ACRM 2021 Longitudinal Data Analysis Workshop

Keith Lohse, PhD, PStat, and Allan Kozlowski, PhD, BsC (PT)

Practical Session 2: Conditional Models and Hypothesis Testing

In Practical Session 1, we saw how we could improve the fit of our mixed-effects model by adding polynomials to our linear regression. As researchers, however, we are typically interested in whether or not trajectories differ between groups, not simply determining what the best trajectory is overall. This involves testing **conditional** models where we make the estimate of the various slopes and intercepts depend on which group you are in (i.e., conditioned on groups).

As before, you will need to open the five packages we will be using for this session using the library function:

```
# Loading the essential libraries.
library("ggplot2"); library("lme4"); library("car"); library("dplyr"); library("lmerTest");
```

If you have not already installed these packages, you will need to use the install.packages() function first. This can take some time and will require an internet connection.

```
# If these packages are not installed already, run the following code:
install.packages("ggplot2"); install.packages("lme4"); install.packages("car"); install.packages("dplyr
```

2.1 Reading in the Data and Plotting the Trajectories in each Group

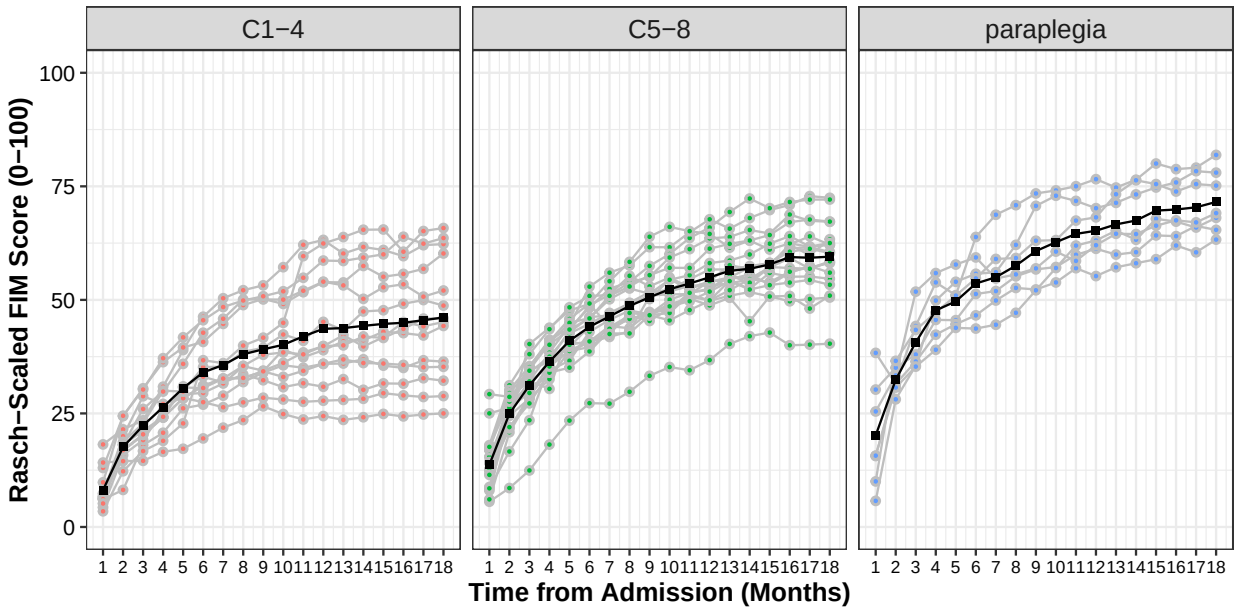
In these hypothetical spinal cord injury data, we have groups of patients classified based on their AIS Grade (where AIS is the American Spinal Injury Association [ASIA] Impairment Scale). AIS Grades are recorded as C1-4 injuries, C5-8 injuries, and patients with paraplegia. In the figure, we present spaghetti plots for each group where the data for each patient are shown in grey, but the group-average data are shown in black. As you can see, there is considerable variability within each group, but in looking at the averages, it seems like there might be some difference between groups as well.

```
# We will read these data in from the web:
DATA <- read.csv("https://raw.githubusercontent.com/keithlohse/LMER_Clinical_Science/master/ACRM_2018/da
                 stringsAsFactors = TRUE, na.strings=c("NA","NaN"," ",""))
# Use the head() function to check the structure of the data file.
head(DATA)
```

```
##   subID month time sex female bmi_z rasch_FIM rasch_FIM_MAR rasch_FIM_MNAR
## 1   s01     1 1.16   f      1 -0.44  12.56000      12.56000       12.56000
## 2   s01     2 2.58   f      1 -0.44  24.85447      24.85447       24.85447
## 3   s01     3 3.69   f      1 -0.44  40.27863            NA       40.27863
## 4   s01     4 4.92   f      1 -0.44  43.86893            NA       43.86893
## 5   s01     5 5.16   f      1 -0.44  47.66534            NA       47.66534
## 6   s01     6 6.63   f      1 -0.44  49.11310      49.11310             NA
```

