Mixed Models

Joshua F. Wiley

2019-06-06

Contents

1 Statistical Inference from Linear Mixed Models (LMMs) 1

2 Random Intercept LMM 2

2.1 Diagnostics and Checks 5

2.2 Sample Write Up 8

3 Fixed Predictor LMM 8

3.1 Diagnostics and Checks 11

3.2 Sample Write Up 16

Download the raw `R` markdown code here `https://jwiley.github.io/MonashHonoursStatistics/LMM_Interpretation.rmd`.

```
## have R round to 2 significant digits
options(digits = 2)

library(data.table)
library(lme4)  ## load first
library(lmerTest)  ## install and load of df and p-values
library(ggplot2)
library(visreg)

## read in the dataset
d <- readRDS("aces_daily_sim_processed.RDS")
```

1 Statistical Inference from Linear Mixed Models (LMMs)

There is ambiguity in terms of how best to calculate degrees of freedom (df) for LMMs. By default `R` does not calculate the df and so does not provide p-values for the regression coefficients (fixed effects) from LMMs.

One easy, albeit imperfect, solution is to use the `lmerTest` package. `lmerTest` use Satterthwaite's method to calculate approximate degrees of freedom and use these for the t-tests and p-values for each regression coefficient. To use `lmerTest` simply make sure that **both** `lme4` and `lmerTest` packages are installed and that you load the `lmerTest` package after `lme4`, by using: `library(lmerTest)`. This is shown in the example above. Once that is done, all regular calls

to `lmer()` function used to fit LMMs will automatically have df estimated and p-values. This is done throughout this interpretation guide.

2 *Random Intercept LMM*

There are two main uses of intercept only models:

- To calculate the intraclass correlation coefficient (ICC)
- As a comparison to see how much better a more complex model fits. Note that for model comparisons, we need to use ML estimation, by setting `REML = FALSE`.

To calculate the ICC, we use this equation:

$$ICC = \frac{\sigma^2_{intercept}}{\sigma^2_{intercept} + \sigma^2_{residual}}$$

Following is an example of an intercept only model, where there is both a fixed effects intercept and a random intercept. The outcome variable is `PosAff`. All predictors come after the tilde, ~. In this case, the only "predictors" are the fixed and random intercept, represented by `1`. The random intercept is random by `UserID`. The function to fit linear mixed models is `lmer()` and comes from the `lme4` package. It also requires a dataset be specified, here `d`. Finally, there are two estimation approaches, both based off of Maximum Likelihood (ML) estimation. The default as it provides the least biased estimates is Restricted Maximum Likelihood (REML), chosen by default or by explicitly setting `REML = TRUE`. We can get a summary using `summary()`.

```
ri.m <- lmer(PosAff ~ 1 + (1 | UserID), data = d,
    REML = TRUE)

summary(ri.m)
```

```
## Linear mixed model fit by REML. t-tests
##   use Satterthwaite's method [lmerModLmerTest
## ]
## Formula: PosAff ~ 1 + (1 | UserID)
##    Data: d
##
## REML criterion at convergence: 14795
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -4.345 -0.647 -0.034  0.617  4.058
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  UserID   (Intercept) 0.629    0.793
##  Residual             0.529    0.727
## Number of obs: 6399, groups:  UserID, 191
##
## Fixed effects:
##             Estimate Std. Error       df
## (Intercept)   2.6787     0.0581 189.8310
##             t value Pr(>|t|)
## (Intercept)    46.1   <2e-16 ***
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1
```

There are four main "blocks" of output from the summary.

1. A repetition of the model options, formula we used, and dataset used. This is for records so you know exactly what the model was. In *this* model, it shows use that we fit a LMM using restricted maximum likelihood (REML) and that the degrees of freedom were approximated using Satterthwaite's method. The outcome variable is positive affect (`PosAff`) and there are only intercept predictors, `1`. The REML criterion at convergence is kind of like the log likelihood (LL), but unfortunately cannot be readily used to compare across models as easily as the actual LL (e.g., in AIC or BIC).
2. Scaled Pearson residuals. These are raw residuals divided by the estimated standard deviation, so that they can be roughly interpretted as z-scores. The minimum and maximum are useful for identifying whether there are outliers present in the model residuals. In *this* model, we can see that the lowest residual is -4.35 and the maximum residual is -4.35 which while a bit large, given there are thousands of observations are not so extreme if interpretted as z-scores as to be concerning. Absolute residuals of 10 or 20 would be large enough that they are extremely unlikely by chance alone and likely represent outliers.
3. Random effects. These show a summary of the random effects in the model. Random effects are basically always also fixed effects, so the random effects only shows the standard deviation and variance of random effects, plus, if applicable, their correlations. The means are showon in the fixed effects section. In the case of a random intercept only model like this one, there are only two random effects: (1) the random intercept and (2) the random residual. We

