MATH 6642 – Longitudinal Data Analysis

Sample Exam Questions

Georges Monette

June 5, 2019

Useful formulas:

- $(A + UDV)^{-1} = A^{-1} - A^{-1}U(D^{-1} + VA^{-1}U)^{-1}VA^{-1}$

1. Write a brief discussion with illustrations (using ellipses in beta space if you wish) explaining the following issues concerning the interpretation of p-values in the output produced by the 'summary' method applied to a mixed model for longitudinal data analysis.
   a) [5] Why should p-values for main effects that are marginal to an interaction be interpreted with caution. Are there any situation in which it is legitimate to use them?
   b) [5] If two main effects that are not marginal to any interaction have non-significant p-values, is it appropriate to drop both terms?
   c) [5] If a time-varying variable is included as a raw variable and as a contextual variable, what is the possible consequence of dropping the contextual variable? When would dropping it result in no change in the estimated coefficient of the raw variable?
2. [10] Consider a random coefficients model for a model with time as a single predictor with a random intercept and a random slope. Let the G matrix have the decomposition $G = AA'$ where

$$A = \begin{bmatrix} a_{00} & 0 \\ a_{10} & a_{11} \end{bmatrix}$$

   is lower triangular. Explain and justify the interpretation of the elements of A in terms of the properties of the random true regression lines of Y on T as they vary from subject to subject. Draw a sketch to illustrate your answer.
3. Consider a longitudinal random slopes model where the model for ith subject is given by

$$Y_i = X_i\gamma + X_i u_i + \epsilon_i$$

   with X_i a $T_i \times 2$ full column rank matrix with a first column of 1's and a second column containing the times at which the vector of responses Y_i was observed, $u_i \sim N_2(0, G)$, with G positive definite, independent of $\epsilon_i \sim N(0, \sigma^2 I)$.
   Let $\beta_i = \gamma + u_i$. In the following questions, treat the macro level parameters γ, G, σ^2 as if they were as they were fixed and known.
   a) [5] Give an expression for the BLUE, $\hat{\beta}_i$, of β_i.
   b) [5] Give an expression for the BLUP, $\tilde{\beta}_i$, of β_i as a function of the BLUE.
   c) [5] Describe how the BLUP is a shrinkage estimator based on the BLUE.
   d) [5] For large T_i and relatively fixed variance for the values of time, how will the BLUP behave relative to the BLUE?
   e) [5] What are the implications for the BLUP if G is highly concentrated?
   f) [10] Show that $Var(\tilde{\beta}_i) < G < Var(\hat{\beta}_i)$. Note: for two symmetric matrices, A and B, $A < B$ if and only iff $B - A$ is positive definite.
4. The following questions refer to the output on the last page for a mixed model fitted on the full high school math achievement data set. The model uses SES, SES.School which is the mean SES in the sample in each school, 'female' which is an individual level indicator variable and 'Type' which is three-level factor with levels "Coed", "Girl" and "Boy" with the obvious definition and Sector which is a 2-level factor with levels "Catholic" and "Public".
   a. [5] Consider two Public coed schools, one with mean SES = 0 and the other with mean SES = 1. Suppose the values of SES in the former school range from -1 to 1 and in the latter school from 0 to 2. Draw a graph showing the predicted MathAch in these two schools over the range of values

of SES in each school. On your graph identify the value and location of the contextual effect of SES, the within school effect of SES and the compositional effect of SES.

b. [5] Suppose you were to refit the model without SES.School. What would you expect would happen to the coefficient for SES? Would it stay roughly the same, get bigger or get smaller, or is the change unpredictable? Explain.

c. [5] Suppose you want to perform an overall test of the importance of gender in the model, either within schools or between schools. Specify a hypothesis matrix that would perform this test.

d. [5] Estimate the difference between the predicted math achievement of a boy in a boys' school versus a girl in a girls'school. If you suspected that this is affected by the SES of the school, how would you modify the model to test this hypothesis?

