Longitudinal Homework 2

Tim Vigers

26 September 2019

1. Estimability

If you have a matrix X which contains an intercept and each level of a categorical variable, then this matrix must be less than full rank (LTFR) because a column will have to be linearly dependent on another column. For example, take a matrix for a class variable with three levels:

$$X_{ltfr} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

When a matrix has linear dependencies like this, the specific inverse of X doesn't exist, which means that $\hat{\beta} = (X^t X)^{-1} X^t Y$ is not unique. However, if you drop one level of the class variable, you can get a matrix with no linear dependencies:

$$X_{fr} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

This means that $X_{fr}^t X_{fr}$ is full rank and the specific inverse can be computed. When fitting a LTFR model, SAS would use $X_{fr}^t X_{fr}$ to create the generalized inverse (essentially by filling $X_{fr}^t X_{fr}$ in with 0s so that it matches the dimensions of $X_{ltfr}^t X_{ltfr}$). This allows you to estimate the parameters for the class variable.

2. $Var(\hat{\beta})$ for full-rank linear mixed model

The algebraic form of $\hat{\beta}$ from ML estimation is:

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y$$

Using the general property $Var(AY) = AVar(Y)A^t$ we can set $A = (X^t V^{-1} X)^{-1} X^t V^{-1}$ and get:

$$Var(\hat{\beta}) = AVar(Y)A^t = (X^t V^{-1} X)^{-1} X^t V^{-1} Var(Y)((X^t V^{-1} X)^{-1} X^t V^{-1})^t$$

Because V^{-1} is symmetric, A^t can be re-written as

$$((X^t V^{-1} X)^{-1} X^t V^{-1})^t = V^{-1} X (X^t V^{-1} X)^{-1}$$

Leaving us with:

$$Var(\hat{\beta}) = (X^t V^{-1} X)^{-1} (X^t V^{-1} Var(Y) V^{-1} X)(X^t V^{-1} X)^{-1}$$

Finally, because $Var(Y) = V$, we get:

$$Var(\hat{\beta}) = (X^t V^{-1} X)^{-1}(X^t V^{-1} V V^{-1} X)(X^t V^{-1} X)^{-1}$$
$$= (X^t V^{-1} X)^{-1}(X^t V^{-1} X)(X^t V^{-1} X)^{-1}$$
$$= (X^t V^{-1} X)^{-1}$$

3. $Var(L\hat{\beta})$

It follows from the above equations that if X does not have full rank, $Var(\tilde{\beta}) = (X^t V^{-1} X)^-$ (see course notes page 143):

$$\tilde{\beta} = (X^t V^{-1} X)^- X^t V^{-1} Y$$
$$Var(\tilde{\beta}) = (X^t V^{-1} X)^- X^t V^{-1} V((X^t V^{-1} X)^- X^t V^{-1})^t$$

Using the M-P general inverse allows you to simplify and get:

$$(X^t V^{-1} X)^- X^t((X^t V^{-1} X)^- X^t V^{-1})^t = (X^t V^{-1} X)^-$$

Because $(X^t V^{-1} X)^- X^t$ is unique in a LMM, any generalized inverse will give you the same result.

So:

$$Var(L\tilde{\beta}) = L Var(\tilde{\beta}) L^t = L(X^t V^{-1} X)^- L^t$$

4. GLM vs. LMM

The general linear model (GLM) is a special case of a linear mixed model (LMM). Basically, a GLM is a LMM with no random effects and a simple covariance structure ($R = \sigma^2 I$). So you can write the GLM as $Y = X\beta + \epsilon$ and the LMM as $Y = X\beta + Zb + \epsilon$, where b is the random effect term and Z is the associated design matrix of the random effects. Including the random effects can account for correlation within clusters (e.g using a random intercept to account for correlations between multiple measures on the same person). Adjusting the correlation structure in a linear model can also allow you to fine-tune the model based on how you believe the correlation between measures is affected by clustering (e.g. using AR(1) to model equally spaced time points).