have both the standard deviation and variance of both. We will use the variances to calculate ICCs. In *this* model, the standard deviation of the random intercept, tells us that the average or typical difference between an individual's average positive affect, and the population average positive affect is 0.79. The standard deviation of the residuals tells us that the average or typical difference between an individual positive affect score and the predicted positive affect score is 0.73. The random effects section also tells us how many observations and unique people/groups went into the analysis. In *this* model we can see that we had 191 people providing 6399 unique observations.

4. Fixed effects. This section shows the fixed effects. It is a table, where each row is for a different effect / predictor and each column gives a different piece of information. The "Estimate" is the actual parameter estimate (i.e., THE fixed effect, the regression coefficient, etc.). The "Std. Error" is the standard error of the estimate, which captures uncertainty in the coefficient due to sampling variation. The "df" is the Satterthwaite estimated degrees of freedom. As an estimate, it may have decimals. The "t value" is the ratio of the coefficient to its standard error, that is: $t = \frac{Estimate}{StdError}$. The "Pr(>|t|)" is the p-value, the probability that by chance alone one would obtain as or a larger absolute t-value. The vertical bars indicate absolute values and the "Pr" stands for probability value. Note that `R` uses scientific E notation. The number following the "e" indicates how many places to the right (if positive) or left (if negative) the decimal point should be moved. For example, 0.001 could be written 1e-3. 0.00052 could be written 5.2e-4. These often are used for p-values which may be numbers very close to zero. In *this* model, we can see that the fixed effect for the intercept is 2.68 which is the like the mean of the random intercept and tells us the average level of positive affect, in this instance since there are no other predictors in the model.

Profile likelihood confidence intervals can be obtained using the `confint()` function. These confidence intervals capture the uncertainty in parameter estimates for both the fixed and random effects due to sampling variation. They do not capture indivdiual differences directly. Note that you only get confidence intervals for random effects when using the profile method, not when `method = "Wald"` although the Wald method is much faster.

```
ri.ci <- confint(ri.m, method = "profile", oldNames = FALSE)
```

```
## Computing profile confidence intervals ...
```

```
ri.ci
```

```
##                          2.5 % 97.5 %
## sd_(Intercept)|UserID  0.72   0.88
## sigma                  0.71   0.74
## (Intercept)            2.56   2.79
```

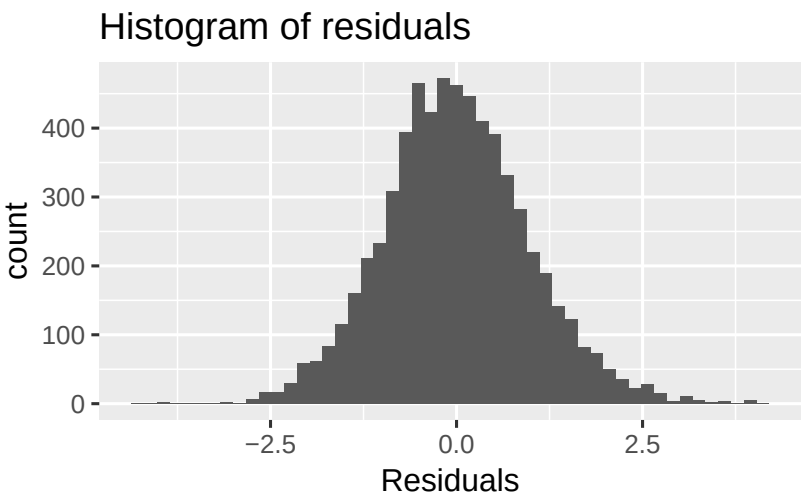
2.1 *Diagnostics and Checks*

Typical diagnostics and checks include checking for outliers, assessing whether the distributional assumptions are met, checking for homogeneity of variance and checking whether there is a linear association between predictors and outcome. With only an intercept, there is no need for checking whether a linear association is appropriate.

