My amazing title

Your R. Name
April DD, 20YY

Submitted to the Department of
Mathematics and Statistics
of Amherst College in partial fulfillment
of the requirements for the degree of
Bachelor of Arts with honors.

Advisor:
Advisor F. Name

Abstract

The abstract should be a short summary of your thesis work. A paragraph is usually sufficient here.

Acknowledgments

Use this space to thank those who have helped you in the thesis process (professors, staff, friends, family, etc.). If you had special funding to conduct your thesis work, that should be acknowledged here as well.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

In standard undergraduate curricula, there is a strong focus on cross sectional data, and thus no emphasis on how time-sequence data is analyzed. However, a significant portion of data that we encounter in the real world is dependent on time. If we want to track trends and changes over time, such as an effect of a certain drug on the body or growth of a company, longitudinal data and analysis will help us examine those points of interest. For example, the Chinese Longitudinal Healthy Longevity Survey from Duke University assessed physical and mental well-being of Chinese elders for over almost 2 decades and re-interviewed survivors every few year. This follow up in data collection allowed researchers to investigate the aging process over time and identify risk factors and causes leading up to death.

Not only can we observe change over time in individuals, but we can look at higher-level grouping, such as change in schools, counties, and organizations. It should be emphasized that only longitudinal data can capture changes within a subject or group; cross-sectional data contain responses that are captured at only one occasion that are then compared to other subjects. Ultimately, it cannot provide information about changes over time.

One key aspect of longitudinal data is that there needs to be repeated measurements of the same individuals across multiple periods of time. If there aren't repeated observations, then it is not possible to make any comparisons between two or more time points. Having repeated measurements of the same individual allows for re-

moval of potential confounding effects, such as gender or socioeconomic status, from the analysis. Since we assume that these confounding variables are fixed effects that do not vary from measurement to measurement, all changes from an individual cannot be attributed to these effects.

The measure that captures the observed changes within an individual is referred to as a response trajectory. There are different ways of comparing response trajectories. For example, it is possible to compare the post-treatment vs baseline changes across multiple treatment groups, or it is also possible to compare the rate of change. The method chosen depends on the specific question of the study.

Apart from comparing just the response trajectories, it is also of interest to compare individual differences in the relationship between covariates and the response trajectory. This can be captured using various different statistical models. The choice of model depends on several characteristics of the data.

1.1 Characteristics of longitudinal data

While the only requirement of longitudinal data is that there is more than one observation for a given individual, there are other components that affect the model chosen. Data can be unbalanced or balanced: *balanced* data refers to when all individuals have the same number of repeated measurements taken at the same occasions. In addition, data can also be missing, resulting in automatically unbalanced data. This affects the accuracy of how changes over time are analyzed depending on if there are any patterns to the missing data or not.

Another unique characteristic of longitudinal data is that repeated measurements of each individual are typically positively correlated. This feature violates conditions of other common statistical methods such as linear regression, where measurements

are assumed to be independent. This positive correlation allows for more accurate estimates of the model coefficients and response trajectories since there is reduced uncertainty knowing that a previous measurement can help predict the next one.

In longitudinal analysis, a covariance matrix is calculated for each individual and all of their measurements. The diagonals of this matrix represent the variance of each of the measurements, which are not constant over time. The off-diagonals of the matrix are non-zero to account for the lack of independence between measurements, and are usually not constant because correlations between measurements decrease over time. While these values are rarely 0, they are also rarely 1. There are different covariance pattern structures that are imposed that account for these features.

These features of the covariance of longitudinal data serve as the underlying premise to the idea that variation can be separated into three distinct parts: 1) between-individual variation, 2) within-individual variation, and 3) measurement error.

Between-individual variation helps explain why measurements from the same individual are more likely to be positively correlated than measurements to a different individual. Within-individual variation helps explain why correlations decrease with increasing time differences, and measurement error explains why correlations are never one. These three types of variation may contribute to total variation in unequal amounts, but may not need to be differentiated depending on the type of longitudinal analysis desired.

1.1.1 Notation

Throughout the rest of the text, we will use a standard set of notation for all parameters and variables. Y_{ij} represents the response variable for the i^{th} individual at the j^{th} measurement. When we have repeated n_i measurements for an individual, we can

construct a vector,

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ .. \\ Y_{in_i} \end{pmatrix}.$$

We use μ_{ij} as the conditional mean response at the j^{th} measurement, where conditional entails a dependence of the mean response on the covariates.

1.2 Estimation and Inference

Regression coefficient values β and the covariance matrix Σ_i can be estimated using maximum likelihood estimation, which identifies values of β and Σ_i that maximize the joint probability of the response variable occurring based on the observed data; the probability is known as the likelihood function. These values are estimates that are denoted by $\hat{\beta}$ and $\hat{\Sigma}_i$. When observations are independent of one another, maximizing the likelihood function for β is equivalent to finding a value of $\hat{\beta}$ that minimizes the sum of the squares of the residuals. However, since there are repeated measurements of each individual that are not independent of one another we use the generalized least squares (GLS) estimator:

$$\hat{\beta} = \{\Sigma_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\}^{-1}\Sigma_{i=1}^{N}(X_i'\Sigma_i^{-1}y_i).$$

In addition, the sampling distribution of $\hat{\beta}$ has mean β and covariance:

$$\hat{Cov}(\hat{\beta}) = \{\Sigma_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\}^{-1}.$$

The GLS estimator assumes that Σ_i is known. However, since this isn't usually

the case, we can substitute Σ_i with a maximum likelihood estimate $\hat{\Sigma}_i$.. It can be shown that the properties of $\hat{\beta}$ still hold using an estimate of the covariance.

While the maximum likelihood estimate of Σ_i is adequate, a modified method known as restricted maximum likelihood (REML) estimation is suggested to reduce bias in finite samples. The bias originates from the fact that β itself is also estimated from data, but is not accounted for when estimating covariance. In REML estimation of Σ_i, β is removed from the likelihood function. This REML estimation of Σ_i can be used in the GLS estimator for $\hat{\beta}$ mentioned above, and is recommended in place of the ML estimator.

Now that we have estimates for β, we can make inferences through construction of confidence intervals and hypothesis testing. For example, using the ML estimate $\hat{\beta}$ and $\hat{Cov}(\hat{\beta})$, we can construct a Wald statistic to test for significance of $\hat{\beta}_k$:

$$Z = \frac{\hat{\beta}_k}{\sqrt{\hat{Var}(\hat{\beta}_k)}}.$$

In later chapters we will explore how inference may be impacted in smaller sample sizes.

1.3 Linear models for longitudinal data

As mentioned previously, there are multiple ways to model longitudinal data. When the response variable is continuous, we can consider a model that relates the mean response and the covariates in a linear way. In a linear model all components can be represented using vectors and matrices. The most general form of the linear model can be represented as:

$$E(Y|X_i) = X_i\beta$$

, where β is a vector of regression coefficients and X_i is a vector of covariates. We will discuss three methods for linear models: 1) response profile analysis, 2) parametric time model, 3) linear mixed effect model.

1.3.1 Response profile analysis

In response profile analysis, we allow for arbitrary patterns in the mean response over time. A sequence of means over time is known as the mean response profile. The main goal of this analysis is to identify differences in pattern of change in mean response profile among 2 or more groups. This method requires that the data be balanced.

There are three effects of interest when analyzing response profiles in longitudinal analysis: 1. $group \times time$ interaction effect (are the mean response profiles different in groups over time?) 2. time effect (assuming mean response profiles are parallel between groups, are the means changing over time?) 3. Group effect (do the mean response profiles differ?)

However, the first question is the primary interest. The goal is to find whether the change in mean response over time differs across groups.

To test for significance of the $group \times time$ effect, we have a null hypothesis that the difference in means between the n groups is constant over time, which in other words entails that mean response profiles between the groups have parallel slopes. We can implement the general linear model $\mu_i = X_i\beta$ to test our hypotheses, using comparison of β slope parameters to determine whether there is a $group \times time$ effect.

