

LMER Methods in Longitudinal (Repeated Measures) Analysis

2018-03-18

Preparation

We will use the data in the *seasonal-depression.csv* file. These data include the Beck depression scores in four different seasons for 14 males under 35 years of age. The source of these data is: Myers & Well (2003). We will use these data to explore seasonal depression. The attributes in the dataset include:

- subject: The subject ID number for each male
- s1: The Beck depression score in winter (season 1)
- s2: The Beck depression score in spring (season 2)
- s3: The Beck depression score in summer (season 3)
- s4: The Beck depression score in fall (season 4)

The data are displayed below:

subject	season	beck
1	s1	7.50
1	s2	11.55
1	s3	1.00
1	s4	1.21
2	s1	7.00
2	s2	9.00
2	s3	5.00
2	s4	15.00
3	s1	1.00
3	s2	1.00
3	s3	0.00
3	s4	0.00

To carry out a longitudinal analysis, we fit a mixed-effects regression model to the data. The equivalent LMER model to the RM-ANOVA model we fitted in the previous notes is simply an unconditional means model. Before we fit this model, let's examine the data structure needed to fit this model.

Data Structure: Tidy/Long Data

To fit this model, we need the data in a *tidy* or *long* format. In the tidy format the outcome variable (beck) is in a single column as is the time predictor (season). Because we have repeated measurements on subjects, there is more than one row for each subject. For example in our data, each subject has four rows of data. The data object *seasonal* is in the proper tidy format for fitting a LMER model.

When longitudinal data is being entered in Excel (or some other program), it is often stored in a *wide* or *untidy* format. In this format, each row corresponds to a different subject. Below is an example of the seasonal data in the *wide* format.

subject	s1	s2	s3	s4
1	7.50	11.55	1.000	1.21
2	7.00	9.00	5.000	15.00
3	1.00	1.00	0.000	0.00
4	0.00	0.00	0.000	0.00
5	1.06	0.00	1.097	4.00
6	1.00	2.50	0.000	2.00
7	2.50	0.00	0.000	2.00
8	4.50	1.06	2.000	2.00
9	5.00	2.00	3.000	5.00
10	2.00	3.00	4.208	3.00
11	7.00	7.35	5.877	9.00
12	2.50	2.00	0.009	2.00
13	11.00	16.00	13.000	13.00
14	8.00	10.50	1.000	11.00

Notice how each row corresponds to a different subject. This is not tidy data because the outcome we are interested in, Beck Depression Scores, are not in a single column. In the wide format, the outcome is spread out over multiple columns. Often there are not columns for the predictor either; they are typically the column name.

The wide format is easier to enter the data in, but the tidy (or long) format is how we need to analyze the data. The library **tidyr** has two functions, `gather()` and `spread()`, that convert data between these two formats. Below, I show the code for going from the wide format (data object is `seasonal_wide`) to the tidy format.

```
seasonal_long = seasonal_wide %>%
  gather(key = season, value = "beck", s1:s4) %>%
  arrange(subject, season)

head(seasonal_long, 12)
```

subject	season	beck
1	s1	7.50
1	s2	11.55
1	s3	1.00
1	s4	1.21
2	s1	7.00
2	s2	9.00
2	s3	5.00
2	s4	15.00
3	s1	1.00
3	s2	1.00
3	s3	0.00
3	s4	0.00

For more information about using these functions, google “tidyr” and read through any number of great tutorials or vignettes; for example [here](#) or [here](#). You can also read Hadley Wickham’s original [paper on tidy data](#).

Unconditional Means Model

Once the data are in the tidy format, we can fit a linear mixed-effects model to predict variation in the outcome; Beck depression scores in this example. Here I use the `display()` function from the **arm** package to simplify the output. (You could also use `summary()`.) To get the variance components, square the standard deviation values.

```
lmer.0 = lmer(beck ~ 1 + (1|subject), data = seasonal, REML = FALSE)
arm::display(lmer.0)

## lmer(formula = beck ~ 1 + (1 | subject), data = seasonal, REML = FALSE)
##   coef.est  coef.se
##      4.13     1.00
##
## Error terms:
##   Groups   Name      Std.Dev.
##  subject (Intercept) 3.52
##  Residual              2.46
## ---
## number of obs: 56, groups: subject, 14
## AIC = 296.9, DIC = 291
## deviance = 290.9
```

Fitting the unconditional means model gives us our baseline comparison model. The variance components suggest that there is residual variation ($\hat{\sigma}_\epsilon^2 = 6.07$) at level-1 (time; within-student) and at level-2 (between-student; $\hat{\sigma}_{\text{Subject}}^2 = 12.40$). Most of the unexplained variation seems to be between-student variation (67.1%).

