Prepared by Braden Hansma and Birsen Donmez, University of Toronto

PART III – Build Linear Mixed Effect Models

In Part II, we imported the data Donmez.csv to R. As a reminder, Experience and SecondaryTask are between subject variables that are crossed. Participants (denoted by Participant_ID) completed four scenarios so they have four data points each. Thus, participant is a random factor.

In this section we will analyze the variable TTFG (time until first glance) through linear mixed effects modeling.

Linear mixed effect models can be built hierarchically. Below, we start with a fixed intercept model, followed by a random intercept model, and then add fixed effects. At each step, the **anova()** function is used to compare the different models. This hierarchical model fitting is done here as an exercise. Normally when we analyze an experimental dataset, we would build the model to include all effects of interest rather than performing forward or backward selection. Further, with each model built, assumptions have to be checked.

Use the **plot(**your_model_name**)** and **qqnorm(**your_model_name**, abline = c(0,1))** functions to create residual plots. We skip this step below but this has to be done for every model built.

You can read more about each function mentioned below using **help(**function_name**)**.

7. Fixed intercept model

> F.Int <- gls(TTFG ~ 1, data = data, method = "ML",na.action = na.omit)
> summary(F.Int)

The **gls()** function fits linear models using generalized least squares. In this function, a single mean is fit to the data (i.e. grand mean). **method** specifies the type of estimation the function will use (Maximum Likelihood Estimation in this case), and the argument **na.action = na.omit** will allow you to omit missing data in the calculation of the model parameters. The **Summary()** function will allow you to view the model fit and estimated parameters.

8. Random Intercept model

> R.Int <- lme(TTFG ~ 1, random = ~1|Participant_ID, data = data, method = "ML", na.action = na.omit)
> summary(R.Int)

The **lme()** function fits linear-mixed effects models. Here, we are regressing TTFG on an intercept, that is set to vary across participants (i.e. estimating a mean glance time for each participant, as well as a grand mean). The random component is specified by the argument **random = ~1|Participant_ID.** This lets intercepts vary across participants (i.e., participant is introduced as a random factor). Figure 2 shows the **summary()** of this model.

The **anova()** function can be used to test whether intercepts are random (i.e., participant is a random factor). This is comparable to testing the intraclass correlation; if participant is significant, it means between participant variance is large. You can examine the residual term to see the within person variance. Here, the between group variance is approx. 1.044, and the within group is approx. 25.40. Based on the equation for the intraclass correlation coefficient (1.044/25.40 + 1.044), the intraclass correlation coefficient is small (0.04). Therefore, random slopes may not be significant, which is confirmed in Figure 3. However, this finding does not mean that we should drop the random participant term from our model since we have not yet built the model that captures this experiment.

```
Linear mixed-effects model fit by maximum likelihood
 Data: data
       AIC      BIC      logLik
  770.1731 778.6581 -382.0866

Random effects:
 Formula: ~1 | Participant_ID
         (Intercept) Residual
StdDev:    1.022747 5.046603

Fixed effects: TTFG ~ 1
                Value Std.Error DF  t-value p-value
(Intercept) 5.371554 0.4883678 93 10.99899       0

Standardized Within-Group Residuals:
        Min         Q1        Med         Q3        Max
-1.0968331 -0.8564662 -0.2579190  0.6524451  3.7999043

Number of Observations: 125
Number of Groups: 32
```

Figure 2: Summary(R.Int). The model parameters for the random intercept model. The highlighted box shows the standard deviation estimates for the random components (intercepts, i.e. participants, and the error term).

> **anova (F.Int, R.Int)**

```
        Model df      AIC      BIC    logLik   Test   L.Ratio p-value
F.Int       1  2 768.4453 774.1019 -382.2226
R.Int       2  3 770.1731 778.6581 -382.0866 1 vs 2 0.2721107  0.6019
```

Figure 3: anova() function comparing the fixed intercept to random intercept model; the random intercept model is not a significant improvement, meaning participants are not a significant random effect.

Another function available to you to explore your model is **coef()**, which will display the estimated parameters for each effect for each participant.

> **coef(R.Int)**

9. Mixed effects model (random intercept + fixed effects)

> **Mixed <- lme(TTFG ~ Experience*SecondaryTask, data = data, random = ~1|Participant_ID, method="ML", na.action = na.omit)**
> **anova(R.Int_F.Eff, type ="m")**
> **summary(R.Int_F.Eff.Int)**

The fixed effects, experience (Experienced = 0, Novice = 1), secondary task (0 = No, 1 = Yes) and their interaction are entered in the above model. Our summary shows that the Experience:SecondaryTask interaction is not significant ($p < 0.69$); the following model omits the interaction variable.

> **R.Int_F.Eff<-lme(TTFG ~ Experience + SecondaryTask, data = data, random = ~1|Participant_ID, method="ML", na.action = na.omit)**
> **summary(R.Int_F.Eff)**

```
Linear mixed-effects model fit by maximum likelihood
 Data: data
       AIC      BIC     logLik
  760.0169 774.1585 -375.0085

Random effects:
 Formula: ~1 | Participant_ID
         (Intercept) Residual
StdDev: 0.0002547661 4.860441

Fixed effects: TTFG ~ Experience + SecondaryTask
                  Value Std.Error DF  t-value p-value
(Intercept)    3.018244 0.7562168 93 3.991241  0.0001
ExperienceNovice 2.674752 0.8803704 29 3.038212  0.0050
SecondaryTaskY   2.036185 0.8803704 29 2.312873  0.0280
 Correlation:
                 (Intr) ExprnN
ExperienceNovice -0.573
SecondaryTaskY   -0.563 -0.024

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3        Max
-1.5902223 -0.8386748 -0.2232123  0.6592308  4.0442842

Number of Observations: 125
Number of Groups: 32
```

Figure 4: Results from a random intercept, fixed effect model.

Results show that experience and secondary task, are both significant at a $p < 0.05$. The beta weights indicate that novice drivers took 2.67 seconds longer (on the average) than experienced drivers to look at an anticipatory cue after it became available. Further, the beta weight for secondary task indicated that drivers in the secondary task condition took 2.04 seconds longer (on the average) to look at an anticipatory cue compared to drivers not in that condition.

10. Mixed effects model (different variance/covariance structures for within group variance)

Lastly, the variance-covariance structure for the repeated measures (i.e. scenarios) within the model can be specified. Below we try CS and AR(1). The default for the lme() function is to use an independence model.

> **Mixed.CS<-lme(TTFG ~ Experience + SecondaryTask, data = data, random = ~1|Participant_ID, correlation = corCAR1(form = ~ 1|Participant_ID), method="ML", na.action = na.omit)**
> **summary(Mixed.CS)**

> **Mixed.AR<-lme(TTFG ~ Experience + SecondaryTask, data = data, random = ~1|Participant_ID, correlation = corCAR1(form = ~ 1|Participant_ID), method="ML", na.action = na.omit)**
> **summary(Mixed.AR)**

Compare the three models you built (Mixed, Mixed.CS, Mixed.AR) in terms of which one is the best fit to this data? Don't forget to check the residual plots for each model. Are there any assumptions violated? If so, how can this be handled.