```
##   rasch_FIM_LOCF inpatient_time los AIS_grade therapy_hours PT_hours OT_hours
## 1       12.56000           5wks  32      C5-8       14.6000   7.3000   7.4000
## 2       24.85447           5wks  32      C5-8       14.7000   7.3000   7.4000
## 3       40.27863           5wks  32      C5-8       14.6000   7.3000   7.3000
## 4       43.86893           5wks  32      C5-8       14.5000   7.2000   7.3000
## 5       47.66534           5wks  32      C5-8        8.3366   4.1683   4.2254
## 6       47.66534           5wks  32      C5-8            NA       NA       NA
##   SUM_PT_hours SUM_OT_hours AVE_PT_hours AVE_OT_hours
## 1      33.2683      33.6254      6.65366      6.72508
## 2      33.2683      33.6254      6.65366      6.72508
## 3      33.2683      33.6254      6.65366      6.72508
## 4      33.2683      33.6254      6.65366      6.72508
## 5      33.2683      33.6254      6.65366      6.72508
## 6      33.2683      33.6254      6.65366      6.72508
```

```
ggplot(DATA, aes(x = month, y = rasch_FIM)) +
  geom_line(aes(group=subID), col="grey") +
  geom_point(aes(fill=AIS_grade), col="grey", pch=21, size=1, stroke=1) +
  stat_summary(fun=mean, geom="point", shape=15, col="black")+
  stat_summary(fun=mean, geom="line", col="black")+
  facet_wrap(~AIS_grade) +
  scale_x_continuous(name = "Time from Admission (Months)", breaks=c(0:18)) +
  scale_y_continuous(name = "Rasch-Scaled FIM Score (0-100)",limits=c(0,100)) +
  theme_bw() + theme(axis.text.x=element_text(size=8, colour="black"),
                     axis.text.y=element_text(size=10, colour="black"),
                     axis.title=element_text(size=12,face="bold")) +
  theme(strip.text.x = element_text(size = 12))+ theme(legend.position="none")
```




```
## ------------------------------------------------------------------------
```

2.2 Creating a Conditional Model

Recall from Practical Session 1 that the best fitting unconditional model was the **quadratic random slopes** model. We will recreate that model below and make it our starting point. After creating the model, we will pass the model object to the *anova()* function to get an analysis of deviance table with Type III calculations of the change in deviance. Note that by default the *anova()* gives Type I calculations, but with the **lmerTest** package installed, we will get Type III calculations with F-values that use Satterthwaite's method to estimate the approximate denominator degrees of freedom.