For example, to express the model for response profile analysis for G groups and n occasions of measurement, we have $G \times n$ parameters for the G mean response profiles. For two groups measured at three occasions, we have 6 slope parameters. if $\beta_1 - \beta_3$ represent slope parameters for mean responses in group 1 and $\beta_4 - \beta_6$ represent slope parameters for mean responses in group 2, our null hypotheses would be that

$(\beta_1 - \beta_4) = (\beta_2 - \beta_5) = (\beta_3 - \beta_6)$.

An unstructured covariance model is typically assumed for response profile analysis. "Unstructured" means that there is no explicit structure or pattern imposed on the covariance for the repeated measures, so each of the variances and covariance pairs are estimated using restricted maximum likelihood estimation (REML). For n repeated measures, there are n variances and $n \times (n-1)/2$ covariances to be estimated. In a study where there are 10 repeated measurements, there 55 total covariance parameters to be estimated, which can become computationally intensive.

One other aspect to consider when conducting analysis on mean response profiles is how to adjust for the baseline measurement. The baseline value is important when we want to calculate measures that compare mean response to the baseline. How we adjust depends on whether the study is randomized or observational. When the study is randomized and baseline measurement is taken before treatment assignment, the mean response at occasion 1 is independent of the group, and assumed to be equal. One possible method is to treat the baseline measurement as a covariate, and use response measurements 2 through n as the dependent measures. This is referred to as the analysis of covariance approach. Additionally, this method only works for randomized studies because using the baseline measurement as a covariate for observational studies may produce confounding effects. For an observational study, it is recommended to subtract the baseline response to create a change score. For both types of longitudinal studies there are various methods to account for the baseline value, and should be considered carefully before implementing the method.

Overall, response profile analysis is a straightforward method in investigating differences between groups for longitudinal data. Since both the covariance and mean responses have no imposed structure, the analysis is more robust and immune to inaccurate results due to model misspecification. However, there are drawbacks as well.

Response profile analysis does not consider time-order of the measurements and does not distinguish between between-individual variation and within-individual variation. In addition, it can only provide a broad analysis of whether there are differences across groups and time, but does not provide the amount of detail usually needed to answer research questions, such as how exactly measurements taken towards the end of the study compare to measurements taken at the beginning. In this method, time is treated as a categorical covariate rather than a continuous one. Another method that addresses the issue of examining time order of the data is parametric time models.

1.3.2 Parametric Time Models

Parametric time models are able to capture time order of the data by fitting linear or quadratic curves to capture an increasing or decreasing pattern over time. Time is treated as a continuous covariate rather than a categorical one. In addition, unlike response profile analysis, parametric time models are able to handle unbalanced and missing data. Rather than fitting a complex and perfect model onto the observed mean response profile, parametric time models fit simple curves that produce covariate effects of greater power. This is because in mean response profile we are testing a wider range of hypotheses since we are looking for inequality between two groups; however, in parametric time models, we are testing more specifically whether the data follow a linear trend, which results in more power.

Additionally, while in the mean response profile analysis an unstructured covariance pattern is assumed, here there is flexibility in choice of the covariance model; there are several options such as Toeplitz or compound symmetric that impose various structures on the model. For example, a Toeplitz model:

$$Cov(Y_i) = \begin{pmatrix} 1 & \rho_1 & \rho_2 & ... & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & ... & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & ... & ... \\ \rho_3 & & & & \end{pmatrix}$$

structures the covariance matrix such that any pair of responses that are equally separated in time have the same correlation.

It is possible to choose an unstructured covariance model as well, but can be computationally intense if there are a large number of measurements.

We can use parametric time models in two ways: through polynomial trends and linear spines.

1.3.3 Polynomial Trends

Using polynomial trends such as linear or quadratic, we can model longitudinal data as a function of time. Linear trends are the most common and interpretable ways to model change in mean over time. In an example comparing a treatment group to a control group, we can fit a linear trend using the following equation:

$$E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 Group_i + \beta_4 Time_{ij} \times Group_i.$$

If $\beta_4 = 0$, then the two groups do not differ in terms of changes in the mean response over time.

For quadratic trends, the changes in mean are no longer constant since the rate of change depends on the time. Thus, we fit an additional parameter to express the rate of change. Using the previous example of treatment vs. control group, we have

the model:

$$E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 Time_{ij}^2 + \beta_4 Group_i + \beta_5 Time_{ij} \times Group_i + \beta_6 Time_{ij}^2 \times Group_i.$$

As we can see from the models above, the inclusion of an additional parameter $Time_{ij}^2$ changes the mean response rate. One problem that may arise from using quadratic trends is that there is collinearity between $Time_{ij}$ and $Time_{ij}^2$, which can affect the estimation of β. To account for this, we can center the $Time_{ij}$ variable around the mean time value for all individuals, instead of centering it around zero as done in normal analysis. For example if we have a set of times $Time = 0, 1, 2, ...10$, then the mean time value is five. Thus time zero would be recentered as -5. The interpretation of the intercept changes to represent the mean response at that recentered mean time value.

1.3.4 Linear splines

In instances where responses cannot be adequately fit by polynomial trends, such as when the responses fluctuate between increasing and decrease at different extents, we can employ a linear spline model. This model consists of piece-wise line segments that have unique slopes for a given set of time measurements. The point at which different line segments meet are called knots, and the number of knots depends on the context of the data and researcher discretion.

Drawing again from our treatment vs control group design, a linear model for the mean responses of the control group is:

$$E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij} + \beta_3 (Time_{ij} - t^*)_+.$$

The $()_+$ indicates a truncated line function and is positive when $Time_{ij} - t^*$ is greater

than 0, and otherwise is equal to 0. In this case, the function depends on the specified time t^*. If the mean response is before t^*, then the mean response is modeled by:

$$E(Y_{ij}) = \beta_1 + \beta_2 Time_{ij}.$$

If the mean response is after t^*, it is modeled by

$$E(Y_{ij}) = (\beta_1 - \beta_3 t^*) + (\beta_2 + \beta_3) Time_{ij}.$$

There are benefits to parametric models that make them a more appealing choice compared to response profile analysis. Parametric time models are able to capture time order, and can be used with unbalanced data. However, they do not differentiate between subject and within subject variation. If further analysis of individual variation is desired, linear mixed effects models can be employed.

1.3.5 Linear Mixed Effects

In both response profile analysis and parametric time models, the regression parameters are considered to be universal for each population group. However, in instances where we want to account for heterogeneity within a population, we can use a linear mixed effects model and consider a subset of the regression parameters to be random. This model distinguishes between fixed effects, which are population characteristics shared by all individuals, and subject specific effects, also known as random effects, which pertain to each individual. These subject specific effects mean that parameters are random, which induces a structure onto the covariance model.

In addition, distinguishing between fixed and random effects allows for differentiation between within-subject and between-subject variation.

One example of the linear mixed effects model is the random intercept model,

which is the simplest version of the linear mixed effects model:

$$Y_{ij} = X'_{ij}\beta + b_i + \epsilon_{ij}$$

This model is very similar to the general linear model with a few additions. b_i is the random subject effect and ϵ is the measurement error. Both effects are random, with mean 0 and $\text{Var}(b_i) = \sigma_b^2, \text{Var}(\epsilon_{ij}) = \sigma^2$.

$X'_{ij}\beta$ is the population mean, and b_i represents the differing subject effect that is unique to each individual. b_i is interpreted as how the subject deviates from the population mean while accounting for covariates.

As mentioned previously, the random effects are responsible for inducing a structure on the covariance model. This structure is not to be confused with the covariance structures that can be chosen when using parametric time models. For a given individual, it can be shown that variance of each response is:

$$\text{Var}(Y_{ij}) = \sigma_b^2 + \sigma^2$$

and the covariance between two measurements Y_{ij} and Y_{ik} is equal to σ_b^2. The resulting covariance matrix $\begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 & ... & \sigma_b^2 \\ \sigma^2 & \sigma_b^2 + \sigma^2 & \sigma_b^2 & ... & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma^2 & ... & ... \end{pmatrix}$\$ implies correlation between measurements, and also highlights the role played by the random effects in determining the covariance.

Extending beyond the random intercept model, multiple random effects can be incorporated.

A linear mixed effects model can expressed as

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i.$$

Where: β is a $p \times 1$ vector of fixed effects b_i is a $q \times 1$ vector of random effects X_i is a $n \times p$ matrix of covariates Z_i is a $n \times q$ matrix of covariates

The subset of regression covariates that vary randomly are found in Z_i. We assume that b_i comes from a multivariate normal distribution with mean 0 and covariance matrix G. We also assume that ϵ_i are independent of b_i, and come from multivariate normal distribution with mean 0 and covariance matrix R_i.

