BIOS 755: Model fit and profile analysis

Alexander McLain

General Linear Model

- The general linear model can be written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{e}_i$$

  where $\boldsymbol{X}_i = \{\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}\}$ and $\boldsymbol{e}_i \sim MVN(\mathbf{0}, \Sigma)$.
- Recall that Σ is a covariance matrix of the residual error terms.

Introduction to Maximum Likelihood (ML)

- When we fit a model to longitudinal data, the main aim is to estimate the unknown parameters, $\boldsymbol{\beta}$ and Σ
- The main idea behind Maximum Likelihood (ML) is simple: use as estimates of β and Σ the values that are most probable (or "likely") for the data that we have observed.
- Choose values of β and Σ that maximize the probability of the response variables evaluated at their observed values (or that best predict the observed data).
- The resulting values, $\hat{\beta}$ and $\hat{\Sigma}$ are called the maximum likelihood estimates (**MLE**)of β and Σ, and have many appealing properties.

MLE Bias

- MLE of has well-known bias in small samples. For ordinary regression with independent errors gives the usual least squares estimator of $\boldsymbol{\beta}$, but the ML estimator of σ^2 is

$$E(\hat{\sigma}^2) = \left(\frac{n-k}{n}\right)\sigma^2$$

- As a result, the ML estimate of σ^2 will be biased in small samples and will underestimate σ^2 (on average).

Restricted Maximum Likelihood

- The theory of restricted (or residual) maximum likelihood (REML) estimation was developed to address this problem.
- REML uses a penalty term that eliminates the biases in the variance parameter estimates.
- Estimates of Σ obtained by the REML are preferred in practice. All (I think) softwares will use this as the default.

Restricted Maximum Likelihood

- **REML should not be used to compare different regression models (i.e., different mean models)** since the penalty term depends upon the regression model specification.
- Instead, the standard ML log-likelihood should be used for comparing different regression models for the mean.
- That is why we have to use `method="ML"` when we compare different models for the mean.
- **The REML can be used to compare different models for the covariance structure.**

Akaike information criterion

- Akaike information criterion (AIC) judges a model by the value of the negative log-likelihood ($-\log\{L(\beta)\} > 0$).
- $-\log\{L(\beta)\}$ will increase as more parameters are added to the model so we penalize by the number of parameters.
- AIC is a very general method of evaluating model fit that can be used in a wide variety of statistical procedures.
- The formula for AIC is

$$-2\log\{L(\hat{\beta})\} + 2q$$

  where q is the number of parameters in the model.
- We choose the model with the lowest AIC

Bayesian information criterion

- Bayesian information criterion (BIC) is very similar to AIC
- BIC penalizes more heavily for adding additional parameters.
- The formula for BIC is

$$-2\log\{L(\hat{\beta})\} + q\log(N^*)$$

where N^* is the number of **effective subjects**.

- We choose the model with the lowest BIC.
- The penalty for AIC is $2q$ the penalty for BIC is $\log(N^*)q$

$$2 < \log(N^*) \quad N^* \geq 8$$

Likelihood Ratio Tests

- Two models are nested (bigger v.s smaller)
- Under the null hypothesis, the difference of $-2\log\{L(\hat{\beta})\}$ between the full and reduced model follows a χ^2 distribution with df=Δdf
- A likelihood ratio test is done by:
  1. $D = -2\log\{L_{Reduced}(\hat{\beta})\} - (-2\log\{L_{Full}(\hat{\beta})\})$
  2. Δdf=Number of parameters in the full model-Number of parameters in the reduced model
  3. Reject H_0 :"models are the same" if $D > \chi^2_{\Delta df, 1-\alpha}$, and use the "full" model.
- When doing a likelihood ratio test with covariance parameters use $\alpha = 0.1$ instead of $\alpha = 0.05$ (the reasons are beyond our scope).

Summary of main points

- When doing a likelihood ratio test the models **must** be nested.
- Some people use the REML function in place of the usual likelihood to form likelihood ratio tests and the AIC and BIC criteria. If the test concerns different mean models, this is generally **not** recommended.
- Use of the AIC and BIC criteria based on the REML objective function to choose among covariance models for the same mean model is often used.

Models for the Mean

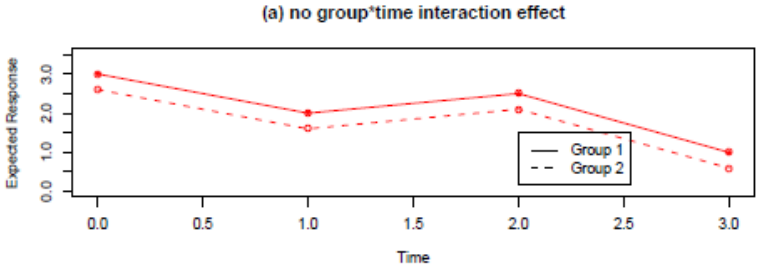
- Two basic strategies for modeling the time trend:
  - Arbitrary Means (Profile Analysis)
  - Parametric Curves (Linear and Nonlinear Trend)

Profile analysis of the LBC data

- Risby and associates (Riesby, et al, 1977) ran a study to clinical response in 66 (endogenous and non-endogenous) depressed inpatients.
- The main outcome is the Hamilton Depression Score (HD) and covariate data included:
  - diagnosis ('Endog' where non-endogenous=0 and endogenous=1),
  - gender,
  - Imipramine (IMI) and Desipramine (DMI) plasma levels
- There are three hypotheses of interest:

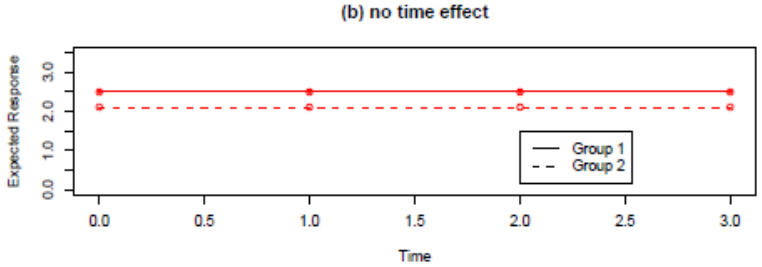
Profile Analysis: Case 1

H_{10}: Are the profiles of means similar in the groups, in the sense that the line segments between adjacent occasions are parallel? This is the hypothesis of no group by time interaction.



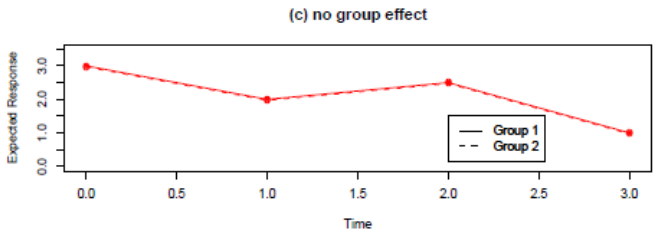
Profile Analysis: Case 2

H_{20}: If the population profiles are parallel, are the means constant over time? This is the hypothesis of no time effect.



Profile Analysis: Case 3

H_{30}: If the population profiles are parallel, are they also at the same level? This is the hypothesis of no group effect.



Although these general formulations of the study hypotheses are a good place to begin, the appropriate hypotheses in a particular study must be derived from the relevant scientific issues in that investigation.

The Model

- If none of those three hypotheses hold, then an interaction term is needed.
- A commonly used model is:

$$Y_{ij} = \beta_0 + \beta_1 Group_i + \beta_2 Time_j + \beta_3 Time_j * Group_i + \epsilon_{ij}$$

Profile coding

- We can include time using dummy coding.
- With dummy coding we choose one time point as the baseline time point.
- The mean for the rest of the time points is modeled relative to the baseline time point.

Profile coding example

- **Three time points** and **two groups** with t_{i1} being the baseline time.
- We'll have two "time coefficients" and on "group coefficient"

$$T_{ij2} = \begin{cases} 1 & \text{if j=2} \\ 0 & \text{otherwise} \end{cases} \qquad T_{ij3} = \begin{cases} 1 & \text{if j=3} \\ 0 & \text{otherwise} \end{cases} \qquad G_{ij} = \begin{cases} 1 & \text{treatment group} \\ 0 & \text{placebo group} \end{cases}$$

- Then the model is

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_{21} T_{ij2} + \beta_{22} T_{ij3} + \beta_{31} G_{ij} T_{ij2} + \beta_{32} G_{ij} T_{ij3} + \epsilon_{ij}$$

where $\boldsymbol{\beta_2} = (\beta_{21}, \beta_{22})$ are the '*Time*' coefficients and $\boldsymbol{\beta_3} = (\beta_{31}, \beta_{32})$ are the '*Group*Time*' coefficients.

Mean differences

- What is the impact of group/treatment when $j = 1$ (baseline), and $j = 3$?
- What is the difference between $j = 1$ (baseline) and $j = 3$ for the treatment group? How about for the control group?
- What is the difference in the difference between $j = 1$ (baseline) and $j = 3$ for the treatment and control groups?

Profile with interaction

When there is an interaction in a profile analysis, you cannot interpret:

- the effect of treatment without specifying the time point, and
- the effect of time without specifying the treatment group.

What you can do:

- give the difference in differences (commonly the main goal),
- discuss the impact of time separately by treatment,
- discuss the impact of treatment separately by time point or
- use least square means (LSMEANS) to get all the mean values and mean differences.

- If there is no group-by-time interaction, what β coefficients will be zero?
- Given there is no interaction, if there is no time effect, what β coefficients will be zero?
- Given there is no interaction, if there is no group effect, what β coefficients will be zero?

T-tests vs type III tests

- Most people focus on the individual t-tests when looking at the results of an analysis which will test, for example,

$$H_0 : \beta_1 = 0$$

- This may be of interest, but commonly, we are interested in a group of coefficients all being equal to zero, for example,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

  where type III tests should be used.
- **This is almost always the case when using dummy coding.**
- In this situation, look at the type III test result first, then at the individual t-tests to see what is driving the results.

SAS Code

- To conduct a profile analysis of data from two or more treatment groups measured repeatedly over time, we can use the following SAS code.
- Which can be done with any kind of covariance matrix

```
proc mixed;
class id group time;
model y=group time group*time /s;
repeated time/type=UN subject=id r rcorr;
run;
```

GO TO SAS EXAMPLE

Summary

- Does not assume any specific time trend
- May have low power to detect specific trends; e.g., linear trends.
- You don't get an "overall" estimate of the time effect.
- How to incorporate mistimed measurements?