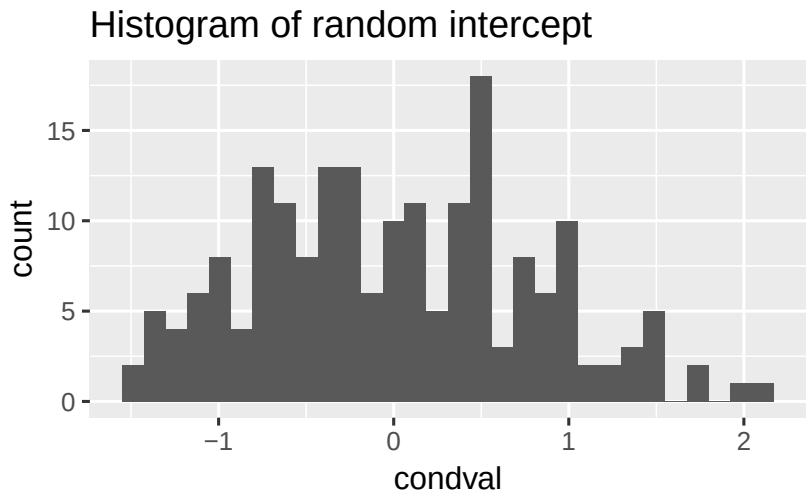
First we check for outliers on the residuals and the random intercept. These plots show some extreme values on the residuals and are somewhat unclear on the random intercept. In this case, using the scaled pearson residuals, which are roughly like z scores, the size of the residual outliers are not too big as to likely be an issue, particularly as we have thousands of observations.

```
res.d <- data.table(Residuals = residuals(ri.m,
    type = "pearson", scaled = TRUE), Yhat = fitted(ri.m))
ran.d <- as.data.table(ranef(ri.m))
```

```
ggplot(res.d, aes(Residuals)) + geom_histogram(bins = 50) +
    ggtitle("Histogram of residuals")
```

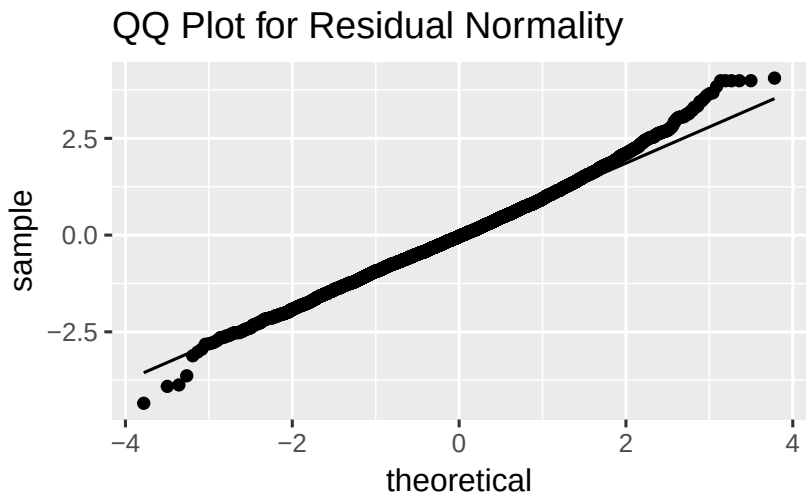



```
ggplot(ran.d, aes(condval)) + geom_histogram(bins = 30) +
    ggtitle("Histogram of random intercept")
```

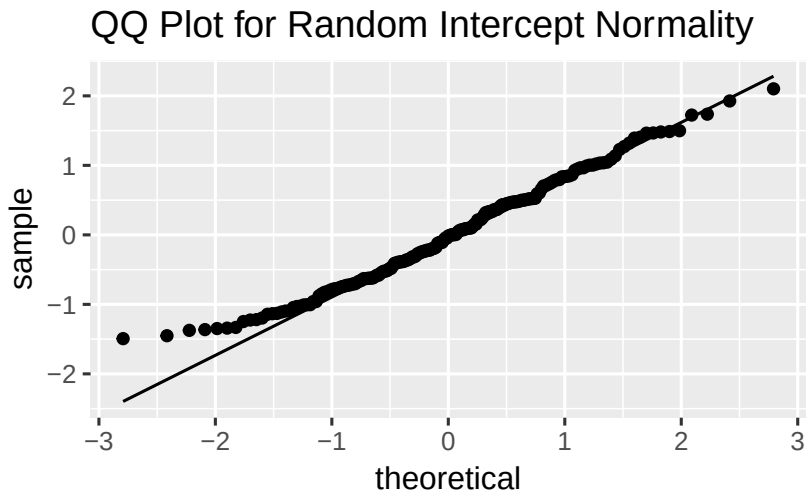


Next, we might check the distributional assumptions. We already have some information on this from the histograms, but QQ plots are helpful as well. The QQ plots indicate some non-normality, but it is not too extreme and probably close enough for inference.

```
ggplot(res.d, aes(sample = Residuals)) + stat_qq() +
    stat_qq_line() + ggtitle("QQ Plot for Residual Normality")
```