Hopefully you are familiar with the difference between Type I, II, and III sums of squared errors in ordinary regression. The exact same concepts apply here, but rather than using the sum of squared errors, this table is based on the **deviance** estimated through maximum likelihood estimation. As a quick reminder, the Type III formulation excludes all shared variance from our predictors and the outcome. Thus, we have to interpret any single predictor as **controlling for** all other predictors in the model. Additionally, this means we need to be concerned about potential **collinearity** between predictors because their shared variance is excluded. If any of those concerns feel unfamiliar, it would be helpful for you to go back and review concepts from multiple regression, especially types of sums of squared error calculations.

```
DATA$year.0 <- (DATA$month-1)/12


raneff_quad_rand<-lmer(rasch_FIM~
                         # Fixed-effects
                         1+year.0+I(year.0^2)+
                         # Random-effects
                         (1+year.0+I(year.0^2)|subID), data=DATA, REML=FALSE)
anova(raneff_quad_rand)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##             Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## year.0      5630.2  5630.2     1 39.998  577.23 < 2.2e-16 ***
## I(year.0^2) 4450.0  4450.0     1 39.998  456.23 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Alternatively, if we want to see the actual coefficients for the fixed-effects and/or see the estimated variances and correlations for the random-effects, we can pass the model object to the *summary()* function.

```
summary(raneff_quad_rand)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: rasch_FIM ~ 1 + year.0 + I(year.0^2) + (1 + year.0 + I(year.0^2) |
##     subID)
##    Data: DATA
##
##      AIC      BIC   logLik deviance df.resid
##   4039.3   4085.0  -2009.6   4019.3      710
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4893 -0.5194  0.0455  0.5442  3.0358
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept)  46.224   6.799
##           year.0      242.058  15.558    0.07
##           I(year.0^2)  44.137   6.644   -0.08 -0.93
##  Residual               9.754   3.123
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   18.103      1.120  39.999   16.17   <2e-16 ***
## year.0        64.044      2.666  39.998   24.03   <2e-16 ***
## I(year.0^2)  -26.956      1.262  39.998  -21.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) year.0
## year.0      -0.030
## I(year.0^2)  0.049 -0.921
```

Looking at the *Estimates* from the fixed-effect output, we can see that the regression equation that defines the overall trajectory is:

$$\hat{y_{ij}} = 18.103 + 64.04(Year.0_{ij}) - 26.956(Year.0_{ij}^2)$$

However, our goal at the start of this was to get an estimate of the trajectory in each group and see if the groups differed from each other! To do that, we will want to add a fixed-effect of AIS_grade to our model as well as its interactions with the time variables. To do this in R, we can use the "*" operator to get the main effects and interactions of both variables. That is, if we write $year.0 * AIS\_grade$ we will get the main-effect of Year.0, the main-effect of AIS_grade, and the interaction between them.

Alternatively, if we wanted to specify all three of these components by hand, we could write $year.0 + AIS\_grade + year.0 : AIS\_grade$, where the ":" operator defines the interaction by itself.

```
cond_mod_01<-lmer(rasch_FIM~
                          # Fixed-effects
                          1+year.0*AIS_grade+I(year.0^2)*AIS_grade+
                          # Random-effects
                          (1+year.0+I(year.0^2)|subID), data=DATA, REML=FALSE)
```

After creating the model, let us first check the analysis of deviance table:

```
anova(cond_mod_01)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                       Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
## year.0                5325.6  5325.6     1    40 546.0008 < 2.2e-16 ***
## AIS_grade              418.4   209.2     2    40  21.4490 4.681e-07 ***
## I(year.0^2)           3967.7  3967.7     1    40 406.7800 < 2.2e-16 ***
## year.0:AIS_grade        37.0    18.5     2    40   1.8979    0.1631
## AIS_grade:I(year.0^2)   15.9     8.0     2    40   0.8154    0.4497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next, let's use the *summary()* function to look at each of the individual coefficients:

```
summary(cond_mod_01)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: rasch_FIM ~ 1 + year.0 * AIS_grade + I(year.0^2) * AIS_grade +
##     (1 + year.0 + I(year.0^2) | subID)
##    Data: DATA
##
##      AIC      BIC   logLik deviance df.resid
##   4010.8   4084.1  -1989.4   3978.8      704
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6725 -0.5150  0.0474  0.5764  3.0172
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept)  20.267   4.502
##           year.0      217.415  14.745   -0.26
##           I(year.0^2)  41.639   6.453    0.14 -0.93
##  Residual               9.754   3.123
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##                                 Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                       12.285      1.315  40.000   9.343 1.32e-11
## year.0                            57.576      4.306  40.000  13.371 2.40e-16
## AIS_gradeC5-8                      6.858      1.733  40.000   3.958 0.000302
## AIS_gradeparaplegia               14.632      2.277  40.000   6.425 1.19e-07
## I(year.0^2)                      -24.867      2.091  40.000 -11.893 1.04e-14
## year.0:AIS_gradeC5-8               8.858      5.675  40.000   1.561 0.126417
## year.0:AIS_gradeparaplegia        12.917      7.458  40.000   1.732 0.091005
## AIS_gradeC5-8:I(year.0^2)         -2.925      2.756  40.000  -1.062 0.294777
## AIS_gradeparaplegia:I(year.0^2)   -3.995      3.622  40.000  -1.103 0.276548
##
## (Intercept)                      ***
## year.0                           ***
## AIS_gradeC5-8                    ***
## AIS_gradeparaplegia              ***
## I(year.0^2)                      ***
## year.0:AIS_gradeC5-8
## year.0:AIS_gradeparaplegia       .
## AIS_gradeC5-8:I(year.0^2)
## AIS_gradeparaplegia:I(year.0^2)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) year.0 AIS_gC5-8 AIS_gr I(.0^2 y.0:AIS_C yr.0:AIS_ AIS_C5-8:
## year.0      -0.352
## AIS_grdC5-8 -0.759  0.267
## AIS_grdprpl -0.577  0.203  0.438
```

```
## I(year.0^2)  0.269 -0.921 -0.204    -0.155
## y.0:AIS_C5-  0.267 -0.759 -0.352    -0.154  0.699
## yr.0:AIS_gr  0.203 -0.577 -0.154    -0.352  0.532  0.438
## AIS_C5-8:I( -0.204  0.699  0.269     0.118 -0.759 -0.921    -0.404
## AIS_:I(.0^2 -0.155  0.532  0.118     0.269 -0.577 -0.404    -0.921     0.438
```

2.3 Unpacking the Conditional Model

Before we consider the statistical significance of these effects, lets first take the coefficients from the full model and then simplify the equation to get the unique equation in each group. By default, R uses treatment coding where the reference group is coded as 0 and other groups are coded as 1. The precise treatment coded variable is identified in the R output. Thus, *AIS_gradeC5-8* is a treatment coded variable where $C5-8$ is coded a 1 and all other groups are coded as 0. Similarly, *AIS_gradeparaplegia* is a treatment coded variable where *paraplegia* is coded as 1 and all other groups are coded as 0. To save space, I will refer to these treatment-coded variables as T1 and T2, respectively. This equation will still get a little long, but I am really just copying the *Estimates* from the fixed-effects section of the output, with some rearranging of terms to keep similar terms together.

$$\hat{y_{ij}} = 12.285 + 6.858(T1) + 14.632(T2) +$$
$$57.576(year.0) - 24.867(year.0^2) +$$
$$8.858(year.0 \times T1) + 12.917(year.0 \times T2) - 2.925(year.0^2 \times T1) - 3.995(year.0^2 \times T2)$$

For individuals in the **reference group** which is *AIS_grade* = C1-4, we can then plug in 0's for the treatment coded variables and simplify the equation to get the curvilinear trajectory in the C1-4 group:

$$\hat{y_{ij}} = 12.285 + 6.858(0) + 14.632(0) +$$
$$57.576(year.0) - 24.867(year.0^2) +$$
$$8.858(year.0 \times 0) + 12.917(year.0 \times 0) - 2.925(year.0^2 \times 0) - 3.995(year.0^2 \times 0)$$

