Analysis of Correlated Data

1. Repeated Measure ANOVA [it excludes obs. with missing data
   Same timepoints for all obs.]
2. Generalized Estimating Equations (GEE)
3. MM
   - Covariance Pattern Model
   - Random coefficient Model.

Types of Missing Data

- Missing completely at Random (MCAR) → available for case analysis
- " at Random (MAR)
- " not " (MNAR)

↓

Basically fill with guess data & then analyze
- Single
- Multiple

] imputation

Identify Sub
↓
Select Fixed effects → Ex - Age x Gender.
↓
Select cov. Structure → correlation = Package nlme function gls()
- compound Symmetry
- Huynh (Type H)
- unstructioned.

change model

↓
Test cov. Parameter
↓
Test FE → change model.
↓
inference

correlation = corCompSymm (form= ~1 | Person)

No. cov. affect Variable component

independent ~~Hace~~ Sub.

Mixed Model → Random effects + Fixed effects

1. Split Plots
2. Repeated Measures
3. Multisite clinical trials
4. Hierarchical linear Models
5. Random cofficients
6. Analysis of covariance

] Special cases of MM

Covariance Structures change based on Missing values or repeated measures

Patterns of covariance Matrix

Types:
- Compound Symmetry
- Variance components
- Unstructured

Compound Symmetry → All the variances are equal to each other & all covariances are also equal to eachother

Variance components → Each Variance is different & all covariances = 0
Ex - if all 4 variables are completely independend of each other & measured on diff^n Scales, that might be a Pattern.

Unstructured → No Pattern at all.
→ Each variance & Covariance is completely diff^n, without any Relation.

2 types of effect size Statistic
- Standardized → Info about Magnitude & direction of difference bet^n 2 groups
- Un "

$$Cov_{x,y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

$\bar{x}$ → Mean of x
$\bar{y}$ → Mean of y
$N-1$ ← No. of data Values

$Cov_{x,y}$ ← It's a measure of how much 2 Random variables vary together but Variance → tells how single variable Varies

if 2 Random variables are independent then covariance = 0

– It indicates the direction of linear Relationship bet^n variables but correlation Measures both Strength & direction of linear Relationship.

Random Coefficient Model : used if Relationship with time or repeated measurements don't occur ~~of~~ at fixed intervals.

if outcome is not "approximately Normal"
- transform the outcome
- Mixture distribution Model
- Non normal outcome

G: Variance / COV matrix of unknown vector of Random effects

Methods for L. Mixed Models (LMM)

Standard LMM
→ Parameters estimated by ML or REML
→ Random effects are estimated by empirical Bays Method.
→ Test for fixed effect Parameters by using t or F Statistics

[Both G & R can be specified by SAS Random & Repeated Statement.]

→ Simple LMM with Random intercept & Fixed effects could fit by repeated Measures ANOVA
→ Model without RE but non indep. error covariance Structure by Generalized least Squares
→ Multivariate GLM like MANOVA with unstructured error cov. St.

3 ways LMM can be expressed
- Subject time level.
- Subject level
- complete "

R: Variance / COV matrix of unobserved vector of Random Errors

How to specify Random components in a model?
- factorial
- Split - Plot
- ANCOVA

What to do when Variance is 0 ?
we can model like,
Fixed factor mode : Day treatment * treatment
Random " " : Subject * Day
Repeated Variance Model : Diagonal
(time)

LMM

$Y = X\beta + Zu + \varepsilon$

Lets assume $u \rightarrow N(0, G)$, $G \rightarrow$ variance - covariance matrix of u

$\varepsilon \rightarrow N(0, R)$, $R \rightarrow$ " of the errors ε

$cov[u, \varepsilon] = 0$

So variance of y i.e

test β (goal of ~~LM~~ MM analysis)

$\uparrow$

unknown vector of fixed effect

Parameters to be estimated

So we need to estimate β, G, R

β Requires estimates of G & R

$$V = Var(y)$$
$$= Var(X\beta + Zu + \varepsilon)$$
$$= 0 + Var[Zu + \varepsilon]$$
$$= ZGZ' + R$$

Z — known design Matrix of Random effects

Fixed effects in a MM is tested by F-test

Variance matrix are obtained by ML & REML

Longitudinal data can be modeled in 2 ways i.e

- covariance pattern models
- Random coefficient models

covariance pattern models: if covariance & correlation bet^n Repeated meas^{ts} are considered; If Repeated measurements occur at fixed intervals & time factor is not imp.

Ex Measuring heart Rate after 2 or few hours

Random coefficient models: to determine Relationship bet^n the Response & time.

How to do?

1. by including the measurement time as a covariate in the model with a corresponding slope.

Here slope vary with subject

So model: a separate intercept & slope for each subject in study.

done by fitting subject variable $\rightarrow$ intercept

subject * time $\rightarrow$ subject to study.

$$G = \begin{bmatrix} \sigma^2_{subject} & \sigma_{subject,\ subject*time} \\ \sigma_{subject,\ subject*time} & \sigma^2_{subject*time} \end{bmatrix}$$

ranef function (check intercept adjustment)

Latent Square Designs

$$y_{ij} = \beta_0 + u_{0i} + \beta_1 SO_{ij} + \varepsilon_{ij}$$

new subject
by subject adjustment to intercept.