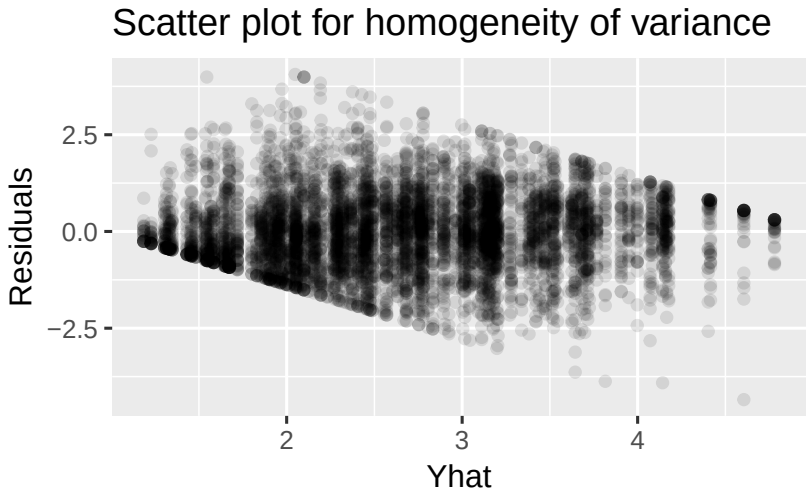



```
ggplot(ran.d, aes(sample = condval)) + stat_qq() +
    stat_qq_line() + ggtitle("QQ Plot for Random Intercept Normality")
```



Finally, we check the homogeneity of variance. The residuals show a characteristic banding when there are floor and ceiling effects. At low predicted values, positive affect cannot be any lower than 1, so you have small or positive residuals. At high predicted values, positive affect cannot be greater than 5 so you have small positive or negative residuals. This is responsible for the straight, angled lines at the extremes. Its not particularly clear whether the residual variance changes much across levels of the predited value (Yhat) so its not terrible evidence against homogeneity of variance. Unless easy alternatives were available (they are not) one would probably proceed.

```
ggplot(res.d, aes(Yhat, Residuals)) + geom_point(alpha = 0.1) +
    ggtitle("Scatter plot for homogeneity of variance")
```



2.2 *Sample Write Up*

An intercept only linear mixed model was fit to 6399 positive affect scores from 191 people. The intraclass correlation coefficient was 0.54 indicating that about half of the total variance in positive affect was between people and the other half is within person due to fluctuations across days. The fixed effect intercept revealed that the average [95% CI] positive affect was 2.68 [2.56, 2.79]. However, there were individual differences, with the standard deviation for the random intercept being 0.79 indicating that there are individual differences in the mean positive affect. Assuming the random intercepts follow a normal distribution, we expect most people to fall within one standard deviation of the mean, which in these data would be somewhere between: 1.89, 3.47.

3 *Fixed Predictor LMM*

Following is an example of a LMM with fixed effects and a random intercept (no random slopes). Although we did not explicitly add a fixed effects intercept by adding 1 to the equation, it is there by default. We still have a random intercept.

```
fp.m <- lmer(PosAff ~ STRESS + (1 | UserID), data = d,
    REML = TRUE)

summary(fp.m)
```

```
## Linear mixed model fit by REML. t-tests
##   use Satterthwaite's method [lmerModLmerTest
## ]
## Formula: PosAff ~ STRESS + (1 | UserID)
##    Data: d
##
## REML criterion at convergence: 13378
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.500 -0.658 -0.034  0.627  4.152
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  UserID   (Intercept) 0.529    0.727
##  Residual             0.423    0.650
## Number of obs: 6399, groups:  UserID, 191
##
```

```
## Fixed effects:
##              Estimate Std. Error        df
## (Intercept)  3.05e+00   5.41e-02  2.01e+02
## STRESS      -1.58e-01   3.97e-03  6.33e+03
##             t value Pr(>|t|)
## (Intercept)    56.5   <2e-16 ***
## STRESS        -40.0   <2e-16 ***
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## STRESS -0.173
```

There are four main "blocks" of output from the summary.