5. Dog data

The model

```
dog_mod <- lme(y ~ group*factor(time),random = ~1|id,data = dogs)
```

Model results

	Value	Std.Error	DF	t-value	p-value
(Intercept)	19.7050000	2.8087380	60	7.0156064	0.0000000

	Value	Std.Error	DF	t-value	p-value
groupcl	-3.8850000	3.9721553	15	-0.9780584	0.3435571
groupco	-3.0333333	3.9721553	15	-0.7636492	0.4569204
factor(time)30	-6.0583333	0.4794934	60	-12.6348617	0.0000000
factor(time)60	-4.4633333	0.4794934	60	-9.3084345	0.0000000
factor(time)90	-2.1416667	0.4794934	60	-4.4665192	0.0000358
factor(time)120	-0.5566667	0.4794934	60	-1.1609474	0.2502614
groupcl:factor(time)30	2.8983333	0.6781061	60	4.2741589	0.0000697
groupco:factor(time)30	5.6900000	0.6781061	60	8.3910169	0.0000000
groupcl:factor(time)60	2.5333333	0.6781061	60	3.7358950	0.0004189
groupco:factor(time)60	4.5033333	0.6781061	60	6.6410450	0.0000000
groupcl:factor(time)90	1.1916667	0.6781061	60	1.7573454	0.0839601
groupco:factor(time)90	1.9850000	0.6781061	60	2.9272704	0.0048248
groupcl:factor(time)120	-0.1350000	0.6781061	60	-0.1990839	0.8428702
groupco:factor(time)120	0.6333333	0.6781061	60	0.9339738	0.3540610

Contrasts with SAS

Contrast code

```
proc mixed data=dogs;
  class id group time;
  model y = group time group*time / solution;
  estimate 'group comp' group 1 -1 0/E;
  estimate 'time comp' time 1 0 0 0 -1 /E;
  contrast 'ch0 vs. ch60 and co0 vs. co60 comp' time 1 0 -1 0 0 group*time 1 0 -1 0 0 0 0 0 0 0 0 0 0 0 0,
                     time 1 0 -1 0 0 group*time 0 0 0 0 0 0 0 0 0 0 1 0 -1 0 0/E;
  contrast 'ch0 vs. ch90 and co0 vs. co90 comp' time 1 0 0 -1 0 group*time 1 0 0 -1 0 0 0 0 0 0 0 0 0 0 0,
                     time 1 0 0 -1 0 group*time 0 0 0 0 0 0 0 0 0 0 1 0 0 -1 0/E;
  random intercept / subject=id(group); run;
```

Contrast results

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > |t|
group comp	2.5873	3.9489	15	0.66	0.5223
time comp	0.3906	0.2768	60	1.41	0.1635

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
ch0 vs. ch60 and co0 vs. co60 comp	2	60	43.33	<.0001
ch0 vs. ch90 and co0 vs. co90 comp	2	60	10.03	0.0002

For this problem I assumed that dogs were unique across groups.

First, I compared the ch group to the cl group averaged across time. The estimated difference in gall bladder volume between the two groups is 2.59, but this difference is not statistically significant (p = 0.52). I also compared time 0 to time 120, to see if there was a significant difference in gall bladder size averaged across the groups. The estimated difference is 0.39, but this was also not statistically significant (p = 0.16).

I also contrasted the difference from time 0 to 60 for the ch group to the difference between 0 and 60 for the co group. The two trends were significantly different (p = <0.0001). Similarly, I compared the difference from time 0 to 90 for the ch group to the difference between 0 and 90 for the co group. This contrast was also statistically significant (p = 0.0002).

Confirm with emmeans

```
emm_group <- emmeans(dog_mod, specs = ~group)
emm_time <- emmeans(dog_mod, specs = ~factor(time))
emm_group_time <- emmeans(dog_mod, specs = ~group*factor(time))
ch <- c(1,0,0)
cl <- c(0,1,0)
t0 <- c(1,0,0,0,0)
t120 <- c(0,0,0,0,1)
contrast(emm_group, method = list("Group ch vs. group cl" = ch - cl))
```

```
##  contrast              estimate   SE df t.ratio p.value
##  Group ch vs. group cl     2.59 3.95 15 0.655   0.5223
##
## Results are averaged over the levels of: time
## Degrees-of-freedom method: containment
```

```
contrast(emm_time, method = list("Time 0 vs. time 120" = t0 - t120))
```

```
##  contrast             estimate    SE df t.ratio p.value
##  Time 0 vs. time 120    0.391 0.277 60 1.411   0.1635
##
## Results are averaged over the levels of: group
## Degrees-of-freedom method: containment
```

Unfortunately I didn't have time to figure out the double row contrasts in R but the ones I was able to do match the SAS estimates, which is a good sign.

6.

a. Model and X matrix

i. LTFR

The statistical model is:

$$Y_{hij} = \mu + \alpha_j + \tau_h + \gamma_{hj} + b_i + \epsilon_{hij}$$

Where i indexes subject, j indexes time, and h indexes group. Also, α represents the effect of time, τ represents the effect of group, γ represents the interaction effect, b represents a random intercept for subject, and ϵ is the error term.