Anything times 0 is 0 of course, so we can simply drop those terms from the model and we are left with:

$$\hat{y_{ij}} = 12.285 + 57.576(year.0) - 24.867(year.0^2)$$

For individuals in the **C5-8** group, we can then plug in 1's for T1 and 0's for T2 and then simplify the equation:

$$\hat{y_{ij}} = 12.285 + 6.858(1) + 14.632(0) +$$
$$57.576(year.0) - 24.867(year.0^2) +$$
$$8.858(year.0 \times 1) + 12.917(year.0 \times 0) - 2.925(year.0^2 \times 1) - 3.995(year.0^2 \times 0)$$

Again, anything with a 0 can be dropped and multiplying by 1 turns some of our variables into constants. We can move some terms around again to show how the slopes and intercepts get updated by the coded variables:

$$\hat{y_{ij}} = 12.285 + 6.858 + 57.576(year.0) + 8.858(year.0) - 24.867(year.0^2) - 2.925(year.0^2)$$

which ultimately simplifies to:

$$\hat{y_{ij}} = 19.143 + 66.434(year.0) - 27.792(year.0^2)$$

Similarly then for individuals in the **paraplegic** group, we can then plug in 0's for T1 and 1's for T2 and then simplify the equation:

$$\hat{y_{ij}} = 12.285 + 6.858(0) + 14.632(1) +$$
$$57.576(year.0) - 24.867(year.0^2) +$$
$$8.858(year.0 \times 0) + 12.917(year.0 \times 1) - 2.925(year.0^2 \times 0) - 3.995(year.0^2 \times 1)$$

Again, anything with a 0 can be dropped and multiplying by 1 turns some of our variables into constants. We can move some terms around again to show how the slopes and intercepts get updated by the coded variables:

$$\hat{y_{ij}} = 12.285 + 14.632(1) + 57.576(year.0) + 12.917(year.0) - 24.867(year.0^2) - 3.995(year.0^2)$$

which ultimately simplifies to:

$$\hat{y_{ij}} = 26.917 + 70.493(year.0) - 28.862(year.0^2)$$

Thus, hidden within our larger model, we actually have three separate trajectories being estimated. **The C1-4 group:**

$$\hat{y_{ij}} = 12.285 + 57.576(year.0) - 24.867(year.0^2)$$

The C5-8 group:

$$\hat{y_{ij}} = 19.143 + 66.434(year.0) - 27.792(year.0^2)$$

The paraplegic group:

$$\hat{y_{ij}} = 26.917 + 70.493(year.0) - 28.862(year.0^2)$$

2.4 Hypothesis Testing in the Mixed-Effects Model

Armed as we now are with the actual coefficients in each group, let us look at the hypothesis tests actually being conducted by both the *anova()* function and the *summary()* function.