1. A repetition of the model options, formula we used, and dataset used. This is for records so you know exactly what the model was. In *this* model, it shows use that we fit a LMM using restricted maximum likelihood (REML) and that the degrees of freedom were approximated using Satterthwaite's method. The outcome variable is positive affect (`PosAff`) and stress is a predictor. The REML criterion at convergence is kind of like the log likelihood (LL), but unfortunately cannot be readily used to compare across models as easily as the actual LL (e.g., in AIC or BIC).
2. Scaled Pearson residuals. These are raw residuals divided by the estimated standard deviation, so that they can be roughly interpretted as z-scores. The minimum and maximum are useful for identifying whether there are outliers present in the model residuals. In *this* model, we can see that the lowest residual is -3.5 and the maximum residual is -3.5 which while a bit large, given there are thousands of observations are not so extreme if interpretted as z-scores as to be concerning. Absolute residuals of 10 or 20 would be large enough that they are extremely unlikely by chance alone and likely represent outliers. We can see there are some more extreme positive than negative residuals. That means that predictions are sometimes too (extremely) low rather than too (extremely) high.
3. Random effects. These show a summary of the random effects in the model. Random effects are basically always also fixed effects, so the random effects only shows the standard deviation and variance of random effects, plus, if applicable, their correlations. The means are showon in the fixed effects section. In the case of a model where the only random effect is the intercept, the random

effects show: (1) the random intercept and (2) the random residual. We have both the standard deviation and variance of both. In *this* model, the standard deviation of the random intercept, tells us that the average or typical difference between an individual's estimated positive affect when stress is 0, and the population average estimated positive affect when stress is 0 is 0.73. The standard deviation of the residuals tells us that the average or typical difference between an individual positive affect score and the predicted positive affect score is 0.65. The random effects section also tells us how many observations and unique people/groups went into the analysis. In *this* model we can see that we had 191 people providing 6399 unique observations.

4. Fixed effects. This section shows the fixed effects. It is a table, where each row is for a different effect / predictor and each column gives a different piece of information. The "Estimate" is the actual parameter estimate (i.e., THE fixed effect, the regression coefficient, etc.). The "Std. Error" is the standard error of the estimate, which captures uncertainty in the coefficient due to sampling variation. The "df" is the Satterthwaite estimated degrees of freedom. As an estimate, it may have decimals. The "t value" is the ratio of the coefficient to its standard error, that is: $t = \frac{Estimate}{StdError}$. The "Pr(>|t|)" is the p-value, the probability that by chance alone one would obtain as or a larger absolute t-value. The vertical bars indicate absolute values and the "Pr" stands for probability value. Note that `R` uses scientific E notation. The number following the "e" indicates how many places to the right (if positive) or left (if negative) the decimal point should be moved. For example, 0.001 could be written 1e-3. 0.00052 could be written 5.2e-4. These often are used for p-values which may be numbers very close to zero. In *this* model, we can see that the fixed effect for the intercept is 3.05 which is like the mean of the random intercept and tells us the average estimated positive affect score when stress = 0. The fixed effect (regression coefficient) for STRESS is -0.16 which tells us how much on average (fixed effect) lower positive affect is expected to be when stress is one unit higher.

Profile likelihood confidence intervals can be obtained using the `confint()` function. These confidence intervals capture the uncertainty in parameter estimates for both the fixed and random effects due to sampling variation. They do not capture indivdiual differences directly. Note that you only get confidence intervals for random effects when using the profile method, not when `method = "Wald"` although the Wald method is much faster.

```
fp.ci <- confint(fp.m, method = "profile", oldNames = FALSE)
```

```
## Computing profile confidence intervals ...
```

```
fp.ci
```

```
##                         2.5 % 97.5 %
## sd_(Intercept)|UserID  0.66   0.81
## sigma                  0.64   0.66
## (Intercept)            2.95   3.16
## STRESS                -0.17  -0.15
```

3.1 Diagnostics and Checks

Typical diagnostics and checks include checking for outliers, assessing whether the distributional assumptions are met, checking for homogeneity of variance and checking whether there is a linear association between predictors and outcome. With only an intercept, there is no need for checking whether a linear association is appropriate.