Unconditional Growth Model (with RE of Intercept)

We can now add the time predictor to the model. In this data set, the time predictor is season, which is a categorical predictor. We could create dummy variables, or simply add season into the model (and let R pick the reference season).

```
## lmer(formula = beck ~ 1 + season + (1 | subject), data = seasonal,
##      REML = FALSE)
##           coef.est coef.se
## (Intercept)  4.29     1.12
## seasons2     0.42     0.84
## seasons3    -1.70     0.84
## seasons4     0.65     0.84
##
## Error terms:
##   Groups   Name      Std.Dev.
##  subject (Intercept) 3.56
##  Residual              2.22
## ---
## number of obs: 56, groups: subject, 14
## AIC = 294.2, DIC = 282
## deviance = 282.2
```

In looking at the output, the reference group is s1 (winter). Although there are no p -values, we can evaluate “significance” by evaluating the size of the t -values. Here, it looks like the mean Beck depression score in s3 (summer) is lower than the mean Beck depression score in the reference season (winter). The mean Beck depression scores in fall and spring do not seem to differ from that in winter.

Since season varies by timepoint (it is a within-student predictor), we expect that we should have explained some of the within-student variation. The within-student variance component in this model is smaller than in the unconditional means model ($\hat{\sigma}_e^2 = 4.93$); a reduction of 18.8%.

The between-student variance component is larger than in the unconditional means model ($\hat{\sigma}_{\text{Subject}}^2 = 12.68$); an increase of 2.26%. This is a mathematical artifact, and does not really signify an increase in unexplained variation.

Replicating the RM-ANOVA using a Mixed-Effects Model

The RM-ANOVA's output can be replicated by fitting the unconditional growth model (with a RE of intercept) using restricted maximum likelihood (REML) estimation.

We want to essentially use output from this model to re-create the F -test to evaluate differences in season from the RM-ANOVA. The ANOVA table from that analysis was:

Table 4: ANOVA Table for the Effects of Seasons and Subjects

Source	df	SS	MS	F	p
Subjects	13	779.0	59.93	11.30	<.001
Seasons	3	47.8	15.93	3.00	.042
Subjects x Seasons	39	207.0	5.31		

In this analysis, we:

- Computed the F -value for this test by dividing the MS for season by the MS for the subject by season interaction. Since there is only one measurement per cell of the subject by season interaction, this is akin to the MS error.
- Evaluate the F -value in an F -distribution with the appropriate df ; 3 and 39.

Remember that Mean Squares are Variance Components. Examining the variance components from the LMER model:

```
arm::display(lmer.1.reml)
```

```
## lmer(formula = beck ~ 1 + season + (1 | subject), data = seasonal,
##      REML = TRUE)
##           coef.est coef.se
## (Intercept)  4.29      1.16
## seasons2      0.42      0.87
## seasons3     -1.70      0.87
## seasons4      0.65      0.87
##
## Error terms:
## Groups   Name      Std.Dev.
## subject (Intercept) 3.70
## Residual                2.30
## ---
## number of obs: 56, groups: subject, 14
## AIC = 288.4, DIC = 288
## deviance = 282.2
```

The variance component for the residual ($\hat{\sigma}_e^2 = 5.31$) is the same as the MS value for error (subject by season interaction) from the RM-ANOVA. To get the variance component for season, we can use the `anova()` function on the fitted LMER model.

```
anova(lmer.1.reml)
```

	Df	Sum Sq	Mean Sq	F value
season	3	47.8	15.9	3

This gives us the MS for season of 15.9. It also computes the relevant F -value of 3 ($\frac{15.9}{5.31} = 3$). Note also that we are given the df value for season (numerator of the F -value). To obtain the denominator (error) df we need to do some quick computation.

- The error df is equivalent to the df in the subject by season interaction. This df can be computed by multiplying the subject df by the season df . Since we have 14 subjects, the subject df is $s-1 = 14-1 = 13$. There are four seasons, so the season $df = 4 - 1 = 3$. The interaction df are then $13 \times 3 = 39$.