The covariance of Y_i can be modeled by

$$\text{Cov}(Z_i b_i) + Cov(\epsilon_i) = Z_i G Z_i' + R_i.$$

This model, which outlines a distinction between G and R_i, allows for separate analysis of between subject and within subject variation. Unlike other covariance models, in linear mixed effects models the covariance is a function of the times of measurement. This allows for unbalanced data to be used for the model since each individual can have their unique set of measurement times. Lastly, the model allows for variance and covariance to change as a function of time. To illustrate, consider the following model:

In an example where individuals can vary both in their baseline response and their rate of change, we have:

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i,$$

where both X_i and $Z_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ ... & ... \\ 1 & t_{in} \end{pmatrix}$. For the i^{th} subject at the j^{th} measurement, the equation is as follows:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{1i} + b_{2i} t_{ij} + \epsilon_{ij}.$$

If $\text{Var}(b_{1i}) = g_{11}$, $\text{Var}(b_{2i}) = g_{22}$, and $Cov(b_{1i}, b_{2i}) = g_{12}$ where these three components represent the G covariance for b_i, then it can be shown that $Cov(Y_{ij}, Y_{ik}) = g_{11} + (t_{ij} + t_{ik})g_{12} + t_{ij}t_{ik}g_{22}$.

Here in the covariance matrix we can see the dependence of the covariance on time. In this example there are four covariance parameters that arise from the two random effects of intercept and time. The number of covariance parameters is represented by $q \times (q+1)/2+1$, where q is the number of random effects. To choose the most optimal model for covariance, we compare two nested models, one with $q+1$ random effects and one with q random effects. We use the likelihood ratio test to make a decision for which model to use.

One additional analysis that is possible with linear mixed effects models is predicting subject-specific responses. Given that b_i is a random variable, we can predict it using:

$$E(b_i|Y_i) = GZ_i(\Sigma)_i^{-1}(Y_i - X_i\hat{\beta}).$$

Because the covariance of Y_i is unknown, we can estimate both G and $(\Sigma)_i^{-1}$ using REML, creating $\hat{b}_i$, also known as the empirical best linear unbiased prediction

(BLUP). Thus, the equation for predicting the response profile is:

$$\hat{Y}_i = X_i\hat{\beta} + Z_i\hat{b}_i.$$

This equation to estimate the mean response profile can be extended to incorporate R_i, which represents within-subject variability. From this extension, we see that the equation and the empirical BLUP account for the weighting of both the within-subject variability and between-subject variability. If there is more within-subject variability, then more weight is assigned to $X_i\hat{\beta}$, the population mean response profile, in comparison to the subject's individual responses, and vice versa.

1.4 Choosing the best model

After presenting three methods of evaluating longitudinal data, the natural question arises of how to choose the most appropriate model. While there is no definite correct answer, there are several factors to consider. If data are unbalanced, response profile analysis should not be considered; rather, parametric time model or linear mixed effect model would be more optimal. If time order is important to the analysis, then only parametric time model and linear mixed effect model should be used. If there is a need to distinguish between the two types of variation that can occur, then only linear mixed effect models are appropriate. The model should ultimately be chosen based on the characteristics and constraints of the data, as well as the specificity of the research question at hand.

1.5 Conclusion

Longitudinal analysis is a valuable method to analyze changes over time. It is important to understand the unique characteristics that come with this analysis and to choose the best model that can capture the salient patterns that arise from the data.

In subsequent chapters we will dive more deeply into how inference in longitudinal analysis is affected when sample sizes are not efficient through both simulation and application.

Chapter 2 Methods for tests of fixed effects in small and nonnormal samples

In chapter 1, we outlined the basics of analyzing longitudinal data and introduced linear mixed models. Next, we will examine inference of linear mixed models, and how methods such as Kenward-Roger (KR) and Satterthwaite can be used in situations where standard procedures for inference may produce questionable results.

2.1 Inference

In statistical inference, the goal is to make conclusions about the underlying characteristics of a set of data and establish a relationship between certain variables. Hypothesis testing is one of the primary examples of inference, and is carried out in order to assess the true value of a population parameter. In linear models, the significance of a slope parameter, β_k, is often assessed, where the null hypothesis, $H_0 $ is $\beta_k = 0$, and the alternative hypothesis H_a is $\beta_k \neq 0$. A test of the null hypothesis involves using a Wald statistic in the form

$$Z = \frac{\hat{\beta}_k}{\sqrt{\hat{Var}(\hat{\beta}_k)}},$$

which is then compared to the normal distribution, and a subsequent p-value is obtained.

Building on foundations of a general linear hypothesis test, given a matrix L of size $q \times p$, where q represents the number of estimable functions of β,

$$(L\hat{\beta} - L\beta)'[L(X'\widehat{Cov}(\hat{\beta})X)^{-}L']^{-1}(L\hat{\beta} - L\beta)$$

is approximately $\chi^2(q)$ (Rencher and Schaalje, 2008). For a null hypothesis $H_0 : L\beta = 0$, the test statistic G is

$$(L\hat{\beta})'[L(X'\widehat{Cov}(\hat{\beta})X)^{-}L']^{-1}(L\hat{\beta}).$$

Aside from using the Wald statistic, likelihood ratio tests are another method to make inferences about β, and involves comparing two models: (1) a nested model, which assumes that β_k is 0, and (2) a full model, that allows β_k to vary without constraint. The difference in the maximized log-likelihood of the two models, $\hat{l}_{reduced}$ and $\hat{l}_{full}$ are compared. This difference is represented by the statistic

$$G^2 = 2(\hat{l}_{full} - \hat{l}_{reduced}),$$

which is compared to a chi-square distribution. The larger the difference, the more likely we are to conclude that the nested model is insufficient, and that β is not zero. While there are benefits to using the likelihood ratio test, the rest of this study will focus on method of using the Wald statistic.

2.1.1 Inference in small sample sizes

One crucial assumption when conducting inference using the ML estimate for β is that the sample size is sufficient enough where it does not affect the estimate for Σ_i. However, what happens when the sample size is too small? This causes $\hat{\Sigma}_i$ to

underestimate the true variance, which in turn causes $\widehat{\text{Cov}}(\hat{\beta})$ to be too small since it relies on covariance estimator. If $\widehat{\text{Cov}}(\hat{\beta})$ is too small, the denominator of the test statistic is inflated, leading to increased Type I error. One can see that the bias of the covariance estimator weakens the entire foundation of estimation and inference.

In very limited cases, where data are complete, balanced, and produce nonnegative values in REML estimation, is it possible to perform exact small-sample inferences. If $[L(X'\widehat{\text{Cov}}(\hat{\beta})X)^{-}L']^{-1}$ with g degrees of freedom can be rewritten such that

$$\frac{(L\hat{\beta})'Q(L\hat{\beta})}{g}\frac{w}{d} = \frac{(L\hat{\beta})'[L(X'\widehat{\text{Cov}}(\hat{\beta})X)^{-}L']^{-1}(L\hat{\beta})}{g},$$

where w is a chi-square random variable with d degrees of freedom. If so, this statistic is F-distributed.

However, in most scenarios, an approximate small-sample method must be used, in which the statistic

$$F = \frac{(L\hat{\beta})'[L(X'\widehat{\text{Cov}}(\hat{\beta})X)^{-}L']^{-1}(L\hat{\beta})}{g}$$

follows a distribution with numerator degrees of freedom g, and unknown denominator degrees of freedom (DDF). There are several ways to approximate the DDF.

Both Satterthwaite and KR are proposed methods of reductions to the DDF when conducting tests in order to account for the uncertainty of the covariance estimator. The KR method goes one step forward to also adjust the test statistic itself.