Since we can check whether there is a linear association of stress or not, it can be worth checking first. This is something of a chicken and egg situation, though, because a non-linear association can be driven by outliers, but poor normality or outliers on the residuals also can be driven by the wrong functional form. I normally begin by checking linearity / functional form. For model comparisons, we want `REML = FALSE` and fit consecutive models with increasingly complicated stress polynomials. Note that `poly()` does not allow missing values, so we need to address that. Its easiest to create a base model and then update.

```
fp0.m <- lmer(PosAff ~ 1 + (1 | UserID), data = d[!is.na(STRESS)],
    REML = FALSE)

fp1.m <- update(fp0.m, . ~ . + poly(STRESS, 1))
fp2.m <- update(fp0.m, . ~ . + poly(STRESS, 2))
fp3.m <- update(fp0.m, . ~ . + poly(STRESS, 3))
fp4.m <- update(fp0.m, . ~ . + poly(STRESS, 4))

AIC(fp0.m, fp1.m, fp2.m, fp3.m, fp4.m)
```

```
##       df   AIC
## fp0.m  3 14797
## fp1.m  4 13372
## fp2.m  5 13341
## fp3.m  6 13342
## fp4.m  7 13344
```

```
BIC(fp0.m, fp1.m, fp2.m, fp3.m, fp4.m)
```

```
##       df   BIC
## fp0.m  3 14817
## fp1.m  4 13399
## fp2.m  5 13374
## fp3.m  6 13383
## fp4.m  7 13391
```

In *this* case, the model shows that `fp2.m` is the best based on BIC and is close but still best by AIC (for both AIC and BIC, lower values are better). Let's look at another summary and confidence intervals. However, as REML estimates are less biased, for reporting, we might use those.

```
fp2.m <- update(fp2.m, REML = TRUE)
summary(fp2.m)
```

```
## Linear mixed model fit by REML. t-tests
##   use Satterthwaite's method [lmerModLmerTest
## ]
## Formula:
## PosAff ~ (1 | UserID) + poly(STRESS, 2)
##    Data: d[!is.na(STRESS)]
##
## REML criterion at convergence: 13332
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.584 -0.655 -0.032  0.633  4.204
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  UserID   (Intercept) 0.526    0.725
##  Residual             0.421    0.649
## Number of obs: 6399, groups:  UserID, 191
##
## Fixed effects:
##                   Estimate Std. Error
## (Intercept)         2.6830     0.0531
## poly(STRESS, 2)1  -32.1557     0.7968
## poly(STRESS, 2)2    4.1826     0.7195
##                         df t value Pr(>|t|)
## (Intercept)       189.6495   50.54  < 2e-16
## poly(STRESS, 2)1 6328.8002  -40.36  < 2e-16
## poly(STRESS, 2)2 6269.6259    5.81  6.4e-09
##
```

```
## (Intercept)      ***
## poly(STRESS, 2)1 ***
## poly(STRESS, 2)2 ***
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) p(STRESS,2)1
## p(STRESS,2)1 -0.002
## p(STRESS,2)2  0.001 -0.062
```

```
fp2.ci <- confint(fp2.m, method = "profile", oldNames = FALSE)
```

```
## Computing profile confidence intervals ...
```

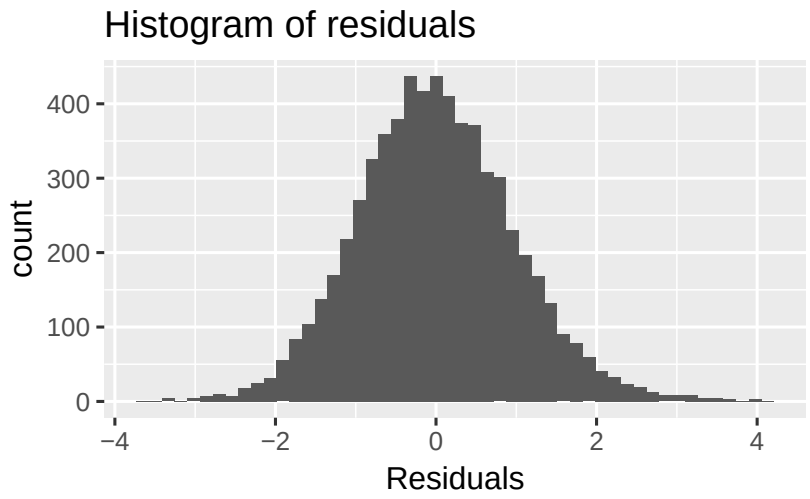
```
fp2.ci
```

```
##                          2.5 % 97.5 %
## sd_(Intercept)|UserID     0.65   0.80
## sigma                     0.64   0.66
## (Intercept)               2.58   2.79
## poly(STRESS, 2)1        -33.72 -30.59
## poly(STRESS, 2)2          2.77   5.59
```