We can now compute the p -value associated with the F -value of 3.

```
1 - pf(3, df1 = 3, df2 = 39)
```

```
## [1] 0.0421
```

This is the same p -value we obtained using the RM-ANOVA!

Better and Easier p -Value for Season

So long as the assumptions of the LMER model have been met, we can obtain a more powerful p -value for the effect of season. As a side benefit, this is also easier to compute. The way we do this is by taking advantage of the fact that the unconditional means model is a nested model of the unconditional growth model. If we have nested models, they can be compared using a *Likelihood Ratio Test*. To carry out this test, we use the `anova()` function and input the two mixed-effects models we want to compare. (Note: Both models need to be fitted with ML.)

```
anova(lmer.0, lmer.1)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
lmer.0	3	297	303	-145	291	NA	NA	NA
lmer.1	6	294	306	-141	282	8.72	3	0.033

The null hypothesis being tested is that the reduced model (`lmer.0`) and the full model (`lmer.1`) fit the data equally well. The way that we measure fit is via the deviance. The deviance of the reduced model is 291 and that for the full model is 282. If the two models fit equally well, we would expect the difference in deviances to be zero. The actual difference in deviances is 8.72. (This is often referred to as ΔG^2 , for goodness-of-fit, or as χ^2 .) This indicates that the fuller model fits the sample data slightly better (smaller deviance).

As with any difference, we wonder whether this is within what would be expected because of sampling (chance) variation. To test this, we evaluate ΔG^2 in a χ^2 -distribution with df equal to the difference in K between the two models (K is the df for each model). This difference should be the difference in the complexity between the two models; the difference in the estimated number of parameters. Our reduced model has three parameters being estimated ($\hat{\beta}_0$, $\hat{\sigma}_\epsilon^2$, and $\hat{\sigma}_{b_0}^2$), and our full model has six parameters being estimated ($\hat{\beta}_0$, $\hat{\beta}_{\text{Spring}}$, $\hat{\beta}_{\text{Summer}}$, $\hat{\beta}_{\text{Fall}}$, $\hat{\sigma}_\epsilon^2$, and $\hat{\sigma}_{b_0}^2$). The difference in complexity between these models is $6 - 3 = 3$.

```
1 - pchisq(8.72, df = 3)
```

```
## [1] 0.0333
```

Note that all of these results are given in the `anova()` output. This is typically reported as something like:

A likelihood ratio test indicated that the model that included the fixed-effects of season fitted the data significantly better than the unconditional means model, $\chi^2(3) = 8.72$, $p = 0.033$.

Why is this called a likelihood ratio test? Remember that the deviance is equal to -2Log-Likelihood . Thus the difference in deviances can be written as:

$$\Delta G^2 = -2 \ln [\mathcal{L}(\text{Reduced Model})] - \left[-2 \ln [\mathcal{L}(\text{Full Model})] \right]$$

Pulling out the -2 we get

$$\Delta G^2 = -2 \left[\ln [\mathcal{L}(\text{Reduced Model})] - \ln [\mathcal{L}(\text{Full Model})] \right]$$

The difference between two logarithms, e.g., $\log(A) - \log(B)$ is the logarithm of the quotient ($\log(\frac{A}{B})$). Thus, we can re-write this as,

$$\Delta G^2 = -2 \ln \left[\frac{\mathcal{L}(\text{Reduced Model})}{\mathcal{L}(\text{Full Model})} \right]$$

Now it should be a little more apparent why this test is called a likelihood ratio test. Note that the only way the difference in deviances between the two models turns out to be zero is if the two models' likelihoods are both equal. This would indicate that both models fit the data equally, and thus we should adopt the reduced model (Occam's Razor).

LMER: A More Flexible Model for Repeated Measures Data

One advantage to using LMER to analyze repeated measures data over RM-ANOVA is that the regression model allows for both categorical and quantitative variables. For example, rather than code our seasons categorically (as s1, s2, s3 and s4), which was a necessity in days of yore, we could have simply coded them as 0, 1, 2, and 3. Then we could have fitted the LMER using this numerical predictor. To illustrate this, I will first convert s1 to 0, s2 to 1, etc.