2.2 Satterthwaite

Sattherthwaite approximation was developed by Fai & Cornelius (1996), with the F statistic following the form:

$$F = \frac{1}{l}\hat{\beta}' L' (L\widehat{\text{Cov}}(\hat{\beta})L')^{-1} L\hat{\beta}.$$

For the denominator degrees of freedom we perform spectral decomposition on $L'\widehat{\text{Cov}}(\hat{\beta})L = P'DP$, where D is a diagonal matrix of eigenvalues and P is an orthogonal matrix of eigenvectors. When r represents the r^{th} row of $P'L$, we have $v_r = \frac{2(d_r)^2}{g_r' W g_r}$, where g_r is a gradient vector, d_r is the r^{th} diagonal element of D, and W is the covariance matrix of $\hat{\sigma}^2$. The denominator degrees of freedom is calculated by:

$$\frac{2E}{E-l},$$

where $E = \sum_{r=1}^{l} \frac{v_r}{v_r - 2} I(v_r > 2)$ if $E > l$, otherwise $DF = 1$.

When $l = 1$ the KR and Satterthwaite approximation will produce the same denominator degrees of freedom. However, since the statistic used for the two methods are not the same, the results for inference will not be the same. It is important to note that both methods are only valid when using REML.

2.3 Kenward-Roger

In Kenward-Roger (1997), a Wald statistic is proposed in the form of:

$$F = 1/l(\hat{\beta} - \beta)^T L (L^T \hat{\Phi}_A L)^{-1} L^T (\hat{\beta} - \beta),$$

where l represents the number of linear combinations of the elements in β, L is a fixed matrix, and $\hat{\Phi}_A$ is the adjusted estimator for the covariance matrix of $\hat{\beta}$. As mentioned previously, $\widehat{\text{Cov}}(\hat{\beta})$ is a biased estimator of $\text{Cov}(\hat{\beta})$ when samples are small, and underestimates. This adjusted estimator is broken down into $\hat{\Phi}_A = \widehat{\text{Cov}}(\hat{\beta}) + 2\hat{\Lambda}$, where $\hat{\Lambda}$ accounts for the amount of variation that was underestimated by the original

estimator of covariance of $\hat{\beta}$. The value Λ is approximated using a Taylor series expansion around σ, to be

$$\Lambda \text{Cov}(\hat{\beta})[\sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}(Q_{ij} - P_i \text{Cov}(\hat{\beta}) P_j)]\text{Cov}(\hat{\beta}),$$

where $P_i = X^T \frac{\partial \Sigma^{-1}}{\partial \sigma_i} X$ $Q_{ij} = X^T \frac{\partial \Sigma^{-1}}{\partial \sigma_i} \Sigma \frac{\partial \Sigma^{-1}}{\partial \sigma_j} X$, and W_{ij} is the (i,j)th element of $\text{W} = V[\hat{\sigma}]$.

This Wald statistic that uses the adjusted estimator is scaled in the form:

$$F^* = \frac{m}{m+l-1}\lambda F,$$

where m is the denominator degrees of freedom, and λ is a scale factor. Using the expectation and variance of the Wald statistic, F Both m and λ need to be calculated from the data, such that:

$$m = 4 + \frac{l+2}{l\rho - 1},$$

where $\rho = \frac{V[F]}{2E[F]^2}$ and $\lambda = \frac{m}{E[F](m-2)}$. This statistic will ultimately follow an exact $F_{l,m}$ distribution.

2.4 Other methods

Residual DDF: The DDF is calculated as $N - rank[X]$, where N is the total number of individuals in the dataset. This method is only suitable for data that are independent and identically distributed, so it is not typically used in linear mixed models.

Containment Method In the containment method, random effects that contain the fixed effect of interest are isolated. The smallest rank contribution to the $[XZ]$ matrix among these random effects becomes the DDF. If there are no effects found, then the

DDF is equal to the residual DDF.

Between-Within Method Schluchter and Elashoff (1990) propose a DDF method where residual DDF are calculated for both between-subject and within-subject subgroups. If there are changes in the fixed effect within subjects, then the within-subject DDF is used, otherwise the between-subject DDF is used.

2.5 Existing literature

Both KR and Satterthwaite methods are frequently used and compared, and its performance is highly dependent on the structure of the data. A majority of studies focusing on DF method comparison in mixed models use split-plot design, as small sample sizes are more common in agricultural and biological fields. Schaalje, et al. (2002) found that in comparison to other degrees of freedom-adjusting methods like Satterthaite, KR was the most suitable for small sample data. Using factors such as imbalance, covariance structure, and sample size, they demonstrated that the KR method produced simulated Type I error rates closest to target values. However, their focus was primarily on complexity of covariance structure, and they found that more complicated structures, such as ante-dependence, produced inflated error rates when coupled with small sample size. Arnau (2009) found that KR produces more robust results compared to Satterthwaite and Between-Within approaches, especially in cases where larger sample size was paired with covariance matricies with larger values.

These studies are conducted with data drawn from normal distributions. However, real-world data used in fields such as psychometrics have distributions that are nonnormal. In Arnau et. al's 2012 paper, the authors extend their evaluation of KR for split-plot data that follow a log-normal or exponential distribution, and for

when the kurtosis and skewness values are manipulated. They found that, compared to normal distribution, the test is less robust for log-normal distributions, but that there is no signficant difference in performance between exponential and normal distributions. In addition, they suggest that skewness has a bigger effect on robustness of KR compared to kurtosis.

Existing research evaluating the performance of methods that reduce Type I error rate in small samples are thorough, however, the differences in simulation setup and structure of data used make generalizations difficult. Although the KR method has been shown as a viable option for analysis of small samples in many occasions, it should continue to evaluated against other methods. To date, there is no literature on the performance of Satterthwaite for nonnormal longitudinal data design. Given the prevalence of nonnormal and small data samples, it is important to continue exploring methods that ensure robust results.

2.6 Goals of this study:

In this study, we aim to expand on previous simulations, evaluating how methods for evaluated fixed effects perform under different nonnormal distributions and sample sizes. The aforementioned studies often use a split-plot design and impose a covariance structure, but goal of this study will be to compare performance of KR and Satterthwaite methods for repeated measures longitudinal data fitted with a linear mixed effects model, and no imposed covariance structure. Since most mixed models use unstructured covariance structure, it would be beneficial to see how these methods perform without considering covariance structure as a factor.

2.7 Simulation Set up:

2.7.1 Generating data: Sample size

In this study, we consider a linear mixed effects model with two discrete covariates: time and treatment. The range of possible values that time takes on depends on how many number of measurements per individual, which can be 4 or 8. The treatment covariate takes on values of 0 or 1, and each assigned to half of the sample. The number of individuals take on possible values of 10, 18, and 26. These were chosen to reflect possible samples that would not hold under the common assumption that the sample size must be at least 30 for it to be considered sufficient enough for the Central Limit Theorem to hold.

2.7.2 Generating data: Fixed Effects

We have three fixed effects: the intercept value and the covariates time and treatment. The intercept, an arbitrary value, is set at 3.1. Time and treatment have a value of 0, and the Type I error rates of treatment will be evaluated.

2.7.3 Generating data: Random effects

In order to generate a continuous response variable that is nonnormal, we generate our random effects values from nonnormal distributions, which are either exponential or lognormal. Previous research shows that many data used in social and health sciences follow nonnormal distributions (Limpert, Stahel, & Abbt, 2001). More specifically many follow lognormal distributions, such as age of onset of Alzheimer's disease (Horner, 1987), or exponential distribution to model survival data. In order to cover a wide range of exponential and lognormal distributions, parameters were chosen to model distinct distributions. For exponential distributions, lambda values of .2, and

.9 were used, (DO I NEED TO INSERT GRAPH?). For lognormal distribution, mean and standard deviation parameter combinations were (0,.25), and (1,.5).

Using the `SimMultiCorrData` package, we derive kurtosis and skewness values based on the distributions specified above. The table below shows the range of skewness and kurtosis values for the Lognormal distribution. In the intercept only model, only one non-normal continuous variable is generated for the random effect, so the function `SimMultiCorrData::nonnormvar1` is used. Values are generated through Fleishman's method for simulating nonnormal data by matching moments using mean, variance, skew, and kurtosis and then transforming normally distributed values.