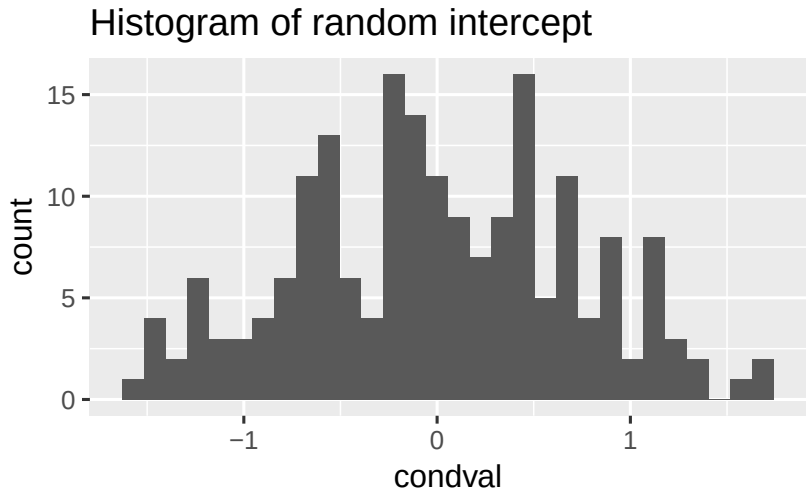
Since a different model is "optimal", we will proceed with that for testing. We check for outliers on the residuals and the random intercept. These plots show some extreme values on the residuals and are somewhat unclear on the random intercept. In this case, using the scaled pearson residuals, which are roughly like z scores, the size of the residual outliers are not too big as to likely be an issue, particularly as we have thousands of observations. There is a small positive tail, which potentially we could seek to exclude or winsorize, but in this case I would not.

```
res.d <- data.table(Residuals = residuals(fp2.m,
    type = "pearson", scaled = TRUE), Yhat = fitted(fp2.m))
ran.d <- as.data.table(ranef(fp2.m))
```

```
ggplot(res.d, aes(Residuals)) + geom_histogram(bins = 50) +
    ggtitle("Histogram of residuals")
```

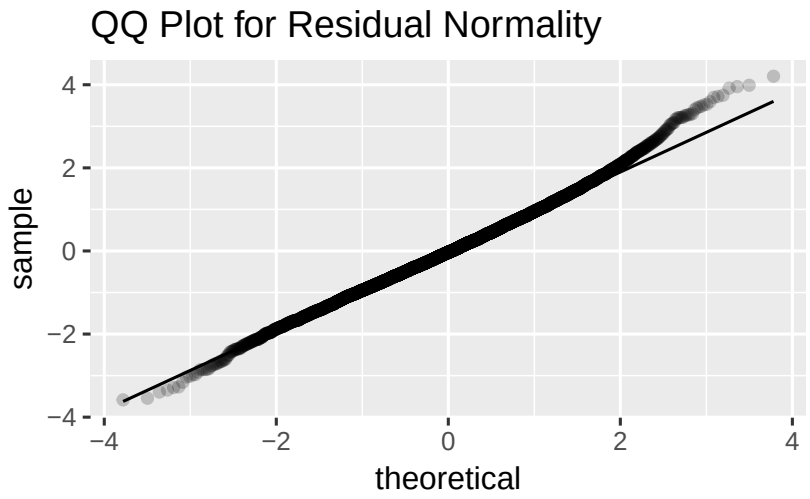



```
ggplot(ran.d, aes(condval)) + geom_histogram(bins = 30) +
    ggtitle("Histogram of random intercept")
```