The X matrix is:

$$X = \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
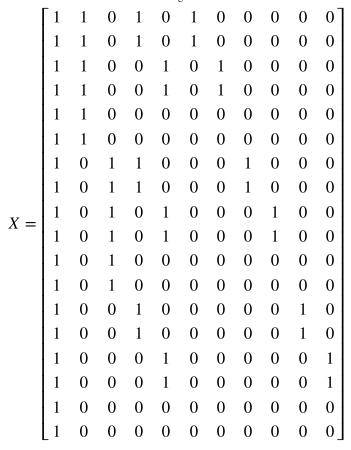1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

ii. With a set-to-0 restriction

The statistical model is:

$$Y_{hij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ih} + \beta_4 x_{4ih} + \beta_5 x_{5hij} + \beta_6 x_{6hij} + \beta_7 x_{7hij} + \beta_8 x_{8hij} + b_i + \epsilon_{hij}$$

Again, i indexes subject, j indexes time, and h indexes group. In this model x_{pij} is a dummy variable indicating whether or not a subject is in a given group at a given time.

The X matrix is:

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This is the same as the matrix in i, but with the highest values of group and time removed.

b. Linear trend for one group compared to another.

The estimate for group 3, time 1 is:

$$L_{13} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

And the estimate for group 1, time 1 is:

$$L_{11} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The linear trend is just the difference between the two:

$$L_{trend} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Using the LTFR design matrix (since the contrasts above are for the LTFR model), show that $L = LH$:

```
# Construct the design matrix
x <- matrix(c(1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
              1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
              1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
              1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
              1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
              1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
              1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
              1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
              1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
              1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
              1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
              1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
              1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
              1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
              1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
              1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
              1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,
              1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1),
            nrow = 18,byrow = T)
# Construct the hat matrix
H <- ginv((t(x)%*%x))%*%(t(x)%*%x)
# Construct L
L <- c(0,0,0,0,-1,0,1,-1,0,1,0,0,0,0,0,0)
# Test if L = LH (round to 1 digit because the ginv() function produces very
# small numbers that are essentially 0 but not quite, I think due to
# machine precision)
round(L %*% H)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## [1,]    0    0    0    0   -1    0    1   -1    0     1     0     0     0
##      [,14] [,15] [,16]
## [1,]     0     0     0
```

```
round(L %*% H) == L
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## [1,] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  TRUE  TRUE  TRUE  TRUE
##      [,14] [,15] [,16]
## [1,]  TRUE  TRUE  TRUE
```

So, the linear trend above is estimable because L=LH.

c. AR(1)

An AR(1) structure for R would allow the correlation between time 1 and time 3 to be lower than the correlation between time 1 and time 2. The distribution of the error term is $\epsilon \sim N(0, R_i)$, so the major change to the model would be the form of R_i which alters the distribution of the errors.

d. Time as continuous

The statistical model is:

$$Y_{hij} = \mu + \alpha + \tau_h + \gamma_h + b_i + \epsilon_{hij}$$

Where α is the effect for time, τ is the effect for group, and γ is the interaction effects between time and group.

And the X matrix is:

$$X = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 2 & 1 & 0 & 0 & 2 & 0 & 0 \\
1 & 2 & 1 & 0 & 0 & 2 & 0 & 0 \\
1 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\
1 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\
1 & 2 & 0 & 0 & 1 & 0 & 0 & 2 \\
1 & 2 & 0 & 0 & 1 & 0 & 0 & 2 \\
1 & 3 & 1 & 0 & 0 & 3 & 0 & 0 \\
1 & 3 & 1 & 0 & 0 & 3 & 0 & 0 \\
1 & 3 & 0 & 1 & 0 & 0 & 3 & 0 \\
1 & 3 & 0 & 1 & 0 & 0 & 3 & 0 \\
1 & 3 & 0 & 0 & 1 & 0 & 0 & 3 \\
1 & 3 & 0 & 0 & 1 & 0 & 0 & 3
\end{bmatrix}$$

e. Unequally spaced times

i. Class variable

With unequally spaced time periods, it would be appropriate to treat time as a categorical variable, as long as you were only interested in the difference between each time point and baseline.

ii.

In order to account for the unequal spacing of time, you could use a spatial power structure for R_i:

$$R_i = \sigma^2 \begin{bmatrix}
1 & \phi & \phi^6 \\
\phi & 1 & \phi^5 \\
\phi^6 & \phi^5 & 1
\end{bmatrix}$$