Kurtosis and skew values for the distributions used in this simulation are shown below. In the case of the linear model that has both random effects for intercept and slope, we want to generate random effects values that are correlated. Using `SimMultiCorrData::rcorrvar`, we use a similar process for generating one nonnormal continuous variable, but extend it to generating variables from multivariate normal distribution that take in to account a specified correlation matrix, and are then transformed to be nonnormal. We use a correlation value of -.38 to generate the random effects, which is based off the correlation observed when fitting a linear mixed effects model from the dataset used in the application portion of this study.

mean	sd	skew	kurtosis	fifth	sixth
1.03	0.262	0.778	1.1	2.3	6.48
3.08	1.642	1.750	5.9	31.4	240.00
5.00	5.000	2.000	6.0	24.0	120.00
1.11	1.111	2.000	6.0	24.0	120.00

***** FIX THIS ***** Lastly, to account for measurement/sampling error, we assume that the error is random and drawn from a $N \sim (0, .2)$. The standard deviation

value was chosen to minimize the variation of the errors in relation to the random effects of the intercept and the covariate.

2.7.4 Linear mixed effects model

In a linear mixed effects model, the amount of random effects that will be modeled depends on the research question at hand. Here, we will examine both a random intercepts-only model, where the intercept of the model is assumed to have a random effects structure, as well as a random intercept and slope model, where in addition to intercept, the covariate time will also have a random effects structure.

We use the `lmerTest` package to fit the linear mixed effects model, and evaluate the significance of the covariate in the model. To evaluate significance, we compare both the KR and Satterthwaite method for adjusting denominator degrees of freedom and its resulting p-value. Because the value of the covariate in our model is fixed at 0 in order to identify Type I error, we expect to see that the p-value for the covariate time to not be significant ($p > .05$) in an ideal scenario.

2.8 Evaluating and Results

After performing 1,000 replications of each condition at a significance level of .05, we evaluate robustness using Bradley's criterion,which considers a test to robust if the empirical error rate is between .025 and .075. In the following section, we will compare Type I error rates produced from KR and Satterthwaite methods as well as t-as-z and using the standard DF formula, further stratified by distribution and other manipulated parameters. T-as-z and standard DF formula are not adjustments to account for smaller sample sizes, and are used as comparison to Satterthwaite and KR, since they are expected to be anti-conservative.

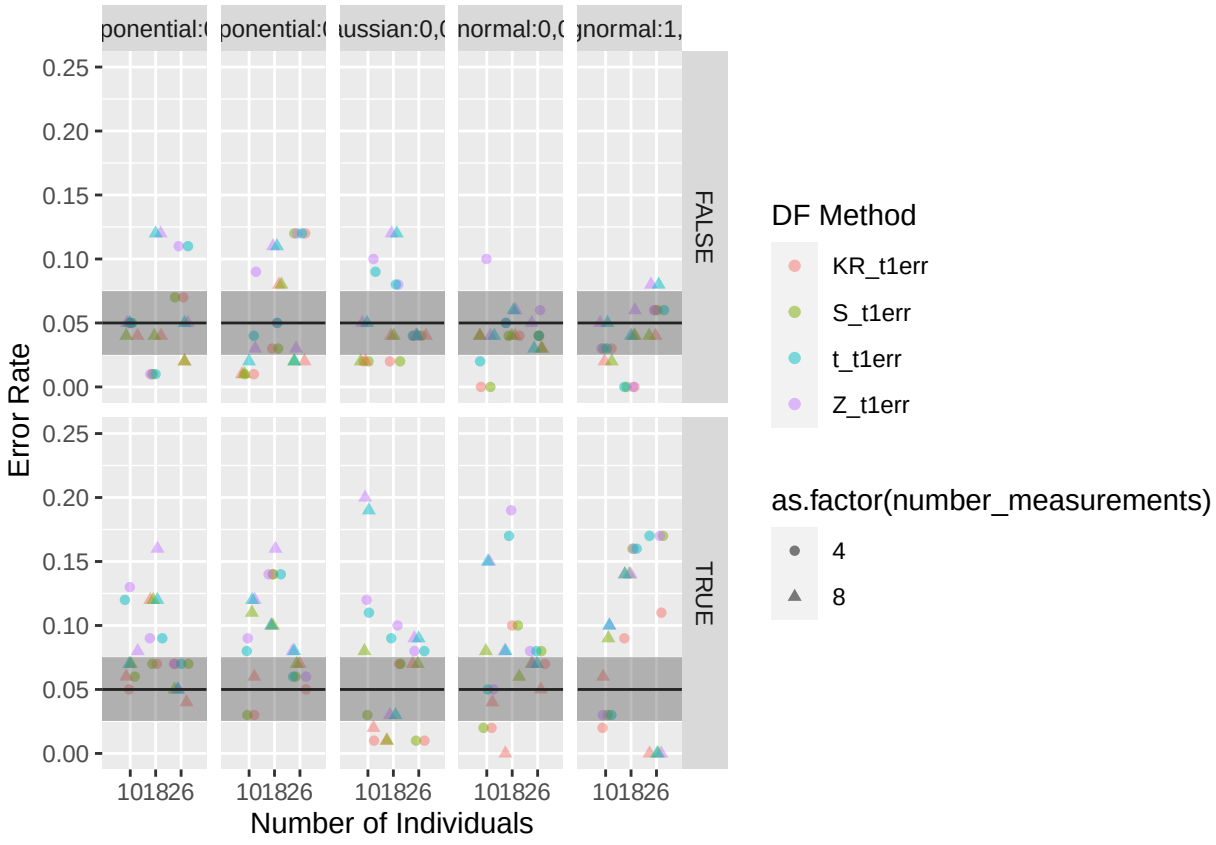


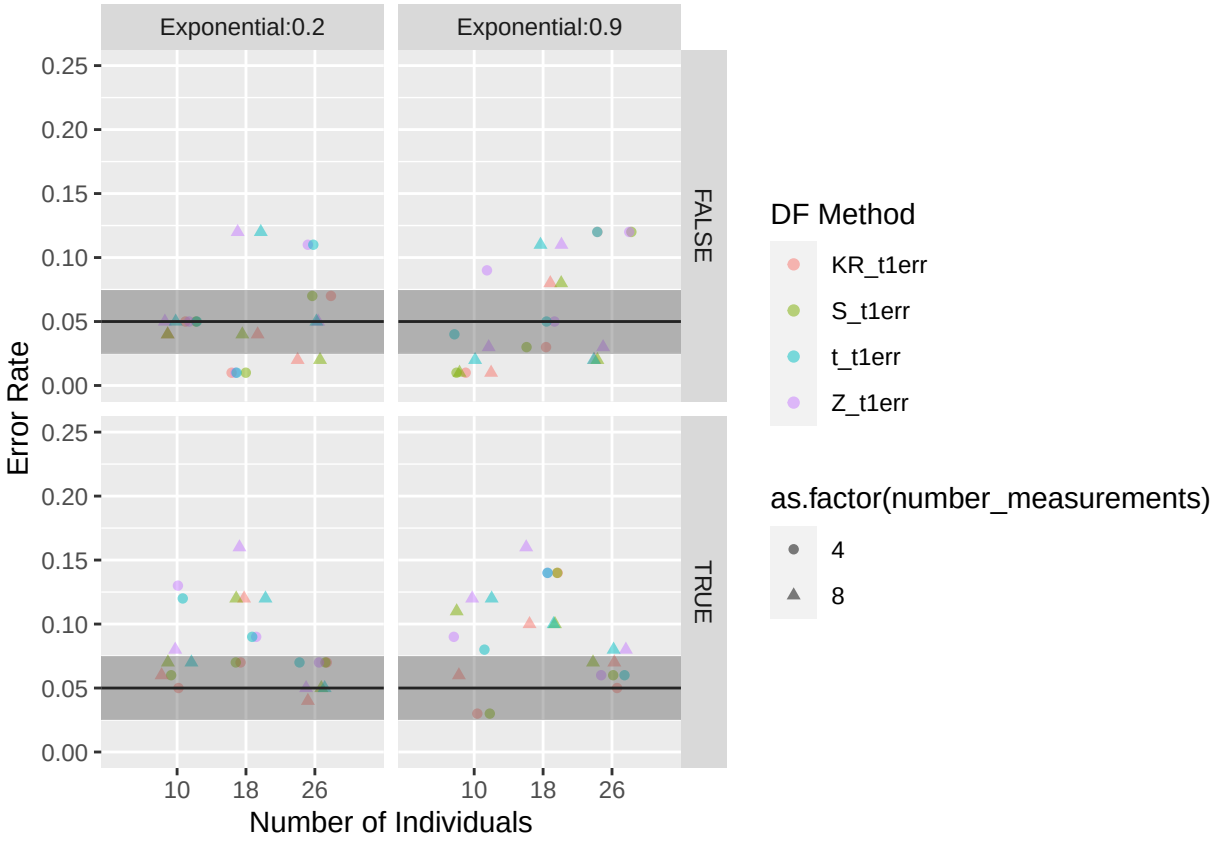
FIGURE 1 displays error rates from all 4 degrees of freedom methods by distribution, parameters, complexity of random model, number of measurements, and number of samples. The shaded region indicates error rates that are considered robust by Bradley's criterion. It is evident that there are varying patterns of performance by distribution. The common conception that larger sample sizes or large number of measurements can improve robustness is not necessarily evident across all distributions, for example in the case of the exponential distribution. While the random intercept model, a more structurally simple model, yields more robust error rates in the lognormal and gaussian distribution, that is not the case in the exponential distribution.