Subj, item id.

G1 G2 G3 G4 →

Sub1
⋮
Sub 20.

i

(4) item

j

3

LMM → By subject adjustment. Normal distribution with some std. deviation.

Intercept " also ND.

linear Model : intercept β_0 (for all subject)
Mixed Model : Separate intercept $(\beta_0 + b_i)$ for individual subject.

Each Subject has a different intercept which is Normally distributed

y_{ij} β_0 - grand mean intercept

β_1 → " slope indexed by relative clausetype SO_{ij}

Adjustment term to Intercept assumed to come from Normal distribution with mean 0 & Standard deviation σ as 0

Residuals (epsilon ε, ND with std. dev σ)

Model have to estimate β_0, β_1, Sigma, Sigma$_0$.

Within Subject = how a given individual varies over time.

Repeated Measures data = Repeated Measures Multivariate analysis of Variance

MANOVA → Covariance Structure is unstructured

$\beta_{Days} = 0$ [Reaction time is independent of amount of sleep depreviation]

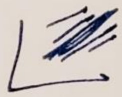
H_a $\beta_{Days} \neq 0$ [Amount of Sleep deprivation does effect reaction time]

For this problem,

we only have within subject fixed effect

lower AIC, better Model

From Random intercept model



It appears the assumption that all Subjects have Same Slope that means,

Same Response to Progressive Sleep deprivation is unrealistic

So Random coefficient model,

this allows Subjects to have both their own intercept & own slope

Pseudoreplication [failure to recognize lack of Statistical independence in data]

Subjects → Random effects (& we don't know their levels)

→ unstructures

→ EN in R. , G matrix what's the next way other than Random effect?

Basic G Structure

Unstructured & Variance Component.

for correlated,

$$y_{ij} = \beta_0 + u_{0i} + w_{0j} + \Big(\beta_1 + \underset{Sub}{u_{1j}} + \underset{item}{w_{1j}}\Big) So_{ij} + \varepsilon_{ij}$$

Subject i & j

There is a grand mean β_0, intercept adjustment by Subjects, intercept adjustment by items j +
for slopes (grand mean slope + Slope adjustment by ~~item~~ Sub + slope adjustment by item) + Slope coding i.e So_{ij}

Variance components

crossed Random effects (no correlation) → be ‖ independent

correlation

$\varepsilon \sim N(0, \sigma)$, mean 0, Standard dev. σ

$\boxed{\begin{array}{l} u_0 \sim N(0, \sigma_{u_0}) \\ u_1 \sim N(0, \sigma_{u_1}) \end{array}}$

$\boxed{\begin{array}{l} w_0 \sim N(0, \sigma_{w_0}) \\ w_1 \sim N(0, \sigma_{w_1}) \end{array}}$

for varying intercept
Varying slope
with correlation

Larger the intercept adjustment of Particular Sub, the larger the slope " " meany Slower Sub. tend to have larger relative clause effect

~~Bivar~~ Bivariate distribution. →

Var-cov matrix =

$$G = \Sigma_U \begin{bmatrix} \sigma^2_{sub} & \rho_{sub, Sub*time} \\ \rho_{Sub, Sub*time} & \sigma^2_{subject*time} \end{bmatrix}$$

Groups	Name	Std.Dev	corr
Sub	interup.	0.31	
		0.11	0.58
item.			

$$G = \begin{bmatrix} 0.31^2 & 0.31 \times 0.11 \times 0.58 \\ 0.32 \times 0.11 \times 0.58 & 0.11^2 \end{bmatrix}$$

Varying Subject & slope

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1j}) \times So_{ij} + \varepsilon_{ij}$$

varying lope

(1 + So || Subject) varying interupt.

Simple linear Model + ~~do~~ adjustment to intercept by Subject mean + Adding an adjustment to slope by subject

$\boxed{\beta_1 + u_{1j}}$ → Slope adjustment term, it can be + or -

Grand Mean slope

We need to know for every Subject & every item i.e for every i & j we can code based on Subject, object relative ± 1

we can know from So_{ij}

Now we will add another term i.e Variance Component →

$u_0 \sim$ Normal $(0, \sigma_{u0})$

$u_1 \sim N(0, \sigma_{u_1})$

$\varepsilon \sim N(0, \sigma)$

bet^n^ Subject variability adjusted intercept & bet^n^ Sub variability in slopes

[LMM, Less variability in slope due to Shrinkage]

We Shrink the estimates of each Subjects towards the Grand mean

As we know 42 Subject

Why? We have sparse or exaggerated estimates so lmer, shrinkage method will drag those Subject estimates as grand mean is more realistic

So any extreme value deviating from grand mean will need to be conservatively brought closer grand mean.

