STA 303 / 1002

Note Title 3/26/2012

Course evaluations : next Monday

Repeated Measures Analysis Using Mixed Models

What is Repeated Measures?

- more than one observation per subject or experimental unit

For example:

(1) measures taken on all members of a cluster (eg. family, class, etc.)

(2) cross-over study - experiment where each subject gets each treatment

(3) "Longitudinal" - subjects followed over time, observations at regular intervals

We're going to look at (3).

"Between subjects effects" - factors on which there are no repeated measures

- constant within a subject

eg. gender, in our example: treatment group

"Within subjects effects" - are repeated measures

eg. in our example: time

(in a cross-over study: treatment)

Issue: observations not independent; expect observations on one subject to be correlated

Example : Carbohydrates in Diabetes

- diet study for people with Type 2 diabetes
- assigned to 1 of 3 treatment groups :
  1. low GI (glycemic index) (LG)
  2. high GI (HG)
  3. high in monounsaturated fats (MUFA) (HM)

  for 6 months
- measurements at 0, 3, 6 months

Data file:
- id - patient identifier
- diet - treatment assigned (LG, HG, HM)
- season - ignore
- time - 1 = 0 months, 2 = 3 months, 3 = 6 months

Outcomes:
- weight
- hemo - measure of diabetic control
- gluc - " " " "
- chol - total chol.
- HDL
- TG - triglycerides
- LDL

The outcome we'll analyze → HDL

We're interested in time $\times$ diet interaction. Does differences among diets change over time?

We'll time as categorical variable:

- only 3 time points
- don't have to assume that the relationship with time is linear

Typical <u>longitudinal</u> setting:

n subjects, $i = 1, \ldots, n$

J treatments, $j = 1, \ldots, J$

<u>Our example</u>

$n = 71$

$= n_{HG} + n_{LG} + n_{HM}$

$J = 3$

measured K times, $k = 1, \dots, K$ $\quad K = 3$

HDL

outcome: Y_{ijk}

Y_{ijk} continuous

Normal distribution model for Y_{ijk}:

$J-1$ indicator variables for treatments

$K-1$ indicator variables for time

$(J-1)(K-1)$ interaction terms for time * treatment interaction

"FIXED" EFFECTS

- coefficients of these terms are assumed unknown constant parameters, β's

Still have to worry about lack of independence of the K observations on each subject

<u>So</u> Treat subject as "RANDOM" <u>EFFECT</u>

- subjects as a random sample from a population of interest
- model the random effect of subject as a $N(0, \sigma_u^2)$ r.v. U

(If wanted to test if there exist differences among subjects, test $H_0: \sigma_u^2 = 0$)

Additional benefits of adding random effect for subject:

- inference extends beyond subjects measured to entire population of interest

key $\rightarrow$ - induces correlation between observations on the same subject

"Mixed" model - has both fixed and random effects

Model for diet in diabetes example:

Y_{ijk} response at time k on subject i in treatment group j

$$
\begin{aligned}
Y_{ijk} = \underbrace{\Big\{ }_{\text{fixed effects}} & \beta_0 \\
& + \beta_1 I_{[diet=HG],j} + \beta_2 I_{[diet=HM],j} \\
& + \beta_3 I_{[time=1],k} + \beta_4 I_{[time=2],k} \\
& + \beta_5 I_{[diet=HG]} * I_{[time=1],jk} + \beta_6 I_{[diet=HG]} * I_{[time=2],jk} \\
& + \beta_7 I_{[diet=HM]} * I_{[time=2],jk} + \beta_8 I_{[diet=HM]} * I_{[time=2],jk} \\
& + u_{ij} \rightsquigarrow \text{random effect due to } i^{th} \text{ subject on } j^{th} \text{ treatment} \\
& + e_{ijk} \rightsquigarrow \text{random error}
\end{aligned}
$$

u_{ij} i.i.d $N(0, \sigma_u^2)$ – assume subjects independent

e_{ijk} i.i.d. $N(0, \sigma_e^2)$

Assume u_{ij}, e_{ijk} are independent

According to model!

$$Var(Y_{ijk}) = \sigma_u^2 + \sigma_e^2, \quad \forall\, i, j, k$$

$$Cov(Y_{ijk}, Y_{lmn}) = \begin{cases} \sigma_u^2 + \sigma_e^2, & i = l,\ j = m,\ k = n \\ 0, & i \neq l,\ (j \neq m) \end{cases}$$

↳ different subjects are independent

For $k \neq n$, (same subject, different time)

$$\begin{aligned}\mathrm{Cov}(Y_{ijk}, Y_{ijn}) &= \mathrm{Cov}(u_{ij}, u_{ij}) \\ &\quad \left.\begin{aligned} &+ \mathrm{Cov}(u_{ij}, e_{ijn}) \\ &+ \mathrm{Cov}(e_{ijk}, u_{ij}) \\ &+ \mathrm{Cov}(e_{ijk}, e_{ijn}) \end{aligned}\right\} 0 \\ &= \sigma_u^2\end{aligned}$$

$$\underset{k \neq n}{\mathrm{Corr}(Y_{ijk}, Y_{ijn})} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Result of adding random effect to model:

- observations on the same subject are correlated

- the correlation is the same for any pair of observations on the same subject
- the variance is the same for all observations

"COMPOUND SYMMETRY" (CS) (or "exchangeable") covariance structure

<u>Variance - covariance matrix for</u> example:

There are 71 subjects
<u>so</u> 213 observations (ignoring missing values)

Variance - covariance matrix of $\underset{\sim}{Y}$ is 213×213 matrix

For each subject, there is a 3x3 variance-covariance matrix of $(Y_{ij1}, Y_{ij2}, Y_{ij3})$ for that subject

$$D = \begin{pmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 + \sigma_e^2 \end{pmatrix}$$

$$Var(\underset{\sim}{Y}) = \begin{bmatrix} D & 0 & 0 & & 0 \\ & D & & 0 & \\ & & D & & \\ & 0 & & \ddots & \\ & & & & D \end{bmatrix}_{213 \times 213}$$

0's are because subjects are independent

Is the compound symmetry covariance structure suitable for our data?

Is it reasonable to assume:

- corr between obs'ns on same subject at time 1 & time 2
equals " " " " " " " " 1 & " 3
equals " " " " " " " " 2 & " 3

and - variance is the same for all observations?

Yes based on scatterplot matrix and output (correlations & s.d.'s) from proc corr.

Fit model in SAS using proc mixed

- model statement - fixed effects only
  - same notation as proc glm

<u>AND</u> - random statement - specify random effect

id (diet)

↳ id is nested within diet since each subject only gets 1 diet

<u>OR</u> - repeated statement - specify form of covariation within subjects

- if you want something more complicated than compound symmetry

How SAS proc mixed fits models (by default):

RESTRICTED MAXIMUM LIKELIHOOD (or Residual ML)

- gives unbiased estimators of variance - covariance parameters

Two steps:

(1) est var-cov. parameters using (profile) maximum likelihood

(2) Using the est for var-cov. parameters from (1), est β's (coef. of fixed effects) using generalized least squares

Generalized least squares: least squares with adjustment because $Var(\underset{\sim}{y}) = V \; (\neq \sigma^2 I)$

GLS estimates:

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}y$$

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

Repeat 1 & 2 until converge.