However, there is one trend that we can identify. When looking at performance of

the 4 methods overall, we can see that the t-as-z and standard DF approach produce significantly more anti-conservative results, regardless of the values of other parameters. These trends align closely with a previous study by Luke (2017) examining only normal distributions.

In order to make more specific observations and identify trends, we will examine performance within each of the three distributions by sample size and number of measurements before attempting to compare across them.

2.8.1 Exponential Distribution




```
# A tibble: 8 x 4
# Groups:   params, rslope [4]
  params rslope number_measurements robustness
  <chr>  <lgl>                <dbl>      <dbl>
```

```
1 0.2    FALSE                    4       0.5
2 0.2    FALSE                    8       0.667
3 0.2    TRUE                     4       0.667
4 0.2    TRUE                     8       0.583
5 0.9    FALSE                    4       0.417
6 0.9    FALSE                    8       0.167
7 0.9    TRUE                     4       0.5
8 0.9    TRUE                     8       0.25
```

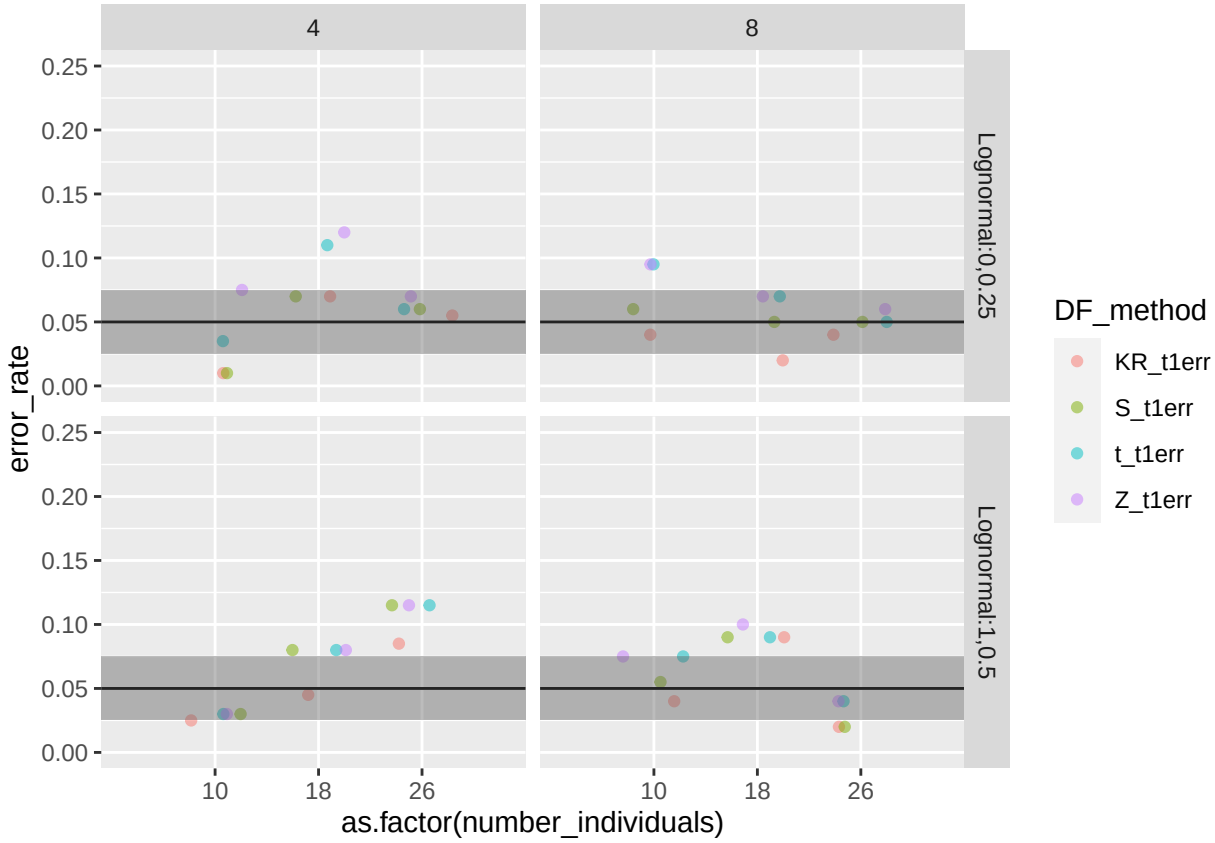
Our simulation results contain two exponential distribution, one with $\lambda = .9$ and $\lambda = .2$. In FIGURE? we identified that in both exponential distributions, random slope models tended to yield more robust error rates, however, we continue to disaggregate by complexity of model to identify how it may work alongside other parameters. At $\lambda = .2$, we can see that in random intercept models the error rates are less anti-conservative at sample sizes of 10 and 18 compared to the random slope model. However, at sample size 26 the random slope model performs significantly better. It is difficult to discern whether increasing the number of measurements reduces robustness overall.

At $\lambda = .9$ and smaller sample sizes, the methods used on the random intercept model do produce more conservative, although not more robust, Type I error rates. However, similar to what was found in the earlier distribution, the random slope model at sample size 26 performs better than the random intercept, and at any size. Increasing the number of measurements to 8 seems to decrease anti-conservatism in DF methods in the random intercept model, but not the random slopes model.

Despite being sampled from the same type of distribution, the application of DF methods to these two exponential distributions produce different trends in error rates. One trend that appears overall is that if the sample size is large enough, fitting a random slopes model will produce more robust error rates compared to a random intercept model in both distributions. Although these two distributions have the same

skewness and kurtosis values, the difficulty in parsing these trends suggests that there are other aspects of this distribution that are affecting the performance of the DF methods, and generalizing to other exponential distributions is not recommended.

2.9 Lognormal



As seen in the first figure, across the lognormal distributions, random intercept models had consistently more robust error rates in comparison to random slope. For ease of interpretability, FIGURE ? displays error rates without disaggregating by complexity of model.

Our first lognormal distribution with parameters $(0, .25)$ has lower values of kurtosis and skewness. At 4 measurements, performance of the 4 DF methods is more

variable, but does converge and become more robust once the sample size increases. At sample size 26, all 4 methods yield robust error rates. In contrast, in 8 measurements, the performance of the methods is relatively stable, with t-as-z and standard DF method being slightly anti-conservative at smaller sample sizes. It appears that once the number of measurements has increased, the effect of increasing sample size is still positive, but less significant in this distribution.