```
lookup_table = data.frame(
  season = c("s1", "s2", "s3", "s4"),
  season2 = c(0, 1, 2, 3)
)

seasonal = left_join(seasonal, lookup_table, by = "season")
head(seasonal)
```

subject	season	beck	season2
1	s1	7.50	0
1	s2	11.55	1
1	s3	1.00	2
1	s4	1.21	3
2	s1	7.00	0
2	s2	9.00	1

Now, I will use the `season2` numerical predictor in the LMER. Before I do, however, I will examine a plot of the data.

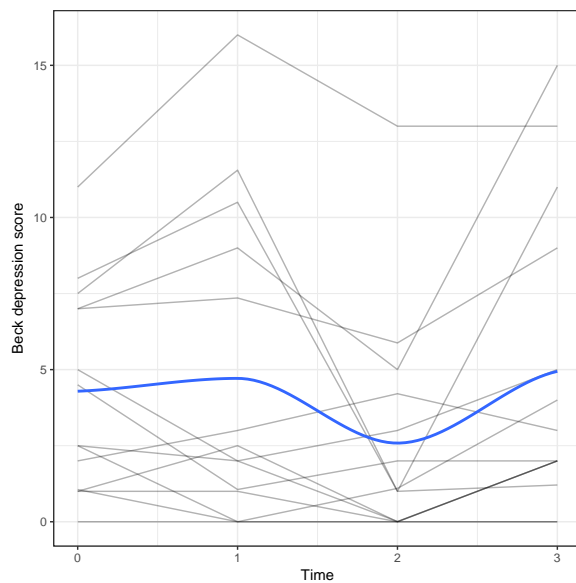


Figure 1: Spaghetti plot of the depression scores by season for each of the 14 subjects. The loess smoother is also displayed.

The smoothed profile suggests that the Beck depression scores change over time, and that the change pattern may be nonlinear (potentially cubic). Because our time predictor is now numeric, we can fit non-linear models quite easily. Below we fit the LMER models to examine a linear, quadratic, and cubic change pattern in the average depression score.

```
lmer.l = lmer(beck ~ 1 + season2 + (1|subject), data = seasonal, REML = FALSE)
lmer.q = lmer(beck ~ 1 + season2 + I(season2^2) + (1|subject), data = seasonal, REML = FALSE)
lmer.c = lmer(beck ~ 1 + season2 + I(season2^2) + I(season2^3) + (1|subject), data = seasonal, REML = FALSE)
```

Since these are nested models, we can perform a series of LRTs to evaluate whether the higher-order polynomial models fit better than the lower-order models. We can also include the unconditional means model in this evaluation

```
anova(lmer.0, lmer.l, lmer.q, lmer.c)
```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
lmer.0	3	297	303	-145	291	NA	NA	NA
lmer.l	4	299	307	-145	291	0.003	1	0.955
lmer.q	5	299	309	-144	289	2.221	1	0.136
lmer.c	6	294	306	-141	282	6.499	1	0.011

The first χ^2 -test (on the `lmer.l` line) compares the linear growth model to the unconditional means model (the model on the previous line). The second χ^2 -test (on the `lmer.q` line) compares the quadratic growth model to the linear growth model, and the third χ^2 -test (on the `lmer.c` line) compares the cubic growth model to the quadratic growth model. This series of tests suggests that of these models, the cubic growth model should be adopted.

Note that if you tried to fit a higher-order polynomial to these data you get an error.

```
lmer.quartic = lmer(beck ~ 1 + season2 + I(season2^2) + I(season2^3) + I(season2^4) + (1|subject), data = seasonal)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

That is because with only four time points the highest-order polynomial model you can fit is a cubic (x^3). When you max out the polynomial fit, we call this a *saturated* model. WARNING: Fitting a saturated model typically overfits the data.

Interpretation

To give an example of the interpretation of a LMER model with continuous (numerical) time predictors, let's go back to the linear growth model (despite its non significant time predictor).

```
arm::display(lmer.l)
```

```
## lmer(formula = beck ~ 1 + season2 + (1 | subject), data = seasonal,
##      REML = FALSE)
##               coef.est coef.se
## (Intercept)    4.16     1.09
## season2       -0.02     0.29
##
## Error terms:
## Groups      Name      Std.Dev.
## subject (Intercept) 3.52
## Residual              2.46
## ---
## number of obs: 56, groups: subject, 14
## AIC = 298.9, DIC = 291
## deviance = 290.9
```

We will focus on the interpretations of the fixed-effects of time, since they are of the most importance when we have longitudinal data. Here the linear effect of time is

- The average Beck depression score when `season2 = 0` is predicted to be 4.16. In other words, the predicted Beck depression score in winter is 4.16.
- Each one-unit difference in `season2` is associated with a decrease in the predicted average Beck depression score of 0.02. Or, each subsequent season, Beck depression score decreases by 0.02, on average.