To get **ombnibus** effects, we can look at the output from the *anova()* function.

```
anova(cond_mod_01)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                       Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
## year.0                5325.6  5325.6     1    40 546.0008 < 2.2e-16 ***
## AIS_grade              418.4   209.2     2    40  21.4490 4.681e-07 ***
## I(year.0^2)           3967.7  3967.7     1    40 406.7800 < 2.2e-16 ***
## year.0:AIS_grade        37.0    18.5     2    40   1.8979    0.1631
## AIS_grade:I(year.0^2)   15.9     8.0     2    40   0.8154    0.4497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From that table, we can see that there is a statistically significant main-effect of $Year.0$, $F(1, 40) = 546.00, p < 0.001$, main-effect of $Year.0^2$, $F(1, 40) = 406.78, p < 0.001$, and main-effect of $AISGrade$, $F(1, 40) = 21.45, p < 0.001$. There are a few things to note here when interpreting the main-effects. First, the effects $Year.0$ and $Year.0^2$ refer to change over time on average across groups **and** each of these tests uses only a single degree of freedom, so there is no need for post-hoc tests to unpack potential differences. Second, the main-effect of $AISGrade$ reflects the difference between groups **at the intercept** before the $Time \times Group$ interactions. Additionally, the main-effect of $AISGrade$ requires two degrees of freedom

(as there are three groups and therefor $K-1$ treatment coded variables required to capture the group differences).

Thus, the main-effects tell us about the effects of time on average and the group differences on average. Potentially complicating these relationships are the $Time \times Group$ interactions. In this case, however, neither the $Year.0 \times AISGrade$ interaction, $F(1, 40) = 1.90, p = 0.163$ nor the $Year.0^2 \times AISGrade$ interaction, $F(1, 40) = 0.82, p = 0.450$ were statistically significant.

To look at the significance of individual coefficients, we can use the output from the *summary()* function. As an important reminder, the way you code your categorical variables should not affect the *anova()* output (provided you are using appropriate dummy codes, contrast codes, etc), but you coding scheme will have a significant influence on the individual coefficients of the regression model, and thus the output of the summary function.

As a thought experiment, consider that if we had made individuals with paraplegia the reference group rather than individuals with C1-4 injury. All of the models predictions would stay the same, because the model would still predict the same values for each group and each time, so the ANOVA table would not change. However, the default intercept, linear slope, and quadratic slope would now be the group with paraplegia, so all of those values (and the t-tests for those coefficients) would be different. Always be careful to consider how your variables were coded when when interpreting a regression output!

```
summary(cond_mod_01)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: rasch_FIM ~ 1 + year.0 * AIS_grade + I(year.0^2) * AIS_grade +
##     (1 + year.0 + I(year.0^2) | subID)
##    Data: DATA
##
##      AIC      BIC   logLik deviance df.resid
##   4010.8   4084.1  -1989.4   3978.8      704
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6725 -0.5150  0.0474  0.5764  3.0172
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept)  20.267   4.502
##           year.0      217.415  14.745   -0.26
##           I(year.0^2)  41.639   6.453    0.14 -0.93
##  Residual               9.754   3.123
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##                                 Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                       12.285      1.315  40.000   9.343 1.32e-11
## year.0                            57.576      4.306  40.000  13.371 2.40e-16
## AIS_gradeC5-8                      6.858      1.733  40.000   3.958 0.000302
## AIS_gradeparaplegia               14.632      2.277  40.000   6.425 1.19e-07
## I(year.0^2)                      -24.867      2.091  40.000 -11.893 1.04e-14
## year.0:AIS_gradeC5-8               8.858      5.675  40.000   1.561 0.126417
## year.0:AIS_gradeparaplegia        12.917      7.458  40.000   1.732 0.091005
## AIS_gradeC5-8:I(year.0^2)         -2.925      2.756  40.000  -1.062 0.294777
## AIS_gradeparaplegia:I(year.0^2)   -3.995      3.622  40.000  -1.103 0.276548
```

```
##
## (Intercept)                     ***
## year.0                          ***
## AIS_gradeC5-8                   ***
## AIS_gradeparaplegia             ***
## I(year.0^2)                     ***
## year.0:AIS_gradeC5-8
## year.0:AIS_gradeparaplegia      .
## AIS_gradeC5-8:I(year.0^2)
## AIS_gradeparaplegia:I(year.0^2)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) year.0 AIS_gC5-8 AIS_gr I(.0^2 y.0:AIS_C yr.0:AIS_ AIS_C5-8:
## year.0      -0.352
## AIS_grdC5-8 -0.759  0.267
## AIS_grdprpl -0.577  0.203  0.438
## I(year.0^2)  0.269 -0.921 -0.204    -0.155
## y.0:AIS_C5-  0.267 -0.759 -0.352    -0.154  0.699
## yr.0:AIS_gr  0.203 -0.577 -0.154    -0.352  0.532  0.438
## AIS_C5-8:I( -0.204  0.699  0.269     0.118 -0.759 -0.921    -0.404
## AIS_:I(.0^2 -0.155  0.532  0.118     0.269 -0.577 -0.404    -0.921     0.438
```