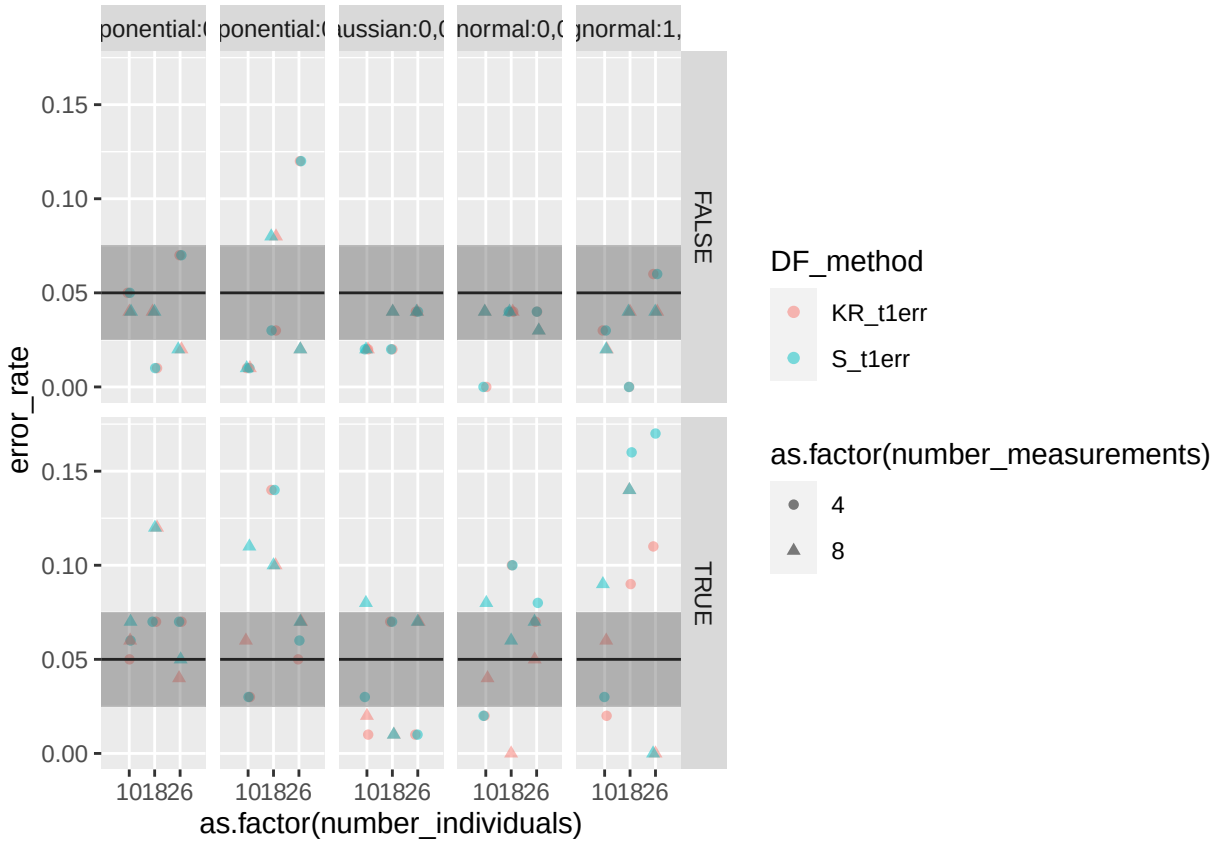
On the other hand, with higher levels of skewness and kurtosis with a lognormal distribution with parameters $(1, .5)$, the effect of number of measurements and sample size is much different. At 4 measurements, the robustness of the 4 measures seems to decrease as sample size increases. At a sample of size 10, the performance is very robust and conservative across all methods. However, at size 26 all methods are anti-conservative. While increasing to 8 measurements increases overall robustness, the difference is not extreme. On a similar note, while increasing the sample size does not lead to worse performance in DB methods as in the 4 measurements condition, it is not signficantly better compared to smaller sample sizes. At sample size 26, the KR and Satterthwaite methods are too conservative to be considered robust, while at sample size 10 they are robust. It appears that in this distribution, sample size and number of measurements appear not to have a strong effect on Type I error rates.

2.9.1 KR vs Satterthwaite

Comparing performance across all 4 methods has yielded signficant evidence that KR and Satterthwaite are superior methods when using linear mixed models on small samples.@luke_evaluating_2017 suggests that both KR and Satterthwaite are comparable solutions to obtain adequate Type I error. The following figure aims to narrow in on differences in performance between the two methods. Looking specifically at the effects of kurtosis and skewness, both KR and Satterthwaite methods tend to

produce more anti-conservative error rates in nonnormal distributions

			Sample Size				
			10		18		
			Random Intercept	Random Slope	Random Intercept	Random Slope	Ra
skew	kurtosis	DF_method	random_intercept_10	random_slope_10	random_intercept_18	random_slope_18	rand
Normal							
0.000	0.0	KR_t1err	0.020	0.015	0.03	0.040	
0.000	0.0	S_t1err	0.020	0.055	0.03	0.040	
Lognormal							
0.778	1.1	KR_t1err	0.020	0.030	0.04	0.050	
0.778	1.1	S_t1err	0.020	0.050	0.04	0.080	
1.750	5.9	KR_t1err	0.025	0.040	0.02	0.115	
1.750	5.9	S_t1err	0.025	0.060	0.02	0.150	
Exponential							
2.000	6.0	KR_t1err	0.028	0.050	0.04	0.108	
2.000	6.0	S_t1err	0.028	0.068	0.04	0.108	



2.9.2 KR Only

While KR method appears to be the most robust adjustment, (TABLE 5?) depicts its relatively variable performance across different conditions. Careful consideration must be used when conducting inference, and if possible, an increase in both sample size and number of measurements appears to ensure more robust results.

				Sample Size				
				10		18		
				Random Intercept	Random Slope	Random Intercept	Random Slope	F
params	number_measurements	skew	kurtosis	10_FALSE	10_TRUE	18_FALSE	18_TRUE	
Exponential								
0.2	4	2.000	6.0	0.05	0.05	0.01	0.07	
0.2	8	2.000	6.0	0.04	0.06	0.04	0.12	
0.9	4	2.000	6.0	0.01	0.03	0.03	0.14	
0.9	8	2.000	6.0	0.01	0.06	0.08	0.10	
Normal								
0,0.2	4	0.000	0.0	0.02	0.01	0.02	0.07	
0,0.2	8	0.000	0.0	0.02	0.02	0.04	0.01	
Lognormal								
0,0.25	4	0.778	1.1	0.00	0.02	0.04	0.10	
0,0.25	8	0.778	1.1	0.04	0.04	0.04	0.00	
1,0.5	4	1.750	5.9	0.03	0.02	0.00	0.09	
1,0.5	8	1.750	5.9	0.02	0.06	0.04	0.14	

2.10 Discussion

Can I compare across distributions?

Ultimately, these results strongly support using either KR or Satterthwaite degrees of freedom adjustments as opposed to methods aimed towards larger sample sizes.

Chapter 3 Application

3.1 Application to Longitudinal Study about Children's Health

In this chapter, we will apply linear mixed models to a longitudinal study about Children's Health, and explore how inference of the effects models can possibly change when using Kenward Roger versus Satterthwaite degrees of freedom approximation.

3.2 Background

The National Longitudinal Study of Adolescent to Adult Health is a longitudinal study spanning 1994 to 2008 that surveyed a U.S sample of students in 7-12th grade in the 1994-95 school year. Four waves of data were collected, in which the sample during the last wave was aged 24-32. Questions about mental health, socioeconomic status, and family background were collected, as well as physical measurements of height and weight.

One question of interest to consider is how salient life experiences that occur during adolescence, such as being exposed to alcohol or being in a physical altercation, may impact changes to one's physical health over time.

One way to capture physical health is through BMI, which is known to follow a skewed nonnormal distribution. With this in mind, we can employ methods of degrees

of freedom adjustment. While this dataset is large and encompasses approximately 5,000 students, the scope of this application will be narrowed in order to examine the performance of KR and Satterthwaite.

3.3 Intercept only model

3.4 Intercept and random slope

Chapter 4 Tables, Graphics, References, and Labels

4.1 Tables

In addition to the tables that can be automatically generated from a data frame in **R** that you saw in [R Markdown Basics] using the `kable` function, you can also create tables using *pandoc*. (More information is available at http://pandoc.org/README.html#tables.) This might be useful if you don't have values specifically stored in **R**, but you'd like to display them in table form. Below is an example. Pay careful attention to the alignment in the table and hyphens to create the rows and columns.

Table 4.1: Correlation of Inheritance Factors for Parents and Child

Factors	Correlation between Parents & Child	Inherited
Education	-0.49	Yes
Socio-Economic Status	0.28	Slight
Income	0.08	No
Family Size	0.18	Slight
Occupational Prestige	0.21	Slight

We can also create a link to the table by doing the following: Table 4.1. If you go back to [Loading and exploring data] and look at the `kable` table, we can create a

reference to this max delays table too: Table **??**. The addition of the `(\#tab:inher)` option to the end of the table caption allows us to then make a reference to Table `\@ref(tab:label)`. Note that this reference could appear anywhere throughout the document after the table has appeared.