Fixed effect = applied to every individual regardless of cluster or item.
(common slope & common intercept)

Model a Random effect, there will be a fixed effect
Each individual cluster might have diffⁿ slope,
but there's alway going to be an avg slope accross all cluster
called Fixed effect

Random effect = we include RE when we expect that, within clusters they might deviate from the avg. slope or intercept

Do u expect each cluster to have their own slope? add ~~Rand~~ RE
" "all clusters to have diffⁿ intercepts? "

for categorical data, don't add RE

i — individual
j — cluster (item)

Random Effect ANOVA
means u don't have any Predictors
- modeling differences of mean

Random Intercept Model

the intercepts are allowed to vary but slopes are fixed

Random slope Model

· Originated at same point but slopes are allowed to vary

in lme() fun.,
random = argument (to specify G & Z)
correlation : (" R)

Ex lme (y~1, random = ~1|f, data
correlation = corAR1())

4 measurement for 30 subs, unstructured covariance
How many cov. Parameters are estim?
Y full data 120x120 (block diagonal)
30 blocks (4x4)
each block 10 cov. Param [4 var, 6 cov]

for LMM, common optimization routine is α, β (Matrix algebraic form)

(Modeling Random Effects $G_{sub} = q \times q$ [q = no. of random effect for each subj]

G Matrix)

How to model?

- by adding Random Slopes

If all subjects start with time 0, then ~~slope~~ intercept is not needed.

When G is used then R is simply $\sigma^2 I$

How to choose a covariance structure?

- if measures taken over time, consider AR(1) or Spatial
- if there are limited no. of observations taken over space / accross treatment then UN or Compound Symmetry (CS)

13 Subject

$$x_i = z_i = \begin{bmatrix} 1 & 1 \\ \vdots & 2 \\ \vdots & 3 \\ \vdots & 6 \\ 1 & 12 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$b_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}$$

$$\varepsilon_i \sim N(0, R_i)$$

$$R_i = \sigma^2 I_{n_i \times n_i}$$

$$b_i \sim N(0, G_i)$$

$$G_i = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$$

or

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$$

(Modeling Error Covariance Structure

R Matrix)

How to choose?

- 1st ask How Random effects are specified
- More RE, less complex R matrix

What does it mean to have +ve covariance?

Assets or two ~~groups~~ variable move positively or same direction.

Structures of G

Diagonal

$$\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

used when covariance = 0 → Random effect models

unstructured.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \rightarrow$$ used for Random coefficient Model.

Covariance Structures

$\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$ Simple (1 cov. Parameter)

Compound symmetry

$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}$ (2 cov).

Toeplitz) = correlation = corARMA(form = ~1|id, p=3, q=0)):

↓

4 cov. $= \begin{bmatrix} \sigma^2 & & & \\ \sigma_1 & \sigma^2 & & \\ \sigma_2 & & & \\ \sigma_3 & & & \end{bmatrix}$

for AR(1)

corARMA(form = ~1|id, p=1, q=0), Type = AR(1)

Ex - 2 cov, $\sigma^2 \begin{bmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & & 1 & \\ \rho^3 & & & 1 \end{bmatrix}$

1st order AR moving average (Type = ARMA(1,1), correlation = corARMA(form = ~1|id, p=1, q=1)

3cov → $\sigma^2 \begin{bmatrix} 1 & & & \\ r & & & \\ r\rho & 1 & & \\ r\rho^2 & & 1 & 1 \end{bmatrix}$

Compound symmetry

M <- gls(y ~ x * as.factor(x1), correlation = corCompSym(form = ~1|Person), data)

Unstructured

in nlme, correlation = corSymm(form = ~1|id)

weights = varIdent(form = ~1|time)

10 cov. parameter

$k + (k-1) + \cdots + 1$

$k = 4$, $4+3+2+1 = 10$

Repeated measure

cov. Parameter.

$\begin{bmatrix} \sigma_{11} & \sigma_{21} & & \\ \sigma_{12} & \sigma_{22} & & \\ \sigma_{13} & & & \\ \sigma_{14} & & & \sigma_{44} \end{bmatrix}$

Toeplitz

gls(distance ~ Gend * as.factor(age), correlation = corARMA(form = ~1|person, p=3, q=0), data = dental)

↓

no. of covariance parameter

RMatrix (vcov matrix of Random errors)

* If repeated measurement times are random, $R = \sigma^2 I$

Ex- 5 time points for each Sub measured: 1hr, 2hr, 5hr, 10 & 24

$n_1 = 5$, so $R_{sub} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \sigma_4^2 & \\ & & & & \sigma_5^2 \end{bmatrix}$

but if 2nd Sub measured as 1, 5, 24 & 2, 10 is missing they.

$R_2 = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_3^2 & \\ & & \sigma_5^2 \end{bmatrix}$, $n_2 = 3$

Structures: Diagonal

Homo $\begin{bmatrix} \sigma^2 & \\ & \sigma^2 \end{bmatrix}$ Het $\begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix}$ correlation $\begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$

compound symmetry:

Homogenous $\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$ Hetero $\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ Correlation $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

AR(1)

$\begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$ $\begin{bmatrix} \sigma_1^2 & \rho^2\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \\ \sigma^2\sigma_3\sigma_1 & & \sigma_3^2 \end{bmatrix}$ $\begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$

So, G = 21 sub i.e 21×21

group = 7

So R = 21×7 = 147