
;
random Day/group=Lab;
random Lab/group=Material;

```

```
repeated / group=Material;
LSMEANS Material
;
run;
```

5.1.6 SAS Codes for Model LB5, between-Lab and between-Material heteroscedasticity Model

```
proc mixed data=SASUSER.DATA1;
class Lab Material Day Type;
model 'ln(Unts/kg)'n= Material
/
HTYPE =3
DDFM=betwithin
INTERCEPT
Residual
OUTP = sasuser._resbyMatdaybylabinter
;
random Day/group=Lab;
random Lab Lab*Material;
repeated / group=Material;
LSMEANS Material ;
run;
```

5.1.7 Log Ratio Likelihood Test code: LB3-;LB4

```
data test;
n2logL_A=1124; n2logL_B=1155; df=7;
G_AtoB=n2logL_B-n2logL_A;
P_AtoB=1-probchi(G_AtoB,df);
run;
proc print data=test; run;
```

5.1.8 Normal distribution analysis of conditional residual of LB5

```
PROC SORT DATA=SASUSER._RESLABBYMATDAYBYLAB(KEEP=Resid Material)
OUT=WORK._egtemp_;
BY Material;
RUN;
ods EXCLUDE MODES QUANTILES MOMENTS EXTREMEOBS ;
goptions htext=1 cells;
SYMBOL v=SQUARE c=BLUE h=1 cells;
PATTERN v=SOLID;
PROC UNIVARIATE DATA=WORK._egtemp_
CIBASIC (
ALPHA=0.050 )
MU0=0
;
BY Material;
VAR Resid;
HISTOGRAM / Normal( W= 1 L= 1
COLOR=BLUE
mu=EST
sigma=EST
)
CFRAME=CXA8A8A8 CAXES=BLACK WAXIS=1
CBARLINE=BLACK
CFILL=BLUE
PFILL=SOLID
;
QQPLOT / Normal( W= 1 L= 1
COLOR=BLUE
mu=EST
sigma=EST
)
CFRAME=CXA8A8A8 CAXES=BLACK WAXIS=1
;
RUN;
symbol;
```

5.2 One way homogeneity test by Lab of conditional residual of LB5 for different materials

```
PROC SORT DATA=SASUSER._RESLABBYMATDAYBYLAB(KEEP=Resid Material LAB)
OUT=WORK._egtemp_;
BY Material;
RUN;
PROC ANOVA DATA=WORK._egtemp_;
BY Material;
CLASS LAB;
MODEL Resid = LAB;
MEANS LAB / HOVTEST=BARTLETT;
RUN; QUIT;
TITLE;
```

5.2.1 SAS output of LB5

5.2 One way homogeneity test by Lab of conditional residual of LB5 for different materials111

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Group	Estimate
Day	LAB 1	0
Day	LAB 2	3.34E-22
Day	LAB 3	0.001204
Day	LAB 4	0.002740
Day	LAB 5	0
Day	LAB 6	0.000282
Day	LAB 7	0.000047
Day	LAB 8	0.003254
Day	LAB 9	0.000521
Day	LAB 10	0.002829
Day	LAB 11	0.000399
Day	LAB 12	0.000323
Day	LAB 13	0.000247
Day	LAB 14	0.000300
LAB		0.002097
LAB*Material		0.001770
Residual	Material 1	0.01092
Residual	Material 2	0.008929
Residual	Material 3	0.009667
Residual	Material 4	0.01100
Residual	Material 5	0.003849
Residual	Material 6	0.000491
Residual	Material 7	0.01404
Residual	Material 8	0.007689

Fit Statistics

-2 Res Log Likelihood -1151.0

Figure 5.1: SAS output of LB51

5.3 SAS output of LB3 conditional residual homogeneity one way test by lab for 8 materials

5.3 SAS output of LB3 conditional residual homogeneity one way test by lab for 8 materials 113

```
One-Way Analysis of Variance
Results

The ANOVA Procedure

Sample number=1

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 13 13.8924 0.3815

The ANOVA Procedure

Sample number=2

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 13 9.7484 0.7144

The ANOVA Procedure

Sample number=3

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 11 20.4614 0.0394

The ANOVA Procedure

Sample number=4

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 12 25.3712 0.0132

The ANOVA Procedure

Sample number=5

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 10 10.6394 0.3863

The ANOVA Procedure

Sample number=6

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 11 19.2 47 0.0573

The ANOVA Procedure

Sample number=7

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 13 11.6517 0.5564

The ANOVA Procedure

Sample number=8

Bartlett's Test for Homogeneity
of Resid Variance
Source DF Chi-Square Pr > ChiSq
LAB 11 14.0208 0.2318
```

Figure 5.2: SAS output of LB51

5.4 Normality Checking for Random Effect Lab in LB3:

```
proc means data=sasuser..reslabbyMatdaybylab noprint;
by Material Lab;
var 'ln(Unts/kg)'n;
output out=a1 mean=ln;
run;
proc univariate data=a1 noprint;
by Material;
var ln;
probplot / normal( mu=est sigma=est noprint);
inset normal ;
run;
quit;
```