4.2 Figures

If your thesis has a lot of figures, *R Markdown* might behave better for you than that other word processor. One perk is that it will automatically number the figures accordingly in each chapter. You'll also be able to create a label for each figure, add a caption, and then reference the figure in a way similar to what we saw with tables earlier. If you label your figures, you can move the figures around and *R Markdown* will automatically adjust the numbering for you. No need for you to remember! So that you don't have to get too far into LaTeX to do this, a couple **R** functions have been created for you to assist. You'll see their use below.

In the **R** chunk below, we will load in a picture stored as `amherst.png` in our main directory. We then give it the caption of "Amherst logo", the label of "amherstlogo", and specify that this is a figure. Make note of the different **R** chunk options that are given in the R Markdown file (not shown in the knitted document).

Figure 4.1: Amherst logo

Here is a reference to the Amherst logo: Figure 4.1. Note the use of the `fig:`

code here. By naming the **R** chunk that contains the figure, we can then reference that figure later as done in the first sentence here. We can also specify the caption for the figure via the R chunk option `fig.cap`.

Below we will investigate how to save the output of an **R** plot and label it in a way similar to that done above. Recall the `flights` dataset from Chapter 2. (Note that we've shown a different way to reference a section or chapter here.) We will next explore a bar graph with the mean flight departure delays by airline from Portland for 2014. Note also the use of the `scale` parameter which is discussed on the next page.

Here is a reference to this image: Figure **??**.

A table linking these carrier codes to airline names is available at https://github.com/ismayc/pnwflights14/blob/master/data/airlines.csv.

Next, we will explore the use of the `out.extra` chunk option, which can be used to shrink or expand an image loaded from a file by specifying `"scale= "`. Here we use the mathematical graph stored in the "subdivision.pdf" file.

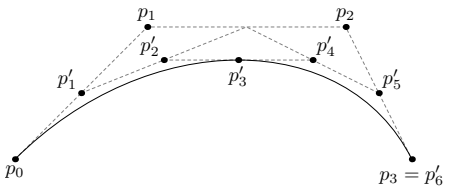


Figure 4.2: Subdiv. graph

Here is a reference to this image: Figure 4.2. Note that `echo=FALSE` is specified so that the **R** code is hidden in the document.

More Figure Stuff

Lastly, we will explore how to rotate and enlarge figures using the `out.extra` chunk option. (Currently this only works in the PDF version of the book.)

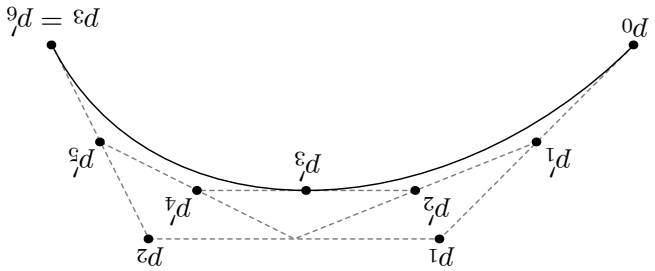


Figure 4.3: A Larger Figure, Flipped Upside Down

As another example, here is a reference: Figure 4.3.

4.3 Footnotes and Endnotes

You might want to footnote something.[1] The footnote will be in a smaller font and placed appropriately. Endnotes work in much the same way. More information can be

[1] footnote text

found about both on the Reed Thesis site https://www.reed.edu/cis/help/latex/thesis.html or feel free to reach out to Prof. Bailey at bebailey@amherst.edu.

4.4 Bibliographies

Of course you will need to cite things, and you will probably accumulate an armful of sources. There are a variety of tools available for creating a bibliography database (stored with the .bib extension). In addition to BibTeX suggested below, you may want to consider using the free and easy-to-use tool called Zotero. The Amherst librarians have created Zotero documentation at https://www.amherst.edu/library/find/citation/zotero. In addition, a tutorial is available from Middlebury College at http://sites.middlebury.edu/zoteromiddlebury/.

R Markdown uses *pandoc* (http://pandoc.org/) to build its bibliographies. One nice caveat of this is that you won't have to do a second compile to load in references as standard LaTeX requires. To cite references in your thesis (after creating your bibliography database), place the reference name inside square brackets and precede it by the "at" symbol. For example, here's a reference to a book about worrying: (Molina & Borkovec, 1994). This `Molina1994` entry appears in a file called `thesis.bib` in the `bib` folder. This bibliography database file was created by a program called BibTeX. You can call this file something else if you like (look at the YAML header in the main .Rmd file) and, by default, is to placed in the `bib` folder.

For more information about BibTeX and bibliographies, see the Reed College CUS site (http://web.reed.edu/cis/help/latex/index.html)[2]. There are three pages on this topic: *bibtex* (which talks about using BibTeX, at http://web.reed.edu/cis/help/latex/bibtex.html), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at http://web.reed.edu/cis/help/

[2]Reed College (2007)

latex/bibtexstyles.html) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at http://web.reed.edu/cis/help/latex/bibman.html). The last page will not be useful unless you have only a few sources.

If you look at the YAML header at the top of the main .Rmd file you can see that we can specify the style of the bibliography by referencing the appropriate csl file. You can download a variety of different style files at https://www.zotero.org/styles. Make sure to download the file into the csl folder.

Tips for Bibliographies

- Like with thesis formatting, the sooner you start compiling your bibliography for something as large as thesis, the better. Typing in source after source is mind-numbing enough; do you really want to do it for hours on end in late April? Think of it as procrastination.
- The cite key (a citation's label) needs to be unique from the other entries.
- When you have more than one author or editor, you need to separate each author's name by the word "and" e.g. `Author = {Noble, Sam and Youngberg, Jessica},`.
- Bibliographies made using BibTeX (whether manually or using a manager) accept LaTeX markup, so you can italicize and add symbols as necessary.
- To force capitalization in an article title or where all lowercase is generally used, bracket the capital letter in curly braces.
- You can add a Reed Thesis citation[3] option. The best way to do this is to use the phdthesis type of citation, and use the optional "type" field to enter "Reed thesis" or "Undergraduate thesis."

[3]Noble (2002)

4.5 Anything else?

If you'd like to see examples of other things in this template, please contact Professor Bailey (email bebailey@amherst.edu) with your suggestions.

Conclusion

If we don't want the conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

A.1 In the main file 4:

A.2 In Chapter 4:

Appendix B The Second Appendix

R code

Corrections

A list of corrections after submission to department.

Corrections may be made to the body of the thesis, but every such correction will be acknowledged in a list under the heading "Corrections," along with the statement "When originally submitted, this honors thesis contained some errors which have been corrected in the current version. Here is a list of the errors that were corrected." This list will be given on a sheet or sheets to be appended to the thesis. Corrections to spelling, grammar, or typography may be acknowledged by a general statement such as "30 spellings were corrected in various places in the thesis, and the notation for definite integral was changed in approximately 10 places." However, any correction that affects the meaning of a sentence or paragraph should be described in careful detail. The files samplethesis.tex and samplethesis.pdf show what the "Corrections" section should look like. Questions about what should appear in the "Corrections" should be directed to the Chair.

References

Angel, E. (2000). *Interactive computer graphics : A top-down approach with opengl*. Boston, MA: Addison Wesley Longman.

Angel, E. (2001a). *Batch-file computer graphics : A bottom-up approach with quicktime*. Boston, MA: Wesley Addison Longman.

Angel, E. (2001b). *Test second book by angel*. Boston, MA: Wesley Addison Longman.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in r. *Behavior Research Methods*, *49*(4), 1494–1502. http://doi.org/10.3758/s13428-016-0809-y

Molina, S. T., & Borkovec, T. D. (1994). The Penn State worry questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265–283). New York: Wiley.

Noble, S. G. (2002). *Turning images into simple line-art* (Undergraduate thesis). Reed College.

Reed College. (2007). LaTeX your document. Retrieved from http://web.reed.edu/cis/help/LaTeX/index.html