Methods for tests of fixed effects in small and nonnormal samples

In chapter 1, we outlined the process for conducting inference for models with repeated measures. When sample size is small, both Kenward-Roger (KR) and Sattherthwaite approximations have been implemented to reduce Type I error rates.

Section on inference for a coefficient

- likelihood ratio test

Satterthwaite

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

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

Note in this approximation we use the original Φ as the variance of $\hat{\beta}$. For the denominator degrees of freedom we perform spectral decomposition on $L'\Phi L = P'DP$, where D is a diagonal matrix of eigenvalues and P is an orthogonal matrix of eigenvectors. When r represents the rth row of $P'L$, we have $v_r = \frac{2(d_r)^2}{g_r'Wg_r}$, where g_r is a gradient vector, d_r is the rth 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.

Kenward-Roger

- Change variables to match chapter 1
- What does this look like for a single coefficient? (Not L, not multiple contrasts)

Kenward-Roger (1997) propose a Wald statistic 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 in chapter 1, $\hat{\Phi}$ is a biased estimator of Φ when samples are small, and underestimates. This adjusted estimator is broken down into $\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda}$, where $\hat{\Lambda}$ accounts for the amount of variation that was underestimated by the original estimator of covariance of $\hat{\beta}$. 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.

Existing literature

Both 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.

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.

Simulation Set up:

In this simulation, inference for fixed effects in a random intercepts and random intercepts and slopes model were conducted. For simplicity, only one discrete covariate, time, is used. The fixed effects included were the intercept value and time. Nonnormal data were generated using the `SimMultiCorrData` package. Using lognormal and exponential parameters (insert params here), skewness and kurtosis values were calculated, and used by the Fleishman method to calculate coefficients to transform normal data into the desired distribution. Parameters for the two distributions were chosen to cover a wide range of possible shapes and spreads. The nonnormal data generated are used as random effects values in order to create a response variable that also has a nonnormal distribution.The correlation between the two nonnormal values was set

at -.38, based on the linear mixed model fit on the application data used in CHAPTER 3. The number of individuals were 10, 18, and 26. The number of repeated measurements was either 4 or 8. In total, there are *48* different crossing conditions that were simulated.

After performing 10,000 replications of each condition at a signficance 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 Kenward-Roger and Sattherthwaite methods, stratified by distribution and other manipulated parameters.