PubH 5483 – Fall 2004

# EXAM – SOLUTIONS

---

For some questions, the solutions below give MORE details than were required to get full credit (all points). Questions are worth 10 points each; only 10 of the 12 questions were to be answered.

## APPLICATION

1. The top figure on page 8 shows us that, on average, SBP for the placebo group was higher than for the treatment group. The figures on page 7 show us that the slope across day differs on average between the placebo and treatment group, and thus there may be a treatment by time interaction.

2. (a) The figures on pages 9-12 show us that the trend across time for each rat is approximately linear, and thus a General Linear Mixed Model with random intercepts and slope may be reasonable. On the other hand, the figure on page 14 shows an approximately horizontal trend in the squared residuals, and thus a General Linear Mixed Model with only random intercepts may be appropriate.

   (b) The figure and correlation matrix on page 13 shows us that compound symmetry and unstructured may be reasonable covariance structures to consider for a General Linear Model. (Toeplitz is also an acceptable answer.)

3. (a) The null hypothesis is that the variance of the random slopes is zero: $H_o : \sigma_1^2 = 0$. (It is also acceptable to write this null hypothesis as $H_o : \sigma_1^2 = 0$ and $\sigma_{01} = 0$.)

   (b) The likelihood ratio test statistic is $T^* = 1355.6 - 1354.1 = 1.5$.

   (c) 1.5 is less than 5.14, the 0.05 critical value for the mixture of $\chi_1^2$ and $\chi_2^2$. We thus conclude that `COVARIANCE MODEL` #1 is not superior to `COVARIANCE MODEL` #2 and we prefer to continue with `COVARIANCE MODEL` #2.

4. Rats on oregano have significantly lower SBP on average than rats on placebo (LS mean difference -2.75 mm Hg, st.error 0.68, p<0.0001). Rats on oregano have a significantly faster decline in SBP by 1.05 mm Hg per day than rats on placebo (st.error 0.06, p<0.0001).

   NOTE: Just reporting the tests and p-values is not a sufficient answer here since the researchers want to know BY HOW MUCH the groups differ for both questions.

5. (a) The figures on page 20 both indicate unimodal symmetric distributions in the residuals with a bell curve shape in the top figure. The figures on page 23 indicate no strong violations of normality in the estimated random effects, although the estimated random intercepts are somewhat skewed. There are no strong indications of violations of normality.

(b) None of the figures on pages 20, 21, and 23 show any indication of outlying clusters or outlying observations.

(c) The figure on page 22 shows that the estimated variances from `PROC MIXED` are an adequate summary of the variances in the OLS residuals. The estimated correlation matrix in the `PROC MIXED` output on page 15 (`Estimated V Correlation Matrix for ratid 1`) also is very similar in structure and magnitude to the OLS residual correlation matrix. These both indicate good model fit for the covariance structure.

## CONCEPTS

6. (a) `UN(2,1)` corresponds to the covariance between the random intercepts and the random slopes.

   (b) Often those clusters which start out with low outcome values (small cluster-specific intercept) will increase faster (large cluster-specific slope) across time.

7. (a) Population-averaged fitted values are computed as $\hat{Y}_{ij}^{PA} = X_{ij}\hat{\beta}$ and cluster-specific fitted values are computed as $\hat{Y}_{ij}^{CS} = X_{ij}\hat{\beta} + Z_{ij}\hat{b}_i$.

   (b) Population-averaged fitted values indicate overall trends in the mean outcome across clusters. Cluster-specific fitted values indicate trends in the mean outcome for one specific cluster in particular.

8. (a) $Var[Y_{ij}] = \sigma_0^2 + (day)^2\sigma_1^2 + 2(day)\sigma_{01} + \sigma_e^2$

   (b) $Cov[Y_{ij}, Y_{ij'}] = \sigma_0^2 + (day_j)(day_{j'})\sigma_1^2 + (day_j)\sigma_{01} + (day_{j'})\sigma_{01}$

9. (a) $\sigma_0^2$ corresponds to between-person variability. $\sigma_e^2$ corresponds to within-person variability.

   (b) We gain power in testing $\beta$.

10. (a) This tests whether or not the group-specific correlations are the same for the two groups, i.e., $H_o : \rho_{trt} = \rho_{placebo}$

    (b) The larger model (with the group effect) has four parameters (correlation and variance for each group), while the smaller model (without the group effect) has two parameters (correlation and variance), so the degrees of freedom for the likelihood ratio test are $4 - 2 = 2$.

11. (a) average SBP across the 7 days

(b) slope in SBP across the 7 days

(c) There are many possible answers. Those that make the most sense in this context are: (i) Two models would have to be fit to answer the two questions of interest, instead of fitting only one model to answer both questions. (ii) With only seven data points per rat, the derived variables are themselves likely to have large variances, which may mean reduced power to test the treatment effect of interest. (iii) Because derived variables are data summaries, predicted values from a derived variable model will have smaller variability than predicted values from a model using the original data; this can be misleading if the reporting of the results do not clarify this.

12. (a) The two models being compared must be nested.

(b) The null hypothesis being tested cannot be of the form $\sigma^2 = 0$.

**EXTRA CREDIT**

1. This is testing whether the two groups have the same average SBP on day 0. This is a test of whether or not the randomization was successful in getting two randomized groups with the same average baseline SBP.

2.
```
proc mixed data=repdat;
   class trt ratid;
   model sbp = trt day trt*day / s;
   random int day / sub=ratid type=un vcorr v;
   lsmeans trt / diff;
run;
```