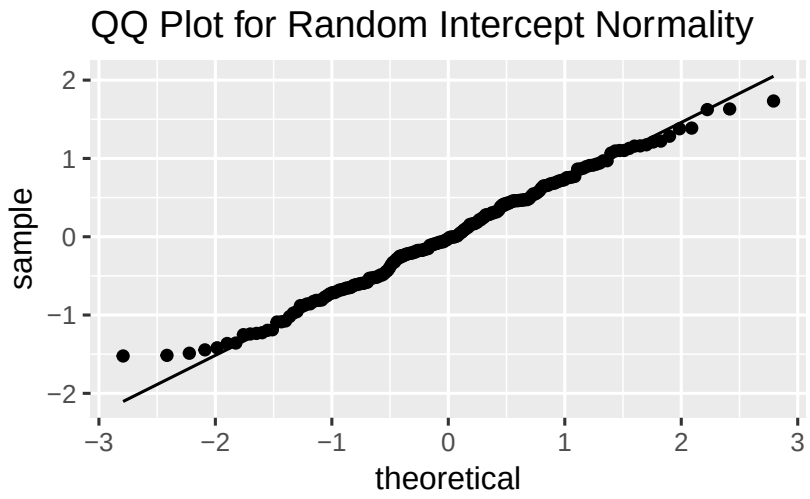


Next, we might check the distributional assumptions. We already have some information on this from the histograms, but QQ plots are helpful as well. The QQ plots indicate only very modest non-normality, but it is not too extreme and probably close enough for inference.

```
ggplot(res.d, aes(sample = Residuals)) + stat_qq(alpha = 0.2) +
    stat_qq_line() + ggtitle("QQ Plot for Residual Normality")
```



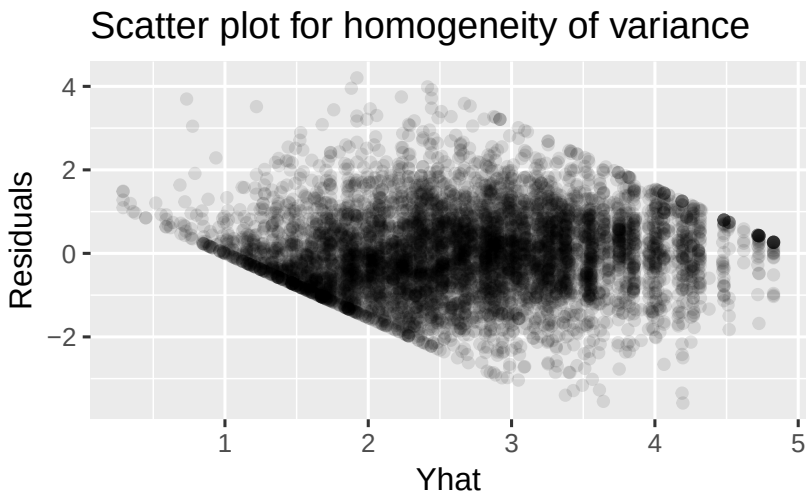

```
ggplot(ran.d, aes(sample = condval)) + stat_qq() +
    stat_qq_line() + ggtitle("QQ Plot for Random Intercept Normality")
```



Finally, we check the homogeneity of variance. The residuals show a characteristic banding when there are floor and ceiling effects. At low predicted values (particularly below 1), positive affect cannot be any lower than 1, so you *must* have positive residuals. At high predicted values, positive affect cannot be greater than 5 so you have small positive or negative residuals. This is responsible for the straight, angled lines at the extremes. Its not particularly clear whether the residual variance changes much across levels of the predited value (Yhat) so its not terrible evidence against homogeneity of variance. Unless easy alternatives were available (they are not) one would probably proceed.

```
ggplot(res.d, aes(Yhat, Residuals)) + geom_point(alpha = 0.1) +
```

```
ggtitle("Scatter plot for homogeneity of variance")
```



3.2 *Sample Write Up*

To examine the association of stress and positive affect, a linear mixed model was fit. As the nature of the stress and affect relationships was not known, we used the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) to compare models with orthogonal polynomials of stress with degrees 1 to 4. Both BIC and AIC pointed to the two degree polynomial as the best fit, indicating that there is a quadratic association between stress and positive affect. The final model included 6399 positive affect scores from 191 people. The fixed effect intercept revealed that the average [95% CI] positive affect when stress is 0 was 2.68 [2.58, 2.79]. However, there were individual differences, with the standard deviation for the random intercept being 0.72 indicating that there are individual differences in the mean positive affect. Assuming the random intercepts follow a normal distribution, we expect most people to fall within one standard deviation of the mean, which in these data would be somewhere between: 1.96, 3.41. Using Satterthwaite's approximation for degrees of freedom revealed that both the linear and quadratic aspects of stress were statistically significantly associated with positive affect (both $p < .001$). As it is difficult to interpret coefficients from orthogonal polynomials, a graph showing average (fixed effect) association of stress with positive affect is shown below. The graph shows that higher stress is associated with lower positive affect scores. There is a slightly faster drop in positive affect when stress is low and it begins to plateau at higher levels of stress, although the difference across the observed range of stress (0 to 10) is modest.

```
visreg(fp2.m, xvar = "STRESS", partial = FALSE,
    rug = FALSE, xlab = "Stress scores", ylab = "Predicted Positive Affect")
```