To interpret the cubic model, we would plot the predicted values over time and interpret the pattern in the plot.

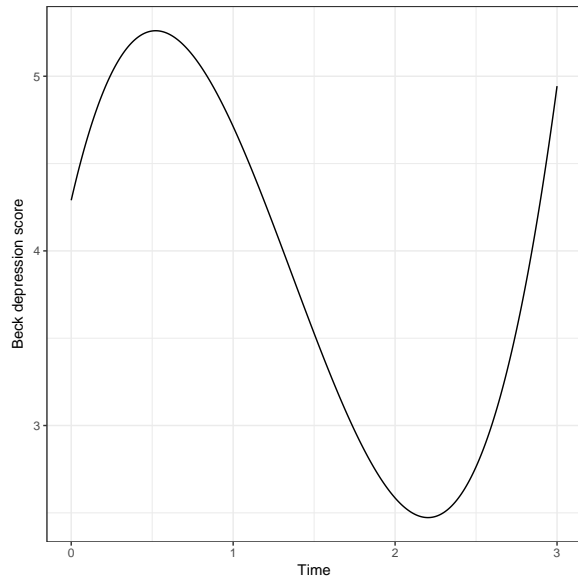


Figure 2: Spaghetti plot of the depression scores by season for each of the 14 subjects. The loess smoother is also displayed.

Changing the Time Metric

What if we had measured the time metric using weeks instead of just indexing them as 0, 1, 2, and 3? For example:

- Winter (index = 0): Week 0
- Spring (index = 1): Week 13
- Summer (index = 2): Week 26
- Fall (index = 3): Week 39

```
lookup_table = data.frame(
  season2 = c(0, 1, 2, 3),
  season3 = c(0, 13, 26, 39)
)
```

```
seasonal = left_join(seasonal, lookup_table, by = "season2")
head(seasonal)
```

subject	season	beck	season2	season3
1	s1	7.50	0	0
1	s2	11.55	1	13
1	s3	1.00	2	26
1	s4	1.21	3	39
2	s1	7.00	0	0
2	s2	9.00	1	13

```
lmer.l2 = lmer(beck ~ 1 + season3 + (1|subject), data = seasonal, REML = FALSE)
summary(lmer.l2)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: beck ~ 1 + season3 + (1 | subject)
## Data: seasonal
##
##      AIC      BIC    logLik deviance df.resid
##      299      307     -145     291      52
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.5322 -0.4491 -0.0656  0.3169  2.6619
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## subject (Intercept) 12.40     3.52
## Residual              6.06     2.46
## Number of obs: 56, groups: subject, 14
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.15757    1.09027    3.81
## season3     -0.00128    0.02264   -0.06
##
## Correlation of Fixed Effects:
##          (Intr)
## season3 -0.405
```

The only thing that has changed in the linear growth model when we use weeks rather than index, is the fixed-effect of time. Since we now measured in weeks, the interpretation is the average depression reduction in a one-week difference (rather than a one-season difference). Note that the t -value associated with this effect, however is exactly the same!

LINEAR TRANSFORMATIONS OF THE TIME METRIC DOES NOT CHANGE WHETHER YOU WILL FIND AN EFFECT; IT ONLY CHANGES THE INTERPRETATION.

Other Advantages of the Numerical Time Predictor

There are other advantages of using a numerically coded time predictor as well. One is that you can account for measurement waves that are not equally spaced. For example, say we had measured depression scores at

- Week 0
- Week 10
- Week 25
- Week 45

If we use index, these would be 0, 1, 2, and 3. This would not capture the unequal time points. Since the fixed-effect in a regression model would be interpreted as a one-week difference, it doesn't matter if the time-points are equally spaced. With the categorical coding of Winter, Spring, Summer and Fall, we don't know if the waves are equally spaced or not. The best we can use is indexing and make the assumption that they are equally spaced (Danger Will Robinson).

We also gain statistical power because we use fewer predictors. With the categorical predictors we needed three dummy variables to account for time. With numerical coding, we only need a single predictor (unless

we want to model non-linearity). Using fewer predictors increases the error df , which increases statistical power. It also leads to a much simpler interpretation of the effect of time.

References

Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.