5. [10] Consider a mixed model for data with a response variable Y and regression on an intercept and a single predictor X. Suppose that each cluster has size 3 and X is observed twice at the value 0 and once at the value 1; i.e. the observed values of X are 0, 0, 1 within each cluster. Show whether it is possible to identify the variance parameters of a random slope model?
6. Data is obtained on the effects of two treatments (A and B). Data was recorded on symptoms weekly throughout 10 weeks of active treatment and then for 8 more weeks
following termination of the treatment. A linear spline model is fit to the data and the following results were obtained:

```
sp  <- function( x ) {
      gsp( x,
                    knots = c( 10 ),      # 1 knots => 2 intervals
                    degree = c( 1 , 1 ) , # linear in each interval
                    smooth = c( 0 )       # continuous at the knot
      )
}

> fit <- lme( Symptom ~ sp(Weeks)*tx, random=~1+Weeks|id, data = d)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: d
       AIC      BIC    logLik
  15099.44 15153.18 -7539.719

Random effects:
 Formula: ~1 + Weeks | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 37.921194 (Intr)
Weeks        2.395329 -0.218
Residual    23.285420

Fixed effects: Symptom ~ sp(Weeks) * tx
                         Value Std.Error   DF   t-value p-value
(Intercept)          131.84484  6.450470 1516  20.43957  0.0000
sp(Weeks)D1(0)       -11.21889  0.508022 1516 -22.08346  0.0000
sp(Weeks)C(10).1      18.24584  0.571104 1516  31.94835  0.0000
txB                   -2.61621  9.122343   78  -0.28679  0.7750
sp(Weeks)D1(0):txB     1.55017  0.718452 1516   2.15765  0.0311
sp(Weeks)C(10).1:txB  -4.83791  0.807664 1516  -5.99001  0.0000

. . .

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-4.109811205 -0.593593851 -0.002546668  0.654503092  3.122385555

Number of Observations: 1600
Number of Groups: 80
```

a) [5] Interpret each fixed effects coefficients in the model drawing an appropriate graph.
b) [5] In a separate graph, sketch the predicted trajectory for `tx` equal to A.

7. [5] What is Simpson's Paradox and how is it related to the inclusion of contextual variables in hierarchical models?
8. [5] In the model used in question 5, why would one want to add the average SES of each school, 'SES.School', to the model?
9. [5] Let Σ be symmetric. Show that Σ is positive-definite if and only there exists a non-singular matrix A such that $\Sigma = AA'$.

Output for question 4

```
Linear mixed-effects model fit by REML
 Data: hs
       AIC      BIC    logLik
  46477.96 46581.14 -23223.98

Random effects:
 Formula: ~1 + SES + female | School
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.6500478 (Intr) SES
SES         0.2661823  0.755
female      0.9507088 -0.606 -0.979
Residual    6.0343349

Fixed effects: MathAch ~ SES * Sector + SES.School + female + Type
                      Value Std.Error   DF  t-value p-value
(Intercept)       13.633578 0.3315175 7022 41.12476  0.0000
SES                1.436456 0.1620660 7022  8.86340  0.0000
SectorPublic      -0.888000 0.3711059  155 -2.39285  0.0179
SES.School         3.082877 0.3662818  155  8.41668  0.0000
female            -1.078197 0.1902275 7022 -5.66793  0.0000
TypeGirl           0.258757 0.4770260  155  0.54244  0.5883
TypeBoy            1.358493 0.5415658  155  2.50845  0.0132
SES:SectorPublic   1.283763 0.2121825 7022  6.05028  0.0000
 Correlation:
                 (Intr) SES    SctrPb SES.Sc female TypGrl TypeBy
SES              -0.010
SectorPublic     -0.807 -0.010
SES.School       -0.273 -0.199  0.374
female           -0.381 -0.023  0.020  0.006
TypeGirl         -0.550  0.037  0.561  0.206 -0.134
TypeBoy          -0.591  0.036  0.465  0.088  0.233  0.321
SES:SectorPublic  0.056 -0.733  0.012 -0.002 -0.001 -0.058 -0.045

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.10239407 -0.73144171  0.02590791  0.75673178  2.99519661

Number of Observations: 7185
Number of Groups: 160
```